Apport de l'apprentissage automatique pour la modélisation et l'analyse des changements d'occupation du sol

Valentine Judge

▶ **To cite this version:**

Valentine Judge. Apport de l'apprentissage automatique pour la modélisation et l'analyse des changements d'occupation du sol : Application au développement urbain de la zone frontalière franco-allemande. Géographie. Université Bourgogne Franche-Comté, 2019. Français. NNT : . tel-02983188

HAL Id: tel-02983188
https://hal.science/tel-02983188

Submitted on 29 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE L'ETABLISSEMENT UNIVERSITE BOURGOGNE FRANCHE-COMTE
PREPAREE AU LABORATOIRE THEMA

École doctorale
Société, Espace, Temps

Doctorat de géographie et aménagement du territoire

Présenté et défendu par
Mme Judge Valentine

Apport de l'apprentissage automatique pour la modélisation et l'analyse des changements d'occupation du sol

Application au développement urbain de la zone frontalière franco-allemande

Présenté à Besançon le 16 juillet 2019

Composition du Jury :

M. Escobar Francisco	Professeur à l'Université d'Alcala, Espagne	Rapporteur
M. Fusco Giovanni	Chargé de recherche HDR au CNRS, France	Rapporteur
M. Engelen Guy	Chercheur émérite VITO Institute, Belgique	Examinateur
M. Köppen Bernhard	Professeur à l'Université de Koblenz - Landau, Allemagne	Président du jury
M. Antoni Jean-Philippe	Professeur à l'Université de Bourgogne Franche-Comté	Directeur de thèse
M. Klein Olivier	Chargé de recherche LISER Institute, Luxembourg	Codirecteur de thèse

Remerciements

Ces remerciements sont le résultat d'un fils rouge d'une durée rétroactive de 4 semaines à compter de la date de rendu. La liste suivante n'est, volontairement, pas hiérarchisée. En effet, l'importance de chaque élément a pu varier suivant l'instant précis de la thèse dont il s'agit. Merci à :

- Vivaldi, Hans Zimmer et tous les autres musiciens qui m'ont accompagné alors que je couchais les mots sur le papier.
- Mes directeurs Jean-Philipe Antoni et Olivier Klein pour leur encadrement et les discussions dont le résultat apparaît noir sur blanc (et parfois en couleur) tout au long de ce manuscrit.
- Mon laboratoire ThéMA, pour l'accueil et la découverte du monde Bisontin de la géographie théorique et quantitative. À tous ses membres qui ont de près ou de loin facilité les démarches administratives, et apporté un support technique optimal.
- Au LISER, pour la découverte d'un monde de la recherche éclectique et puisque ce projet n'aurait pas pu être initié sans son implication.
- Mes parents qui m'ont toujours permis et aidé à réaliser mes études comme je le souhaitais, en France ou ailleurs.
- Gilles, pour avoir implémenté ma méthode dans LucSim et évité ainsi les dizaines d'aller-retour Arcgis, R du premier essai manuel.
- Celles et celui qui ont eu la gentillesse de relire ces quelques pages avec efficacité (Léa, ma Maman, Julie, Thibault)
- Maud et Pauline qui ont facilité la finalisation de ce manuscrit par leur intervention !
- Mes ami.e.s qui, à travers un monde sans frontières, ont contribué dans une plus ou moins grande mesure à cet achèvement.

Merci à mes deux structures d'accueils pour le financement de cette recherche et des multiples déplacements qui m'ont permis d'agrandir le champ des possibles en matière de réflexion.

Merci à tous mes collègues doctorants, ceux avec qui j'ai commencé et fini : Vincent, Mehdi, Maud, Nico, Max, Joe, Yohan, Boulou, Emilie, Sophie, Claire, …

Merci à Anke pour les discussions, et l'accueil lors de mon passage éclair à l'Université d'Ottawa !

Merci à Anaïs pour son aide précieuse, ainsi qu'à Marion et Laure pour toujours plus de discussions motivantes !

Et un immense merci à Léa.

Avant – propos

Cette recherche doctorale a été financée par le Fond National de la Recherche Luxembourgeoise.

Introduction Générale

L'étude des changements d'occupation du sol constitue un enjeu majeur tant sur un plan environnemental que sur le plan de la gestion d'une ressource : l'espace terrestre présente un caractère fini. Cet enjeu constitue désormais un champ de recherche scientifique à part entière en raison de la prise de conscience de l'impact des changements anthropiques sur les macrocycles environnementaux comme celui de l'eau par exemple (Rindfuss et al., 2004; Turner et al., 1993). Le « *Land Change Science* » (Turner et al., 2007) vise ainsi à comprendre les dynamiques d'occupation du sol comme un système relevant des interactions du couple « Homme – Environnement ». Les changements et les contraintes, initiées tant par l'Homme que par les dynamiques environnementales, se matérialisent différemment suivant leur localisation. La différence de forme qu'ils adoptent s'explique notamment par les divers contextes environnementaux dans lesquels ils s'installent (climat, topographie, etc.), mais également par la variété des contextes anthropiques émergeant suivant les continents, les pays ou les régions. Ainsi, la forme que prennent les changements d'occupation du sol dépendent également du contexte socio-économique, politique et technologique développé par les sociétés humaines à travers le monde. A. M. Hersperger (2018) définit ces trois composantes comme des caractéristiques culturelles. L'influence culturelle serait donc observable à travers la forme et la configuration des changements d'occupation du sol (Koceva et al., 2016) particulièrement dans le cadre de l'étude des espaces urbains (Couclelis, 2005). En effet, le développement urbain apparaît distinct selon sa localisation géographique et le contexte culturel dans lequel il s'ancre. Pour autant, et c'est là le paradoxe, l'action d'urbaniser s'appuie toujours sur un seul et même principe : construire des bâtiments pour l'habitation et les activités humaines. Par conséquent, l'objectif principal de ce travail de recherche est de déterminer si une part universelle du processus de développement urbain peut se distinguer d'une part spécifique, liée à une localisation géographique. Pour ce faire, notre travail propose d'étudier plus spécifiquement le développement urbain à travers une démarche de modélisation.

Si certaines recherches se sont déjà penchées sur l'étude des processus de changement d'occupation du sol, notamment à travers l'élaboration de modèles, ce travail de thèse propose s'intéresser plus spécifiquement à l'influence de la composante culturelle intervenant dans la caractérisation des processus au regard de leur localisation. Autrement dit, nous proposons d'étudier l'incidence de la culture locale sur l'homogénéité et/ou l'hétérogénéité d'une zone frontalière à travers la modélisation du développement urbain :

Est-il possible de dissocier une composante locale du développement urbain du processus d'urbanisation, soutenue par des pratiques différentes au regard de la variation culturelle intrinsèque aux espaces séparés par une frontière étatique, d'une composante plus générale qui caractériserait le processus indépendamment de sa localisation ?

Changement d'occupation du sol et processus d'urbanisation

Les relations Homme-Environnement constituent un facteur prépondérant des changements d'occupation du sol (Turner et al., 1993; Rindfuss et al., 2004). La littérature scientifique dédiée à cette question met clairement en lien l'influence anthropique sur les changements d'occupation du sol, et l'émergence multi-scalaire et multi-domaines de leurs impacts (Rounsevell et al., 2012; Seto et al., 2011; Verburg et al., 2013). En effet, des changements initiés en un lieu spécifique peuvent influencer les macrocycles environnementaux, comme les conditions sociaux-économiques à un niveau plus local. Les interactions entre la multitude de parties prenantes induit l'émergence d'un système dont la complexité est étudiée depuis plusieurs décennies (White et al., 2015). Les observations rendues possibles par l'acquisition d'image satellite ou de photographies, et l'analyse de publications scientifiques, ont permis de mettre lumière l'aspect protéiforme de la matérialisation de l'urbanisation dans l'espace. La diversité des formes peut donc être expliquée, en partie, au regard du contexte environnemental dans lequel elle s'insère (climatique, topographique, etc.), ou du contexte économique qui ne permet pas toujours l'élaboration des structures urbaines identiques. Par exemple, la périurbanisation autour de Paris ne ressemble pas aux bidonvilles autour de New-Delhi, ce en dépit du fait qu'elles remplissent les mêmes fonctions : héberger des individus ou abriter des activités. Pour autant, des divergences sont également notables quand ces deux paramètres sont plus similaires. Ainsi l'urbanisation autour de Paris ne prend pas, elle non plus, la même forme que celle de Barcelone. Dans ce cas, une des différences qui pourrait expliquer cette variation est celle de la culture de française ou espagnole.

La littérature et les observations permettent donc de poser le postulat suivant : les caractéristiques culturelles locales influencent la forme de développement urbain et par conséquent les caractéristiques de l'environnement favorables à ce processus. Or, l'observation de la forme prise par le développement urbain à proximité d'une frontière permet de se questionner sur la portée de ce postulat. Il est parfois difficile de faire intuitivement la différence entre l'urbanisation de part et d'autre d'une frontière. Pour autant, la différence culturelle est fondamentalement présente, de manière plus ou moins importante selon la frontière envisagée. Elle peut être linguistique, économique, monétaire, politique ou porter sur l'organisation structurelle des institutions de chaque pays. Nous nous interrogeons donc sur la portée de cette influence culturelle dans la caractérisation du processus spatial de développement urbain, mais surtout sur la part d'universalité qu'il peut présenter, indépendamment de sa localisation géographique.

Application dans un contexte frontalier

C'est à travers l'étude d'un espace frontalier **intra-**européen que sera abordée l'hypothèse de l'existence d'une part universelle permettant de caractériser le développement urbain. En effet, la frontière marque une limite dans l'espace, notamment culturelle. Pour autant,

l'évolution des zones frontalières européennes à travers l'histoire permet également de s'interroger sur la prégnance de son rôle contemporain en tant que limite. La littérature définit la frontière comme un objet évolutif. Son rôle et la perception qu'en ont les institutions, comme les individus, ont donc eux aussi évolué. Il devient dès lors pertinent de s'interroger sur l'homogénéité d'un espace frontalier à travers l'étude du développement urbain pour savoir si, comme semble l'indiquer la littérature, ce dernier est influencé par la culture locale (Hersperger et al., 2018).

L'analyse du processus de développement urbain sur un espace frontalier nécessite une réflexion sur les données mobilisées, notamment au regard de l'harmonisation nécessaire à leur exploitation. L'utilisation de données géographiques spatialisées est ici privilégiée. Cette démarche a pour but de favoriser une description objective et harmonisée de l'espace transfrontalier. En effet, pour être en mesure d'étudier une différence émergente dans le processus de développement urbain, il est nécessaire d'introduire un minimum de données présentant une différence intrinsèque en raison de la localisation de l'un ou l'autre des côtés de la frontière. La zone d'étude est donc décrite uniquement à travers l'occupation du sol et la caractérisation de l'accessibilité aux pôles urbains majeurs. Dès lors, sur la base de ces données, l'objectif consiste à d'observer si le processus de développement urbain se différencie de part et d'autre de la frontière. Par ailleurs, l'utilisation de données libres d'accès favorise également le caractère transposable de la méthodologie développée dans ce travail.

Une approche de la modélisation par l'apprentissage automatique (Machine Learning)

Partir des données pour construire une la connaissance nouvelle, tel est l'objectif du processus de modélisation et de l'utilisation de modèles, en géographie comme ailleurs. Un modèle est en effet souvent mis au point pour capturer et résumer l'information cachée dans un jeu de données plus ou moins conséquent. Nous évoluons aujourd'hui dans un flot continu d'informations, dont la plupart peuvent être enregistrées et analysées dans le but de parfaire la compréhension du monde dans lequel nous vivons. Une quantité importante de données peut donc être mobilisée. Cependant, le problème reste, justement, d'en extraire l'information intéressante et adéquate pour traiter la question étudiée. La profusion de données peut en effet également induire une nouvelle complexité technique dans la gestion de leur volume, ou du temps de traitement, mais aussi dans la compréhension de ce que chacune représente. Il devient de ce fait nécessaire de sélectionner judicieusement la méthode de modélisation la plus appropriée.

Notre travail s'est orienté sur l'utilisation d'un automate cellulaire. Cet outil, utilisé dans un premier temps par les mathématiciens de Princeton pour l'étude de l'émergence de processus dynamiques complexes sur la base de la définition de « règles » simples, a très largement été mobilisé en géographie, notamment pour la modélisation des changements d'occupation sur sol. Pour autant, la question de l'élaboration de « bonnes règles » reste et

demeure l'une des principales difficultés pour la modélisation des systèmes urbains. Dès lors, ce travail de thèse propose de coupler un automate cellulaire à un arbre de décision pour élaborer de manière automatique et objective les règles de développement urbain nécessaires à l'étude et à la simulation de ce processus sur la zone transfrontalière. Les travaux de recherche présentés dans cette thèse s'inscrivent donc dans une démarche modélisatrice et tentent de contribuer à la réflexion portant sur l'influence de la frontière franco-allemande sur la zone d'étude à travers la modélisation du développement urbain.

Pour ce faire, la thèse est divisée en trois parties à l'image de l'élaboration du raisonnement dont fait l'objet cette introduction.

La première partie proposera de revenir sur des éléments de littérature qui mettent en lien l'influence anthropique sur les changements d'occupation du sol, et l'émergence multi-scalaire et multi-domaines de leurs impacts. Après avoir illustré le caractère complexe de l'étude des changements d'occupation du sol il s'agira de s'attarder sur le lien entre le développement urbain et le contexte culturel dans lequel il s'intègre. Enfin nous présenterons plus en détails la zone frontalière retenue dans l'espace franco-allemand de Strasbourg et Kehl, ce contexte étant par la suite privilégié pour l'étude de l'influence culturelle sur la caractérisation du processus d'urbanisation.

La deuxième partie sera consacrée à la modélisation du processus de développement urbain sur la zone frontalière. Nous présenterons dans un premier temps l'automate cellulaire, outil de modélisation sélectionné pour ses atouts. Ce choix s'appuie notamment sur sa composante spatiale intrinsèque et sur la simplicité de son paradigme d'un point de vue thématique mais aussi technique. Ces deux points présentent certes un avantage pour notre travail, mais ils impliquent l'élaboration des règles de transitions adaptées à la modélisation du processus d'urbanisation de part et d'autre de la frontière. Il s'agira donc dans un second temps de présenter la méthode sélectionnée pour définir et calibrer ces règles pour simuler le processus dynamique d'urbanisation. Le principe de l'élaboration automatique du jeu de règles à travers l'utilisation d'une méthode d'apprentissage automatique par arbre de décision a donc été retenu. Ce choix présente un axe novateur dans le couplage d'une méthode d'apprentissage automatique pour définir et calibrer les règles indispensables à l'automate cellulaire. Pour finir, nous présenterons le protocole méthodologique mis en place pour définir le paramétrage de l'algorithme d'apprentissage automatique par arbre de décision et obtenir le jeu de règles le plus performant au regard de nos objectifs de modélisation.

Enfin, la troisième partie propose un protocole pour analyser le jeu de règles obtenu par apprentissage automatique et permet une analyse thématique des facteurs favorisant le développement urbain sur la zone d'étude franco-allemande. Ainsi nous proposons une analyse de la relation entre la localisation du processus de développement urbain au sein de la zone d'étude, notamment dans l'un ou l'autre des pays, et ses caractéristiques. Enfin, il s'agit également de discuter l'ensemble des résultats obtenus, afin de proposer de nouvelles

pistes de recherche ainsi que des perspectives d'application des méthodes utilisées, notamment vers l'aide à la décision en aménagement.

Première partie

Étudier le développement urbain en zone frontalière

Introduction de la première partie

Cette recherche doctorale s'inscrit dans le cadre théorique de l'étude des changements d'occupation du sol, en général, et de celle du processus d'urbanisation, en particulier. De toute évidence, l'occupation du sol change ; c'est un fait établi et observable à différentes échelles de temps. Parmi les facteurs contribuant au changement, le climat et les catastrophes naturelles ont une influence non négligeable sur ce processus. Ainsi, de manière brutale, une coulée de lave peut remplacer une forêt. Par ailleurs, un changement de température influencera sur une temporalité plus longue l'évolution de la couverture végétale, et à terme la présence ou non d'un certain type de végétation. Au demeurant, cette influence est aussi rétroactive puisqu'un changement majeur de l'occupation du sol, initié par un changement climatique, impactera à son tour le climat. Tel est le cas du réchauffement des pergélisols ou de l'accroissement de l'exploitation des zones forestières impulsant une déforestation majeure. Aussi, l'être humain lorsqu'il aménage l'espace, impacte l'occupation du sol et influence donc indirectement, par ce biais, la mise en place de changement plus globaux (industrialisation, déforestation, etc.). C'est la relation entre la forme que prennent les changements d'occupation du sol initiés par l'être humain, suivant leurs localisations dans un contexte environnemental ou culturel spécifique qui nous intéresse plus précisément dans cette recherche.

Après avoir discuté des enjeux relatifs à l'étude des changements d'occupation du sol, en s'appuyant sur la relation du couple « Homme – environnement » décrite dans la littérature, nous aborderons dans un premier temps les changements d'occupation du sol comme un élément pouvant être le reflet d'une culture locale à travers, notamment, les pratiques dont émerge le développement urbain. Dans ce cadre, nous nous intéresserons à l'étude des espaces frontaliers qui présentent l'avantage de renforcer les disparités culturelles au sein d'un même espace, dont la limite est matérialisée par la frontière.

C'est à travers l'élaboration d'un protocole méthodologique basé sur une approche modélisatrice qui s'inscrit dans le cadre de la géographie théorique et quantitative que nous aborderons la question de l'influence culturelle sur la forme du changement d'occupation du sol dans un second temps. La donnée spatiale sera décrite comme un support permettant d'étudier le phénomène de développement urbain. Dans ce cadre, nous présenterons et argumenterons le choix de l'approche de modélisation retenue et les objectifs qui lui incombe. Un des enjeux majeurs concerne le choix des données mobilisées et leur pertinence au regard de ces objectifs.

Une fois le cadre établi, nous présenterons de manière plus détaillée la zone frontalière franco-allemande qui fera l'objet de notre étude. Les éléments sur lesquels notre choix s'est appuyé seront présentés dans un troisième temps. Il s'en suivra la définition et la description des limites de notre zone d'étude et des données utilisées.

Chapitre 1

Développement urbain : un changement d'occupation du sol

La caractérisation de l'occupation du sol et l'étude de son évolution est un vaste champ de recherche qui ne se limite pas à la géographie. Les sciences du changement d'occupation du sol (« Land use Change Science », LCS) ont émergé en raison de l'enjeu majeur que représente l'analyse et l'observation de l'appropriation anthropique de l'espace terrestre (Turner *et al.*, 2007). Ces dernières ont pour objectif principal de comprendre l'organisation spatiale de l'occupation du sol et son évolution dans le but d'évaluer l'impact des changements (Rindfuss *et al.*, 2008; Turner *et al.*, 2007). Pour autant, les changements d'occupation du sol en général, et le développement urbain en particulier, « résultent des interactions politiques et institutionnelles, économiques, culturelles, technologiques ainsi que de facteurs naturels et spatiaux additionnées de l'influence des différents acteurs » (Hersperger *et al.*, 2018). Il convient alors de s'interroger sur la spécificité géographique des processus à l'œuvre puisqu'ils sont, en partie tout du moins, influencés par un contexte culturel spécifique. C'est à travers l'étude du développement urbain que nous proposons d'aborder la question. En effet, si l'aspect culturel influence ce processus, son objectif n'en reste pas moins similaire quelle que soit sa localisation : la construction de bâtiments visant à accueillir des habitations ou d'autres activités humaines.

Dans ce chapitre, nous nous intéresseront à l'essor des sciences de l'occupation du sol, impulsé par un intérêt croissant pour la compréhension des processus de changement qui impactent directement l'environnement humain. Parallèlement, nous présenterons l'influence anthropique les processus de changements d'occupation du sol, notamment en matière de développement urbain. Ces deux entrées permettront d'aborder la question de l'influence culturelle à travers l'étude du développement urbain, dans l'objectif d'analyser l'ampleur des divergences au regard de la localisation géographique dans un contexte culturel spécifique. Cette question présente un intérêt particulier pour les zones frontalières, dont la frontière établie une différence culturelle intrinsèque. Néanmoins, l'évolution de la notion de frontière à travers l'Histoire et celles différentes politiques mises en place par les États, rend également cette limite poreuse. Ce terrain est donc privilégié pour analyser l'influence culturelle à travers l'étude du développement urbain.

1. CHANGEMENTS D'OCCUPATION DU SOL

L'étude des changements d'occupation du sol présente un enjeu majeur de la recherche contemporaine. En effet, l'intérêt particulier pour la compréhension du phénomène de changement est nécessaire dans la mise en œuvre de la gestion de d'une ressource finie ainsi que dans la mesure de l'impact anthropique sur l'environnement. Nous discuterons dans une première sous-section de l'émergence des sciences de l'occupation du sol, avant de revenir sur des éléments de définition dans la deuxième sous-section. La troisième sous-section permettra de discuter des facteurs de changement aussi bien que de leurs impacts.

1.1. Émergence des Land use Change Sciences

Les Sciences de l'Occupation du Sol représentent un domaine qui s'est largement enrichi au fil du temps, à mesure qu'a été identifiée l'influence de l'homme sur son environnement. Ces sciences apparaissent dans la littérature depuis plusieurs dizaines d'années. En 1976, Anderson établissait déjà la nécessité d'intégrer les changements d'occupation du sol dans l'analyse des processus environnementaux (Anderson, 1976). Le développement de ce champ coïncide en effet avec la mise en lumière de la relation entre les changements d'occupation ou d'utilisation du sol et des modifications à une échelle globale. C'est par exemple l'étude de l'albedo[1] qui a permis d'établir l'influence d'un changement d'occupation du sol sur les flux d'énergie échangés entre la surface de la Terre et l'atmosphère, ce qui impacte le climat global (Lambin & Geist, 2006; Otterman, 1974; Turner *et al.*, 2007). L'impact de ces mêmes changements a été observé sur les « macrocycles » environnementaux, comme celui du carbone, principalement à travers l'exploitation des puits que sont les forêts (Turner, 1990). Ainsi, « une étape fondamentale de développement des Sciences de l'Occupation du Sol fut l'acceptation du rôle de l'humain dans la transformation du paysage et la distinction entre occupation du sol et utilisation du sol » (Rindfuss *et al.*, 2008; Turner *et al.*, 1994). Ce domaine de recherche apparaît donc d'autant plus important qu'il met en lumière les modifications essentielles générées par l'homme sur l'environnement naturel, son cadre de vie quotidien. Il importe dès lors de comprendre ces interactions afin d'améliorer, ou simplement de préserver, la qualité et les standards de vie des populations.

Les sciences de l'occupation du sol ont également bénéficié d'avancées technologiques dans les domaines de l'informatique ou de l'industries spatiale, lesquelles ont permis une analyse plus approfondie et précise des processus spatiaux en améliorant l'acquisition des données. Par exemple, le développement des Systèmes d'Information Géographique (SIG) à partir de

[1] L'albédo est une valeur physique qui permet de connaître la quantité de lumière solaire incidente réfléchie par une surface. Concernant le climat, cette variable est importante car elle exprime la part de rayonnement solaire qui va être renvoyée par l'atmosphère et la surface terrestre vers l'espace et qui ne servira donc pas à chauffer la planète (CNRS, 2O18).

années 1960s favorise l'exploitation des données spatiales (Tomlinson, 1962). Le développement des techniques de télédétection a également joué un rôle considérable dans la documentation des changements d'occupation du sol (Mollicone *et al.*, 2003). La puissance du traitement des données a augmenté, de même qu'a évolué la précision d'acquisition, notamment grâce à des capteurs plus performants. Si la richesse de ces évolutions est significative, ces dernières soulèvent de nouvelles interrogations, notamment sur la façon de manipuler une telle quantité de données (Turner *et al.*, 2007). Ce point sera abordé au cours du deuxième chapitre, puisqu'il est au cœur de la réflexion sur le choix des données nécessaires au présent travail.

Ces avancées technologiques ont été suivies par la mise en place de projet internationaux de *monitoring* et d'exploration à différentes échelles. Plusieurs projets comme le *Global Land Project*[2] (Ojima *et al.*, 2005) à l'échelle globale, *Corine Land Cover*[3] et l'*Urban Atlas*[4] à l'échelle européenne ou encore la *BD Carto*[5] pour le territoire national français ont ainsi vu le jour. Cette activité, tant internationale que nationale, souligne les enjeux majeurs que recouvre ce champ de recherche ainsi que son aspect spatial multi-échelle. Le rayonnement des études sur l'occupation du sol est favorisé par la diffusion de connaissances dans des journaux scientifiques dédiés, traduisant l'intérêt accordé à cette discipline dans la recherche internationale. Des revues académiques réputées telle que *Journal of Land Use Science, Land Use Policy, Journal of Transport and Land Use* en sont les principaux exemples.

L'importance de l'acquisition de données spatiales a été relevée par de nombreux auteurs dans la littérature, et particulièrement soulignée depuis une vingtaine d'années (GIEC, 2000; Lambin & Geist, 2006; Meinel & Winkler, 2002; Turner *et al.*, 2007). La démarche, qui était dans un premier temps, orientée vers l'acquisition de connaissances pour la réalisation d'états des lieux de la ressource sol, s'est progressivement transformée en démarche d'évaluation des impacts anthropiques sur l'environnement, notamment par des analyses rétrospectives. La prise de conscience de leur ampleur a poussé les chercheurs à s'interroger sur le futur, et à mettre en place des méthodes de modélisation et d'évaluation des changements à des fins prospectives. Ces démarches permettent d'évaluer les impacts futurs afin de mener une réflexion sur la mise en place d'actions dans le présent.

L'occupation du sol, en tant que ressource fait donc l'objet d'une attention particulière. L'étude des changements est appréhendée par des domaines de recherche variés, dans le cadre de projets nationaux ou internationaux. La compréhension des phénomènes de changement est un enjeu majeur pour la gestion de la ressource en elle-même, ainsi que pour l'évaluation ou la mitigation des impacts à différentes échelles.

[2] http://globallandcover.com/GLC30Download/index.aspx

[3] http://land.copernicus.eu/pan-european/corine-land-cover/clc-2012

[4] https://land.copernicus.eu/local/urban-atlas

[5] http://professionnels.ign.fr/bdcarto

1.2. Occupation du sol et utilisation du sol

Avant de développer les intérêts et enjeux sous-jacents à l'étude de l'occupation ou de l'utilisation des sols, la définition des termes employés dans le cadre de ce travail de recherche paraît indispensable à la clarté du discours.

Nous définissons l'espace terrestre comme le lieu où les activités des êtres vivants se développent. Son utilisation comme une ressource, notamment par l'être humain, fait émerger les notions de sol (espace terrestre émergé), d'occupation du sol et de son utilisation. Le sol est, d'après la définition proposée par l'Organisation des Nations Unies (ONU) pour l'alimentation et l'agriculture, une « zone délimitée de la surface terrestre, qui englobe tous les attributs de la biosphère situés immédiatement au-dessus ou en dessous, y compris le climat près de la surface, le relief, l'hydrologie avec les lacs peu profonds, les rivières, les marécages et les marais, les couches sédimentaires de surface, les nappes d'eau souterraines et les réserves hydrogéologiques associées, la flore et la faune, l'habitat humain et les marques physiques de l'activité humaine présente et passée (aménagement de terrasses, réserves d'eau ou structures de drainage, routes, bâtiments, etc.) ». Toujours d'après l'ONU pour l'alimentation et l'agriculture, le concept de sol est étendu « aux eaux intérieures et aux bancs découverts à marée basse » (European Union & Eurostat, 2000).

Dès le début des années 1960, Burley suggéra la distinction entre « occupation du sol » et « utilisation du sol », en anglais respectivement *Land Cover* et *Land Use* (Burley, 1961). La frontière dans la définition des deux termes reste discutable et flexible suivant la littérature consultée. Dans le cadre de cette thèse, ils sont définis de la sorte :

- L'occupation du sol est « une description physique de l'espace », de ce que l'on retrouve à la surface du sol. Elle est donc observable. Les moyens d'observation peuvent être divers (télédétection, photographies aériennes, observation directe à travers l'œil humain) et vont permettre d'établir une classification servant de support à la description de l'espace. D'une certaine façon, la classification de l'occupation du sol est la première étape de modélisation de l'espace terrestre.
- L'utilisation du sol, quant à elle, présente une dimension plus fonctionnelle et induite par la présence humaine. Le niveau de détail peut être considéré comme plus fin, mais aussi plus spécifique à l'observation d'un événement donné. Si on prend l'exemple d'une occupation du sol de type étendue d'eau, l'utilisation pourrait être définie en tant que réserve naturelle, ou plan d'eau à vocation de loisirs abritant une base nautique, ou encore une source de pompage, etc. De même qu'une occupation du sol « forêt » peut avoir une vocation forestière type bois de construction ou bois de chauffage, mais aussi simple zone de promenade.

En résumé, l'utilisation du sol n'est pas forcément observable. Elle est façonnée par les besoins humains qui induisent des actions sur un environnement biophysique caractéristique du sol en lui-même (Verburg *et al.*, 2016). On ne peut parfois pas la définir sans un apport externe d'information en complément de l'information visuelle. Au contraire, l'occupation du sol s'observe. Ce travail de recherche, étudiant le développement urbain, se focalise sur l'occupation du sol. L'utilisation du sol ne sera abordée qu'exceptionnellement, et ce dans des contextes très spécifiques.

1.3. Impacts des changements d'occupation du sol.

L'acquisition et l'analyse des données d'occupation du sol font l'objet de classifications différentes suivant les motifs d'exploitation de la donnée. H. Briassoulis (2000) distingue ces motifs selon six catégories d'utilisation : la description, l'explication, la prédiction, l'évaluation d'impact, la prescription et l'évaluation générale.

La description est en général une première étape indispensable à toute étude concernant l'analyse des changements d'occupation du sol. Elle énonce les caractéristiques observables d'un objet spatial, c'est-à-dire l'occupation du sol dans le cas présent. La description est, dans la plupart des cas, faite tant d'un point de vue qualitatif que quantitatif. Elle est conditionnée par les objectifs de l'étude, par l'échelle à laquelle l'analyse est réalisée ainsi que par les données utilisées. Cette étape fournit une matière de base pour toute analyse portant sur les causes ou les impacts des changements. Ces derniers peuvent alors être observés et quantifiés. La description doit cependant être complétée par l'explication pour s'avérer utile à la compréhension des processus et phénomènes.

L'explication, qui vise à répondre à la question « pourquoi ? », lie l'observation à des informations externes. Elle permet de saisir ou d'évaluer les conséquences de la prise de certaines décisions ou de l'application de certaines règlementations. L'explication a pour objectif de comprendre les raisons pour lesquelles les changements observés apparaissent et de dévoiler les forces plus ou moins visibles qui les orientent. Elle peut au contraire mettre en lumière un manque de données explicatives. La précision d'explication est liée à la résolution choisie et aux objectifs de l'étude, mais elle peut différer de la description et de la résolution des données. En effet une évolution locale peut être le fruit d'un processus plus global, et réciproquement.

Les deux étapes décrites ci-dessus peuvent servir de base à la prédiction de futurs changements. On différenciera deux types de prédiction : inconditionnelle et conditionnelle (Turner *et al.*, 1993b). La première s'appuie sur une projection des tendances observées[6], enrichie ou non par l'application d'une théorie particulière. La seconde repose sur l'exploration de différentes alternatives basées sur divers scénarios ou hypothèses. Elle est

[6] L'utilité de l'analyse d'une projection de tendance dans le cadre de l'élaboration de modèle prédictif est souvent discutée dans la littérature (Lambin *et al.*, 2001; Veldkamp & Verburg, 2004).

souvent utilisée pour observer spatialement les résultats de l'application de différentes réglementations, notamment dans le cadre d'une approche normative de l'aménagement. Cette approche normative vise à répondre à la question de ce que devrait être le futur, et permet d'explorer les possibilités pour atteindre un objectif fixé. Elle n'est donc plus descriptive mais bien prescriptive. L'objectif est donc de préconiser les changements nécessaires pour atteindre, par exemple, des objectifs de préservation environnementale, de prospérité économique, d'équité sociale ou d'accessibilité à des pôles urbains.

Enfin, l'analyse de l'impact des changements d'occupation du sol est un pan relatif à l'exploitation des données acquises et permet une évaluation générale des processus. L'intérêt porte donc moins sur le changement de d'occupation du sol en lui-même que sur son incidence socio-économique ou environnementale. Par exemple, le développement durable induit une évaluation des impacts liés au développement dans le but de le rendre « durable ».

L'étude de l'évolution de l'occupation ou de l'utilisation des sols présente donc un intérêt majeur, ce quels que soient les motifs d'exploitation de la donnée précédemment exposés (Rounsevell *et al.*, 2012). En outre, le rapport 2014 du Programme Environnemental des Nations Unies (UNEP) insiste sur la qualification du sol comme une *ressource* ayant un caractère *fini* (UNEP, 2014). La gestion de cette ressource, son caractère complexe et néanmoins essentiel, est un véritable défi. Le Secrétaire Général et Directeur Exécutif de l'UNEP, A. Steiner, souligne l'importance de sa connaissance, qualifiée de « pouvoir » : connaissance des ressorts à l'œuvre, des facteurs et forces en jeu. Cela apparaît d'autant plus fondamental dans un contexte où l'étendue des changements environnementaux n'est pas nécessairement prévisible. Les questions qui se posent sont donc les suivantes (Moser, 1996) :

- Quels sont les facteurs moteurs du changement d'occupation du sol ?
- Quels impacts – direct ou indirect, dans le présent ou dans le futur – ces changements ont-ils sur l'environnement ou les sociétés humaines ?

Les facteurs de changements sont regroupés en deux catégories (Turner *et al.*, 1993b) :

- Les facteurs biophysiques. Ils dépendent des caractéristiques et des processus environnementaux (partie supérieure de la Figure 1. 1). Ils peuvent englober, au sens large, les variations climatiques (globales ou locales), les changements topographiques (glissements de terrain, effondrements, séisme, processus géomorphologiques divers), les inondations, les changements de végétations liés aux phénomènes précédemment évoqués, etc. Les facteurs biophysiques n'engendrent que rarement des changements d'utilisation.

- Les facteurs socio-économiques. Ils représentent la part de l'influence humaine sur les changements environnementaux globaux et reposent sur quatre piliers (Meyer & Turner, 1992) : le changement de population, les changements technologiques, l'organisation socio-culturelle et socio-économique (institutions économiques et marché global, institutions politiques), et l'idéologie.

Moser (1996) présente à ce titre, un couple « Humain – Environnement » comme interdépendant (Figure 1. 1). L'Humain agit sur l'environnement, lequel contraint en retour et dans une certaine mesure, les actions humaines. Le développement agricole est une illustration intéressante de cette relation d'interdépendance : une parcelle forestière peut par exemple faire l'objet d'un changement d'occupation du sol en effectuant une transition agricole (besoin socio-économique). La parcelle est alors utilisée en tant que pâturages puis accueille des cultures. Le type de culture est en partie contraint par l'environnement (climat). D'autres conversions permettent d'accueillir des infrastructures, ou encore des habitations, dont le type et les fondations seront également contraints par l'environnement (géologie) (Briassoulis, 2000). Les agriculteurs sont alors délocalisés vers une autre parcelle forestière où le processus de transition se répètera. Les changements d'occupation du sol sont donc contraints par l'environnement, lui-même affecté par les activités anthropiques.

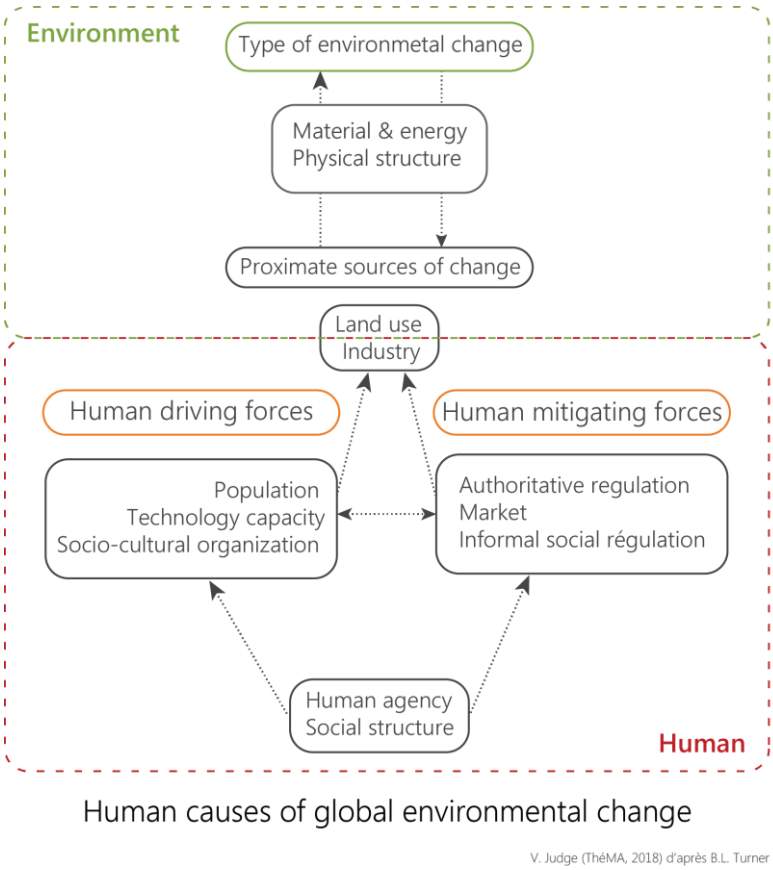


Figure 1. 1 : Couple Humain-Environnement – Impact – interactions – rétroactions

Les impacts humains ont la particularité de dépendre d'une volonté ou d'un besoin (Verburg *et al.*, 2013). Ils peuvent de la même manière être atténués selon les initiatives et orientations de ces besoins. Prenons l'exemple de la combustion des énergies fossiles. Une évolution technologique a permis un changement dans l'utilisation des ressources énergétiques en passant progressivement du « tout charbon » au pétrole ou au nucléaire. La diminution très nette de l'exploitation du charbon a engendré une diminution significative des émissions de gaz liées à sa combustion et de la pollution générée par les exploitations. Dès lors, dans un premier temps, l'abandon des sites d'extraction a atténué la pollution locale. Par ailleurs, l'abandon, partiel ou total suivant les régions, a localement changé l'utilisation, voire l'occupation du sol. En effet, les sites sont devenus inactifs et ont parfois fait l'objet d'une reconversion. Certains sites ont été réhabilités à des fins touristiques, orientés vers un devoir de mémoire, ce patrimoine culturel ayant marqué les régions. Il est alors seulement question de changement d'utilisation du sol puisque l'occupation reste majoritairement la même. La transition a cependant impacté tout le système local provoquant une nécessité de transition économique plus ou moins rapide. Les « gueules-noires », tel que l'on nomme les travailleurs miniers, ont dû s'adapter au changement en trouvant une autre activité : l'exploitation agricole les terrils par exemple (terrils viticoles). Par ailleurs, si l'impact sur la pollution atmosphérique s'en est trouvé atténué, l'abandon de cette ressource n'a été rendue possible que par l'utilisation d'énergies nouvelles : le pétrole ou le nucléaire. Or leur effet présente également une contrainte environnementale, même différente.

Cet exemple illustre l'impact de l'Humain sur l'environnement, les atténuations déclenchées par les changements de son comportement, et les conséquences environnementales ou sociales qui en découlent. Il démontre par ailleurs que les transitions économiques et culturelles, impulsées par une évolution technologique, laissent des traces visibles sur l'utilisation et l'occupation du sol. Les figures suivantes permettent également de réaliser l'ampleur, ne serait-ce que visuelle, de ces évolutions.

La Figure 1. 2montre l'évolution de la couverture arborée sur une vingtaine d'années en Amérique du Sud. La forêt est exploitée pour son bois, mais les cultures de soja, de palmes, ou encore les besoins de zones de pâturages présentent un impact significatif visible sur les images satellites, exponentiel durant les trois décennies. La Figure 1. 3 permet quant à elle d'observer le développement des zones urbanisées autour de Paris au cours du 20ème siècle. Ce développement, observable sur les cartes, n'est pas un processus aussi rapide et important en matière d'espace consommé, que la déforestation contemporaine, facilitée notamment par le développement des technologies de l'industrie agroalimentaire par exemple. Il reste un phénomène d'urbanisation important, qui modifie de manière pérenne l'environnement.

Figure 1. 2 : Images satellites de la forêt amazonienne acquises respectivement en 1986, 1998 et 2005. ©CNES – SPOT

L'ampleur des changements d'occupation du sol, malgré une influence sur les grands processus environnementaux, reste rarement observable à l'échelle globale, dans une mesure aussi prégnante que les phénomènes de déforestation intense (Figure 1. 2) ou d'urbanisation à plus long terme (Figure 1. 3). Pour autant, une multitude de changements, de moindre envergure, s'opèrent à l'échelle globale. Turner parle donc d'effet global « cumulé ». Il s'agit d'un cumul de changements régionaux ou locaux. Ces derniers n'ont pas, pris isolément, d'incidence directe sur les processus globaux, mais viennent à en avoir par la multiplicité de leurs occurrences autour du monde (Turner, 1990). Ainsi, des phénomènes de déforestation, d'expansion agricole, ou d'urbanisation de l'ampleur de ceux observés Figure 1. 2 ou d'une amplitude moindre, mais multipliés et disséminés à travers le monde, influencent les processus environnementaux globaux.

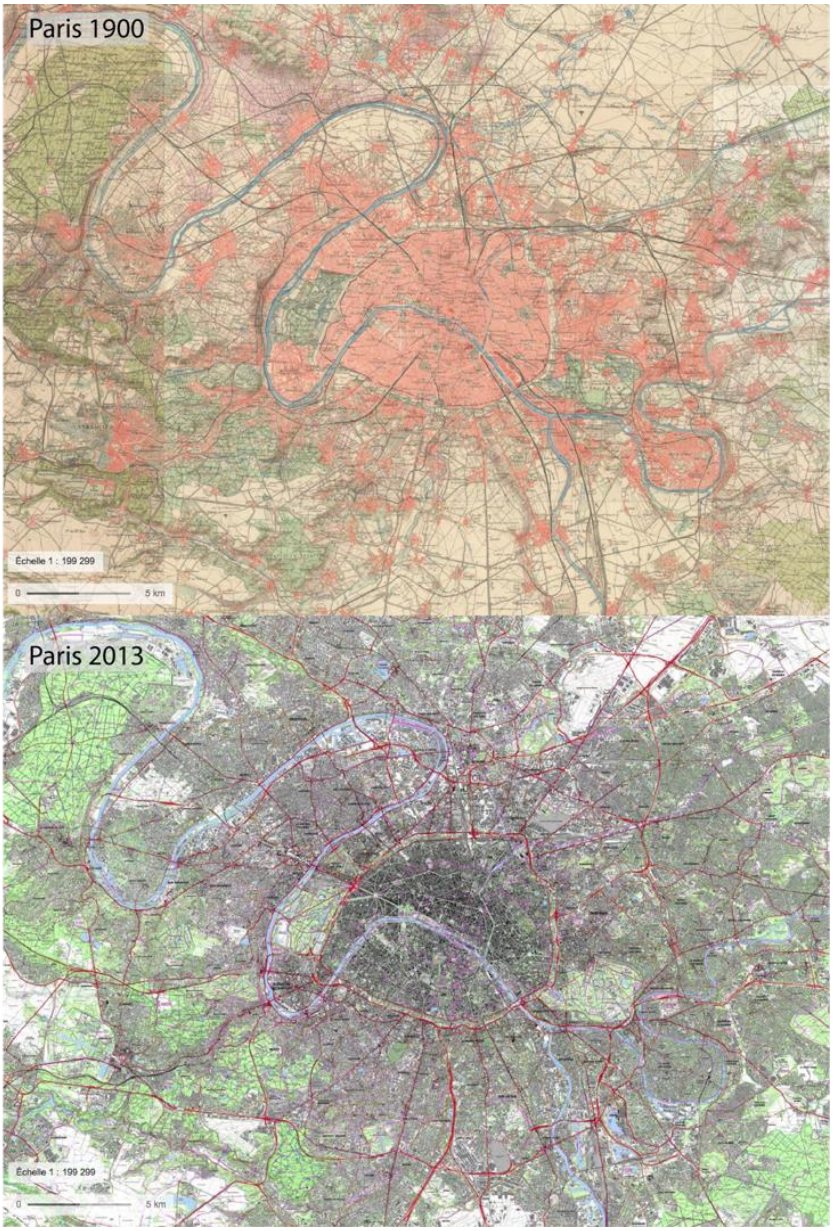


Figure 1. 3 : Développement de l'urbanisation autour de Paris entre 1900 et 2013(Source : Géoportail – Carte Topographique 1900 – 2013).

Comme mentionné précédemment, les facteurs biophysiques sont intrinsèquement liés aux facteurs socio-économiques. Ces changements environnementaux, induits ou accentués par l'humain, soulèvent des questions de sécurité alimentaire, de pénurie d'eau, de vulnérabilité des populations en général, soit des risques naturels et technologiques. *In fine*, certains contextes peuvent engendrer des tensions socio-économiques voire politiques (Biermann *et al.*, 2010). Des organismes internationaux tels que la FAO (Food and Agriculture united nation Organisation) ou l'IHDP (International Human Dimension Programme) tentent donc d'évaluer l'impact des politiques mises en place par les différents pays ou groupement de pays à l'échelle globale.

À une échelle plus fine, les impacts environnementaux sont listés par H. Briassoulis[7] (2000). Parmi eux : l'eutrophisation des milieux aquatiques, l'acidification des écosystèmes qu'ils soient aquatiques ou terrestres, les inondations, la pollution par les nitrates ou l'appauvrissement et la désertification du sol, la pollution des nappes d'eaux souterraines, etc. Ces manifestations peuvent être locales ou régionales, voire dissociées spatialement et temporellement de leur source (pluies en amont provoquant des inondations en aval, pluies acides, pollution chimique mobilisée *a posteriori*). Ces impacts sur différents milieux sont directement ou indirectement liés aux changements d'occupation ou d'utilisation du sol. Par exemple, les problèmes locaux de sécurité alimentaire ou de pénurie d'eau peuvent provenir du développement de l'urbanisation qui nécessite une production/extraction plus importante de ces denrées. Paradoxalement, cette urbanisation nécessite de l'espace et ne peut se faire qu'au détriment des zones agricoles avoisinantes, donc de la production de nourriture. *A contrario*, si les zones agricoles ou industrielles sont étendues, la demande en eau augmentera. Pour autant, ces deux secteurs sont intrinsèquement liés au développement des populations, et donc à l'urbanisation sur laquelle ils s'appuient.

D'autres exemples spécifiques d'enjeux et d'impacts liés aux changements d'occupation du sol apparaissent dans la littérature (Plieninger *et al.*, 2016; Rounsevell *et al.*, 2012; Seto *et al.*, 2000; Turner *et al.*, 1993a, 2007; Verburg *et al.*, 2013). Toutefois cette recherche n'a pas pour objectif de fournir une description exhaustive des processus et interactions, mais bien d'illustrer globalement le phénomène et ses implications. Les enjeux plus spécifiques au développement urbain, qui constituent l'élément central de ce travail de recherche, seront abordés dans un second temps (Section 1.2).

À travers les enjeux liés à l'étude des changements d'occupation du sol, cette section a présenté deux axes de recherche majeurs :

- l'acquisition et l'analyse de données concernant la gestion d'une ressource ;
- l'utilisation de ces données dans un but prospectif en aménagement.

Les sous-sections précédentes mettent également en évidence le lien existant entre l'action d'aménagement initiée par les divers acteurs et les besoins socio-économiques, commerciaux ou encore les volontés politiques qui sous-tendent ces aménagements. Les changements d'occupation du sol reflètent donc une dynamique humaine. C'est pourquoi, le présent travail de recherche s'oriente vers l'analyse des changements d'occupation du sol dans l'objectif d'apporter une meilleure compréhension du phénomène, reflet de l'action humaine. C'est particulièrement le développement urbain qui fera l'objet d'une étude approfondie. En effet, l'urbanisation représente un des trois axes majeurs de changement avec celui du développement agricole et de l'évolution des surfaces forestières.

[7] Elle s'appuie entre autres sur les travaux (Blaikie & Brookfield, 1986; Brouwer & Chadwick, 1991; Jongman, 1995)

2. DÉVELOPPEMENT URBAIN

« Une étape fondamentale de développement des Sciences de l'Occupation du sol fut l'acceptation du rôle de l'humain dans la transformation du paysage » (Rindfuss *et al.*, 2008; Turner, 1994). Parmi les changements d'occupation du sol, le développement urbain en particulier, « résulte des interactions politiques/institutionnelles, économiques, culturelles, technologiques ainsi que de facteurs naturels et spatiaux additionnés de l'influence des différents acteurs »(Hersperger *et al.*, 2018). Il est donc un élément central de l'étude, sa forme d'expression paraissant intrinsèquement liée au contexte de l'action humaine dans lequel il s'insère.

2.1. Les enjeux du développement urbain

Au sein des changements d'occupation du sol, l'urbanisation est l'une des marques anthropiques les plus pérennes dans le temps (Seto *et al.*, 2011). Elle concurrence les autres types d'occupation, notamment dans les pays développés, où elle intervient par exemple au détriment des terres agricoles (Seto *et al.*, 2000). Pour autant, l'étude du développement urbain n'occupe pas la première place dans les domaines de recherche liés à l'occupation du sol. Les changements agricoles ou forestiers restent en effet étudiés prioritairement à l'urbanisation (Magliocca *et al.*, 2015; Seto *et al.*, 2011). Mais s'il est aisé de comprendre leurs impacts, ceux du développement urbain sont tout aussi conséquents pour le climat, la biodiversité, les systèmes hydrologiques, etc. (Seto *et al.*, 2011).

« Il est communément admis, en particulier dans les régions urbaines, que l'aménagement de l'espace (*spatial planning*) – une action dont les objectifs variés incluent des projets d'aménagements, la planification et la stratégie d'occupation du sol – influence la configuration et le motif d'occupation ou d'utilisation du sol » (Couclelis, 2005; Hersperger *et al.*, 2018). Dans l'étude de ces processus d'aménagement, A.M. Hersperger (2018) fait état de deux paradigmes distincts voire opposés. Les experts en aménagement tendent à prioriser l'influence du contexte et de la construction sociale de l'espace au sein de ces processus. Pour les experts en changement d'occupation du sol, il s'agit plutôt de réalités pouvant être objectivement décrites, mesurées, voire même généralisables à différents espaces (Briassoulis, 2008; Hillier, 2007). Une recherche qui lierait ces deux paradigmes reste rare. C'est donc précisément à cette interface que se situe le travail développé par la suite. **Cette recherche vise à étudier l'émergence de caractéristiques socio-culturelles lors de la description objective et généralisable du processus de développement urbain et résidentiel.**

Les facteurs influençant le développement urbain sont similaires à ceux responsables des changements d'occupation du sol en général (Figure 1. 4). La dimension culturelle est définie d'après trois composantes interdépendantes : socio-économique, politique, technologique (Hersperger *et al.*, 2018). Si la nature est un élément objectif contraignant physiquement les tendances de développement, les facteurs culturels divisés selon les trois composantes ci-dessus, sont également prépondérants à un niveau équivalent aux données naturelles.

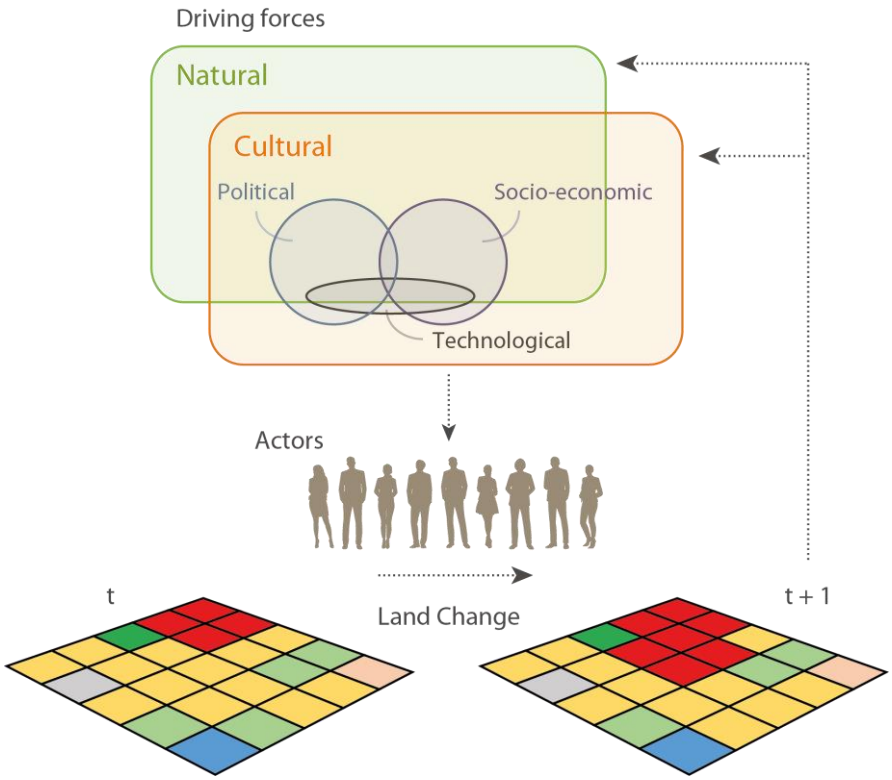


Figure 1. 4 : Le développement spatial dans les régions urbaines(Hersperger et al.*, 2018)*

Le développement urbain reflète le comportement et les attentes des sociétés. En effet, le choix d'une résidence est, par exemple, directement lié aux affinités et aux objectifs de vie des populations au niveau individuel ou à celui des ménages, dans un contexte politique et socio-économique imposant un certain cadre. Ces éléments sont interdépendants. Le choix résidentiel est ainsi effectué selon des caractéristiques personnelles : cycle de vie, revenus, catégories socio-professionnelles par exemple. Il est également contraint par un contexte : la disponibilité des logements et celle des terrains à bâtir (Ben- Akiva & de Palma, 1986; Griffond-Boitier *et al.*, 2012; Houssemand *et al.*, 2018). La localisation de ces derniers est dépendante des possibilités liées à l'occupation du sol mais aussi de la volonté politique en matière d'aménagement puisque les conséquences de l'urbanisation ont une portée sociétale (étalement urbain). Les choix politiques eux-mêmes, comme les autres facteurs, sont alors encadrés rétroactivement par le contexte culturel et environnemental (Rounsevell *et al.*, 2012). Des outils de planification (plans d'urbanisme) ou d'orientation sont ainsi mobilisés pour organiser l'espace en fonction des besoins et des volontés locales, même si, selon

plusieurs auteurs, le développement urbain effectif n'est pas nécessairement en adéquation avec ces plans que ce soit au niveau de sa localisation spatiale, de son contenu, ou de l'intensité du contenu (Loh, 2011; Tesema & Dinsa Sori, 2017; van Karnenbeek & Janssen-Jansen, 2018). En effet, un développement résidentiel prévu peut en réalité devenir un développement industriel ; une zone prévue pour accueillir une faible densité de bâtis peut en réalité accueillir des immeubles de tailles conséquentes. Ces divergences et inadéquations varient là encore suivant la localisation géographique à travers le monde. Par ailleurs, sur certains espaces où le développement urbain n'est pas contraint réglementairement par l'existence d'une planification, il y a néanmoins urbanisation. Dès lors, dans l'ensemble de ces espaces, par-delà les volontés précises et attentes locales des sociétés, qu'il existe ou non des outils de planification, la tendance commune est bien celle d'une urbanisation. Le seul outil réglementaire n'apparaît donc pas suffisant pour comprendre l'intégralité de ces processus.

Si l'impact humain sur les changements d'occupation du sol est indéniable, et particulièrement prépondérant quand il s'agit du développement urbain, la question de son organisation, contrainte par le contexte culturel tel que défini par A. M. Hersperger (Figure 1. 4), reste ouverte.

Dans ce contexte, ce travail oriente la réflexion sur l'influence de l'aspect culturel dans le développement urbain, et ce à travers les formes variées qu'il peut prendre selon la localisation géographique.

2.2. Les formes du développement urbain

Si les facteurs de changement d'occupation restent globalement communs malgré les différentes localisations géographiques, la forme prise par le changement apparaît dans une certaine mesure façonnée par le contexte et l'espace considéré. Un résultat différent peut parfois aisément s'expliquer en raison des contraintes imposées par le cadre environnemental du développement. Ainsi, l'exploitation agricole en Europe (champs), ne prendra pas la même forme que l'exploitation agricole en Asie (rizières). Pour cause, les climats et caractéristiques environnementales (topographie, géologie, etc.) divergent profondément. L'exploitation agricole est ici intrinsèquement liée à l'environnement naturel dans lequel elle s'intègre, indépendamment de la volonté des acteurs de la chaîne d'exploitation. Néanmoins, au-delà de ces données naturelles, les moyens technologiques eux-mêmes pourraient également, du moins en partie, expliquer la différence observée. En effet, l'industrie agricole n'est pas développée pareillement dans certaines parties d'Europe (agriculture intensive) et dans certaines parties d'Asie (agriculture traditionnelle manuelle).

Dans le cas où les contraintes environnementales sont similaires, l'étude du développement urbain paraît tout aussi intéressant : son résultat peut en effet diverger en termes de structures ou de formes. Les photographies aériennes permettent de se rendre compte des différences visuelles d'un point de vue urbanistique, notamment dans la forme ou l'emprise du

développement urbain dont les réseaux de transport dessinent les contours. L'échantillon ci-dessous (Figure 1. 5) a pour objectif de montrer les variations de développement à divers endroits du globe. Ces localisations présentent des contextes environnementaux similaires en termes de topographie. S'il existe des différences climatiques, par exemple, entre la banlieue de Londres et celle de New Delhi, ce sont les variables socio-économiques, culturelles voire historiques que l'on peut immédiatement souligner. En effet, Paris et Buenos Aires sont toutes deux des capitales soumises à une certaine pression démographique. Toutefois, leurs réseaux adoptent des formes respectivement radiale et quadrillée. La banlieue de Londres quant à elle, n'est pas similaire dans sa forme à celle de Los Angeles pourtant également située à la périphérie d'une agglomération majeure. Ce constat s'explique potentiellement par des besoins et contraintes socio-économiques dissemblables des deux espaces. La différence entre l'urbanisation nord-américaine et européenne peut également être soulignée. L'histoire rentre alors en compte. Le développement des villes majeures européennes commence avant l'Antiquité romaine et s'adapte aux besoins à l'instant de leur construction. Sun City en Arizona est en revanche un exemple de ville récente (1960). Sa structure répond donc à la demande de maisons individuelles privilégiant l'utilisation de la voiture (voies larges et parking). S'agissant de New Delhi, la visualisation du réseau routier n'est pas rendue difficile par une vue aérienne éloignée, mais bien par l'absence d'organisation identifiable. Quant aux villes européennes, c'est la forme et la géométrie des îlots qui varie entre Paris, capitale française, et Valence, localisée sur la côte sud-est de l'Espagne. Celle de Valence est très similaire à la forme d'une autre ville espagnole, Barcelone.

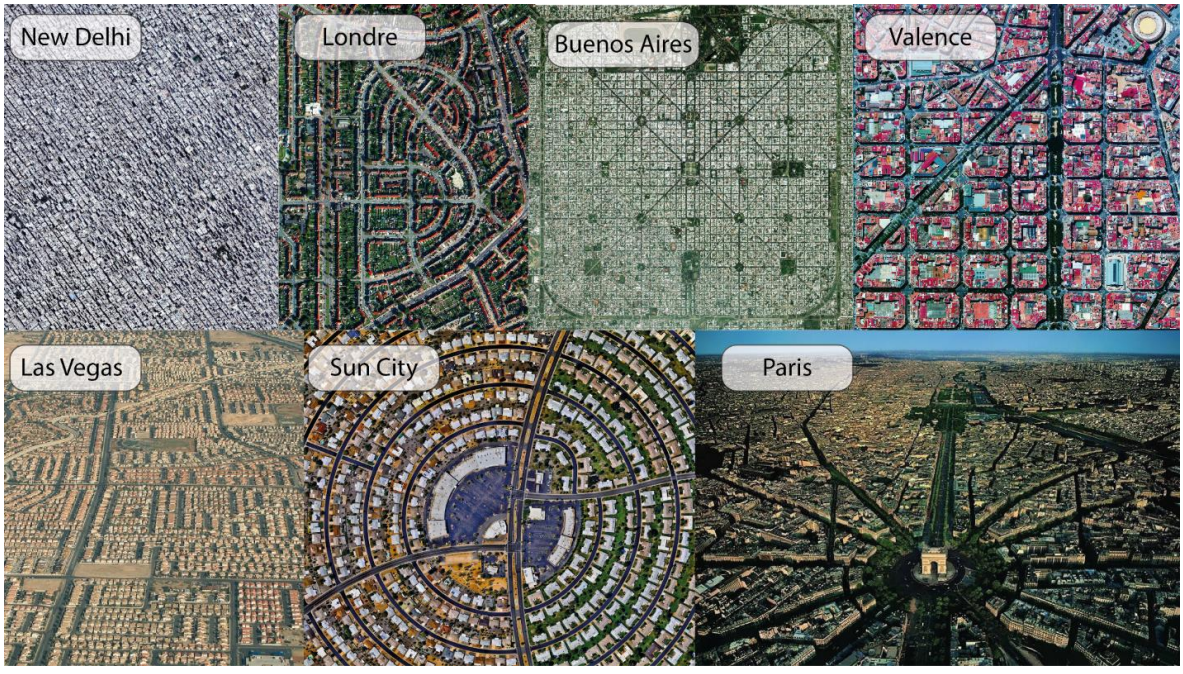


Figure 1. 5 : Physionomie du développement urbain à travers le monde (Source : Pinterest)

Le développement urbain s'exprime donc différemment dans l'espace, pour un même objectif : abriter des logements ou des activités humaines. Il est également dépendant de composantes culturelles et environnementales comme tout changement d'occupation du sol.

L'objectif de ce travail est d'étudier l'expression de l'aspect culturel du développement urbain dans la description du processus. Les exemples internationaux présentent une diversité particulièrement intéressante, et dont il convient d'interroger les causes. Pour autant, il est possible d'observer des développements urbains similaires dans des contextes culturels pourtant différents, notamment en Europe. Ce point est développé dans la section suivante.

2.3. L'Europe face au développement urbain

À l'échelle mondiale, il existe deux types principaux de changements d'occupation du sol : d'une part, une extension des zones bâties, à vocation urbaine, industrielle ou commerciale, au détriment des zones agricoles, et d'autre part, une extension des zones agricoles au détriment des prairies, forêts et espaces naturels (Gibbs *et al.*, 2010; Holmgren, 2006). Holmgren estime qu'à l'échelle mondiale 80% de l'extension urbaine s'effectue sur des zones agricoles. Toutefois la totalité des changements d'occupation du sol en Europe, d'après les données Corine Land Cover (CLC) acquises entre 1990 et 2006, s'élève à près de 7 Mha, et ne concerne donc que 0,05% de la surface émergée mondiale et 0,7% de la surface européenne (1018 Mha). Les changements d'occupation du sol en Europe ne sont donc pas aussi intenses que dans d'autres parties du monde (Chine, Inde, etc.).

De nos jours, les pays en voie de développement contribuent majoritairement à l'extension des zones urbanisées à travers le monde[8]. C'est au 19ème siècle, pendant la révolution industrielle, que le phénomène d'urbanisation a pris de l'importance en Europe, notamment en raison d'un exode rural conséquent. En 2014, 72,8% de la population européenne vit en ville ou en zone périurbaine. Pour autant, les morphologies du développement urbain varient à travers l'Europe. Le rapport sur l'urbanisation européenne produit par Eurostat (Koceva *et al.*, 2016), souligne par exemple, que la répartition des populations change suivant les territoires des différents pays (Figure 1. 6).

[8] En 2008, les Nations Unies soulignent que plus de 50% de la population mondiale vit en zone urbaine (Koceva *et al.*, 2016)

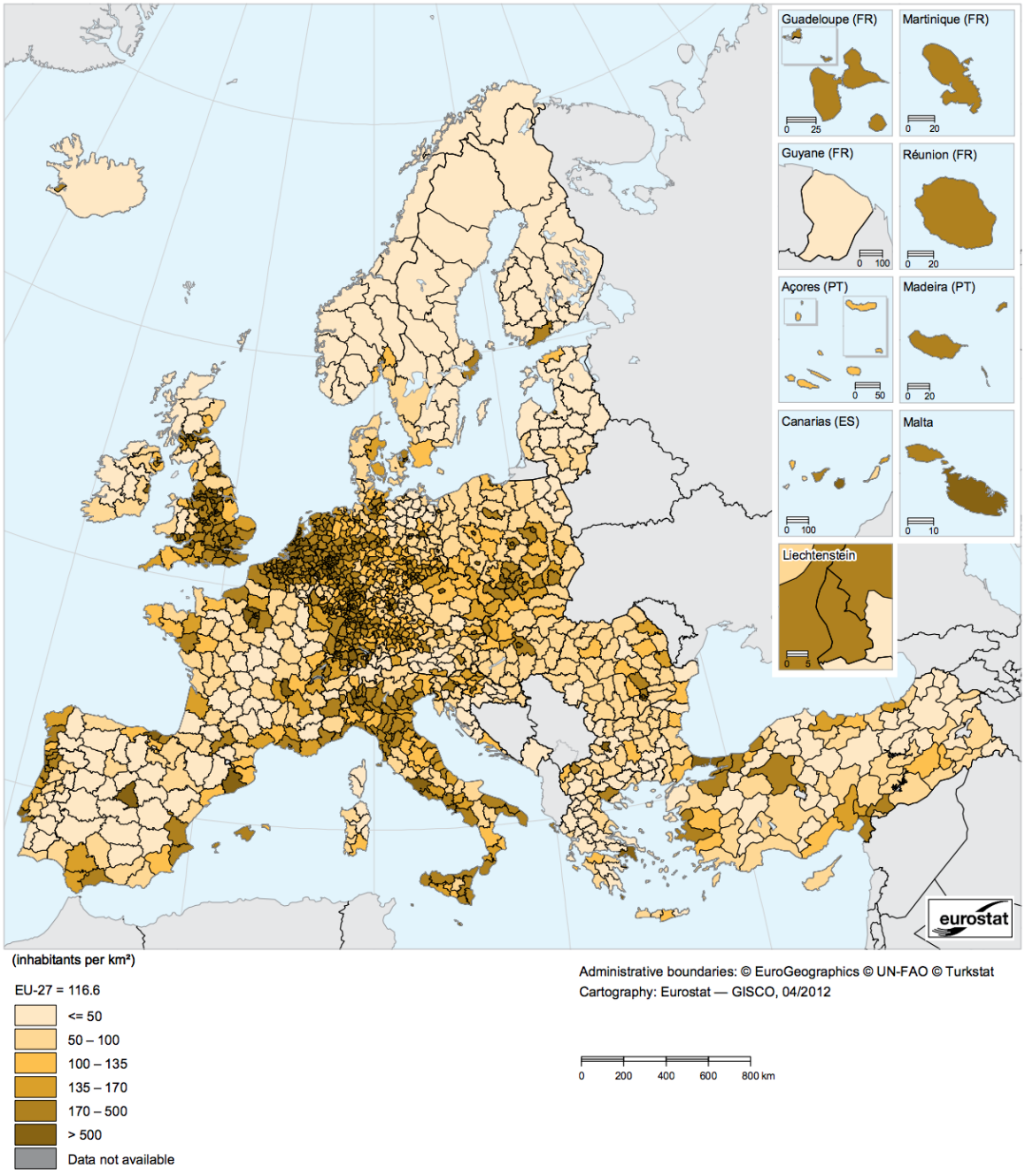


Figure 1. 6 : Densité de population en 2010 par NUTS de niveau 3 (habitant par km²).

Certains pays présentent une forte densité de population (Pays-Bas), d'autres une densité plus faible (Pays Nordiques). Si la population en 2010 continue de croître, la densité de population, elle, n'évolue pas non plus de manière uniforme sur la totalité du territoire européen (Figure 1. 7). Ces points peuvent en partie soutenir certains choix ou la forme des aménagements urbain, mais pas seulement.

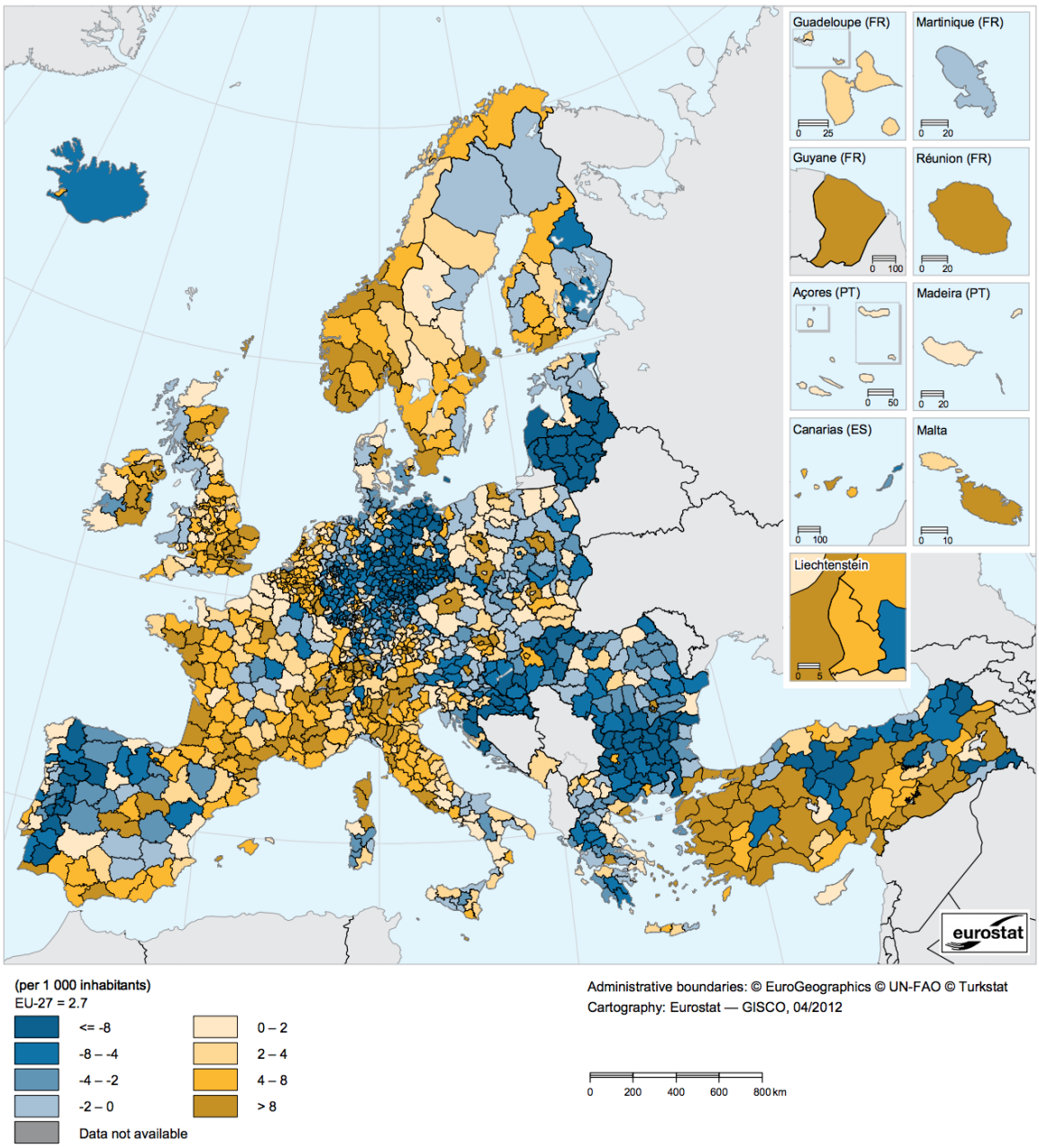


Figure 1. 7 : Variation de population en 2010 par NUTS de niveau 3 (pour 1000 habitants).

Les différences de morphologie urbaine peuvent faire l'objet de plusieurs explications. Le contexte environnemental (topographie, climat, ressources locales, etc.) explique en partie la variation. Le contexte culturel (socio-économique, politique, technologique, etc.) joue également un rôle important. « La distribution spatiale des zones urbaines dans les pays membres [de l'Union Européenne] dépend de leur géographie et de leur histoire » (Koceva *et al.*, 2016). La politique au niveau national ou à des échelles inférieures influence également l'attractivité et le potentiel en termes d'urbanisation d'un territoire. « Ces différences reflètent, au moins en partie, les structures politiques et institutionnelles » (Koceva *et al.*, 2016). Le rapport prend l'exemple de l'Allemagne, présentant généralement un développement polycentrique composé de villes de tailles moyennes, reflétant

potentiellement l'importance de l'état fédéral. Il s'appuie également sur un second exemple, celui de la France, présentant majoritairement une urbanisation monocentrique, centrée sur sa capitale, Paris.

Chaque pays européen possède donc une culture qui lui est propre, laquelle influence localement la morphologie urbaine, en dépit d'un objectif identique dans la construction.

En portant un intérêt plus particulier à la morphologie du développement urbain dans certains des pays membres de l'UE, suivant les mêmes modalités qu'à l'échelle mondiale (Figure 1. 5), la manifestation des différences dans la forme ou la morphologie de l'urbanisation devient moins évidente. La Figure 1. 8 est constituée de zones urbaines de tailles similaires en France et en Allemagne. Les deux pays membres de l'Union sont pris en exemple dans le rapport « *Urban Europe* » cité précédemment. La première ligne (Figure 1.8.a et 1.8.b) localise deux villes d'importance secondaire l'une en France, l'autre en Allemagne. Les proportions d'occupation du sol sont similaires avec une légère prépondérance des zones industrielles et commerciales visibles sur seconde image satellite (Figure 1.8.b). La seconde ligne (Figure 1.8.c et 1.8.d) présente un espace agricole et ses villages, localisés à proximité d'une ville moyenne. La troisième ligne (Figure 1.8.e et 1.8.f) présente deux villes moins étendues et denses que celles présentées Figure 1.8.a et 1.8.b. Le constat est néanmoins similaire à celui qui les concerne : proportion d'occupation comparable, avec prépondérance d'un type industriel. Pour finir, la dernière ligne (Figure 1.8.g et 1.8.h) présente des villages dont la morphologie semble équivalente. Ces « villages-tas » typiques des structures agraires de type openfield, sont généralement constitués d'une agglomération de maisons formant un noyau villageois historique, souvent autour d'un monument culturel (clocher), près d'une intersection. Nous pouvons d'ailleurs distinguer des axes de circulations. Ils ne sont majoritairement pas rectilignes.

Le rapport « *Urban Europe* » soulignait les différences institutionnelles et politiques des deux puissances, lesquelles peuvent engendrer une variation de forme ou de morphologie dans le développement urbain effectif (Koceva *et al.*, 2016). Il est néanmoins difficile d'attribuer intuitivement l'une ou l'autre des zones urbaines à la France ou à l'Allemagne. Si la culture influence la morphologie urbaine (Figure 1. 5), l'observation, au niveau régional, d'une morphologie urbaine similaire dans des pays aux caractéristiques culturelles différentes est donc également possible à proximité d'une frontière (Figure 1. 8).

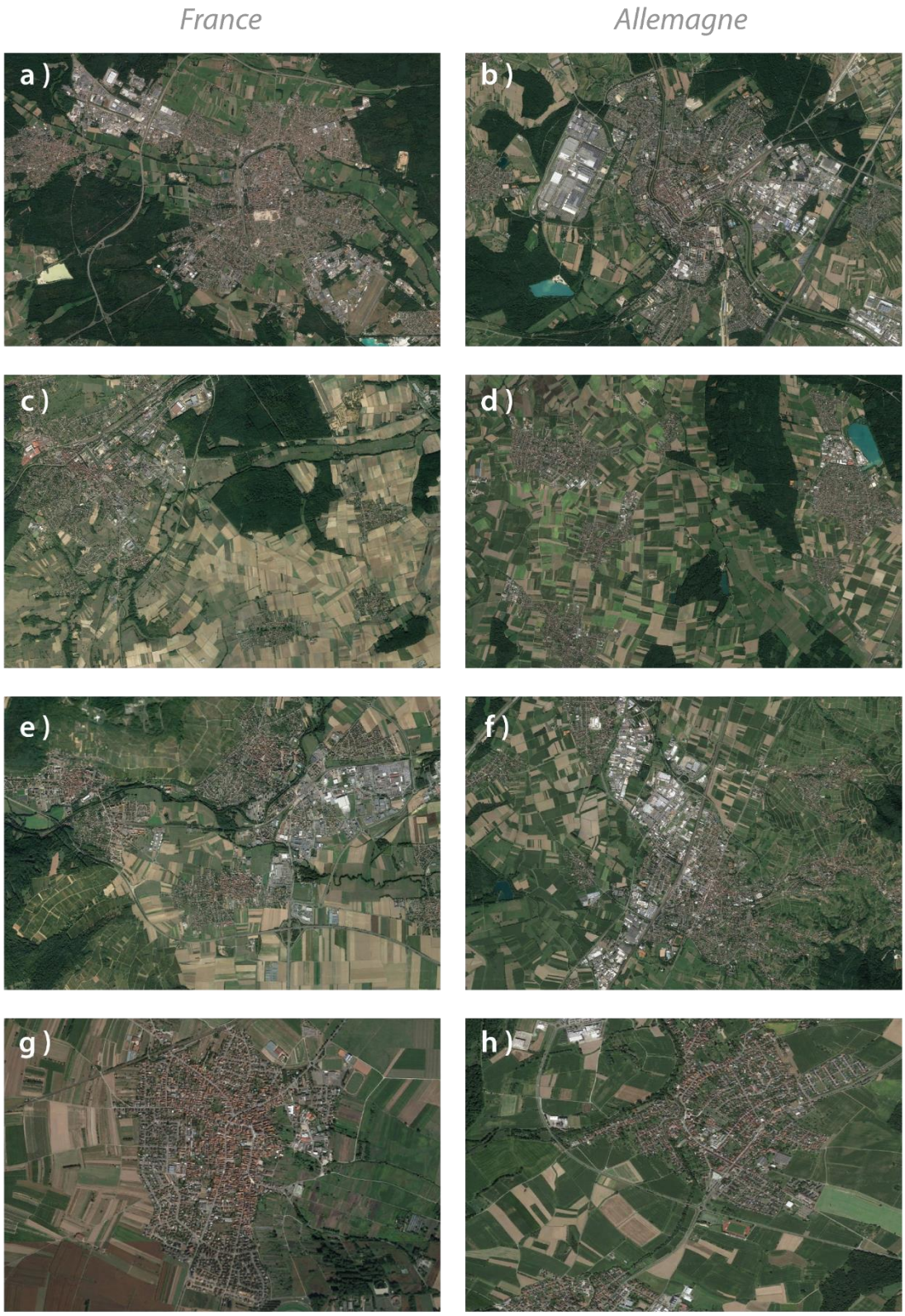


Figure 1. 8 : Urbanisation à proximité de la frontière franco-allemande (source : GoogleEarth – 2009)

La localisation et la composante culturelle locale se traduisent par une différenciation du processus de développement urbain (forme, morphologie, etc.). Pour autant, la comparaison visuelle du développement urbain de part et d'autre de la frontière franco-allemande, présentant une différence culturelle intrinsèque dont la limite est matérialisée dans l'espace par la frontière, ne révèle pas de distinction franche et évidente. Cette raison nous pousse à privilégier la frontière franco-allemande, au niveau de l'agglomération strasbourgeoise dans notre recherche. **Ce travail a donc pour objectif de rechercher une part non réductible à une divergence culturelle des facteurs de développement urbain, et donc d'évaluer un éventuel degré d'universalité du processus de développement urbain.** L'intérêt porté aux zones frontalières en Europe est présenté dans la section suivante.

3. FRONTIÈRE

Ce travail de recherche vise à étudier le développement urbain dans un espace frontalier, ce dernier regroupant l'ensemble des enjeux préalablement évoqués : un développement urbain aux objectifs communs, prenant forme dans des localisations nationales distinctes. Il s'agit dès lors d'une zone d'étude intéressante pour l'observation des variations de son expression, des facteurs à l'œuvre.

Il ressort de la littérature que les changements d'occupation du sol, comme le développement urbain, sont généralement liés aux caractéristiques culturelles locales. Pour autant, il est parfois possible d'observer une urbanisation similaire malgré des différences culturelles évidentes. C'est le cas de certaines zones frontalières. Elles peuvent être considérées comme intrinsèquement hétérogènes d'un point de vue culturel, selon la définition du terme « culture » d'après A. M. Hersperger, rappelée dans notre section 1.2 (Hersperger *et al.*, 2018). La frontière représente actuellement une limite administrative qui marque spatialement une différence politique, institutionnelle, juridique, réglementaire, linguistique. Cependant, depuis la fin de la 2^{e} Guerre Mondiale, l'importance des frontières internes européennes en tant que séparation a diminué avec la consolidation des accords et des coopérations entre les pays membres de l'Union Européenne (traité de Rome, accords de Schengen, traité de Maastricht, etc.). La frontière peut alors être considérée comme un lien, une barrière poreuse susceptible de s'effacer pour ne former qu'une zone homogène (Amilhat Szary & Fourny, 2006). À ce titre, l'urbanisation elle-même semble s'exprimer selon une morphologie similaire (Figure 1. 1). Un questionnement sur l'homogénéité de ces zones particulières, à travers l'étude des processus de développement urbain, apparaît donc pertinent.

Notre hypothèse initiale pose donc le postulat de l'existence d'une composante universelle dans le développement urbain, à travers une urbanisation similaire malgré les différences culturelles.

La présente section porte sur des zones frontalières internes à l'Union Européenne. La première sous-section étudie la notion de frontière : objet progressivement défini,

matériellement localisé dans l'espace comme limite d'un État. La deuxième sous-section s'attache à la polysémie de la notion et de son statut, à travers la perception des institutions et des individus. Enfin nous revenons sur la caractérisation des différentes phases, dans la littérature, pour décrire l'état de frontière dans sa perception en tant que limite.

3.1. Une limite de l'État

La thématique de la frontière et des zones transfrontalières suscite un intérêt grandissant depuis les années 80 (Medina-Nicolas, 2004). Cette tendance est soulignée par l'émergence des « *border studies* », vaste champ de recherche qui nécessite la mobilisation de plusieurs disciplines et notamment de la géographie (Kolosov, 2017). La création et le financement d'associations (*Association for Borderlands Studies*), de symposiums (*World Conference Border Studies, Border Regions in Transition Conference*), de réseaux de chercheurs (*UniGR-Center for Border Studies*) mais aussi de programmes financements (INTERREG) ou encore de revues scientifiques spécialisées (*Journal of Borderlands Studies*) sur les frontières attestent de cet intérêt.

La frontière européenne est désormais plus qu'une « marque dans l'espace et une marque de l'espace » (Nordman, 1998) : il s'agit d'une construction historique évolutive et d'un objet spatial en mutation (Ratti & Schuler, 2013). D'après la littérature, les deux derniers siècles ont vu évoluer le statut et le rôle des frontières intra-européennes. Auparavant utilisée pour mettre de la distance dans la proximité (Arbaret-Schulz, 2013), la frontière s'est transformée en une zone privilégiée d'échange et de rapprochement : « de coupure elle devient couture » (Kolosov, 2017). Le 19ème siècle voit même apparaître la notion d'obsolescence des frontières (Amilhat Szary & Fourny, 2006).

Par son étymologie le mot « frontière » dérive du terme militaire « front ». La frontière indique sans ambiguïté la notion de limite fortifiée, ou à fortifier, à défendre, mobile, faisant face à l'ennemi et délimitant l'appartenance à un espace particulier. De ce sens initial, étymologiquement figuré, le terme s'est éloigné. La sédentarisation de l'Homme, « action de s'établir de manière continue à un endroit précis » (Dictionnaire Larousse), de posséder un habitat fixe par opposition au nomadisme, a engendré l'apparition de zones d'influence où le sentiment d'appartenance et de propriété deviennent omniprésents. Les frontières spatiales et limites de territoires qui existaient avant la sédentarisation, ont vu leur définition modifiée. Auparavant dessinée suivant l'espace parcouru au gré des saisons, la limite du territoire devient fixe, basée sur le village, la ville, point de référence puisque qu'il est aussi le point de retour régulier voire quotidien. Cette limite peut être tracée dans les esprits mais se matérialise souvent dans un paysage, des objets géographiques de diverses natures (montagnes, fleuves, rivières, etc.).

La signature des traités de Westphalie (1648) affirme la fixation des frontières européennes modernes en marquant leurs tracés. Plusieurs auteurs s'accordent pour prendre cet événement comme le point de repère permettant de situer l'apparition de la notion de

frontière linéaire tracée (Amilhat Szary & Fourny, 2006; Brunet *et al.*, 1993; Velasco-Graciet, 2008). En effet, ces traités, s'ils ne donnent pas à proprement parler naissance à la notion de frontière, favorisent la reconnaissance mutuelle de ces limites d'influence des États. Ils en font des repères politiques communément admis. Il faudra d'ailleurs attendre cette période du 18e siècle pour que ces repères commencent à être matérialisés de manière systématique sur les cartes (Di Napoli-Marchetti, 2009). Ces outils de visualisation spatiale délimitent alors un espace associé à l'influence d'un État, et dont les normes sont juridiques et entérinées (Leloup & Gagnol, 2017).

En dépit de la signature des traités de Westphalie en 1648, il faut souligner que moins de 50% des frontières européennes actuelles existaient au début du 20ème siècle (Moullé, 2017). Au-delà de l'évolution de la notion en termes politiques et institutionnels, la frontière est donc un objet fluctuant physiquement et géographiquement.

Ainsi progressivement définie et reconnue, la frontière peut remplir plusieurs fonctions. C'est un repère spatial délimitant l'identité, un espace de pouvoir politique servant de repère pour les différentes populations. Elle est utilisée comme marqueur d'une limite, localisant une différence plus ou moins abstraite : culturelle (langue, etc.) et politique, mais aussi économique (prix du foncier, taxes, devises, etc.) ou réglementaire (législation, etc.). Elle sert des objectifs plus concrets : la sécurité d'un espace, la régulation des flux de biens ou de personnes. C'est alors une barrière, intrinsèquement liée à la notion de mobilité des personnes. En plus d'être un marqueur géographique, la frontière recoupe alors des enjeux géopolitiques et historiques essentiels.

Pour résumer, « l'idée de frontière est historiquement liée à l'idée d'état » et à l'histoire de ces états au cours du temps (Hamez *et al.*, 2013). Elle est donc, par l'ensemble de ses fonctions, une notion plurielle, utilisée de différentes manières à travers les âges et au gré de l'évolution des états. En réalité, au regard de celle de ses fonctions qui semble privilégiée, selon le contexte historique et international, c'est en réalité le caractère multiforme de sa perception qui se dégage. Elle est perçue et utilisée par des institutions, par un pouvoir politique qui s'exerce dans l'espace qu'elle délimite, mais également par des individus qui peuvent la vivre différemment. La combinaison de ces échelles de perception, impacte alors sa matérialisation physique et symbolique, sa réalité dans un espace. C'est l'objet de la sous-section suivante.

3.2. Une barrière et une suture

La frontière est une notion protéiforme selon qu'elle est envisagée dans un milieu politique, institutionnel et réglementaire, ou à l'échelle individuelle. Ces deux visions seront successivement évoquées.

Approche institutionnelle

L'approche institutionnelle est définie ici comme la perception de la frontière par l'État. La sous-section précédente présentait l'émergence de la frontière comme délimitation réciproque des espaces d'influence de différents pouvoirs politiques. Cette fonction première, dans une acception tant militaire que géopolitique en fait un objet physique, que l'on a pu voir comme étant même associé à des éléments naturels. Dans cette acception géopolitique, et selon les éléments du contexte international, la frontière n'est pas seulement un objet (de sécurisation d'un territoire par exemple), mais également un symbole. La présente sous-section propose d'envisager la frontière perçue par l'État suivant deux dimensions : une dimension « objet », une dimension « symbole ». C'est dans une dimension objet, la limite de sa souveraineté politique, mais dans sa dimension symbole, la représentation du type de relation que ledit État entretient avec celui dont il est limitrophe.

La dimension « objet » de la frontière est particulièrement intéressante. Elle est souvent associée à l'image de la « barrière » laquelle s'exprime alors dans la réalité de manière plus ou moins forte et évidente. Il s'agira d'une barrière douanière laquelle conduit à la construction de zones de douanes, matérialisées dans l'espace. L'histoire a démontré que cette matérialisation peut se faire plus franche et évidente, ainsi en est-il de la construction de murs de séparation : le mur d'Hadrien (127), le mur de Berlin (1961). La notion de fortification pour se protéger des peuples voisins est prégnante si ce n'est prioritaire. Si l'objet qu'est la frontière, limite géographique, instrument sécuritaire semble alors figé et pérenne, elle renvoie aussi à l'idée du conflit, à la temporalité de ce dernier, et à l'évolution des relations entre les États. Cet objet vraisemblablement figé peut d'ailleurs être modifié en fonction du début ou de la fin du conflit. La frontière, d'outil de souveraineté et de clôture de ce territoire, devient alors symptomatique d'une tension, d'une relation inter-étatique, d'une problématique géopolitique. C'est alors sa perception symbolique qui est également mise en exergue. Cette dernière rétroagit sur les manifestations physiques de la frontière : elle détermine le type d'objet qu'elle doit être (un mur qui fait barrière, ou au contraire, un espace d'échange).

Parallèlement à sa dimension d'objet, la frontière et ses alentours peuvent donc revêtir une dimension symbolique. Une frontière très fortifiée, qui parait infranchissable peut véhiculer une impression de pouvoir puissant et servir l'image extérieure d'un pays dans l'hypothèse d'un conflit. Lors du déplacement d'une frontière après un affrontement, il n'est pas rare d'observer des velléités politiques d'effacement de la précédente appartenance et d'incarnation de la victoire en un lieu symbolique. Lors d'une annexion, le pays contraint de céder une partie de son territoire peut avoir à cœur de récupérer ses terres, non pas uniquement dans une démarche économique et pour la rentabilité de l'espace mais bien parce que cet espace est devenu de part et d'autre, le symbole du conflit, de sa résolution, et donc de la puissance victorieuse.

En l'absence de conflit, et si la frontière n'est à nouveau qu'une ligne, rattachée à la délimitation d'une sphère de pouvoir institutionnel, sa perception en est-elle pour autant dénuée de symbolique ? Une frontière apaisée, ce qui est notamment propre à l'évolution des frontières intra européennes à l'issue des périodes de guerre, ne présente plus les symboles précédents rattachés à la belligérance. En l'absence de contact particulier avec le pays frontalier dont elle fait le départ, la zone frontière apparaît n'avoir aucun intérêt distinct de celui des autres régions du pays. La frontière est ramenée à la dimension préalablement examinée « d'objet ». Pour autant, son utilisation géopolitique et institutionnelle n'apparaît pas entièrement restreinte.

C'est l'évolution des relations internationales et l'émergence du capitalisme qui permettent de poursuivre le processus d'évolution des zones frontières et de la frontière elle-même. En Europe la reconstruction d'après-guerre nécessite l'implication des différents États. La création de la Communauté du Charbon et de l'Acier (traité de Paris, 1952), impulse une coopération qui fait progresser les États du statut d'ennemis vers celui d'alliés. L'exemple des deux grandes puissances que sont la France et l'Allemagne apparaît pertinent : au centre des deux conflits mondiaux, ces pays opposés par les guerres deviennent moteurs du rapprochement économique européen, *plus encore que ne l'ont été certaines puissances victorieuses*, et ce par la signature d'un accord de mise en commun de ressources.

Ces évolutions des relations internationales ont ainsi façonné des territoires initialement ennemis, pour en faire des partenaires économiques voire des pays « amis ». Nous pouvons citer en exemple « l'amitié franco-allemande » développée sur la base de traités comme celui de l'Élysée (1963) ou plus récemment d'Aix-la-Chapelle (2019). A cet égard, la zone frontalière est à nouveau le terrain d'enjeux symbolique : la proximité qu'elle représente entre les deux pays est à l'image du lien qui se tisse. La construction, du Pont de l'Europe (1960) pour franchir le Rhin entre Strasbourg et Kehl incarne cette réalité, le symbole d'une réconciliation dont la toile de fond est celle du développement de l'Union européenne. Nous pouvons également citer la Passerelle des Deux Rives pour piétons et cyclistes (2004), ou le Jardin des Deux Rives (2004), un parc public transfrontalier. La frontière devient ainsi un outil de cette coopération : c'est une zone d'accessibilité qui peut rétablir voire renforcer la liberté de mouvement, de circulation et d'échange. Ces idées géopolitiques s'incarnent alors dans une perception réglementaire, tant législative qu'économique, de la zone concernée.

Approche réglementaire

Les différentes perceptions institutionnelles ainsi développées apparaissent intrinsèquement liées à une vision réglementaire, laquelle peut être présentée selon deux aspects interdépendants : législatif ou économique. Ils sont tous deux essentiels à la compréhension d'une frontière vécue moins comme une barrière que comme un espace de « suture » et, d'échanges.

Comme mentionné au titre de « l'amitié franco-allemande », l'apaisement des frontières européennes a mis en place d'une dynamique d'échange, spontanée ou soutenue par les projets politiques émergents (Moullé, 2017). L'existence de flux économiques transfrontaliers concernant les biens ou les personnes est antérieure à la création de l'Union Européenne. Plusieurs motifs peuvent inciter la mobilité transfrontalières : tourisme, achats, visite chez des amis, travail, visite de la famille, recours à des services publics (Decoville & Durand, 2019; Klein, 2017). Il est toutefois intéressant de souligner que l'évolution des relations entre les pays membres, par l'adoption de textes et de réglementations vise à encourager plus ou moins directement ces échanges. Ainsi, la ratification d'accords européens (traité de Rome, accords de Schengen, traité de Maastricht, etc.) diminue l'influence de la frontière d'État en tant que barrière. Si la signature des accords de Schengen et du traité de Maastricht ne font pas mention explicitement de l'aspect économique pouvant impacter les frontières, ils offrent de nouvelles possibilités dans les zones frontalières, auparavant sujettes aux contrôles douaniers, à une limitation de la circulation des biens et des personnes. Ces nouvelles dynamiques augmentent la porosité de cette ancienne barrière et impactent de fait les zones frontalières.

Ces volontés institutionnelles et réglementaires peuvent également viser directement ces zones. Les projets transfrontaliers renforcent cette idée d'une porosité. C'est par exemple en 1981 que la Convention-cadre européenne sur la coopération transfrontalière des collectivités ou autorités territoriales rentre en vigueur. « La Convention a pour but d'encourager et de faciliter la conclusion d'accords entre régions et communes, de part et d'autre d'une frontière, dans les limites de leurs compétences. De tels accords pourront s'étendre entre autres au développement régional, à la protection de l'environnement, à l'aménagement des infrastructures et des services publics, etc. allant même jusqu'à la création de syndicats ou d'associations de collectivités locales transfrontalières. » (Conseil de l'Europe, 1980). La volonté de coopération entre les pays membres place les zones frontalières au centre d'un intérêt à la fois symbolique mais aussi pratique. Le traité de Rome mentionnait déjà dans son préambule l'objectif « de renforcer l'unité de leurs économies et d'en assurer le développement harmonieux en réduisant l'écart entre les différentes régions et le retard des moins favorisées » (Traité de Rome, 1957). Si par le passé l'objectif prioritaire était d'éviter une nouvelle guerre, ce dernier a lui aussi évolué. L'initiative à travers le développement de la coopération transfrontalière est d'encourager les échanges et a pour finalité de limiter les effets de coupures dans les espaces situés à proximité d'une frontière.

La coopération transfrontalière peut être envisagée à différentes échelles et suivant différents thèmes. Des programmes Européens ont été développés dans le but de donner l'occasion aux espaces, notamment transfrontaliers, d'initier de nouveaux projets ou d'entretenir les liens existants. L'union Européenne intervient dans l'aménagement des territoires par le biais du Fonds Européen de Développement Régional (FEDER) créé en 1975 et du Fonds Social Européen (FSE). Les fonds des programmes opérationnels FEDER peuvent être attribués aux territoires présentant sur une grande échelle des enjeux de préservation et de

développement communs (massifs, bassins fluviaux, etc.). Ces fonds servent surtout à renforcer la cohésion économique (contre les déséquilibres régionaux). Les territoires qui appartiennent à différents pays membres, mais présentent des enjeux communs sont ainsi, de fait, souvent situés aux alentours d'une frontière. Il existe également des fonds spéciaux de soutien au développement au développement rural (FEADER) et à l'activité halieutique (FEAMP). Pour finir, l'Union Européenne soutient encore la modernisation des infrastructures des derniers pays ayant adhérés, par le biais du Fond de Cohésion. La coopération transfrontalière est financièrement soutenue par l'UE depuis 1990 à travers l'initiative INTERREG. Celle-ci porte sur l'approche transversale du développement spatial. Elle est la plus significative et met l'accent sur une intégration du développement des régions frontalières, mais aussi plus largement sur les aires de coopération transnationales.

L'initiative aboutissant à la signature des différents accords et traités européens ou encore celle de la monnaie unique ont un réel impact sur la situation européenne en général mais tout autant voire plus, sur les frontières des états membres. Ces espaces auparavant séparés par des limites physiques (douanes), institutionnelles, ou monétaires, ne le sont plus. La frontière, telle qu'elle fût, évolue pour s'effacer progressivement, en tout cas d'après les points évoqués précédemment.

Approche sociale

Ces évolutions réglementaires, initiées par un contexte historique et géopolitique (construction de l'Union Européenne) ont également des conséquences directes sur les individus ayant la citoyenneté européenne et sur leur rapport à la frontière. Ces derniers peuvent, de jour comme de nuit, traverser la frontière, habiter ou travailler d'un côté ou de l'autre, sans même avoir le souci d'une conversion monétaire (salaire ou dépenses).

F. Leloup (2017) se posent la question d'une transformation de l'espace transfrontalier en « territoire transnational réunifié » lors d'un passage de la « frontière barrière » à une « frontière trait d'union ». Cette notion de « territoire transnational réunifié » dépendra de la manière dont les individus s'approprient quotidiennement l'espace de vie que représente pour eux la zone frontalière. À ce titre, leur vision sociale de la frontière est nécessairement influencée par une vision politique et par les relations internationales entretenues à un moment donné. Toutefois, si la position du pouvoir est claire et univoque s'agissant de la ligne de conduite dans ces relations internationales, les comportements individuels et la vision sociale peuvent être plus contrastés.

De la même manière que le sens étymologique de la frontière peut conduire à examiner sa vision institutionnelle militaire voire belligérante, la frontière peut être perçue par les individus comme une barrière de protection face à l'altérité. Quand la frontière peut marquer une division binaire voire manichéenne entre deux puissances politiques, deux idéologies, les comportements sociaux ne s'approprient pas forcément cette perception univoque. Le

mur érigé, après la seconde guerre mondiale, entre la RFA et la RDA est un exemple de frontière binaire entre Europe de l'Est et Europe de l'Ouest. Le pouvoir matérialisait clairement cette frontière (mur, poste de garde, etc.), encadrant drastiquement le passage entre deux zones jugées radicalement opposées. Pour autant, cette opposition binaire n'est pas vécue et acceptée par tous les habitants de la zone : certains allemands voient cette frontière diviser leur lieu de vie voire les séparer des membres de leur famille. Si une partie de la population, sensible à l'idéologie et à l'objectif défendu par cette frontière, a pu accepter cette différenciation du reste de la ville, une autre partie de la population s'est trouvée désemparée par cette construction, allant jusqu'à transgresser cette limite.

Si la frontière n'est plus, comme dans l'exemple précédent, une barrière imposée, les individus peuvent au contraire « l'utiliser », à travers leur domiciliation au sein de l'espace frontalier par exemple (Ray *et al.*, 2014). Cette limite crée en effet un différentiel politico-juridique, réglementaire, qui devient exploitable. Les chercheurs (Ray *et al.*, 2014) ont par exemple observé une mobilité résidentielle initiée par le différentiel frontalier (réglementation, taxes, etc.). Les individus « optimisent » leur niveau de vie en choisissant de résider d'un côté ou de l'autre de la frontière, suivant la phase de leur cycle de vie. L'exemple de la frontière Ontario-Québec montre l'utilisation du différentiel par certains couples cherchant à optimiser leurs revenus en s'installant en Ontario, État présentant de plus faibles taxes. Leur accession à la propriété y est simplifiée, en dépit d'une couverture sociale moindre. Le phénomène s'inverse lors de la naissance d'enfants, puisque cette fois, la couverture sociale est favorisée (Gilbert & Brosseau, 2011). Des comportements similaires sont observés en Europe (Carpentier *et al.*, 2013). Le phénomène du travail transfrontalier rendu possible par les différents accords permet également à un Français de vivre en Allemagne tout en travaillant en France si la situation lui semble plus avantageuse en raison des réglementations fiscales, des avantages sociaux ou simplement en raison du coût du logement.

La vision symbolique de la frontière influence également ces phénomènes « d'exploitation » personnelle de la zone. Quand certains individus privilégient des modalités pratiques et rationnelles et optimisent leur niveau de vie en conséquence, d'autres ne franchissent pas cette limite. Quand bien même une frontière serait apaisée, elle reste chargée d'histoire, et marque toujours une séparation entre des cultures et systèmes différents. « Globalement, la dématérialisation de certaines frontières entraine le renforcement de la frontière symbolique » (Moullé, 2013) : si elle existe moins comme une barrière, qu'elle est même un outil à disposition de l'organisation quotidienne des individus, la frontière marque aussi leur appartenance à une nation. Ce symbole, à travers le lieu de vie, est donc un élément de la perception qui influence les choix d'un habitant de la zone : est-il prêt, ou non, à amoindrir la force de ce symbole pour optimiser certains éléments de son cadre de vie ? Veut-il au contraire, en renforcer la dimension symbolique ? La frontière de la langue doit également être prise en compte, et peut l'être dans le cas de la frontière Ontario – Québec comme dans celui de la frontière Franco-Allemande. Là encore, des projets sont mis en place pour réduire

l'impact de cette différence en favorisant notamment le bilinguisme. Ces actions visant à réduire le différentiel frontalier sont développées ultérieurement (Chapitre 3).

« La construction culturelle et identitaire pose des problèmes plus redoutables que la construction institutionnelle car elle s'inscrit à l'encontre de nos modes de conceptualisation du territoire » (Amilhat Szary & Fourny, 2006). Quels que soient les objectifs politiques et les aménagements réglementaires, les réactions individuelles au tracé de la frontière, telles qu'évoquées ci-dessus, constituent une part irréductible et diversifiée de l'évolution de la frontière, de son usage et de sa porosité.

Les trois aspects de la perception de la frontière tendent à démontrer une forme d'évolution. Le tracé des frontières politiques modernes a été légitimé par l'identité, la culture, la structure des sociétés et des peuples qu'elles séparent. Pour autant, la frontière a également été érodée par l'évolution des politiques et la création d'accords Européens. « Mon pays c'est la frontière »[9] (« Mon pays c'est la frontière », 2017) titre d'un reportage diffusé sur LCI en 2017, illustre parfaitement l'évolution du rôle et de la perception de cette dernière. Autrefois limite militaire empêchant le passage, lieu de conflits armés, devenue limite institutionnelle associant la gestion et l'organisation d'une partie de la zone traversée à un État, aujourd'hui les frontières, ou plutôt les zones frontières sont devenues l'image des similarités ou des différences. Elles permettent l'union ou le tissage de liens soulignant la complémentarité et l'intérêt commun de pays, plutôt que des différences favorisant les conflits. C'est même de ces différences que peut émerger la complémentarité et l'intérêt d'une coopération.

L'ensemble de ces évolutions historiques, ces perceptions plurielles et divergentes, font l'objet, dans la recherche, d'un travail de conceptualisation, lequel envisage le « cycle de vie » d'une frontière. Il convient de s'intéresser dans la prochaine sous-section, aux outils conceptuels utilisés pour retracer les notions envisagées précédemment. Si une tendance globale, celle d'une porosité grandissante de la frontière, peut se dégager, il est intéressant de constater que ces repères conceptuels sont eux-mêmes diversifiés et ne permettent pas une mesure précise des tendances décrites.

3.3. Un « cycle de vie »

La recherche fait le parallèle entre l'évolution du statut ou du type de frontière, et l'évolution que l'on peut observer dans la vie d'un objet, voire d'un être vivant. O.J. Martinez (1994) s'appuie sur quatre phases pour décrire cette évolution du point de vue des interactions sociales : aliénation, coexistence, interdépendance, intégration. R. Ratti (2013) préfère résumer l'évolution en 3 phases. Ces dernières sont plus orientées sur l'aspect économique des échanges et consistent à voir la frontière comme une barrière dans un premier temps, puis comme un filtre, pour finir par une intégration totale de la zone. M. Baud et W.Van

[9] Le reportage porte sur les pratiques des individus autour de la frontière franco-espagnole.

Schendel (1997) présentent l'évolution de la frontière en cinq phases, calquées sur le cycle de la vie humaine : enfance, adolescence, adulte, vieillissement (déclin), disparition (relique). B. Reitel (2017) capture cette évolution à travers l'utilisation de trois termes, traduit de l'anglais : frontiérisation, défrontiérisation, refrontiérisation (*bordering, debordering, rebordering*). Une prise de recul tant spatiale qu'historique permet d'observer ce cycle sur un certain nombre de frontières à travers le monde. Il précise néanmoins que cette vision reste très « euro-centrée ».

B. Reitel (2017) souligne que les approches fonctionnelles laissent supposer que l'augmentation des flux renforce l'intégration spatiale, et remet donc en cause l'utilité de la frontière. La conséquence est la dévaluation de la frontière qui revient à être une « simple limite administrative ». Cette vision fait l'objet de critiques puisque la focalisation sur l'aspect fonctionnel mène à occulter l'aspect symbolique et identitaire. L'auteur insiste sur l'importance de ces derniers aspects. L'augmentation des interactions ne signifie pas forcément le lissage et l'intégration des groupes sociaux. Plusieurs auteurs ont, par ailleurs, critiqué les modèles d'évolution du statut de la frontière vers l'obsolescence, en soulignant l'existence d'une phase de refrontiérisation (Reitel, 2017). En effet, certaines parties du monde font l'objet d'une création de nouvelles frontières, d'autres voient la porosité préexistante de leur frontière diminuer au rythme des nouvelles mouvances politiques (Hongrie par exemple). D'après Buchanan (2003), la frontière ne peut faire abstraction des conceptions morales, culturelles, et politiques qui lui sont attribuées. Pour B. Reitel, loin de disparaître, les frontières du 21ème siècle changent de formes et sont rendues mobiles par les innovations mises en place sur le territoire frontalier, que l'on a déjà pu décrire dans la sous-section précédente.

Dans la description de ce cycle de vie, le devenir de la frontière, son rôle et son sens sont également questionnés par la recherche. J. Lévy et M. Lussault (2013) écrivent dans leur *Dictionnaire de la géographie et de l'espace des sociétés* que « dans un monde démilitarisé et ouvert aux échanges, la frontière perd de son sens ». Au contraire, A.L. Amilhat Szary et M.C. Fourny (2006) font l'hypothèse que la frontière, à travers des « territorialités contemporaines post-Schengen [….] continue à faire sens, à spécifier des appartenances et à fonder des artéfacts spatiaux ».F. Leloup et L. Moyart (2006) prennent du recul sur l'influence de la frontière en soulignant que même par le passé, « cette rupture n'a pas mené forcément à un éclatement définitif », que l'augmentation de la porosité de cette dernière n'entraine pas nécessairement une disparition d'un différentiel frontalier ni l'émergence d'une nouvelles région homogène. A.L. Amilhat Szary et M.C. Fourny (2006) pointent la contradiction qui peut émerger du territoire transfrontalier, en considérant que ce dernier est supposé « intégrer l'altérité et donner lieu à un collectif composite ».

L'ensemble des travaux cités constatent une tendance relativement similaire s'agissant de la frontière : fermée, progressivement ouverte, disparaissant et/ou réhabilitée. Toutefois une différence essentielle persiste dans la description de l'évolution des frontières, que ce soit dans le nombre de phases décrites, dans la terminologie utilisée ou dans l'impact précis de

ces évolutions sur la réalité de la frontière. Il apparaît alors opportun de s'interroger sur l'absence d'un référentiel commun permettant de caractériser le stade d'évolution de la frontière étudiée d'un point de vue général. Aucun indice transversal ne permet à ce jour de « mesurer » le stade d'évolution du statut de la frontière. Ce travail de recherche propose donc une méthode pour estimer la porosité de la frontière à travers la modélisation du développement urbain.

Conclusion

L'objet frontière est constamment modifié. Il l'est par les actions militaires qui peuvent mener à sa disparition ou à son déplacement, par les organisations qui la franchissent et dont l'objectif est de ne pas respecter les législations de part et d'autre (ex : contrebande, migration clandestine, etc.). De façon plus douce, les prises de décisions réglementaires et les comportements des populations peuvent aussi amener la séparation représentée par cet objet à disparaître. Il peut également être renforcé, à travers la réglementation, ou encore physiquement par la construction de nouvelles séparations.

Cette section permet de faire ressortir une hétérogénéité dans la manière de percevoir, de définir et d'étudier les frontières, les espaces transfrontaliers et leurs évolutions. Nous avons choisi de nous concentrer sur trois aspects distincts pour présenter la frontière, lesquels permettent de saisir une partie de la complexité qui peut être associée à l'étude des frontières et des espaces transfrontaliers. Si ces trois aspects sont décrits séparément, et si la structure de la section distingue l'émergence historique de la notion de frontière et les différents degrés de sa perception, il est difficile de ne pas lier intrinsèquement cet ensemble. Ces visions sont interdépendantes et permettent d'appréhender la frontière dans une dimension symbolique tout autant que dans une dimension objet. La Figure 1. 9 permet de résumer la démarche entreprise dans la présente section. La frontière, en réalité, se construit et émerge des différents angles selon lesquels on l'envisage.

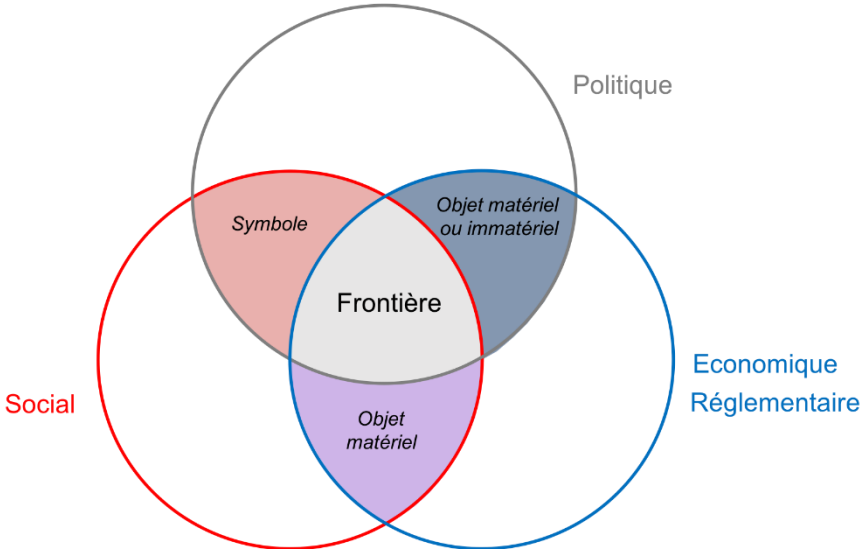


Figure 1. 9 : Émergence de la notion de la frontière comme la construction résultant de différentes visions

Précisons que l'objectif ici n'est pas d'établir la légitimité d'une vision particulière de la frontière, ni de calibrer notre approche sur ces dernières, mais bien d'observer les processus à l'œuvre sur un espace transfrontalier, à travers le développement urbain. Cette analyse basée sur les données spatiales, tentant de maximiser l'objectivité et la transférabilité de la méthode d'analyse à d'autres espaces, frontaliers ou non

.

Conclusion chapitre 1

Face à l'étude des changements d'occupation du sol, et plus particulièrement du développement urbain, une interrogation émerge. Au sein de ces processus globaux, y a-t-il une part irréductible d'influence de facteurs locaux ? Si le présent chapitre avait pour objet de présenter les notions essentielles de l'occupation du sol et du développement urbain, de mettre en lumière l'ancrage culturel de la forme de ces processus, il importait également d'y lier la composante frontalière du présent travail de recherche. Celle-ci, parce qu'elle condense une diversité culturelle, apporte à l'interrogation soulevée ci-dessus, une réelle pertinence : si au sein de cette zone diversifiée, il est possible d'extraire une caractéristique générique des processus de développement urbain étudiés, alors un premier élément de réponse peut être apporté. La porosité de la frontière en termes de développement urbain, contribue à l'hypothèse d'une part universelle de ce même processus.

Plusieurs auteurs ont cherché et cherchent encore à décrire ces espaces frontaliers dans le but d'améliorer la compréhension des processus à l'œuvre. Comme le montre la section 3 du présent chapitre, leurs approches généralement descriptives, ne permettent pas de mesurer entièrement les processus et leurs évolutions d'un point de vue spatial (augmentation de la porosité de la frontière, création de nouveau espaces homogènes, etc.). C'est pourquoi, une entrée par la donnée géographique, focalisée sur la thématique de l'occupation du sol et du développement urbain, a été privilégiée dans ce travail de recherche. Si les différentes approchent s'appuient sur des faits ou des concepts pouvant être amenés à évoluer au cours du temps, voire à disparaître, l'occupation du sol, tout en évoluant, demeure et tout comme les réseaux qui favorisent la mobilité perdurent. Notre démarche a pour objectif de mettre en lumière l'apport d'une approche empirique maximisant l'objectivité, dans l'analyse d'un espace transfrontalier à travers des données spatiales. Ces dernières font l'objet du second chapitre.

C'est à travers l'analyse de données spatiales et plus précisément, l'observation du développement urbain, que cette thèse propose d'explorer l'homogénéité des processus à l'œuvre sur une zone frontalière. Le chapitre 2 permet de développer les raisons supportant cette démarche ainsi que les données sélectionnées dans le cadre de cette recherche.

Chapitre 2
Zone frontalière :
Des données à modéliser

Nous avons souligné précédemment, l'intérêt majeur de l'étude du processus de développement urbain sur une zone frontalière. Dans ce contexte, l'objectif est d'être en mesure de caractériser le phénomène de part et d'autre de la frontière. Il est question d'établir s'il est similaire, ou s'il présente des différences suivant la localisation dans l'un ou l'autre des pays. En effet, dans ce contexte particulier nous nous interrogeons sur l'influence de la différence culturelle dans les caractéristiques du processus d'urbanisation. Le chapitre précédent a permis de mettre en évidence le lien qui existe entre les formes d'urbanisation et le cadre environnemental, ou culturel, sans lequel elles s'expriment. Néanmoins, les espaces transfrontaliers ont vu évoluer le statut de la frontière auparavant appréhendé comme une limite visant à inscrire dans l'espace l'appartenance à une identité mais surtout à se protéger du voisin. Ces anciennes zones de conflits portant sur la représentation de la primauté et faisant l'objet de luttes de pouvoir se sont progressivement pacifiées et apaisées. Dans un premier temps par nécessité (reconstruction spatiale, économique et identitaire dans ces régions) puis par la volonté d'exploitation d'un nouveau potentiel économique et politique sur ces territoires périphériques dont l'importance a grandi parallèlement à la volonté de coopération entre les puissances européennes. Ce constat permet de questionner l'influence de la différence culturelle sur notre zone en dépit de la présence d'une frontière. A savoir, si nous sommes face à des processus transfrontaliers qui s'affranchissent de la frontière étatique ou bien, au contraire, si cette dernière apparaît toujours dans leur organisation spatiale. C'est à travers l'utilisation de données spatiale et la modélisation du processus de développement urbain que la question est abordée dans ce chapitre. Si l'élaboration de modèle est, aujourd'hui, courante lors de l'étude de processus spatiaux dynamiques, il est nécessaire qu'elle soit adaptée aux objectifs et au contexte de la recherche.

Il s'agira dans ce chapitre de revenir dans un premier temps sur le cadre théorique qui entoure l'élaboration des modèles au regard des objectifs de leur application. Après avoir répondu au « Pourquoi ? » modéliser, qui induit la nécessité de répondre à la question du « comment ? » modéliser, nous aborderons la question de l'utilisation des données dans les modèles en géographie, notamment lors de la modélisation en contexte frontalier (Section 1). Ensuite, la nature des données sélectionnées pour la modélisation de l'espace frontalier est abordée dans un second temps (Section 2).

« Nous ne voyons pas le monde tel qu'il est,
mais tel que nous sommes » E. Kant.

1. MODÉLISER UN ESPACE FRONTALIER

Depuis R. Descartes, les êtres humains ont eu à cœur d'essayer de comprendre le monde dans lequel il leur est donné d'évoluer. A. Einstein disait « l'essentiel est de ne pas arrêter de se poser des questions ». La description de leur environnement (ici appréhendé dans son sens le plus général), ou la perception qu'ils en ont, ainsi que celle des processus spatiaux en constante évolution est une étape fondamentale sur le chemin de la compréhension. La première section de ce chapitre revient sur l'utilisation de la modélisation pour améliorer la compréhension des processus spatiaux œuvrant au sein des zones anthropisées.

La modélisation pour étudier les processus dynamiques, notamment en géographie, est une démarche mobilisée depuis plusieurs décennies. J.M. Epstein indique que la pratique de modélisation remonte aussi loin qu'Archimède, mais souligne une certaine incompréhension généralisée de la démarche, comme celle de l'associer exclusivement à un but prédictif (Epstein, 2008). En effet, la création d'un modèle peut être initiée pour différentes raisons, motivées par différents objectifs. Pour Epstein, la question principale n'est pas de savoir s'il faut construire un modèle, « c'est de savoir si ce modèle doit être explicite » donc construit pour représenter une situation détaillée. La sous-section 1 permet de revenir sur les objectifs liés à la démarche modélisatrice.

1.1. Pourquoi modéliser ?

Epstein (2008) insiste sur l'association dans les esprits, de la démarche modélisatrice à la démarche de modélisation prédictive uniquement. Cette dernière, qui vise à étudier l'issue de potentiels scénarios se déroulant dans le futur, n'est néanmoins pas la seule option soutenant l'utilisation de la modélisation. L'auteur s'attache d'ailleurs à énumérer seize autres raisons soutenant la démarche de modélisation.

1. Expliquer (à distinguer de prédire)
2. Guider la collecte de données
3. Mettre en lumière la dynamique principale
4. Suggérer des analogies entre dynamiques
5. Découvrir de nouvelles questions
6. Promouvoir une démarche scientifique
7. Englober certains résultats possibles dans des classes bornées
8. Mettre en lumière les incertitudes majeures
9. Offrir des options de gestion de crise presque en temps réel
10. Monter les échanges / suggérer l'efficacité
11. Questionner la robustesse des théories existantes
12. Révéler la sagesse antérieure comme incompatible avec les données disponibles
13. Entrainer les praticiens

14. Discipliner les dialogues réglementaires
15. Apporter une connaissance générale publique
16. Révéler le « simple » en apparence comme étant « complexe » ou inversement

Les seize points listés permettent d'entrevoir le panel de raisons pouvant mener à l'élaboration d'un modèle quel qu'il soit. Ces derniers ne seront pas tous développés en détails, cependant, la liste ci-dessus permet de mettre en lumière l'apport multiple de modélisation qui n'est pas restreint à une démarche prospectiviste. Plusieurs des points évoqués par Epstein peuvent illustrer le cadre de ce travail de recherche, notamment les points 1, 2, 3, 5, 6, 8 15, 16.

Au cours des décennies qui virent se développer le champ théorique et quantitatif de la géographie, plusieurs types de modèles ont vu le jour pour explorer les processus dynamiques. Ces derniers seront présentés plus en détails dans la sous-section suivante. « Tous les modèles ont une fonction générale de médiation, plus exactement de facilitation d'une médiation dans une enquête de connaissance » (Varenne & Pumain, 2017). Ils servent d'intermédiaires entre diverses capacités cognitives. Trois types de capacités sont retenues par F. Varenne (2017) :

- Pratique (contrôle, décision),
- Théorique (conceptualisation, formulation, prédiction, compréhension, explication)
- Sociale (partage, accréditation, contrôle, décision collective).

« Un modèle apparait comme un objet de substitution facilitant un questionnement de par l'effet de la facilitation d'une médiation entre capacités cognitives distinctes. » Varenne distingue 5 fonctions majeures et 21 sous fonctions permettant d'aider la médiation (Varenne & Pumain, 2017). Ces dernières peuvent être synthétisées de la sorte : faciliter l'expérience, l'observation contrôlée, l'intelligibilité, la théorisation ou encore la co-construction de savoir. En outre le modèle peut assurer plusieurs fonctions comme favoriser l'intelligibilité d'un processus spécifique dans un objectif explicatif tout en visant la théorisation du phénomène ou encore le partage permettant une réflexion collective. La frontière des différents types ou des différentes fonctions d'un modèle est donc poreuse.

Les modèles statistiques se sont invités dans la recherche en sciences sociales. Ils peuvent être divisés en trois catégories d'après G. Shmueli (2010) : explicatif, prédictif ou descriptif. Néanmoins, la frontière entre ces trois types de modèles n'est pas toujours clairement établie (Shmueli, 2010; Varenne & Pumain, 2017). Les définitions suivantes permettre de préciser la portée des différentes démarches.

En sciences sociales, les méthodes statistiques sont principalement utilisées pour tester des hypothèses, et à travers elles, des théories. Les méthodes statistiques sont appliquées à des données pour le test d'hypothèses. Shmueli résume la démarche de la façon suivante : un ensemble de facteurs représentés par les variables X, sont supposés causer l'effet étudié et mesuré par la variable Y. Les modèles sont donc élaborés dans le but d'expliquer l'effet étudié Y d'après les variables X. Dans ce cadre, le rôle de la théorie est extrêmement

prégnant. Le recours des chercheurs à des données spécifiques ou encore à une méthode statistique particulière s'effectue strictement dans le cadre de la théorie à tester. La théorie justifie l'utilisation ou l'application respectivement des données ou des méthodes statistiques.

Un modèle explicatif est élaboré pour tester une explication causale à l'aide de modèles statistiques. Un diagramme causal peut permettre d'illustrer la relation entre l'hypothèse où les hypothèses, et le phénomène étudié. L'objectif d'un modèle prédictif est de pouvoir établir la valeur d'une nouvelle ou future observation dans un jeu de données, en fonction des caractéristiques qui lui sont propres. Pour ce faire, des modèles statistiques ou encore des algorithmes de *Data Mining* sont développés. Les modèles statistiques descriptifs sont quant à eux utilisés pour décrire un jeu de données. L'objectif est de résumer l'information qu'il contient et de représenter sa structure (distribution, etc.). Son application n'est basée sur aucune théorie, ne cherche pas à expliquer un phénomène en particulier et ne permet pas en l'état de faire des prédictions.

À travers l'exploration des processus de développement urbain en contexte transfrontalier, la démarche modélisatrice initiée dans ce travail affiche deux objectifs :

- Le premier objectif est de développer une méthode à la fois descriptive et explicative pour modéliser le développement urbain. En effet, le modèle doit permettre de s'interroger sur le rôle de la spécificité territoriale, notamment celui de la localisation géographique lors de la caractérisation du processus d'urbanisation. Dès lors, il est nécessaire qu'il présente un caractère assez explicite pour permettre d'analyser le processus en détails, et d'observer, le cas échéant, ses différences et ses similitudes sur la zone d'étude. La démarche élaborée permettra donc d'étudier le caractère généralisable du processus d'urbanisation sur la zone frontalière et s'ancre de fait dans une approche théorique.

- Le second objectif est de développer une méthode transposable à d'autres contextes. Pour ce faire, nous explorons le potentiel d'analyse lié à la mise en place d'une méthode simple et rapide. La démarche de modélisation doit être accessible tant par la réalisation technique que par la clarté des étapes qui la compose et des résultats qui sont obtenus. Ce point permet également de rendre le modèle accessible à un large public, qu'il soit expert ou non.

Le modèle créé dans ce travail a donc pour objectif premier d'être explicatif puisqu'il est conçu pour explorer les caractéristiques du processus de développement urbain sur les zones frontalières. Il permettra, dans un premier temps, de mieux comprendre les processus de développement urbain, notamment à travers l'étude du lien entre leur localisation géographique dans un pays et leurs caractéristiques. Néanmoins, une portée prédictive de la modélisation peut aussi être envisagée sur la base de ce travail. Le modèle pourrait permettre, dans un second temps, de proposer et d'étudier les résultats de projections ou de

scénarios prospectifs issus de simulations basées sur le modèle créé précédemment. La modélisation permet également de fournir un support visuel offrant la possibilité aux utilisateurs d'observer concrètement un processus spatial. Il présentera donc également les caractéristiques requises pour s'intégrer dans une approche sociale telle que définie par F. Varenne précédemment, et servir de support à la discussion. C'est en appréhendant la modélisation sous cet angle que le travail de recherche a été réalisé. Après avoir abordé la question du « pourquoi » modéliser, la section suivante s'intéresse aux moyens qui peuvent être mobilisés pour atteindre les objectifs énoncés précédemment.

1.2. Comment modéliser ?

Le mot modèle tire sa racine du latin *modulus*, mesurer. Même s'il présente un caractère polysémique, la notion de mesure est omniprésente quel que soit le contexte de son utilisation. Plusieurs exemples sont présentés ici, pour illustrer ces propos. Il peut caractériser « ce qui est donné pour être reproduit ». Un modèle fourni pour la création de t-shirt par exemple, ou chaque entité aura pour objectif de ressembler exactement au modèle. Dans ce cas, le modèle présente la « mesure » de l'objectif à atteindre. Ce dernier sert de référence. Un modèle peut aussi caractériser un objet qui d'après ses caractéristiques sert de cadre de référence (gamme, qualité, etc.). C'est le cas lorsqu'on parle de modèle de voiture. *À contrario*, quand le terme est utilisé pour caractériser une personne, l'objectif sous-jacent est de souligner son caractère exceptionnel. Kilian Jornet[10] est un modèle dans le domaine de la course nature, il présente les qualités considérées comme optimales et les performances les plus abouties dans ce sport. Socrate est, lui, un modèle par la pensée qu'il a développé de son vivant, ou encore par la force dont il a fait preuve en acceptant la sentence de sa mort. Dans un cadre plus artistique, un modèle est le support de l'inspiration des auteurs. L'objectif peut alors diverger, tendre vers une reproduction la plus exacte possible, vers l'imitation « parfaite », même si cette dernière restera sujette à la vision de l'auteur ou encore tendre vers une interprétation personnelle de l'artiste, faire ressortir une vision particulière, un trait de caractère. La liste des exemples pourrait s'étendre, il n'en reste pas moins que fondamentalement, le modèle permet de mesurer ou d'évaluer un état, qui soit un objectif à atteindre ou une référence permettant de se positionner.

C'est la définition du modèle dans son utilisation en géographie, et permettant la simulation, qui attire notre attention dans la suite de ce paragraphe. L'intérêt porte sur les modèles dynamiques qui offrent la possibilité de simuler un processus spatial. Dans ce contexte, le mot modèle ne déroge pas à la règle. Il sert en effet d'outil de mesure : la mesure de l'évolution d'un ou plusieurs phénomènes par le biais de leurs descriptions. La définition classiquement utilisée en géographie est celle énoncé par P. Haggett (1977). Il définit le modèle comme une « représentation idéalisée du monde réel construite pour démontrer

[10] Sportif professionnel espagnol connu notamment pour ses exploits en Ultra-trail, 7 fois champion du monde de ski-alpinisme, 5 fois vainqueur des Skyrunner World series, record man d'ascensions comme celles du Mont Blanc, du Cervin ou encore du Denali.

certaines de ses propriétés ». Un modélisateur cherche à visualiser ou à représenter ce qu'il ne peut pas voir ou bien ce qu'il a du mal à concevoir, à comprendre ou prédire. D'une certaine façon, le modélisateur, comme l'artiste, peut moduler sa création suivant son objectif : essayer de décrire au maximum le phénomène et/ou son environnement pour capter les moindres détails qu'il peut percevoir, ou au contraire limiter les détails pour faire ressortir ou mettre en lumière un trait spécifique, l'objet d'étude.

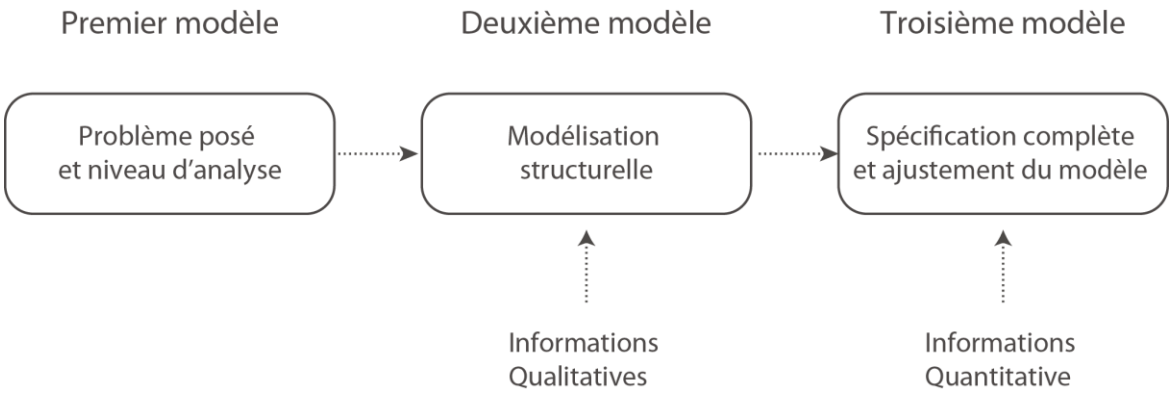


Figure 2. 1 : Du modèle conceptuel au modèle de simulation (Source : C. Tannier (2010))

C. Tannier illustre ce processus à travers la création de trois modèles (Figure 2. 1). Le premier, conceptuel, dépend du problème posé et du niveau d'analyse. La « simplification » de la réalité nécessaire à la création d'un modèle va donc être différente suivant le champ de recherche et de niveau d'analyse. Par exemple la ville, un objet géographique très largement étudié dans différentes disciplines (géographie, économie, urbanisme, sociologie, etc.), pourra être conceptualisée à travers différentes représentations. Pour l'économiste, elle peut être simplifiée en un point, avec une logique binaire. C'est la ville ou ça ne l'est pas. Elle pourra être définie par son nombre d'emploi, faisant d'elle un pôle plus ou moins attractif, qui présentera un potentiel d'échange de flux plus ou moins important (flux de travailleur, flux monétaire, etc.). En géographie, la ville peut être modélisée avec un niveau de détails plus important. Elle pourra être, par exemple, représentée, dans un premier temps, par son occupation du sol (urbain dense, urbain diffus). À un niveau de détails plus fins, elle pourra, dans un second temps, être composée de bâtiments plus ou moins volumineux, plus ou moins haut, et plus ou moins espacés pour un architecte ou un urbaniste. Elle pourra également être représentée de manière linéaire par son réseau de transport (route, train, transports publics collectifs) dans le cadre de l'étude de la desserte ou de la congestion à l'échelle d'une agglomération. Bien d'autres exemples pourraient être mobilisés pour illustrer ces propos.

Dans chacun des cas mentionnés précédemment, la ville est modélisée différemment selon des formes d'implantations différentes (point, ligne, surface). Par exemple, la ville est représentée sous la forme d'un point auquel on associe un cercle plus ou moins large traduisant le nombre d'emplois qu'elle accueille, ou sa population. Dans le cas d'une étude de l'occupation du sol, la ville sera subdivisée en unités spatiales, plus ou moins fines,

régulières ou non, permettant de décrire son occupation du sol. Concernant la desserte, la ville sera modélisée de manière linéaire, avec un ensemble d'arcs et de nœuds, décrivant son réseau de transport et permettant de visualiser les possibilités en termes de trajets. Dans ces différents cas de figures, la ville est schématisée conceptuellement respectivement, selon un signe à petite échelle, une surface (ou un ensemble de surface), une ligne. Chacune de ces visions apporte sa contribution dans la compréhension du système urbain. Il est néanmoins nécessaire de se focaliser sur une des visions, pour « simplifier » l'objet et pouvoir en étudier un aspect spécifique. La question du niveau de détails nécessaire à l'exploration du processus étudier se pose alors.

Comme mentionné précédemment, la modélisation ne peut être totalement objective et l'environnement modélisé sera forcément le fruit de la vision du modélisateur sur la thématique, le processus qu'il cherche à étudier. « Pour un observateur B, un objet A* est un modèle d'un objet A dans la mesure ou B peut utiliser A* pour répondre à des questions qui l'intéressent au sujet de A » (Minsky, 1965). Le modèle est doublement relatif : relatif à l'observateur et relatif au questionnement qui est le sien dans le contexte dans lequel il se trouve (Varenne & Pumain, 2017). L'option retenue dans ce travail est de limiter la quantité de données mobilisées. L'idée développée à travers cette démarche est de limiter par la même occasion le parti pris du modélisateur en généralisant la description. La description est alors moins précise mais elle est aussi moins contrainte. Pour exemple, la caractérisation d'une zone en tant que « zone urbaine » a des implications en matière de description d'occupation du sol. Le choix d'inclure dans cette description la présence d'une école va définir plus précisément la zone urbaine qui possède une école. Ce choix indique cependant implicitement que la présence d'école peut être un facteur important à prendre en compte dans l'étude du processus dont on cherche à modéliser l'évolution et/ou à comprendre les composantes. La question générale que l'on peut légitimement se poser est donc la suivante : Pourquoi les écoles ?

> *Les écoles sont-elles plus pertinentes qu'un autre bâtiment spécifique ? Plus pertinentes que les commerces ? Doit-on inclure les crèches ? Les universités ? Est-il pertinent d'atteindre ce niveau de détail dans une exploration de processus régionaux ? Peut-on trouver la donnée en général, et dans un cadre transfrontalier en particulier ? Si oui quel sera le biais ?*

En considérant la situation dans le cas contraire, le fait de caractériser la zone comme urbaine ne restreint en aucun cas son potentiel à abriter une école ou tout autre bâtiment dont l'usage peut influencer l'attractivité d'une zone et donc son urbanisation. La démarche consiste, ici, à localiser les zones qui semblent favoriser les transitions urbaines en fournissant l'environnement[11] adéquat. Une fois les zones influentes localisées, il est également possible d'analyser plus en détails leur composition pour tenter d'expliquer leur influence positive sur l'urbanisation (commerces, structures institutionnelles, etc.).

[11] L'environnement tel qu'appréhendé dans ce travail, inclus la description de l'occupation du sol dans le voisinage des cellules urbanisées, ainsi que leur accessibilité aux pôles urbains majeurs et la localisation dans l'un ou l'autre des pays frontaliers.

La restriction dans la diversité des données mobilisées est par ailleurs également liée au contexte transfrontalier. En effet, cette localisation impose une réflexion sur les contraintes de disponibilité concernant une donnée transfrontalière harmonisée. Ce point sera développé plus en détails dans la sous-section suivante. Toutefois, même s'il apparaît qu'un contexte particulier puisse entrainer des choix méthodologiques, il est important de souligner que ce choix provient aussi d'une volonté préalable. Il est question, dans les objectifs, de développer un modèle simple et transposable. Dès lors, nous pouvons explorer les avantages inhérents à l'utilisation d'un modèle épuré.

Si l'objectif est de rendre la méthode accessible et transposable, donc simple, un second point à pourtant légitimé l'intérêt de la démarche : celui de l'exploration du potentiel de modélisation du processus de développement urbain par un modèle simple. Cette démarche ne s'inscrit pas dans celles développées récemment, plutôt orientées vers l'implémentation d'une grande quantité de données détaillées. Ce travail de recherche a pour dessein de permettre l'exploration du potentiel de données basiques dans le cadre de la modélisation du processus de développement urbain. Il permet l'élaboration d'une méthodologie transposable pour promouvoir la diversité des zones d'étude dont l'analyse des processus géographiques dynamique serait possible à moindre coût (temps, financier), mais aussi de rendre accessible le processus de modélisation ainsi que l'interprétation des résultats à un public qu'il soit expert ou non. Ce point favorise également l'exploration des performances d'un modèle épuré. Cette approche de la modélisation est par ailleurs particulièrement intéressante à explorer, dans un monde ou la complexité des modèles créés à des fins exploratoires permet d'obtenir des résultats concluant au détriment de la clarté de l'illustration du processus étudié. La littérature y fait parfois référence en tant que « boites noires » (Antoni, 2016).

Après avoir défini et argumenté le choix de l'approche par la modélisation, la question des données doit être abordée. Le choix des données est fondamental puisqu'il aura une incidence sur les résultats produits. Expliciter les choix ne permet pas seulement d'évaluer leur pertinence au vu du contexte, mais permet aussi d'évaluer la pertinence de l'interprétation des résultats, mais aussi les apports et les limites de la méthode utilisée. Ici, l'élaboration de la méthodologie se base sur l'utilisation d'une donnée libre. Les motifs soutenant cette décision sont développés dans la sous-section suivante.

1.3. Avec quoi modéliser ?

Les données en modélisation

La donnée est au cœur des préoccupations d'une multitude d'organismes publics ou privés. Ainsi, le Commissariat général au développement durable, à travers Laurence Monnoyer-Smith (Commissaire générale au développement durable et Déléguée interministérielle au développement durable) réaffiche dans sa lettre n° 21 de février 2018 ses objectifs de production, d'exploitation et de valorisation des données en France (Monnoyer-Smith, 2018). Elle insiste sur leur lien avec les politiques publiques qui doivent s'appuyer sur des diagnostics établis à partir de données fiables. Elle souligne également leur rôle dans l'évaluation de ces mêmes politiques publiques, qui nécessite une information robuste. Pour finir, de son point de vue, les données constituent une richesse à exploiter encore plus en amont, notamment en favorisant leur utilisation dans le cadre de la recherche académique. L'utilisation de la donnée représente donc un enjeu majeur d'après cette institution. La communauté scientifique rejoint ce point de vue. Le développement de la puissance informatique, parallèlement au développement des outils d'acquisitions a fait émerger un nouveau défi pour les chercheurs. En effet, l'exploitation des données produites peut s'avérer compliquée notamment en raison de l'absence d'harmonisation.

La constitution des premières bases de données a soulevé des points qui n'avaient pas été anticipés lors de l'acquisition même de la donnée. Les responsables des bases de données ont dû faire face aux incompatibilités entre données supposées contenir la même information. Pour cause, les définitions thématiques[12] ou techniques[13] de la donnée acquise peuvent différer (Terrier, 2011). Les définitions des classes, les nomenclatures, les échelles d'acquisitions, n'étaient pas harmonisées. L'harmonisation des données dans leur définition et leurs méthodes d'acquisition a été réalisée progressivement. Cette démarche, s'est considérablement répandue mais se heurte parfois encore aux limites administratives. L'Union Européenne, à travers la mise en place de socles communs tend à essayer de gommer ces différences. La directive INSPIRE[14] invite et incite à renseigner en détails les méthodes d'acquisitions et attributs des données crées en Europe. Elle pousse les pays membres à créer et mettre à disposition leurs données suivant des règles communes et à constituer des catalogues de données (métadonnées) afin de favoriser les échanges voire l'interopérabilité lors de leur utilisation. L'objectif, sous-jacent, est également l'acquisition unique pour éviter de dupliquer cette étape dans tous les services ayant recours à l'utilisation d'une donnée particulière. La directive encourage pour finir la mise à disposition gratuite des métadonnées et des données, dans la mesure du possible.

[12] Par exemple, la définition du seuil de densité d'une catégorie d'occupation du sol comme le tissu urbain

[13] Par exemple, une année ou une résolution d'acquisition

[14] Directive 2007/2/EC du Parlement Européen et du Conseil du 14 mars 2007 établissant une Infrastructure de l'information spatiale dans la Communauté. Son objectif est d'assurer l'interopérabilité entre bases de données et faciliter la diffusion, la disponibilité, l'utilisation et la réutilisation de l'information géographique en Europe.

Le programme CORINE Land Cover est un exemple qui illustre parfaitement la mise en application des préconisations d'INSPIRE. Il a permis l'élaboration d'une base de données d'occupation du sol, d'après une nomenclature unique utilisée par chaque pays prenant part au projet (European Environment Agency, 1985). L'acquisition de la donnée, mais aussi sa mise à jour, sont harmonisées dans la mesure du possible.

L'Union Européenne porte donc une attention particulière sur l'acquisition des données au sein des États membres, mais surtout sur leur mise à disposition et leur exploitation à travers l'instigation de modalités de partage. Cette entreprise est également présente à l'échelle internationale comme le montre la charte élaborée par le G8 concernant l'ouverture des données publiques. Six ans après la directive INSPIRE, le groupe de discussion composé de 8 puissances économiques majeures sur la scène internationale à savoir États-Unis, Japon, Allemagne, France, Royaume-Uni, Italie, Canada, et la Russie (suspendu depuis 2014) insiste sur la partie prenante de l'ouverture des données publiques, dans un mouvement mondial. La Chartre présente ainsi des catégories de données à forte valeur ajoutée (« Charte du G8 pour l'ouverture des données publiques », 2013). Parmi celles-ci figurent notamment : les données géospatiales, environnementales, et d'observations de la Terre. Les données géographiques prennent donc places aux côtés des données socio-économiques comme données d'intérêt majeur pour « améliorer le fonctionnement des démocraties et encourager les usages innovants » (« Charte du G8 pour l'ouverture des données publiques », 2013).

Pour autant, l'harmonisation des données en général n'est pas toujours effective suivant l'échelle ou l'objet d'étude. C'est parfois l'interprétation des données par les institutions, qui ne permet pas une comparaison directe des observations. Pour exemple, l'évaluation des tendances majeures de changement d'occupation du sol à partir des données CORINE Land Cover sont disponibles en hectare sur la totalité du territoire pour la France mais sont données en hectare par habitant en Allemagne. Ce problème peut paraître anecdotique et simple à résoudre. Il dénote néanmoins particulièrement bien des divergences en termes de pratiques, pouvant rapidement rendre compliqué l'exploitation de données internationales lors de l'étude d'une zone transfrontalière.

Le caractère transfrontalier de la zone d'étude, associé à la volonté institutionnelle de créer des référentiels harmonisés oriente cette recherche vers l'utilisation de données libres. L'harmonisation des données utilisées pour la modélisation est un point fondamental pour l'interprétation des résultats fournis par le modèle. Leur exploitation rentre aussi dans un cadre qui facilite la création d'une démarche transposable et simple. L'utilisation des données libres est plus amplement discutée dans la section suivante.

La donnée libre (Open data) : avantage et inconvénient

L'*open data* est définie comme une donnée dont la particularité est d'être accessible à tous. Sa production a pour but de remplir deux objectifs :

- faire preuve de transparence dans la mise en œuvre des politiques publiques ;
- ouvrir des données susceptibles de générer de l'activité et créer des richesses.

Elle est produite aussi bien par des organismes publics que privés. La mouvance « données libres » est le plus souvent associé à l'utilisation de données numériques. Elle en promeut l'accès favorisant la diffusion de la connaissance, tout comme les logiciels *Open Sources* favorisent l'exploitation, l'analyse et la création de connaissances à partir de données. Des chercheurs et penseurs d'internet (notamment Lawrence Lessig, Tim O'Reilly) ont défini 8 principes qui doivent être respectés pour qu'une donnée publique soit considérée comme pleinement ouverte. La donnée doit être complète, brute (sans traitement), à jour et accessible à tous (Chignard, 2013). Elle doit permettre un traitement automatisé, sans enregistrement préalable (donc sans discrimination d'accès possible), dans un format non propriétaire et avec une licence libre (sans copyright ou droits d'auteur).

L'utilisation de données libres offre un avantage majeur pour l'utilisateur : son coût d'acquisition est nul. Par ailleurs, comme elles ont été construites dans l'optique d'une exploitation pluridisciplinaire leur utilisation est facilité (interopérabilité des formats, etc.). Contrairement à la donnée produite par les institutions nationales, l'emprise spatiale présente parfois l'avantage de ne pas être limitée à l'intérieur des frontières d'un État. Cependant, la qualité des données doit être étudiée suivant la source mobilisée. En effet, la qualité des données accessibles acquises de manière participative peut varier. Les détracteurs souligneront cet inconvénient potentiel, auquel les pro Open Data répondront par la transparence totale de l'information et le droit de regard qui permet, si besoin de corriger les erreurs ou de compléter les lacunes.
Au-delà des avantages et inconvénients, l'utilisation de données libres peut aussi être considérée comme une démarche visant à augment l'accessibilité des travaux réalisés. Elles facilitent la reproduction des analyses effectuées et permet la confrontation des résultats à un public plus large. Cette approche s'inscrit dans la démarche INSPIRE.

Le choix d'une donnée libre ?

La question de la donnée est toujours centrale en modélisation. Elle apparaît d'autant plus importante dans le cadre de l'analyse de processus sur une zone frontalière en raison de la nécessité de s'appuyer sur des données harmonisées. L'utilisation de bases de données nationales dépendrait d'une l'harmonisation préalable indispensable. Par exemple, dans le cas d'utilisation de données d'occupation du sol, comme celles définies en tant que tissu urbain, certains points doivent être éclaircis comme : les critères ayant servis à définir un type urbain, le pourcentage de surface imperméable, la méthode d'acquisition (photo interprétation uniquement, télédétection, algorithmes, résolution), la méthode d'agrégation pour une donnée agrégée. Pour un travail portant sur des données élaborées dans différents

pays : dates d'acquisition, méthodes d'acquisitions, temporalité, etc. Ces points permettent d'établir la pertinence d'une comparaison ou d'une utilisation simultanée d'une donnée censée être équivalente de part et d'autre d'une frontière. Si les méthodes divergent, les différences et biais doivent être pris en compte et l'influence sur les résultats doit être discuté, quantifié voire corrigé. En résumé, l'utilisation de données nationales officielles, parfois plus précises, nécessite un traitement préalable important. Ces points permettent de s'interroger sur l'intérêt d'avoir recours à une donnée nationale officielle dans le cadre de notre étude. C'est pourquoi l'utilisation d'une donnée libre et internationale a été privilégiée dans ce travail.

L'utilisation d'une donnée libre s'appuie sur trois aspects de la recherche. Le premier est la volonté d'étudier un espace transfrontalier. L'approche par la modélisation spatiale en géographie pose la question de la disponibilité de données harmonisées de part et d'autre de la frontière. Dans un deuxième temps, l'utilisation de données libres produites par des institutions, soutenant démocratisation de cette pratique, semble un choix opportun. Leur mobilisation favorise d'autant plus le caractère transposable du modèle créé, un des prérequis de la démarche. Pour finir, en dépit d'un intérêt croissant pour l'*Open Data* la diversité et la précision des données libres reste moins importante que celle des données propriétaires, quand bien même cette situation pourrait s'inverser dans le futur. Ce point n'est néanmoins pas limitant puisqu'il permet de restreindre le parti pris lors de la démarche de modélisation en généralisant la description (1.2). La description est alors moins précise mais elle est aussi moins contrainte. Le modèle est donc plus simple. Il permet également d'évaluer la performance de la modélisation du développement urbain sur une zone transfrontalière par des données libres.

Les points exposés précédemment soutiennent l'utilisation des données libres. La section suivante présente les données disponibles sur l'espace de l'Union Européenne, et mobilisées pour le développement du modèle.

2. COLLECTER DES DONNÉES FRONTALIÈRE

Comme présenté dans la première section de ce chapitre, la démarche modélisatrice est une démarche enrichissante, en-deçà de l'aspect prospectif recherché pour l'appui des politiques d'aménagements. « Modéliser c'est apprendre » (Banos, 2016). La modélisation du processus de développement urbain sur une zone frontalière est appréhendée dans cette perspective puisque notre objectif est d'analyser l'influence de la localisation dans l'un ou l'autre des pays frontaliers.

Il s'agira dans cette section de présenter les types de données spatiales harmonisées et disponibles à l'échelle des zones frontalières intra-européennes. C'est principalement trois types de données qui retiennent notre attention : les données d'occupation du sol, les données de réseau routier, les données démographiques. Nous présenterons l'aspect fondamental de la donnée d'occupation du sol pour la démarche adoptée dans cette recherche (sous-section 1). Nous aborderons ensuite la question de la prise en compte des réseaux routier est abordée (sous-section 2). En effet, les travaux d'analyse spatiale menés jusqu'à présent montrent que les réseaux structurent le territoire en favorisant les déplacements (Mimeur, 2016). Pour finir, nous discuterons de l'utilisation des données démographiques (sous-section 3) puisque la répartition de la population, comme celle des réseaux, à un effet structurant sur le territoire.

2.1. L'occupation du sol

Qualification et quantification du changement

Pour fournir une donnée riche d'intérêt, cette dernière doit dans un premier temps être organisée de façon adaptée. La définition des types d'occupations du sol et leurs organisations au sein d'une classification, typologie ou encore nomenclature, permet la visualisation spatiale d'unités analytiques thématiques sur une zone donnée. Dans un second temps, les caractéristiques thématiques de chacune des classes doivent être précisées. Elles peuvent, en effet, varier d'une classification à l'autre, et sont influencées par les contraintes techniques lors de l'acquisition (superficie de la zone d'acquisition de la donnée, les moyens en termes de temps et de technologie, etc.). Par ailleurs, les phénomènes et processus territoriaux peuvent apparaître à différentes échelles spatiales et temporelles ; la précision nécessaire dépend donc aussi du cadre d'application.

L'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) ainsi que le Programme Environnemental des Nations Unis (UNEP) ne se sont pas seulement intéressées aux définitions théoriques de l'occupation et de l'utilisation du sol, ils ont aussi porté leur attention sur la question des classifications de données de d'occupation du sol. Anderson insiste déjà sur le besoin de standardisation et de coordination dans l'acquisition des données, mentionnant la duplication superfétatoire du travail réalisé par les différents utilisateurs (Anderson, 1976). David Harcharik (Député Directeur de la FAO en 2005), lui,

souligne l'importance critique d'avoir une classification harmonisée des classes d'occupation du sol, notamment dans le cadre de l'élaboration d'une donnée ayant pour objectif des interactions entre régions et nations à l'échelle mondiale. Pour soutenir cette démarche, en fournissant une donnée mobilisable quel que soit le lieu, le domaine ou l'échelle d'application, l'élaboration d'une nomenclature présentant 22 classes a été réalisée pour décrire l'occupation du sol à l'échelle mondiale (Figure 2. 2). À une échelle plus fine, l'initiative européenne Corine Land Cover (CLC) présente une classification qui définit jusqu'à 44 classes (Figure 2. 3). Une lecture plus détaillée montre qu'en dépit de son nom « Land Cover », la limite entre « occupation » et « utilisation » du sol est franchie dans certains cas. Par exemple, la classe agricole est aussi définie plus précisément d'après le type d'agriculture ou d'exploitation, notamment pâturages ou exploitation viticole. Les classifications sont souvent composées de plusieurs niveaux. Cette hiérarchie permet d'augmenter progressivement le détail dans la description et engendre, la plupart du temps, une évolution progressive dans la définition d'une occupation du sol vers la qualification de son utilisation à mesure de l'augmentation de la précision du niveau de description. C'est le cas de la classification CLC présentée plus en détails par la suite.

Value	Label	Color
0	No Data	
10	Cropland, rainfed	
11	Herbaceous cover	
12	Tree or shrub cover	
20	Cropland, irrigated or post-flooding	
30	Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous	
40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) /	
50	Tree cover, broadleaved, evergreen, closed to open (>15%)	
60	Tree cover, broadleaved, deciduous, closed to open (>15%)	
61	Tree cover, broadleaved, deciduous, closed (>40%)	
62	Tree cover, broadleaved, deciduous, open (15-40%)	
70	Tree cover, needleleaved, evergreen, closed to open (>15%)	
71	Tree cover, needleleaved, evergreen, closed (>40%)	
72	Tree cover, needleleaved, evergreen, open (15-40%)	
80	Tree cover, needleleaved, deciduous, closed to open (>15%)	
81	Tree cover, needleleaved, deciduous, closed (>40%)	
82	Tree cover, needleleaved, deciduous, open (15-40%)	
90	Tree cover, mixed leaf type (broadleaved and needleleaved)	
100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	
110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	
120	Shrubland	
121	Evergreen shrubland	
122	Deciduous shrubland	
130	Grassland	
140	Lichens and mosses	
150	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)	
152	Sparse shrub (<15%)	
153	Sparse herbaceous cover (<15%)	
160	Tree cover, flooded, fresh or brakish water	
170	Tree cover, flooded, saline water	
180	Shrub or herbaceous cover, flooded, fresh/saline/brakish water	
190	Urban areas	
200	Bare areas	
201	Consolidated bare areas	
202	Unconsolidated bare areas	
210	Water bodies	
220	Permanent snow and ice	

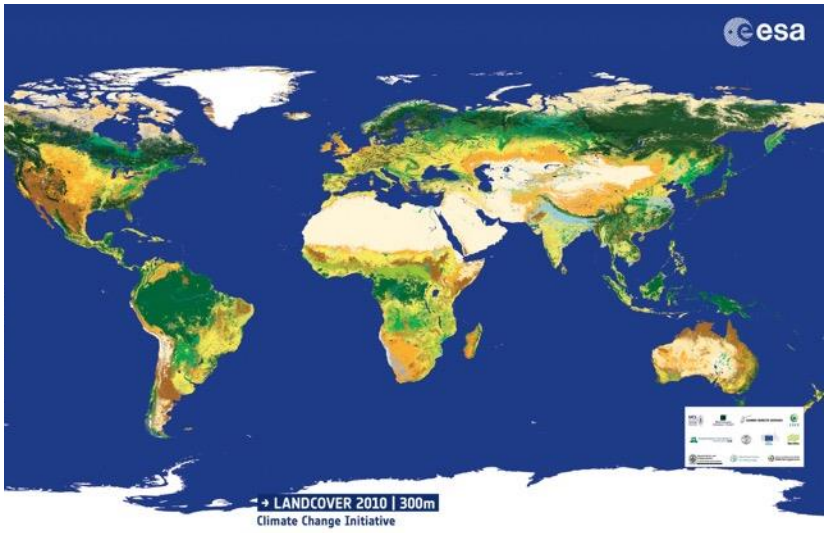


Figure 2. 2 : Global Map Project source ESA/ ESA CCI Land Cover dirigé par UC Louvain

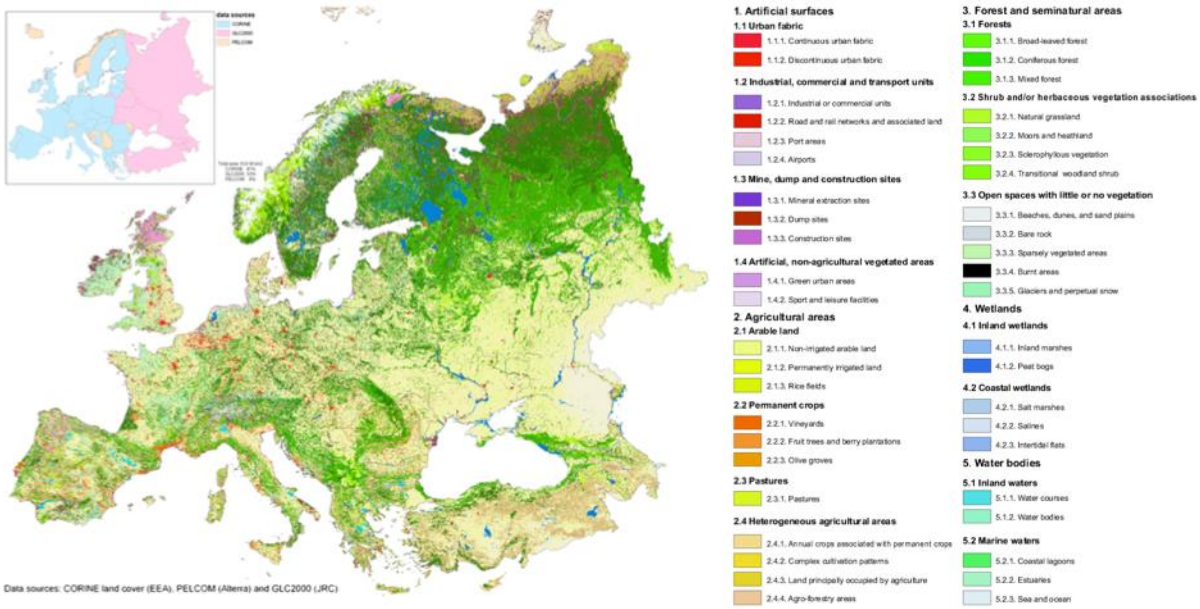


Figure 2. 3 : Corine Land Cover adapté de Mücher 2015

D'après l'observation des classifications précédentes, une même unité spatiale est définie différemment suivant l'objet d'étude. En effet, les zonages GMP et CLC ne présentent pas exactement les mêmes limites, et la définition des classes n'est pas exactement similaire. Pour autant, l'unité spatiale, quel que soit la classification reste intrinsèquement la même. La Figure 2. 4 illustre ce constat à une échelle plus fine. Les cartes sont issues de deux thèses différentes portant sur le même terrain d'étude, à une échelle similaire.

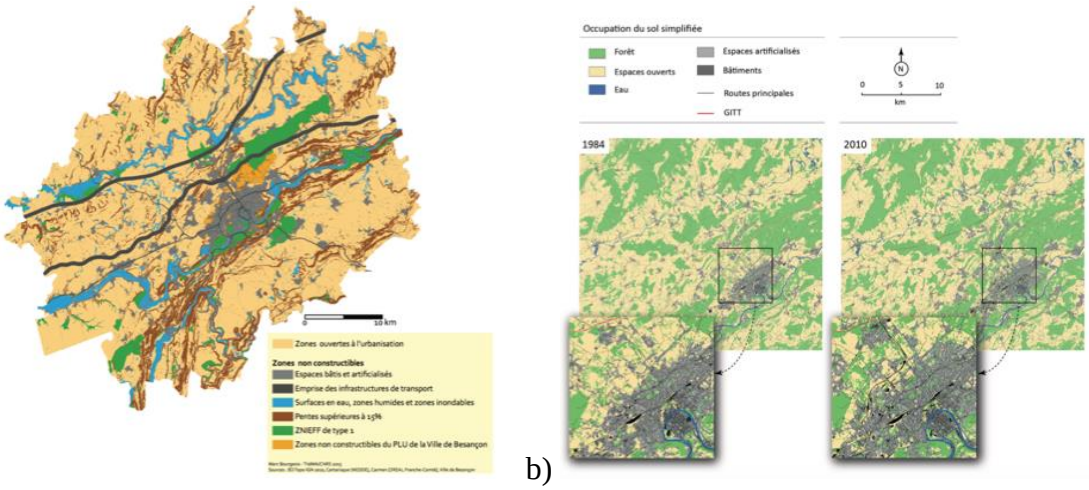


Figure 2. 4 : Illustration portant sur les choix de classifications. A) Carte présentant la zone d'étude centrée sur Besançon dans le cadre de la thèse de M. Bourgeois (2015) intitulée « Zones constructibles de l'air urbaine de Besançon » b) Carte présentant la même zone avec un périmètre différent dans de le cadre de la thèse de Y. Sahraoui (2016) intitulée « Occupation du sol simplifiée du terrain de Besançon en 1984 et 2010 ».

La Figure 2. 4– a) montre une nomenclature composée de 7 classes, permettant de localiser et de différencier les zones constructibles des zones non-constructible. Les espaces non-constructibles voient leur définition se préciser en référence à un bâti préexistant, à un zonage réglementaire restrictif, ou à une contrainte environnementale limitante. Cette classification allie occupation et utilisation du sol. La Figure 2. 4 – b) présente, elle, 5 classes d'occupation du sol simplifié en 1984 et 2010. Elle est donc plus générale que la précédente. Ces deux exemples permettent le constat suivant : sur une même zone, les classifications peuvent être orientées différemment suivant les besoins de l'étude. Pour autant, la zone et ses caractéristiques fondamentales sont exactement identiques.

Les différentes classifications illustrent l'aspect à la fois multi-scalaire et multithématique de la donnée qui, comme les systèmes dont elle tente de capter une certaine essence, peut être considérée à différentes échelles spatiales (étendue et résolution) d'après des thématiques diverses. L'enjeu lié à l'utilisation d'une classification d'occupation du sol réside donc dans la recherche de l'équilibre entre la disponibilité des données, leurs résolutions et l'objectif de l'étude associée. La disponibilité diachronique peut également être un élément décisif dans l'utilisation d'une classification, notamment pour des besoins du calibrage de modèles ou encore pour l'étude du processus de changement en lui-même.

En effet, l'occupation du sol peut être altérée et modifiée par les processus naturels (inondation, sécheresse, éboulement, etc.) ou par les pratiques anthropogéniques (agriculture, étalement urbain, etc.). L'aspect quantitatif est une part importante du travail lié à l'étude de l'occupation puisque l'estimation des surfaces est une condition indispensable à l'estimation des changements. Une meilleure gestion de la ressource sol passe par l'acquisition d'informations sur les configurations d'occupation existantes et leurs évolutions (Anderson, 1976).
Le changement peut être de deux natures différentes : la conversion, qui concerne l'utilisation du sol, et la transition qui elle concerne l'occupation du sol (Turner *et al.*, 1993b). La conversion représente un changement dans l'utilisation du sol. Par exemple, un changement dans l'utilisation agricole d'un sol, qui viserait à faire évoluer une terre cultivée en zone de pâturage ou réciproquement. Il s'agira, dans ce travail, de se focaliser exclusivement sur la transition, c'est-à-dire sur un changement d'occupation du sol, puisque le développement urbain représente la transition d'une classe non-urbaine à une classe urbaine.

L'acquisition d'une donnée d'occupation du sol diachronique est donc primordiale à l'estimation du changement. Sa sélection dépendra du caractère identique de la méthode d'acquisition et de la nomenclature utilisée. Ces deux points doivent être identique pour une estimation optimale du changement. En effet, toute évolution méthodologique lors de l'acquisition, en cas notamment d'améliorations techniques, peut biaiser l'estimation du changement en corrigeant les précédentes imprécisions et biaiser l'interprétation des résultats. Si la précision d'acquisition de la données augment, une correction de l'occupation du sol initiale pourra être confondue avec une transition, induisant une surestimation artificielle des transitions observées.

Fusionner

Pour étudier les changements d'occupations du sol, la question de la définition des différentes classes est primordiale et dépendante du contexte de l'étude ou de la question scientifique. J. Anderson (1996) met l'accent sur la multitude de données acquises au court du temps rendant leurs analyses parfois difficiles. Cette partie a pour objectif de revenir sur ces difficultés de définition ou d'exploitation.

L'utilisateur doit être en mesure de définir ses besoins en amont. Une étape plus ou moins chronophage. Les différentes classes proposent une information qui, dans la plupart des cas, est basée sur le type d'occupation prédominant dans l'unité spatiale à l'étude (Letourneau *et al.*, 2012; Pontius *et al.*, 2001; Sohl & Sayler, 2008; Verburg *et al.*, 2013). La précision de la classification influence la précision des résultats obtenus lors de son analyse ou de son utilisation dans le cadre du développement d'un modèle de simulation.

Par exemple, la classification CLC prétend dissocier l'urbain dense de l'urbain diffus. Dans cette démarche, la définition des classes relève d'une réflexion théorique sur la signification de ces deux termes ainsi que celle d'une réflexion technique. En effet, le moyen technique de classification peut influencer le choix du seuil. La qualité de la donnée initiale doit être prise en compte. Dans le cadre d'une photo-interprétation, il faut s'assurer que la résolution de la photographie permette la distinction au sein d'une même unité thématique des bâtiments, parcs, ou autres types. Si la télédétection est utilisée, il faut être en mesure de s'assurer du potentiel de classification lié à la qualité de l'acquisition de la donnée de base, à la précision des algorithmes ou méthodes de classifications automatiques ou supervisées utilisées, et être conscient des biais. Les paramètres ou critères utilisés doivent être définis et renseignés pour une exploitation optimale de l'information contenue dans la donnée. Ce point est notamment essentiel à son transfert et à la pérennité de son exploitation. En effet, la récupération d'une donnée travaillée non renseignée rend questionnable, voire impossible toute exploitation.

Le renseignement de la donnée a deux objectifs. Le premier mentionné ci-dessus : permettre son utilisation par toute personne ayant droit. La seconde est de pouvoir mesurer l'évolution d'une donnée dans le cas d'une acquisition chronique. Une classification peut être affectée par des changements techniques d'acquisitions permettant une plus grande précision, ou par des changements de positionnement thématiques. De plus, l'agrégation de données peut s'avérer très délicate si les systèmes d'acquisition ou de classification diffèrent. Cette prise de conscience apparaît progressivement dans les démarches d'acquisition depuis les années 2000. Par exemple la directive européenne INSPIRE (2007) impulse l'initiative de la création systématique des métadonnées uniformes (fiche d'identité de la donnée) lors de toute création d'information spatialisée. L'évolution de la qualité de la donnée produite est une information particulièrement importante lors de l'utilisation de données diachroniques et plus particulièrement dans le cadre d'analyses de changements. En effet, un processus de changement ou une intensité peut être biaisé par des différences de mesures liées à l'évolution des moyens techniques ou des procédures. Plus simplement, un changement peut être confondu avec la correction d'une donnée qui était moins précise lors de la précédente acquisition.

En conclusion, une grande variété de données disponibles à différentes échelles spatiales ou temporelles existe. Il est possible de discuter de leur qualité mais « il n'y a pas de classification parfaite d'occupation ou d'utilisation du sol, et il est peu probable que nous puissions en développer une un jour. Il y a différentes perspectives introduites lors du processus de classification, et le processus lui-même tend à être subjectif, quand bien même une approche numérique objective est utilisée » (Anderson, 1976). Le travail développé ici s'intéresse à l'exploitation d'une donnée libre harmonisée à l'échelle de la zone d'étude frontalière. L'objectif est de baser l'élaboration du modèle sur des données présentant ces caractéristiques, et d'évaluer la performance des résultats produits. La donnée d'occupation du sol CLC a été choisie pour remplir ces objectifs et est présentée dans la sous-section suivante.

CORINE Land Cover : une vision européenne diachronique.

Les données CORINE (COoRdination de l'INformation sur l'Environnement) Land Cover sont construites pour fournir un socle d'information commun concernant l'occupation biophysique du sol à l'échelle européenne. L'occupation du sol fait donc l'objet d'une classification commune. Le projet CORINE Land Cover est piloté par l'Agence Européenne pour l'Environnement (AEE) et couvre 39 États. L'Europe est décrite par l'AEE comme étant le continent ou les terres sont le plus intensément utilisées. Cette situation induit une concurrence directe et une compétition entre les différentes occupations ou utilisations du sol. L'étude des divers ressources mises à disposition du public par le biais des sites ministériels ou agences gouvernementales (nationales ou européenne) montre bien une prise de consciences des multiples enjeux liés à l'étude de l'occupation du sol, mais aussi du caractère fini de cette ressource. Une volonté d'étude et de gestion de la ressource qui passe par la mise en place d'objectifs politiques visant, par exemple, à réduire la consommation

de l'espace par l'expansion de l'urbanisation, est clairement affichée. Certains objectifs peuvent paraître antagonistes avec d'une part l'augmentation de la population et le besoin qui consiste à augmenter l'accessibilité, et d'autre part la conservation voire l'augmentation de la protection de l'environnement ou encore du cadre et de la qualité de vie. Les réglementations visent donc une mise en phase entre l'occupation du sol et les intérêts environnementaux et sociaux, qui peuvent parfois être contradictoires. Par exemple, la construction d'un axe routier peut s'avérer nécessaire pour désenclaver un espace d'habitation ou un espace de développement industriel. Néanmoins, cette construction peut réduire sa connectivité écologique, ou bien sera la source d'une nouvelle pollution atmosphérique et sonore. Il en est de même pour le phénomène de mitage[15] qui vise à répondre positivement aux affinités d'une certaine catégorie de la population et est vivement critiqué depuis la seconde moitié du 20ème siècle en raison de l'impact sur les écosystèmes (Herrmann, 2018).

Dès lors, la recherche de l'équilibre entre le contrôle des conséquences sur l'environnement et la production de ressources nécessaires à la santé socio-économique d'un territoire est une priorité des décideurs. Les décisions sont principalement prises à l'échelle locale, la Commission Européenne a cependant un rôle à jouer en s'assurant que les États Membres prennent les mesures nécessaires lors de la planification de leur développement concernant l'occupation ou l'utilisation du sol. Les réglementations au sein de l'union européenne sont établies dans le but de soutenir, d'influencer ou de préconiser l'adaptation des comportements nationaux au regard des changements globaux (ex : le changement climatique) sachant que le secteur économique en est directement dépendant. De manière plus concrète, l'AEE coordonne notamment le pan européen de la mission Copernicus[16] dont l'un des objectifs est de centraliser l'acquisition et la mise à disposition de données d'occupation et d'utilisation du sol sur quatre séries temporelles. Chaque pays reste néanmoins en charge de l'acquisition des données. Ces derniers doivent se conformer aux attentes et contraintes du cadre Européen qui favorise de ce fait l'harmonisation et la disponibilité de la donnée. Le projet met aussi à disposition des données plus spécifique portant sur les changements d'occupation du sol (données SIG, rapport de synthèses, etc.). C'est la base de données Corine Land Cover crée à partir de1985, qui a été développée dans l'objectif de stocker les données d'occupations du sol acquises par chaque membre de l'Union. Dans un premier temps, la photo-interprétation effectuée à partir d'images satellites à haute résolution fût la méthode d'acquisition majoritaire. La couche représentant l'occupation du sol en 1990 a principalement bénéficié de cette méthode. Aujourd'hui, des classifications semi-automatiques assistées par ordinateur sont également utilisées dans certains pays. Ces derniers ont recours à des logiciels pour traiter les images satellites et

[15] Éparpillement, sans plan d'urbanisme réellement cohérent, d'infrastructures, de zones d'habitat, de zones d'activité, dans des espaces initialement ruraux (forestiers ou agricoles). Le phénomène de mitage s'observe en zone périurbaine, sous l'effet de fortes pressions foncières ou/et touristiques et en l'absence d'une réglementation d'occupation du sol suffisamment cohérente et contraignante. (Géoconfluence, 2018).

[16] http://land.copernicus.eu/pan-european

localiser les changements d'occupation du sol (Télédétection, Systèmes d'Information Géographique). Cette donnée est « homogène, strictement comparable pour tous les pays concernés et susceptible d'être mise à jour périodiquement » (Commissariat général au développement durable, 2009). Elle est disponible dans 38 États Européens avec pour année de référence 1990, mise à jour pour 2000, 2006, 2012 et finalement 2018, pour une résolution de 400 m^2. Sa nomenclature présente 44 classes de types de sols au plus bas niveau hiérarchique considéré, le niveau trois (Figure 2. 5). En effet, la classification présente trois niveaux hiérarchiques (Figure 2. 6). Le premier composé de 5 classes et le second de 15. La mise à jour est effectuée en considérant une « unité minimum de cartographie » de 25 ha. Cette information est à prendre en considération lors de l'analyse de changement entre deux dates. La base de changement est quant à elle, plus précise, puisqu'elle répertorie les changements pour une unité minimum de cartographie de 5 ha.

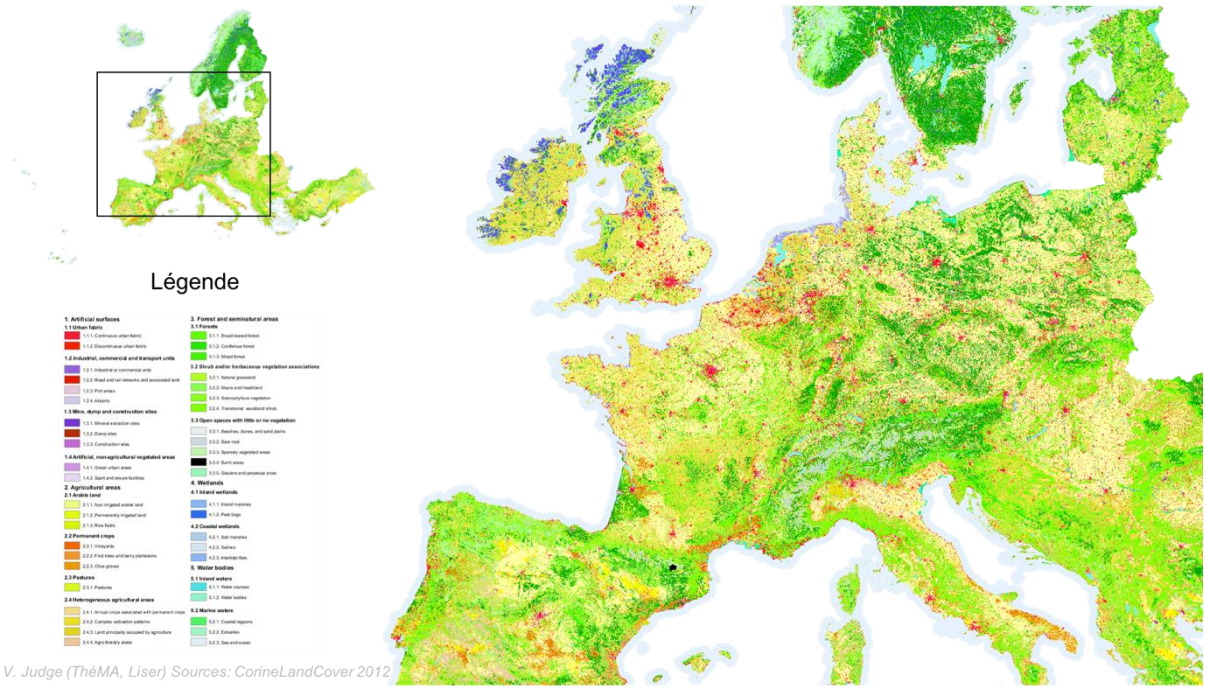


Figure 2. 5 : Nomenclature Corine Land Cover - 44 classes (Source : Copernicus, CLC 2012)

Libellé Niveau 1		Libellé Niveau 2		Libelle Niveau 3
1	Territoires artificialisés	11	Zones urbanisées	111 Tissu urbain continu
				112 Tissu urbain discontinu
		12	Zones industrielles ou commerciales et réseaux de communication	121 Zones industrielles ou commerciales et installations publiques
				122 Réseaux routier et ferroviaire et espaces associés
				123 Zones portuaires
				124 Aéroports
		13	Mines, décharges et chantiers	131 Extraction de matériaux
				132 Décharges
				133 Chantiers
		14	Espaces verts artificialisés, non agricoles	141 Espaces verts urbains
				142 Equipements sportifs et de loisirs
2	Territoires agricoles	21	Terres arables	211 Terres arables hors périmètres d'irrigation
				212 Périmètres irrigués en permanence
				213 Rizières
		22	Cultures permanentes	221 Vignobles
				222 Vergers et petits fruits
				223 Oliveraies
		23	Prairies	231 Prairies et autres surfaces toujours en herbe à usage agricole
		24	Zones agricoles hétérogènes	241 Cultures annuelles associées à des cultures permanentes
				242 Systèmes culturaux et parcellaires complexes
				243 Surfaces essentiellement agricoles, interrompues par des espaces naturels importants
				244 Territoires agroforestiers
3	Forêts et milieux semi-naturels	31	Forêts	311 Forêts de feuillus
				312 Forêts de conifères
				313 Forêts mélangées
		32	Milieux à végétation arbustive et/ou herbacée	321 Pelouses et pâturages naturels
				322 Landes et broussailles
				323 Végétation sclérophylle
				324 Forêt et végétation arbustive en mutation
		33	Espaces ouverts, sans ou avec peu de végétation	331 Plages, dunes et sable
				332 Roches nues
				333 Végétation clairsemée
				334 Zones incendiées
				335 Glaciers et neiges éternelles
4	Zones humides	41	Zones humides intérieures	411 Marais intérieurs
				412 Tourbières
		42	Zones humides côtières	421 Marais maritimes
				422 Marais salants
				423 Zones intertidales
5	Surfaces en eau	51	Eaux continentales	511 Cours et voies d'eau
				512 Plans d'eau
		52	Eaux maritimes	521 Lagunes littorales
				522 Estuaires
				523 Mers et océans

Figure 2. 6 : Classification détaillées des différents niveaux de Corine Land Cover (Source : Copernicus)

Une grande variété de classes est initialement disponible dans la nomenclature. Une reclassification a cependant été nécessaire dans le cadre de la réalisation de ce travail de recherche. Cette reclassification est présentée dans le chapitre 3 puisqu'elle est intrinsèquement liée au choix de la zone d'étude.

Comme spécifié dans l'introduction de cette sous-section, la donnée d'occupation du sol est une donnée fondamentale. Néanmoins, d'autres données libres disponibles à l'échelle de la zone frontalière existent. Ces dernières enrichissent le modèle et peuvent influencer sa performance tout en enrichissant l'interprétation thématique. La sous-section suivante s'attache à présenter une autre donnée libre mobilisable : la donnée liée au réseau routier.

2.2. Le réseau routier : une perception hétérogène de l'espace

La notion de déplacement, donc d'accessibilité, se trouve à la base d'un certain nombre de décisions humaines, dans un cadre personnel ou professionnel. La structure du réseau routier induit une hétérogénéité spatiale du point de vue de l'accessibilité, rendant une zone plus ou moins attractive en termes de développement urbain. Le réseau routier est donc une donnée présentant un intérêt particulier. Le second attribut qui sous-tend l'intégration de cette donnée, est sa disponibilité en tant que donnée libre, à travers entre autres la plateforme OpenStreetMap. Après avoir mis en place la notion de réseaux et apporté des précisions sur les objectifs poussant à intégrer cette donnée dans la modélisation, cette section retracera le lien entre le réseau et la localisation géographique, puis présentera les bases de données réseaux en justifiant la préférence pour OpenStreetMap. Pour finir une présentation de la modélisation spatiale concrète est effectuée.

Les réseaux

L'espace terrestre est composé de lieux « dont la théorie de la localisation, la géographie et la sociologie urbaine permettent de saisir une partie des logiques de répartition » (Antoni, 2016). Leur connexion par le biais d'un réseau permet les déplacement et l'échange de flux. La perception de l'espace se voit alors modifiée par les changements et le développement des réseaux qui le structure. W. Harvey au 17ème siècle associe le mot réseau à sa théorie sur la circulation sanguine, imputant une nouvelle dimension dynamique à ce concept auparavant aux connotations statiques de par son étymologie : du latin « retis » c'est à dire filet (Mimeur, 2016). Le terme réseau, en général, peut recouvrir plusieurs domaines (santé, télécommunication, infrastructure de transports, etc.) qui dans tous les cas sont liés à la communication et favorisent un échange. Ils peuvent être visibles ou non, aériens ou souterrains. Il peut tout aussi bien s'agir de réseaux de transports de personnes (route, train, etc.), de réseaux d'alimentation (gaz, eau potable, eaux usées, électricité, etc.), de réseaux de communication (téléphonique, internet, etc.). Les individus construisent leur représentation de l'espace par le biais d'une appréhension des réseaux qui leur est propre. Cette construction personnelle va influencer leurs actions et choix notamment en matière de mobilité. Ainsi, l'attention des individus est portée sur des notions de distances et de proximité, donc d'accessibilité. Chacun peut de nos jours accéder à l'information géographique localisée grâce au développement des technologies et supports. Ces derniers mettent à disposition quantité d'information auparavant réservées à un public expert. L'accessibilité croissante de l'information pouvant orienter un choix de localisation nous incite à prendre en compte l'hétérogénéité du territoire en matière d'accessibilité par le biais du réseau. En effet, ce dernier comme l'occupation du sol reflète les actions menées par la société en termes d'aménagement tout autant que les modes de vies des individus la composent.

Les réseaux routiers : interaction entre sociétés et réseau routier.

L'interaction spatiale est considérée comme structurante depuis W.Tobler et son énoncé de la première loi de la géographie (1970) basé sur le principe de la gravitation de Newton : « tout interagit avec tout, mais deux choses voisines ont plus de chance d'entrer en interaction que deux choses éloignées ». Ces relations basées sur la proximité vont bien sûre être affectées par le développement des réseaux qui l'impact directement. Pour exemple, la création d'un contournement routier peut réduire le temps d'accessibilité d'un point A à un point B. La distance ne diminue pas puisque dans un contournement, par définition, la distance spatiale augmente. Par contre, la proximité, elle, augmente puisque le temps de parcours, lui, diminue. Ici le développement d'un réseau routier « rapproche » donc deux points malgré l'augmentation de la distance. L'exemple le plus extrême qui pourrait illustrer ces propos est le développement du télétravail rendu possible par le développement du réseau internet. L'accessibilité à la zone d'emploi devient obsolète puisqu'elle se confond avec le logement de l'individu et ne nécessite donc plus l'emprunt du réseau routier. Il est possible d'imaginer que la structure de l'espace géographique pourrait être différente dans un monde ou le télétravail serait démocratisé pour tous les services ne « nécessitant » pas un cadre de travail spécifique mais seulement l'accès à des moyens de communication.

Néanmoins, le recours au télétravail reste encore sporadique de nos jours. Il pourrait également être argumenté que le travail bien qu'occupant une place importante dans le quotidien de la majeure partie de la population, n'est pas la seule activité pouvant influencer un choix résidentiel. L'accessibilité aux aménités localisées majoritairement dans les pôles urbains peuvent aussi être un facteur de décision important. Les espaces abritant des structures en rapport avec l'éducation ou la santé peuvent eux aussi être à la base d'un choix localisé d'habitation. Sur la majorité du territoire français, notamment dans le Grand Est, plus de 83 % des ménages possèdent au moins une voiture (INSEE, 2015). L'importance du parc automobile, qui constitue un transport de masse depuis la fin de la seconde guerre mondiale (Dupuy, 1999), peut aussi intuitivement souligner l'importance de la prise en compte du réseau routier permettant de caractériser l'hétérogénéité d'une zone en termes d'accessibilité aux pôles urbains majeurs regroupant en partie les zones d'emplois mais aussi les aménités pouvant être recherchées par les ménages.

Ainsi, l'accessibilité via le réseau routier reste une donnée dont la mobilisation devrait enrichir l'analyse effectuée sur la zone d'étude. Ce constat nous oriente vers l'exploitation de la base de données OpenStreetMap puisqu'à l'image de CORINE Land Cover pour l'occupation du sol, cette base est disponible librement à l'échelle non pas européenne, mais mondiale.

OpenStreetMap : une vision mondiale.

OpenStreetMap (OSM) est le fruit d'un projet international dont les premiers pas furent initiés en 2004. L'objectif de ce projet est de rendre disponible librement des données de nature diverses, collectées dans le monde entier. Elles peuvent tout aussi bien concerner des données appartenant à la sphère environnementale (rivières, forêts, etc.), urbaine comme le bâti associé à certaines de ses caractéristiques (fonctions) ou encore référencer les réseaux structurant comme les routes ou les voies ferrées. Les données cartographiques collectées sont réutilisables sous licence libre *Open Database Licence* (ODbL), qui oblige les ré-utilisateurs à respecter en retour cette licence et donc à ouvrir leurs propres travaux (depuis le 12 septembre 2012).

OSM fait appel à des contributeurs externes pour remplir les bases de données exploitées lors de la cartographie en ligne. Toute personne est libre de devenir contributeur et d'apporter sa pierre à l'édifice OSM, en suivant les règles établies pour encadrer la contribution (respect du droit de propriété intellectuelle, etc.). La création de la donnée peut, comme dans le cas de l'occupation du sol, être diverse (liens avec des répertoires institutionnels libre, photo-interprétation, etc.). La méthode varie suivant le type de donnée. La photo-interprétation d'images satellites est une méthode d'acquisition majeure pour la création de la donnée réseau. L'étape de digitalisation est effectuée sous forme d'arcs (polylignes) interconnectés par des nœuds (points) illustrant la jonction de deux tronçons d'un même niveau hiérarchique ou de niveaux hiérarchiques différents.

La question de la qualité d'une donnée créée par des contributeurs externes aux institutions et sur la base du volontariat est récurrente. Ce questionnement est de même ordre que celui de l'utilisation de données libres en général et pour toute sortes de contributions volontaires *(Eng. Crowdsourcing)*. Bien que fournissant une donnée libre d'utilisation et créée dans un esprit participatif, OSM reste une association organisée, à laquelle les contributeurs adhèrent en signant un règlement dont l'objectif est d'encadrer la participation. Outre l'utilisation de plus en plus répandue des fonds de cartes OSM (applications téléphoniques, flux wms Arcgis ou Qgis, etc.) des institutions comme l'Institut National de l'Information Géographique et Forestière (IGN) par le biais de leur plateforme Géoportail[17], dont l'emprise des données se limite principalement au territoire français, ont récemment intégré la donnée OSM dans la liste des fonds de cartes exploitables sur leur plateforme. Si ces démarches et exemples ne sont pas un gage objectif de qualité, ils soulignent néanmoins une certaine confiance accordée par les institutions elles-mêmes. De surcroît, le caractère libre de la donnée permet à l'utilisateur d'explorer librement toutes les données et de se faire sa propre opinion relative à la qualité.

[17] Le Géoportail est une plateforme nationale de diffusion des données couvertes par la directive INSPIRE, directive européenne visant à faciliter la diffusion de l'information géographique. Il s'appuie sur les référentiels de l'IGN et sur des producteurs de données institutionnels pour vous proposer une information officielle et fiable sur le territoire français.

Le dernier point à aborder dans l'optique d'une approche parallèle à celle effectuée à propos de l'occupation du sol est celui du changement. Une limite posée par l'utilisation d'une donnée libre pour matérialiser le réseau est celle de l'absence de données datées, entravant une approche diachronique. Le caractère statique de la prise en compte des réseaux dans les modélisations menant à la simulation de scénarios est une question récurrente dans la modélisation des changements d'occupation du sol ou de l'évolution des systèmes urbains. Même si dès 1997, R. White introduit l'effet des réseaux de transport dans son modèle par le biais d'une contrainte intégrée dans l'automate cellulaire utilisé pour modéliser les changements d'occupation du sol à Cincinnati, ce réseau est définitivement stationnaire (White *et al.*, 1997). La question ici n'est néanmoins pas tant de savoir s'il serait préférable d'intégrer le réseau diachronique ou non puisque cette donnée n'existe pas dans un format libre et harmonisé à l'échelle européenne, mais bien de définir s'il est intéressant d'utiliser une donnée réseau stationnaire plutôt que de s'en abstenir. Pour autant il est également important de souligner que l'utilisation de cette donnée est envisagée dans le but d'illustrer l'hétérogénéité de la zone en matière d'accessibilité, un facteur impactant de manière plus ou moins directe les changements d'occupation du sol. Ainsi, une image globale peut être préférable à l'absence de représentation.

La notion de réseaux routier fait implicitement référence à la notion de déplacement. C'est bien du déplacement des individus dont il s'agit ici. Si l'accessibilité d'un territoire n'est pas en tout point homogène en raison notamment de la présence ou non d'un réseau de transport mais aussi de la taille des infrastructures en question, ce point est largement dépendant de la répartition des individus et de leurs habitudes sur le territoire. Toutefois la dépendance est réciproque : un lieu d'importance le sera en raison de l'activité qui s'y déroule, étroitement liée à son importance démographique mais aussi à son accessibilité via le réseau ; le réseau se développera pour rendre plus accessible les pôles présentant un certain intérêt au départ, cette importance se verra favorisée par le développement de l'accessibilité. La sous-section suivante s'attarde de ce fait sur la question de la donnée démographique.

2.3. La répartition démographique : le reflet de la prépondérance d'un lieu.

La sédentarisation de l'être humain, et l'évolution des espaces d'habitations qui s'en suivirent ont participé à la structuration de l'espace tel que nous le connaissons. C'est plus précisément au peuplement que cette sous-section fait référence. En effet, « le terme de peuplement désigne la forme prise par la répartition de la population dans un territoire, c'est donc une configuration spatiale. » (Pumain, 2008). Cette répartition n'a jamais été purement aléatoire, elle est le résultat d'une adaptation aux contraintes du milieu. Si les premiers facteurs d'implantation nécessaires à l'être humain étaient directement liés à l'environnement naturel, lui permettant ou non de s'implanter : ressources alimentaires, ressources matérielles, climat, topographie, occupation du sol, etc. c'est encore et toujours le cas actuellement. Seule la forme de ces ressources a évolué. La ressource alimentaire n'est plus directement récoltée, mais accessible à travers l'emploi fournissant les moyens

financiers d'acquérir à minima le nécessaire pour vivre. La localisation optimale ne sera donc plus influencée par la localisation de l'arbre fruitier ou celle de la parcelle favorable à l'élevage personnel, mais à celle des zones d'emplois, tout du moins pour la majeure partie de la population. Nous assistons aux mêmes phénomènes que par le passé en considérant un contexte actuel dont les facteurs ont évolué. L'occupation du sol va donc influencer les mouvements de population et réciproquement.
Cette sous-section s'attache à présenter l'intérêt d'utiliser les données démographiques dans la démarche de modélisation de l'espace transfrontalier, initié dans cette recherche.

Le peuplement : reflet de l'organisation de la société

« La notion de peuplement, qui interroge les localisations, les concentrations et la, plus ou moins grande, régularité de l'installation des populations, est caractéristique de l'approche géographique par opposition à une approche démographique qui se centre davantage sur les caractères de cette population et les processus de son évolution » (Pumain, 2008). W. Christaller (1933) et A. Lösch (1940) décrivent initialement le phénomène de localisation des activités à travers la théorie des lieux centraux. La centralité est le concept principalement à l'origine de la formalisation de cette théorie. « La notion de centralité justifie le regroupement en un même lieu de la production des services de même niveau et de même portée destinés à la population dispersée » (Pumain, 2004). La hiérarchie provient de la corrélation entre le type ou le niveau d'importance des services et leurs portées spatiales (distance acceptable pour recourir à ce service). Cette théorie a été éprouvée à travers le temps et dans différentes régions du monde (Claval, 1966). L'objectif ici n'est pas d'expliquer la répartition ou la taille des villes sur la zone d'étude, mais d'y définir des pôles pouvant être plus ou moins attractifs et intervenir dans la localisation des processus d'urbanisation. Présentant des faiblesses relatives à la singularité des hypothèses[18] de départ et à l'évolution des systèmes de productions, la théorie des lieux centraux garde cependant un intérêt certain puisque « de nombreuses activités localisent en fonction de la présence des services urbains et renforcent la corrélation entre le niveau de ces services et le rang des villes dans la hiérarchie urbaine, définie d'après la population des villes ou l'importance de leur production » (Pumain, 2004). Ces nombreuses applications permettent d'observer un rendu correct, tenant compte de la différenciation des réseaux urbains aux échelles moyennes, dans des régions relativement homogènes (Pumain, 2008).

[18] 1. L'espace géographique est homogène (isotrope) ; 2. La population et les ressources sont uniformément distribuées ; 3. Les consommateurs sont identiques et maximisateurs ; 4. Les prix sont fixés et aucun producteur ne peut réaliser de marge excédentaire (concurrence parfaite) ; 5. Les coûts de transport des biens urbains sont à charge du consommateur rural ; 6. Il existe des économies d'échelles dans la production (i.e. le coût de production de certains produits décroît avec la quantité produite) ; 7. Un lieu central de niveau t produit également des biens de niveau inférieur (dans un système de profondeur 3 (k = 3) les lieux de niveau 1 produiront ainsi des biens de niveau 1, 2 et 3, tandis que les lieux de niveaux 2 ne produiront que des biens de niveau 2 et 3 et les lieux à la base de la hiérarchie ne produiront que des biens de niveau 3).

La théorie de Christaller attire notre attention puisque la localisation de population au sein d'une zone spécifique et engendrée par l'attractivité d'une zone abritant services, activités, et emplois va elle aussi engendrer une modification de l'occupation du sol, voire créer une boucle de rétroaction positive. La recherche en géographie a montré qu'un pôle plus important a tendance à attirer plus d'individus, aux vues des opportunités plus importantes et plus diversifiées auxquelles il donne accès. Par conséquent il attire plus d'individus. Le recours à des données démographiques semble donc approprié pour définir et localiser les pôles attractifs de la zone, ayant potentiellement un effet sur le développement urbain par le rayonnement des services, emplois, et activités qu'ils abritent. La sous-section suivante permet de présenter la donnée démographique sur le modèle utilisé précédemment pour les données d'occupation du sol et celles du réseau.

La donnée démographique

En géographie, la densité de population était considérée comme un indicateur synthétique de la qualité du milieu et la réussite de la mise en valeur d'une région (Pumain, 2008) Sans faire directement référence à cette « qualité », c'est la concentration comme facteur structurant qui est ici retenu. La donnée démographique dans sa définition la plus simple révèle « l'état quantitatif de la population humaine dans une région ou un pays déterminés » (Dictionnaire Larousse). Le centre national de ressources textuelles et lexicales du CNRS va un peu plus loin : « Science dont l'objet est l'étude statistique des collectivités humaines dans leurs structures fondamentales, sociales, intellectuelles, etc. » et différencie l'aspect qualitatif de cette science : « qui étudie les caractéristiques physiques, intellectuelles, socio-culturelles d'une population » ; et l'aspect quantitatif : « qui s'attache à étudier les structures d'une population − par âge, sexe, profession, etc. − et ses mouvements internes − natalité, mortalité, migrations − en particulier ». L'objet ici est de considérer la population dans son plus simple appareil, à savoir d'après le nombre d'individus qu'elle comporte sur une zone définie. Ce choix résulte de l'effet conjugué de la démarche modélisatrice initié ici et du choix de la zone d'étude : garder le modèle simple, utiliser des données libres et harmonisées. Si les données de recensement sont en général accessibles au sein des pays appartenant à l'Union Européenne, tout du moins sous forme agrégées, elles ne sont pas encore normalisées. Cette absence d'harmonisation limite leur usage brut dans le cadre des études transfrontalières. Ce point est discuté dans la suite de cette section.

INSEE - Destatis : une vision nationale

Comme mentionné précédemment, l'objectif visé à travers l'utilisation d'une donnée démographique dans cette recherche est de localiser des pôles urbains importants à travers leurs poids démographiques. L'hypothèse sous-jacente est que l'importance de ces pôles structure le territoire et influence la localisation des processus d'urbanisation. La zone d'étude localisée à la jonction entre la France et l'Allemagne, et plus amplement présentée par la suite (chapitre 3). Il s'agira ici de présenter la disponibilité des données

démographiques en lien avec le terrain d'étude est présentée dans la suite de cette sous-section.

Le contexte frontalier impose le recours à deux bases de données différentes : l'INSEE (Institut national de la statistique et des études économiques) pour la France et Destatis pour l'Allemagne. Ces deux institutions partagent des objectifs communs, « fournir les informations statistiques nécessaires à la prise de décision dans une société démocratique et garantir la neutralité, l'objectivité et l'indépendance scientifique du travail ainsi que le traitement confidentiel des données individuelles qui nous sont fournies » (Statistisches Bundesamt, [s.d]). Si leurs objectifs sont communs, les méthodes mises en œuvre pour la réalisation de leur mission divergent parfois. Ainsi, les indicateurs et les calendriers ne se superposent pas exactement. La comparaison d'indicateur comme celui du chômage entre la France et l'Allemagne a fait couler de l'encre dans les journaux d'information (*FranceSoir, l'Express, Bulletin de l'observatoire des politiques économiques en Europe, etc*.) soulignant des comparaisons « hasardeuses » entre les taux en question. J. M. Daussin-Benichou et M. Sala respectivement au département des études économiques et au département de la conjoncture à l'INSEE développent 5 facteurs pouvant expliquer la baisse du taux de chômage en Allemagne après 2007 (Daussin-Benichou & Sala, 2013) :

- La population en âge de travailler diminue ;
- Le chômage partiel s'est très fortement développé durant la crise ;
- Les salaires ont nettement moins progressé ;
- La croissance y est plus élevée ;
- Les réformes de la loi Harz modifiant le suivi et l'indemnisation des chômeurs.

Ces articles soulignent les différences structurelles des populations influençant le cet indicateur souvent utilisé comme référence dans l'évaluation de la performance de la gestion d'un pays. Eurostat travaille néanmoins sur l'harmonisation de tels indicateurs à l'échelle européenne, facilitant les comparaisons entre pays membres. La donnée est alors agrégée au NUTS-3[19]. L'exemple du chômage est un exemple parmi d'autres ayant pour objectif de souligner les différences qui peuvent affecter l'analyse des données disponibles ou non disponibles de part et d'autre de la frontière, quel que soit les pays frontaliers considérés. Dès lors, l'utilisation de ces données nécessite, dans la majeure partie des cas, une phase d'harmonisation chronophage.

Ces considérations influencent le recours à une donnée démographique uniquement pour caractériser les pôles urbains majeurs sur la zone d'étude. L'objectif est d'explorer l'émergence de différences frontalières à travers l'élaboration du modèle et la simulation du développement urbain, cela en dépit de l'utilisation de données fondamentalement non-discriminantes en tant que données d'entrées (occupation du sol, accessibilité).

[19] Nomenclature des Unités Territoriales Statistiques : découpage territorial destiné à faciliter les comparaisons entre pays, ou entre régions, d'un même ensemble.

Contrairement aux données de réseaux, la donnée de population pourrait être obtenue à différentes dates. Néanmoins, leur utilisation se limite à fournir une image de la tendance globale de l'importance des pôles urbains sur le territoire de la zone d'étude. Excepté un changement majeur dans la structure de ces pôles sur une 15ène d'années, l'aspect diachronique n'est pas indispensable. En effet, comme mentionnée précédemment, l'urbanisation intrinsèquement liée à l'évolution de la structure de la population, bien que répertoriée comme un des phénomènes de changement d'occupation du sol remarquable à l'échelle mondiale, reste un phénomène plus anecdotique au regard de l'inertie engendrée par le non changement en Europe et plus précisément en France et en Allemagne. Ainsi, la population reste relativement stable et l'évolution démographique ne présente plus d'incidents majeurs comme ceux représentés par les 2 Guerres Mondiales, ou encore le *Baby Boom* qui suivit la seconde.

Conclusion chapitre 2

Suite à la définition des perspectives en matière de modélisation, notamment celles de l'élaboration d'un modèle simple et d'une méthode transposable, nous avons donc opté pour une approche qui s'appuie sur l'utilisation des données libres disponibles à l'échelle des zones frontalière intra-européenne pour modéliser le développement urbain. Notre démarche vise l'exploration du potentiel d'une modélisation qui s'appuie sur un nombre restreint de données. L'objectif à travers ce choix est de conserver un aspect simple qui favorisera la compréhension et l'interprétation des résultats obtenus. Par ailleurs, le cadre frontalier impose également une réflexion supplémentaire sur les données mobilisées et pose la question de l'étape d'harmonisation, chronophage, néanmoins indispensable à leur utilisation. Notre choix s'est donc orienté sur une modélisation mobilisant uniquement des données harmonisées à l'échelle de notre zone d'étude. Pour répondre à notre problématique nous avons choisi la frontière franco-allemande, et centrée sur l'agglomération transfrontalière de Strasbourg-Kehl. Le chapitre suivant permet de présenter la zone d'étude et les arguments soutenant ce choix spécifique.

Chapitre 3
Données libres : La zone d'étude

Nous avons exposé précédemment l'intérêt de focaliser notre attention sur un espace frontalier dans la perspective d'analyser l'influence culturel dans le processus d'urbanisation. L'étude d'une zone frontalière semblent être un choix pertinent puisqu'elle représente par essence, la séparation entre deux cultures différentes. Pour autant, nous avons également souligné que leurs fonctions varient, aussi bien dans le temps et que dans l'espace, à mesure de l'évolution des relations entretenues par les États dont elles délimitent l'emprise. Elles peuvent ainsi être politiques, historiques, administrative, mentale, économique, monétaire, etc. et combiner plusieurs de ces caractéristiques (Hinfray, 2010). Si la frontière est avant tout une limite d'état (Hamez *et al.*, 2013), son statut son statut est évolutif. Nous avons vu que plusieurs auteurs s'attachent à décrire cette évolution, sans pour autant qu'il n'existe aujourd'hui de méthode générique pour mesurer l'influence ou la porosité de la frontière. Dans ce contexte, notre choix s'est porté sur l'étude de la frontière Franco-Allemande pour mesurer l'influence de la frontière à travers la modélisation du développement urbain. En effet, cette frontière est particulièrement intéressante en raison de son évolution notamment à travers l'histoire européenne. La première section permet de revenir sur l'évolution des relations franco-allemande au cours du temps qui ont permis de construire cette frontière physique et symbolique. Les sections 2 et 3 s'attacheront à présenter les données libres mobilisées sur la zone d'étude. La section 2 s'attardera sur la description des données d'occupation du sol. La section 3, quant à elle, est focalisée sur les données permettant une description plus précise de la zone urbanisée, notamment celle de l'accessibilité et de la localisation géographique dans l'un ou l'autre des États séparés par la frontière.

1. ÉVOLUTION HISTORIQUE

Cette section est consacrée à la présentation de la zone d'étude choisie dans le cadre de cette recherche. En effet, les zones frontalières sont des espaces particuliers présentant de forts enjeux à l'échelle européenne (Moullé, 2017). La frontière franco-allemande, incluant l'agglomération de Strasbourg-Kehl, a attiré notre attention pour différents motifs, notamment historiques. Ces motifs sont développés dans la première partie de cette section. La zone et ses limites sont ensuite présentées avant d'aborder les programmes de coopération qui y sont mis en place.

1.1. Une frontière historique mobile

L'histoire illustre l'utilisation par l'Homme des frontières tracées par les éléments naturels. La frontière franco-allemande ne répond que partiellement à ce constat. Bien qu'elle s'appuie de nos jours sur le fleuve Rhin, cette vérité est relative à l'étude de la frontière étatique contemporaine. Cette limite n'a pas toujours été le tracé de référence.

La signature des traités de Westphalie[20] (1648) marque le tracé et la reconnaissance officielle des frontières Européennes. Plusieurs auteurs s'accordent pour prendre cet événement comme le point de repère permettant de situer l'apparition de la notion de « frontière linéaire tracée » (Amilhat Szary & Fourny, 2006; Brunet *et al.*, 1993). Pour cause, ces traités favorisent la reconnaissance mutuelle de la limite de l'influence d'un état. L'existence antérieure des frontières n'est pas remise en cause, cependant, les limites des entités étatiques étaient floues, tout comme les limites spatiales concernant l'exercice du pouvoir de l'état. La Figure 3. 1 présente un tracé des limites disparates entre le Saint-Empire romain germanique[21] et le Royaume de France, avant la signature des traités.

[20] Traités conclus en 1648 entre le Saint Empire Germanique, la France, la Suède et leurs alliés respectifs. Ils présentent 3 composantes majeures signées pour la plupart dans l'année 1648 : i - fin à la guerre de Quatre-Vingts Ans, ii - fin de la guerre de Trente Ans et cession de l'Alsace à la France par l'Empereur Ferdinand III, iii – élaboration deux conventions entre le Saint-Empire et la France, et entre le Saint-Empire et la Suède, relatives à l'exécution de la paix. La signature des traités est aussi considérée comme la création du droit international.

[21] « Désignation officielle (à partir du milieu du XVe siècle) de l'empire fondé par Otton Ier le Grand (962), comprenant les royaumes de Germanie, d'Italie et, à partir de 1032, celui de Bourgogne » (Larousse). La signature des traités de Westphalie fut la base de l'organisation de l'Allemagne jusqu'à la suppression du Saint-Empire romain germanique en 1806.

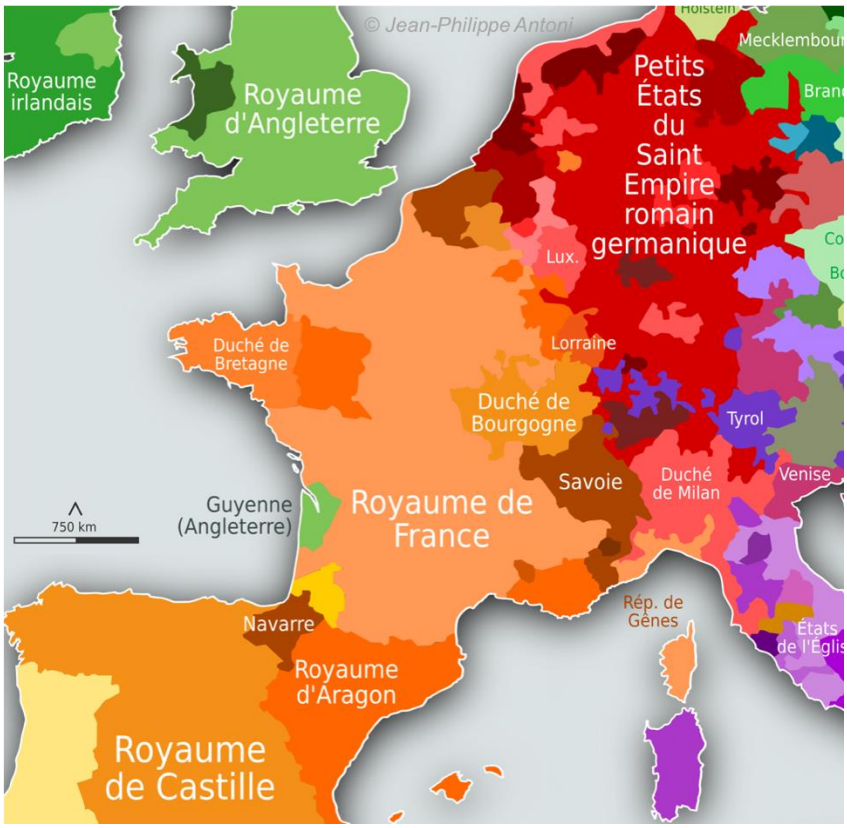


Figure 3. 1 : Limites entre le Royaume de France et le Saint-Empire romain germanique précédent la signature des Traités de Westphalie (Zajec & Antoni, 2017)

La frontière Franco-Allemande suit aujourd'hui le tracé du Rhin. Cependant, il est intéressant de noter que la position de la frontière place l'Alsace, tout comme la Franche-Comté et la Savoie, à l'intérieur des limites du Saint-Empire peu avant la signature des Traités de Westphalie. C'est après la Révolution Française que, pour la première fois, la frontière franco-allemande prend place dans le lit du Rhin. Les révolutionnaires français s'opposent au Saint-Empire romain germanique pendant la guerre de 1792. L'accord conclu le 9 févier 1809 reconnaît l'appartenance à la France de toute la rive gauche du Rhin (Figure 3. 2). Cette situation ne va pas se stabiliser dans le temps puisqu'un peu plus d'un demi-siècle plus tard, l'Alsace ainsi qu'une partie de la Lorraine sont rétrocédée au II[e] Reich nouvellement créé après la victoire de l'armée Prussienne[22] pendant la guerre de 1870-1871. Il est important de souligner une opposition latente des populations intégrées à l'Empire sous de Chancelier Bismarck[23] (Brun, 2015). Cette opposition a fait l'objet de démarches politiques parfois conciliatrices, parfois répressives. La frontière telle qu'elle est localisée

[22] Initialement duché d'Europe de l'Est (localisé au niveau de l'actuelle enclave Russe de Kaliningrad en 1525) il s'étend jusqu'à Brandebourg moins d'un siècle plus tard. L'État prend son essor parallèlement au règne de Louis XIV. C'est Bismarck, qui pour établir l'unité allemande, groupa, sous contrôle prussien, la totalité des États allemands contre la France lors du conflit de 1870.

[23] Otto von Bismarck issue d'une famille bourgeoise berlinoise, occupe successivement les postes de parlementaire et ambassadeur (1847-1862) pour être un membre influent de l'État de Prusse avant de devenir Chancelier du 1[er] Reich (1871), principal artisan de l'Allemagne unifiée au dépend de l'État Prussien.

en 1870, restera en place jusqu'à l'issue de la première Guerre Mondiale (Figure 3. 3 et Figure 3. 4).

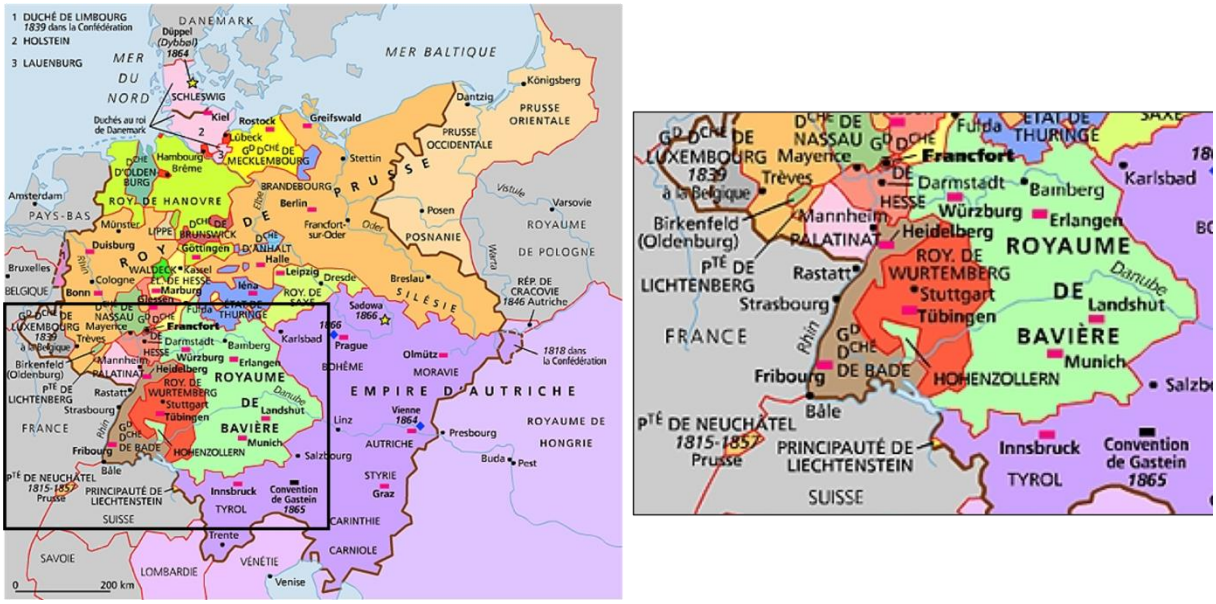


Figure 3. 2 : Frontière franco-allemande en 1809 (Source : Encyclopédie Larousse)

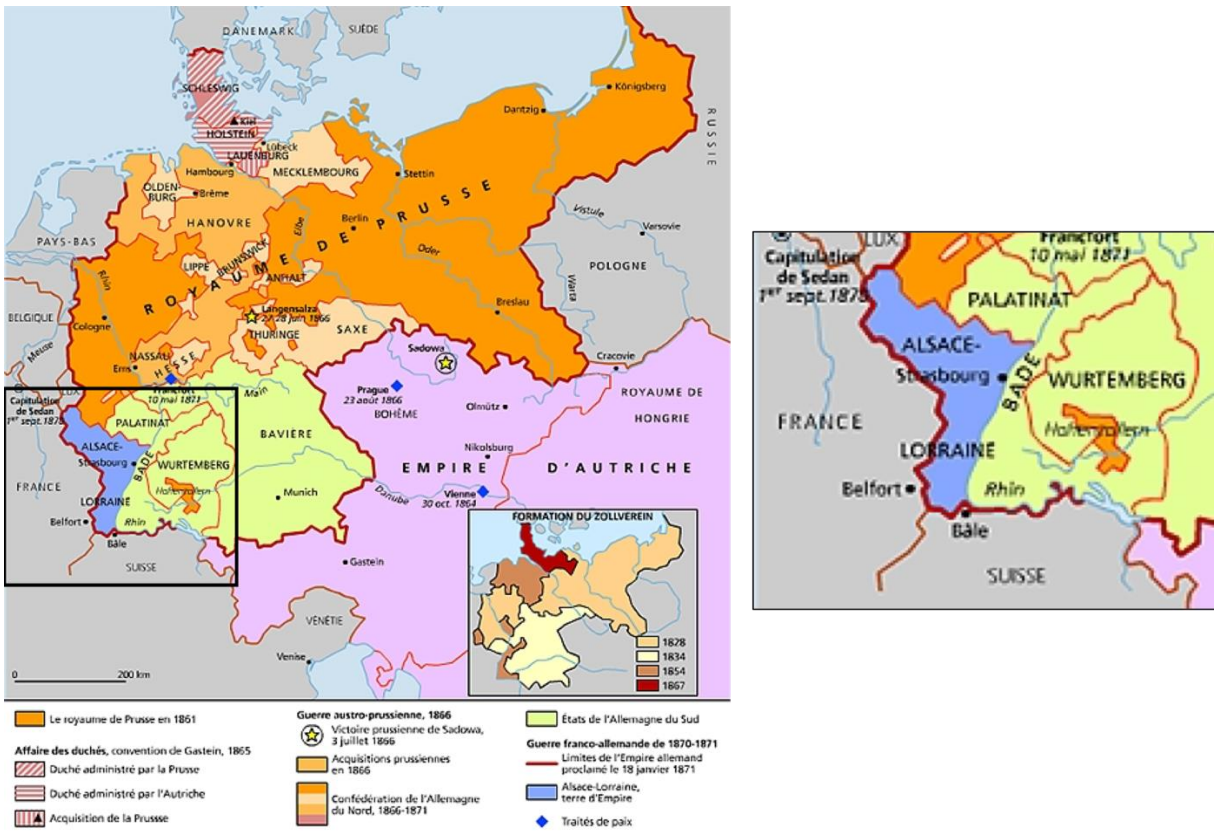


Figure 3. 3 : L'unité Allemande en 1871 inclus l'Alsace et la Lorraine (Encyclopédie Larousse)

Figure 3. 4 : La frontière franco-allemande en 1871 (Brun, 2015)

Le traité de Versailles (28 juin 1919) rend finalement l'Alsace, la Lorraine et la Moselle à l'État français. À peine plus d'un demi-siècle suffit donc pour observer un nouveau déplacement spatial de la frontière qui néanmoins retrouve son empreinte au moment de la Révolution Française, le lit du Rhin. Le maintien de ce tracé est d'autant plus éphémère que la Seconde Guerre mondiale, suivie de l'armistice du 24 juin 1940, marque de début de l'occupation, redessinant une nouvelle fois les limites de l'État français d'une part, du III^{e} Reich d'autre part (Figure 3. 5). L'Alsace, la Lorraine et la Moselle sont, contrairement au reste du territoire français, annexées directement au Reich. Une frontière étatique sépare donc les départements du Haut-Rhin, Bas-Rhin ainsi d'une partie de la Moselle, de la France. Cette dernière est notamment matérialisée dans l'espace par un cordon douanier. Par ailleurs, la littérature mentionne des pressions assimilationnistes (Brun, 2015). Le Reich aurait pour objectif la destruction des traces de double cultures (langue, germanisation de noms, mise en place du système éducatif Nazi, etc.). Il est important de souligner que la démarche assimilatrice en Alsace est engagée quelle que soit la domination (Français ou Nazi). Au-delà du mouvement spatial observé, c'est le transfert des systèmes (politiques, administratifs, éducatifs, etc.) qui est mis en évidence ici. Le déplacement spatial de la frontière induit, et rend légitime l'application du système mis en place par les Nazis. L'assimilation spatiale de ces départements, ainsi que l'histoire (ces derniers ayant déjà été rattachés chronologiquement au Saint-Empire, puis au second Reich) justifie même l'utilisation de la force pour imposer une autre façon de fonctionner. Ce point est particulièrement intéressant au regard de l'influence culturelle impactant le développement urbain. Nous y reviendrons dans la conclusion de cette section.

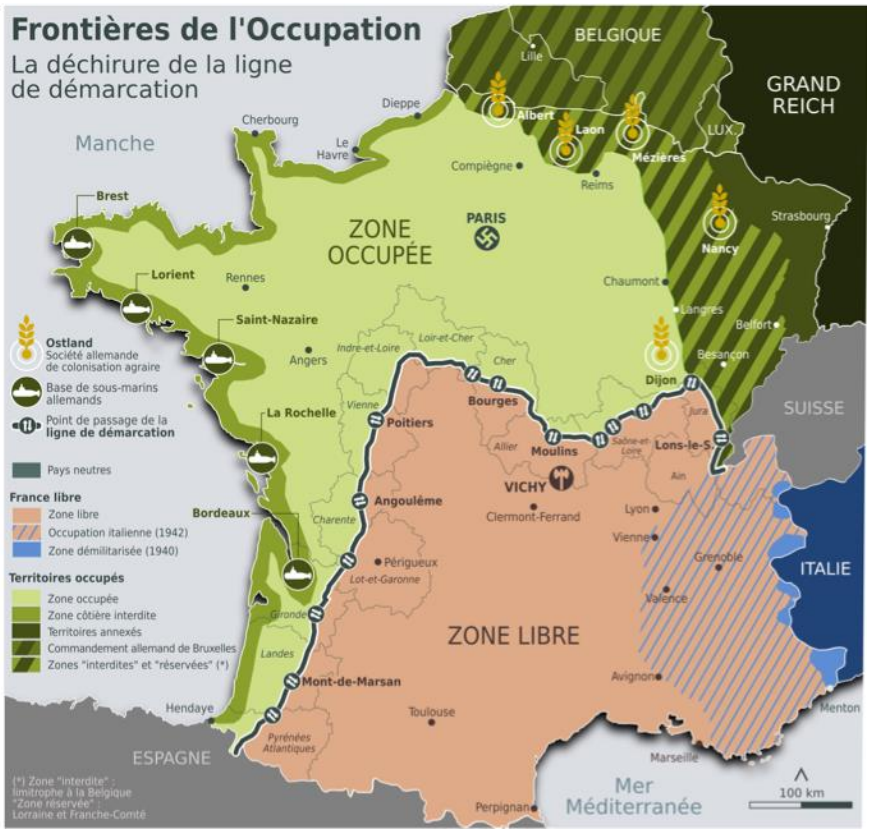


Figure 3. 5 : La France occupée (Zajec & Antoni, 2017)

Le débarquement des Alliés annonce la fin imminente de la domination du Nazi sur les territoires annexés notamment en raison du soulèvement des populations locales qui y est associé. Les frontières telles qu'elles furent dessinées avant l'arrivée d'Hitler à la tête du Reich sont rétablies. La situation s'inverse néanmoins dans une certaine mesure, puisque le Bade Wurtemberg est à son tour sous occupation française. La différence réside dans le statut de ce *Land* allemand qui n'est pas annexé. Le tracé de la frontière égalant celui précédent l'armistice de 1940, atteste de cette situation (Figure 3. 6).

Figure 3. 6 : Retour au tracé de la frontière précédent la Seconde Guerre Mondiale (Source : Encyclopédie Larousse)

La France, comme le Royaume-Uni et les Etats-Unis, assume son rôle en suivant la politique des alliés, notamment sur la partie allemande de notre zone d'étude : « destruction complète de l'armée et extirpation de l'esprit militariste ; dissolution de toutes les organisations hitlériennes et poursuites contre les principaux chefs nazis ; contrôle de l'éducation ; réorganisation complète du système judiciaire dans un esprit démocratique » (Brun, 2015). Cette période, cependant, n'impacte pas de façon aussi durable la zone du Bade-Wurtemberg où l'objectif principal n'est plus d'assimiler le voisin. Au contraire de l'annexion au Reich dont les traces sont toujours visibles aujourd'hui dans l'occupation du sol à l'instar de la Neustadt de Strasbourg. Ce quartier de la ville a vu le jour dans les années 1880, sous autorité allemande. Sa construction, qui triplera la surface de la ville, a plusieurs objectifs. Outres l'augmentation du nombre de logement nécessaire avec une augmentation de l'immigration allemande en raison de l'annexion, « il s'agit d'abord d'incarner dans la pierre le changement de régime et l'avènement du nouveau pouvoir » (Strasbourg Eurométropole, 2019). L'ascendant culturel est donc symbolisé dans le développement urbain. Ce quartier est classé depuis 2017 au patrimoine de l'Organisation des Nations unies pour l'Éducation, la Science et la Culture (UNESCO).

La description de l'évolution du tracé de la frontière permet plusieurs constatations. Dans un premier temps, que le type de frontière en lui-même, ne change pas durant cette période. En effet, qu'elle soit positionnée à la limite Ouest des départements, ou suivant le cours du Rhin, cette dernière permet de marquer fortement la séparation avec « l'ennemi ». Elle agit comme une barrière, dont le franchissement est plus ou moins périlleux, dépendamment de

la période ou de la personne réalisant le passage. Il est important de souligner que chaque changement de rattachement, induit une tentative d'effacement de la présence précédente. Cet espace devient donc un symbole qui transcende la simple extension du territoire d'un état. Il devient le symbole de la différence qui existe vis-à-vis de « l'ennemi » localisé de l'autre côté de la frontière, voire le symbole de la suprématie par l'annihilation de la culture antérieure. Il n'est pas envisageable de laisser des traces rappelant que cet « ennemi » avait à un moment, l'ascendant sur ce territoire.

Cette description de la mobilité du tracé de la frontière Franco-Allemande, permet d'interroger l'ascendant culturel présent sur ces territoires. En effet, l'histoire montre que le tracé a été mobile. Elle montre aussi que lors des changements successifs d'État référent, les pouvoirs locaux ont entrepris d'effacer l'héritage de la précédente appartenance, sans y parvenir complètement quelle que soit la situation. Des réminiscences de leur présence perdurent néanmoins. Pour exemple, une partie des bâtiments de l'actuel centre de Strasbourg ont été construits pendant l'occupation Allemande. La présence de cet héritage de l'histoire lié à la mobilité du tracé frontalier, permet un questionnement sur l'hétérogénéité inhérente à la présence d'une frontière étatique, puisque la frontière n'a pas toujours été localisée dans le lit du Rhin, et que les traces de cette mobilité s'observent encore de nos jours, notamment à travers l'urbanisme et le développement urbain initié à l'époque de l'occupations. C'est alors le rôle de la frontière en tant que marqueur d'une différence culturelle observable à travers le prisme du développement urbain qui est questionné ici.

1.2. Des relations en évolution

La sous-section précédente a permis d'insister sur l'aspect mobile du tracé de la frontière franco-allemande sur les deux derniers siècles. Outre le changement de pouvoir institutionnel, allant de pair avec l'annexion d'un espace, c'est bien l'héritage de ces périodes, observable à travers le développement urbain, qui est souligné. Cette sous-section a pour objectif de revenir sur l'aspect symbolique de la frontière franco-allemande. En effet, si la frontière a été mobile, et est fixée dans le lit du Rhin depuis l'armistice de 1944, son statut et le symbole qui lui est associé a évolué (Chapitre 1 section 3). Dans un premier temps symbole de la différence entre les deux pays, elle devient progressivement le symbole de la coopération de deux des grandes puissances européennes.

La seconde partie du 20ème siècle, permet d'observer une évolution dans la perception du voisin. Auparavant vu comme un ennemi hostile, nécessitant la mise en place d'une barrière de protection entre les deux pays, il est dorénavant considéré comme un voisin « paisible » avec lequel la cohabitation[24] est possible. Les relations sont cordiales, cependant aucun effort particulier n'initie un rapprochement ou un échange. La réminiscence des conflits

[24] Le terme cohabitation est envisagé comme habiter dans une certaine proximité, mais n'inclut pas l'aspect de coopération qui sera discuté par la suite.

antérieurs peut justifier ce positionnement. Pour autant, le besoin d'éviter un nouveau conflit et de focaliser l'énergie disponible sur la reconstruction physique et psychologique est un intérêt commun des pouvoirs en place. Toutefois les fonctionnements étatiques de part et d'autre peuvent aussi être considérés comme un facteur limitant les échanges, notamment dans le cas français.

La France est un État centralisé regroupant les institutions dans la Capitale. La période de transition entre le conflit et la coopération, caractérisée par ce que nous appellerons ici « l'indifférence », peut donc trouver une explication spatiale à travers l'éloignement des régions frontalières et du pouvoir central. Ce dernier est renforcé par une performance plus limitée, des réseaux de transports. Par ailleurs, la zone frontalière ne fait plus l'objet d'une attention particulière puisque le conflit et les tensions préalables entre les deux pays se sont apaisés avec la fin des revendications territoriales. L'indifférence peut aussi trouver sa source dans le besoin « d'oublier » le conflit passé.

Cette phase de transition amène une évolution du statut de la zone frontalière, passant d'une zone de conflit à une zone de coopération. Auparavant symbole des différences entre la France et l'Allemagne, elle devient le symbole de la mise en commun des forces de productions avec la création de la Communauté européenne du charbon et de l'acier[25] dans un premier temps, puis de l'Union Européenne. On soulignera ici le choix de la localisation des institutions européennes. Strasbourg est considérée comme une capitale européenne puisqu'elle accueille le Conseil de l'Europe ainsi que la Cour Européenne des Droits de l'Homme. Le symbole est d'autant plus fort au regard de l'histoire entre la France est l'Allemagne dans cette zone. L'aspect symbolique, tout comme le statut de la frontière est donc en cours d'évolution. La France et l'Allemagne, en tant que piliers, membres fondateurs, doivent montrer l'exemple. La zone transfrontalière est donc particulièrement intéressante puisqu'elle traduit l'état des relations entre ces pays. Les différents gouvernements portent dorénavant un intérêt particulier à cet espace. Il se traduit notamment par la mise en place de projets de coopérations suivant des programmes et des financements favorisant l'intégration de ces espaces. La frontière, à ce titre, est dévaluée en tant que limite, non seulement en tant que barrière symbolique, mais également dans son appréhension physique puisqu'elle est progressivement dématérialisée. Elle incarne, dans ce contexte, une ressource politique mais aussi socio-économique.

La question de l'influence de la frontière contemporaine se pose alors, et plus particulièrement la façon dont elle peut être mesurée. La zone d'étude a été choisie pour l'intérêt particulier qu'elle représente vis-à-vis de l'objectif de cette recherche. En effet, le questionnement sous-jacent nous pousse à nous interroger sur l'influence de la frontière,

[25] La Communauté européenne du charbon et de l'acier (CECA) : organisation internationale fondée sur le traité de Paris (1951) entré en vigueur le 23 juillet 1952 pour une durée de 50 ans. L'initiative de Robert Schuman était de rendre la guerre « non seulement impensable mais aussi matériellement impossible » en instituant un marché unique du charbon et de l'acier, entre les principaux protagonistes européens (Benelux, l'Italie, l'Allemagne de l'Ouest et la France). Ces pays sont alors liés économiquement.

non pas en tant qu'objet, mais en tant que délimitation étatique, séparant des systèmes institutionnels et politiques différents, mais aussi et en premier lieux, des individus dont les cultures et racines diverges en partie. La frontière Franco-Allemande présente une richesse particulière puisqu'elle permet d'étudier le phénomène d'urbanisation, sur une zone transfrontalière pouvant être considérée comme étant simultanément homogène, à plusieurs niveaux, et hétérogène, avec bon nombre de divergences.

Le cadre frontalier Franco-Allemand offre la possibilité d'observer et de caractériser les processus spatiaux sur une zone où les différences culturelles de part et d'autre de la frontière ne font aucun doute (langue officielle, culture, etc.). De plus, les systèmes institutionnels sont organisés différemment, que ce soit à l'échelle nationale (État centralisé vs État fédéral) ou locale (temps d'un mandat, pouvoir et responsabilités associées aux fonctions, etc.) induisant des différences de fonctionnement et de temporalité. La zone représente, qui plus est, particulièrement bien les différences de hiérarchies urbaines. Du côté français, Strasbourg figure comme le pôle urbain principal alors que du côté allemand, un ensemble de pôles secondaires structurent l'espace. La définition des pôles urbains est explicitée ultérieurement (Section 3.1).

L'hétérogénéité (politique, institutionnelle, culturelle, etc.) de cet espace frontalier est pourtant atténuée par la mise en place de projets de coopérations transfrontalières, de zonages transfrontaliers et par des prises de décisions communes. Ces initiatives sont soutenues par l'évolution d'un cadre réglementaire européen incitatif. Les différences historiques liées à l'instabilité du tracé de la frontière sont également, à présent, atténués.

Le choix de la zone d'étude franco-allemande autour de Strasbourg-Kehl a également été motivé du fait de sa relative homogénéité. En effet, cette zone se démarque notamment des zones franco-luxembourgeoise et franco-suisse, par l'absence de différentiels économiques majeurs. Au demeurant, les poids démographiques de part et d'autre de la frontière sont relativement similaires (Section 3). La zone d'étude constitue également une zone de contact entre deux des puissances européennes majeures, à la base de la création de l'Union Européenne. Si la culture diffère de part et d'autre de la frontière, cette zone en particulier a vu, à travers l'histoire, sa frontière se déplacer plusieurs fois. Comme le rappel la Mission Opérationnelle Transfrontalière[26] : « La coopération transfrontalière dans la région revêt une importance symbolique particulière. Dans l'histoire européenne, l'espace Rhénan a été une zone âprement disputée. La région du Rhin supérieur est aussi emblématique de la coopération transfrontalière et plus largement du développement de l'Europe depuis la Seconde Guerre mondiale » (MOT, 2015).

[26] La Mission Opérationnelle Transfrontalière (MOT) est une association créée en 1997 par le gouvernement français à l'initiative de la DATAR. Elle regroupe des partenaires institutionnels (Commissariat Général à l'Égalité des Territoires, etc.) ainsi que des acteurs locaux (Intercollectivités transfrontalières, communes, etc.). Les missions de la MOT : i - veiller aux intérêts des territoires transfrontaliers, ii - assister les porteurs de projets, iii - mettre en réseau les acteurs et les expériences. L'actuel maire de Strasbourg, Roland Ries en fut le premier président.

Cette zone frontalière est également singulière au regard de l'évolution des relations entre la France et l'Allemagne. Au sein de cet espace, la perception de la frontière à changer au cours du temps, passant d'une barrière « hermétique » à une barrière « poreuse ». Elle présente une symbolique antinomique, à la fois symbole d'un conflit séculaire et d'une réconciliation à l'origine de la création de l'Union Européenne. L'ensemble de ces points rend la zone d'étude particulièrement intéressante. Si la frontière marque spatialement une différence, sa mobilité spatiale historique associée à l'évolution de sa perception par les individus et les institutions rendent le questionnement sur l'homogénéité de la zone d'autant plus pertinente.

1.3. Fonctionnement et coopération récente

Comme mentionné précédemment la France présente un pouvoir central localisé dans la capitale, contrairement au fonctionnement fédéral allemand. De même, les fonctionnements institutionnels diffèrent. De ce fait, les deux territoires frontaliers présentent des hiérarchies différentes dans leurs pouvoirs décisionnaires, influençant notamment l'aménagement des territoires. Cette sous-section s'attache à présenter le fonctionnement français puis le fonctionnement allemand, avant de rebondir sur les enjeux de part et d'autre de la frontière soutenant l'élaboration de projets de coopération transfrontalière.

Planification spatiale en France

En France, la planification spatiale ou l'aménagement du territoire est historiquement une compétence d'État qui a été régionalisée dans les années 80. Les lois de décentralisations (1982, 1983) imputent aux Régions la conception des plans État-région renommés en 2007 « contrats de projets État-Région » (CPER). L'État est alors associé aux régions pour déterminer le financement des projets portant sur le développement économique, ou l'aménagement, de leur espace. La dimension nationale s'atténue mais reste présente en tant que cadre (législatif, financier, politique) dans le processus de négociation et de financement (Santamaria, 2016). Ainsi, les différents documents d'aménagement se doivent de respecter la loi telle qu'elle est présentée dans le Code de l'Urbanisme entre autres. Si les lois de décentralisation accordent plus de pouvoir et de flexibilité localement, l'État reste néanmoins présent puisqu'il apporte des financements sous forme de dotations, ou de subventions aux collectivités, pour la réalisation de certains projets d'aménagement. Il relaie également la politique de l'Union Européenne en matière d'aménagement, ou d'objectifs de développement spatial, et par son intervention dans le pilotage des fonds européens. Les décisions d'aménagements relèvent donc d'une action partagée entre l'État et les Régions, tout en étant encadré légalement et cofinancées par l'État.

Différents documents d'aménagement sont élaborés en fonction de l'échelle spatiale considérée. Ces derniers peuvent présenter une portée prescriptive ou non. Il est néanmoins important de souligner que la plupart des documents ne présentant pas une opposabilité directe, ils le sont indirectement puisque le document, opposable à l'échelle inférieure, est

tenu de prendre en compte les objectifs dudit document. C'est le cas du Schéma Régional d'Aménagement, de Développement Durable et d'Égalité des Territoires (SRADDET) dont les objectifs doivent être pris en compte dans le ScoT[27] ou le PLU[28] (en l'absence de ScoT). Il est donc possible de conclure de l'existence d'un certain impact des échelons supérieurs sur l'échelon local, temporisant de ce fait la décentralisation effective initiée au début des années 1980. Les principaux plans d'aménagement et schémas directeurs sont répertoriés dans la Figure 3. 7, suivant leur échelle d'élaboration.

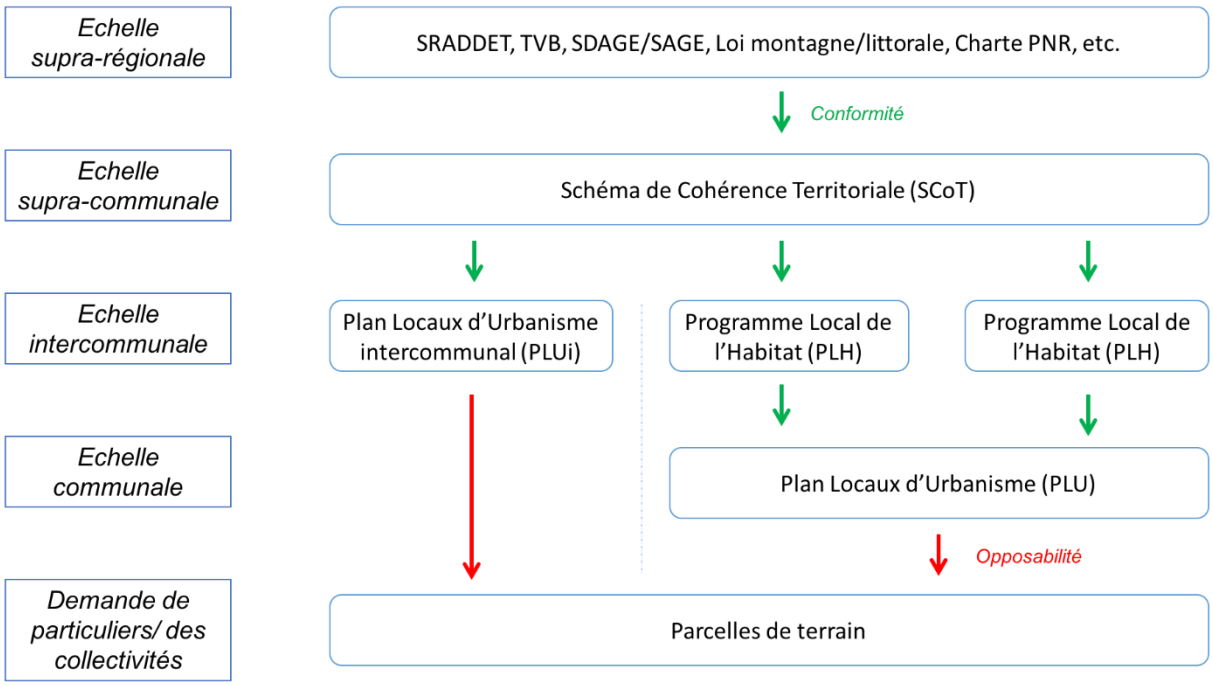


Figure 3. 7 : : Hiérarchie des documents liés à l'aménagement du territoire en France (Sénat, 2018)(d'après Sénat, 2018)

L'aménagement du territoire s'appuie sur des principes comme la recherche d'ordre et d'équilibre (Antoni, 2016; Merlin, 2007), notamment après l'intégration aux différents schémas directeurs des prescriptions de la loi Grenelle II en matière d'environnement. Cette recherche d'équilibre est historiquement basée en France sur un volontarisme politique, « idée qu'il est possible de changer l'organisation de l'espace à partir d'une action humaine » (Santamaria, 2016). Ainsi l'article L.110 du Code de l'Urbanisme précise : « Le territoire français est le patrimoine commun de la nation. Chaque collectivité publique en

[27] Le Schéma de COhérence Territoriale (SCOT) est « l'outil de conception et de mise en œuvre d'une planification stratégique intercommunale, à l'échelle d'un large bassin de vie ou d'une aire urbaine, dans le cadre d'un Projet d'Aménagement et de Développement Durables (PADD). Il est destiné à servir de cadre de référence pour les différentes politiques sectorielles, notamment celles centrées sur les questions d'organisation de l'espace et d'urbanisme, d'habitat, de mobilité, d'aménagement commercial, d'environnement.... Il en assure la cohérence tout comme il assure la cohérence des documents sectoriels intercommunaux : Plans Locaux d'Urbanisme Intercommunaux (PLUi), Programmes Locaux de l'Habitat (PLH), Plans de Déplacements Urbains (PDU), et des Plans Locaux d'Urbanisme (PLU) ou des cartes communales établis au niveau communal. » (Ministère de la Cohésion des Territoires, 2018)

[28] Le Plan Local d'Urbanisme (PLU) est « un document d'urbanisme établissant un projet global d'urbanisme et d'aménagement. » (Ministère de la Cohésion des Territoires, 2018)

est le gestionnaire et le garant ». D'autres pays européens partagent cet objectif, notamment le pays frontalier qu'est l'Allemagne.

Planification spatiale en Allemagne

Après avoir été hiérarchisé du niveau national à celui des municipalités sous le régime National Socialiste, la planification en Allemagne fût différente dans chacun des deux états allemands lors de la séparation Est-Ouest. C'est dans les années 60 qu'un système de planification établi à l'échelle nationale est institué (Reimer *et al.*, 2014). Ce dernier donne un pouvoir important aux municipalités et aux *Länder* (division administrative équivalente aux Régions en France).

Avant de présenter succinctement l'organisation du système de planification en Allemagne, il est important de rappeler que la structure administrative du pays est composée de 16 *Länder* ayant leur propre constitution, leur propre parlement élu ainsi que leur propre gouvernement. La planification est donc basée sur une organisation décentralisée et multi-niveaux, selon une structure fédérale. L'objectif de la planification en Allemagne est « un développement spatial durable, qui rapporte les besoins économiques et sociaux sur une zone dans le respect de sa fonction écologique et, considérant une échelle plus large, qui résulte d'un équilibre stable donnant des conditions de vie équivalentes dans les sous-espaces » ADDIN ZOTERO_ITEM CSL_CITATION {"citationID":"ui0h76ZN","properties":{"formattedCitation":"(Scholl {\\i{}et al.}, 2007)","plainCitation":"(Scholl et al., 2007)","noteIndex":0},"citationItems":[{"id":26,"uris":["http://zotero.org/users/3163049/items/Q2C2T4UH"],"uri":["http://zotero.org/users/3163049/items/Q2C2T4UH"],"itemData":{"id":26,"type":"article-journal","title":"Raumplanung in Deutschland - Formeller Aufbau und zukünftige Aufgaben","container-title":"Schriftenreihe des Instituts für Städtebau und Landesplanung","volume":"35","author":[{"family":"Scholl","given":"Bernd"},{"family":"Elgendy","given":"Hany"},{"family":"Nollert","given":"Markus"}],"issued":{"date-parts":[["2007"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"} (Scholl *et al.*, 2007)(Federal spatial planning act, 2008). Les instruments fondamentaux utilisés sont des « plans » montrant une structure spatiale prospective, que ce soit par le biais de cartographies, ou de descriptions textuelles. Remier (2014) mentionne trois principes fondamentaux structurant de façon hiérarchique le système de planification (Scholl *et al.*, 2007) :

1. Le principe de subsidiarité : chaque décision politique doit être réalisée à l'échelon politique le plus bas possible.
2. Le principe d'autonomie local ou municipale concernant la planification : les municipalités ont le droit de structurer de façon indépendante leur développement local dans le cadre de la planification de l'occupation du sol en accord avec les lignes directrices établies au niveau supérieur.
3. Le principe de retours mutuels : les divers niveaux hiérarchiques de planification doivent prendre en compte les conditions et objectifs des autres niveaux. La

municipalité doit pouvoir s'inscrire dans l'unité supérieure (*Land*) mais cette dernière doit considérer les demandes et objectifs des individus la composant.

Le système de planification en Allemagne se structure comme présenté sur la Figure 3. 8.

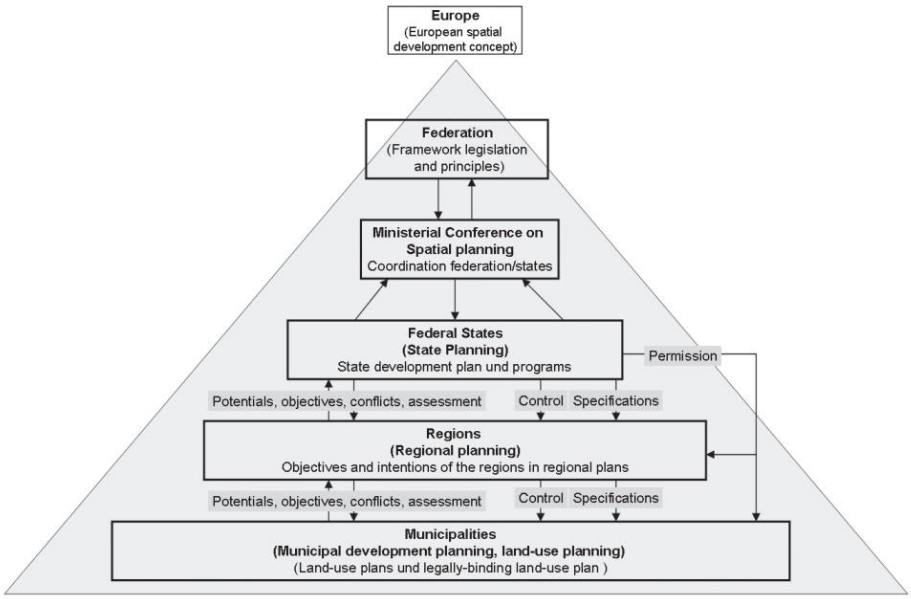


Figure 3. 8 : Système de planification en Allemagne (Scholl et al., 2007)

Dans ce système, chaque niveau de planification peut établir un plan pour son espace, excepté le niveau fédéral qui lui ne peut qu'établir des lignes directrices. L'État Fédéral et les différents états fédérés introduisent des programmes de développement et peuvent soutenir financièrement certains axes de développement dans l'objectif de les promouvoir. B. Scholl soulignent que l'objectif suprême du système de planification allemand est l'aspect durable de ce dernier sur la totalité de son espace (Scholl *et al.*, 2007). L'aspect écologique est donc pris en considération au même titre que les besoins économiques et sociaux. Ce premier objectif est mis en parallèle avec celui d'atteindre des niveaux de vies comparables dans les différentes régions et les différents états fédérés. L'état fédéral est donc en charge de l'élaboration la plus abstraite du plan de développement puisque les lignes directrices établies doivent s'appliquer à l'ensemble des Länder. Ce dernier doit cependant intégrer les objectifs visés par la politique européenne de développement spatial. Les plans établis n'ont pas de portée réglementaire. L'état fédéré doit préparer la stratégie de développement spatial sur son espace, le « plan de développement étatique ». Dans les cadres établis par l'État Fédéral, chaque État fédéré peut établir son plan de développement. C'est au niveau régional qu'incombe la coordination supra-locale du développement urbain. Cette dernière doit être faite en considérant les objectifs et principes exprimés dans le plan établi par l'État fédéré dont la région dépend. La région en collaboration avec ses municipalités doit, à son tour, établir un plan de développement. Elle travaille tout autant avec l'État fédéré dont elle dépend pour évaluer la faisabilité d'un projet de développement au regard des critères de développements présents dans les deux plans, y compris les critères environnementaux.

En Allemagne, la planification du développement spatial est donc régulée par les lois des États fédérés. Elle peut donc varier d'un État à l'autre et présenter des différences. Pour finir, la municipalité est en charge de sauvegarder l'équité à l'échelle locale, et d'assurer l'aspect durable du développement à cette même échelle. Elles sont, comme en France, responsables de la rédaction de plans de développement urbain, en accord avec la réglementation régionale dont elles dépendent (équivalent des PLU français).

Enjeux communs et coopération

Alors que l'organisation et la hiérarchie des institutions entre La France et l'Allemagne sont différentes, les enjeux en termes d'aménagement et de développement des territoires sont identiques. Comme plusieurs pays Européen, France et Allemagne font face, au 21ème siècle, à des changements démographiques importants, qui ne sont pas répartis de façon homogène sur leurs territoires[29]. Le changement s'opère aussi par des comportements individuels influençant les politiques institutionnelles. La France présente, par exemple, une évolution de la surface moyenne des logements individuels qui tend à diminuer alors que le nombre de logements est en augmentation. Ce développement urbain traduit la recherche d'une amélioration du cadre de vie en France. Ces constatations font écho aux objectifs fixés par l'État Fédéral d'homogénéiser son espace, notamment la qualité de vie au sein des différents Länder. Ces nouveaux objectifs impactent le développement spatial, et plus particulièrement le développement urbain.

La recherche d'équilibre concerne aussi celui qui doit exister entre les activités humaines et les besoins naturels au sein d'un espace en termes environnementaux. L'intégration de mesures imposant la prise en compte de l'environnement dans l'exploitation de l'espace est introduite de part et d'autre de la frontière (Loi Grenelle en France, Loi sur la protection de la nature en Allemagne - *Bundesnaturschutzgesetz*). La dimension environnementale, intégrée dans la recherche d'équilibre dans l'exploitation des ressources d'un espace, rend l'étude des zones frontières indispensable. Les limites géographiques des éléments naturels tendent à dépasser les frontières, même si ces derniers sont utilisés pour les matérialiser. L'exemple du Rhin est particulièrement représentatif. Le fleuve sépare la France et l'Allemagne, faisant office de frontière étatique, mais son bassin versant s'étend de part et d'autre. Une action environnementale à un point de ce bassin pourra avoir des répercussions globales à l'échelle de ce dernier, donc au-delà de la frontière étatique (poldérisation, action sur les zones humides, mise en place d'aménagements hydrauliques, etc.). Ce constat souligne l'intérêt du suivi et de l'analyse des évolutions spatiales à l'échelle des zones transfrontalières et de la coopération dans la gestion de ces dernières.

[29] Par exemple le taux de fécondité en Allemagne est de 1.5 en 2015 contre 2,01 pour la France (en dessous ou à peine au niveau du seuil de renouvellement des générations en 2015), perte d'attractivité des régions industrielles, flux migratoires internes, etc.

Cette description du fonctionnement et des objectifs en termes de planification spatiale des États français et allemand nous amène à nous interroger sur les objectifs en termes d'occupation du sol au niveau des frontières. En effet, si les plans d'aménagement sont la plupart du temps limités par des frontières administratives, les problèmes d'aménagement, eux, s'étendent au-delà. L'effet dit « frontière » que l'on peut observer aux limites d'un espace, notamment dans le cadre de la modélisation de processus dynamiques, prend tout son sens dans ce contexte puisqu'il est l'image d'un effet de voisinage dépassant les limites administratives qu'elles soient municipales, régionales ou même étatiques. Si l'objectif des États français et allemand est une planification durable sur leur territoire, comment atteindre cet objectif aux frontières ? La discussion de tels problèmes nécessite la coopération des différents acteurs, au-delà de leurs limites d'influence administrative, mais surtout une analyse spatiale transfrontalière visant à caractériser les processus œuvrant sur ces zones.

La création de l'Union Européenne, la proximité frontalière et le symbole que peuvent représenter les zones frontalières ont favorisé l'émergence de coopérations transfrontalières. La zone transfrontalière franco-allemande de Strasbourg-Kehl révèle plusieurs zonages présentant différents organes de coopération, à différentes échelles (Figure 3. 9).

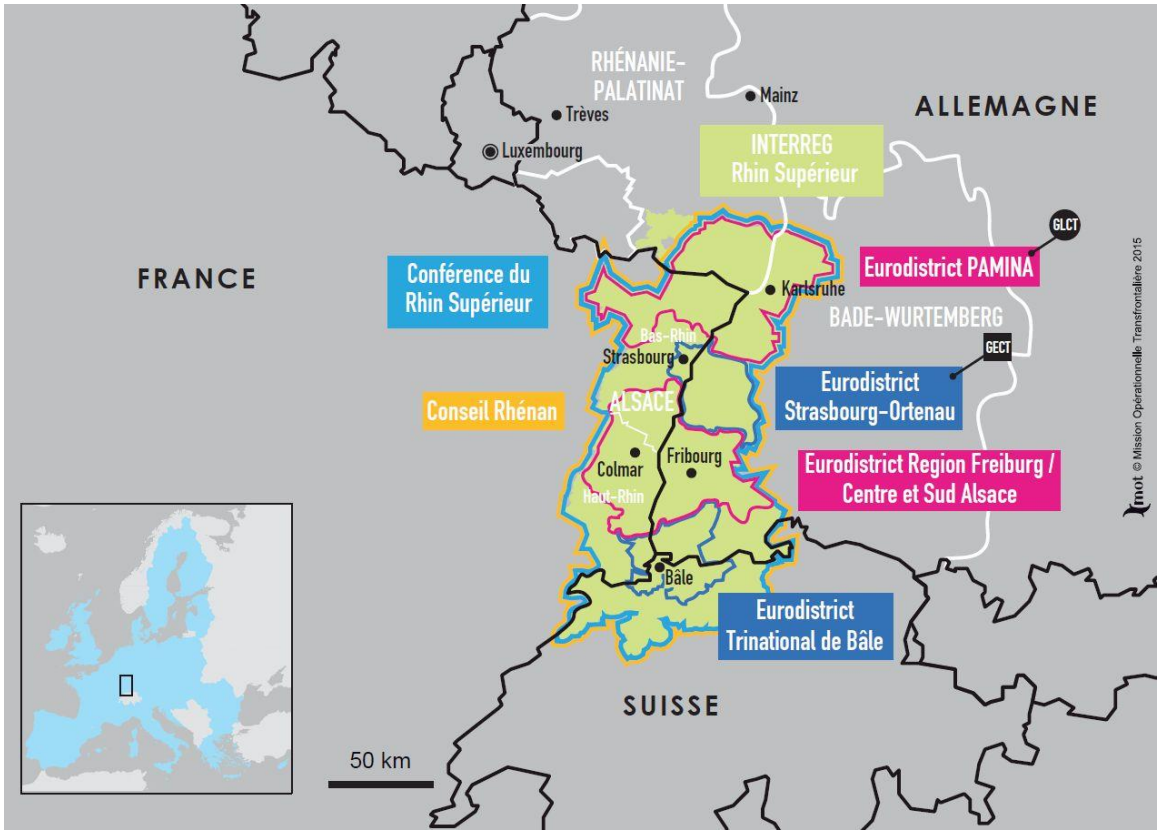


Figure 3. 9 : Projets de territoires et organes institutionnels de coopération (MOT, 2015)

Les différents zonages sont matérialisés suivant un code couleur. Le zonage concerné par le programme INTERREG (Rhin Supérieur – Oberrhein) est le plus vaste (vert) suivit par ceux, identique, du Conseil Rhénan (jaune) et de la Conférence du Rhin Supérieur (bleu clair).

Deux types d'Eurodistricts apparaissent : les Eurodistricts - Groupement Européen de Coopération Territoriale, les Eurodistricts - Groupement Local de Coopération Transfrontalière (Figure 3. 9). Ainsi, la frontière franco-allemande au niveau de Strasbourg fait partie du Conseil Rhénan et de la Conférence du Rhin Supérieur, mais aussi de l'Eurodistrict de Strasbourg-Ortenau. Elle est pour finir concernée par le programme INTERREG Rhin Supérieur – Oberrhein, qui permet de financer des projets transfrontaliers grâce au soutien du Fonds Européen de DEveloppement Régional (FEDER).

Pour commencer avec le zonage le plus étendu, concernant la zone d'étude, la Conférence du Rhin Supérieur a été créé dans les années 1970. Cette commission a pour objectif de réunir les représentants des États allemand, français et suisse autour des thématiques de coopérations transfrontalières. Leur rôle est « d'étudier et de faciliter les échanges transfrontaliers et la vie quotidienne des habitants » (MOT, 2015). Les champs d'actions concernés sont notamment : les transports, l'environnement, la culture, l'éducation et la formation, la jeunesse, la santé, l'aménagement du territoire, le sport, le climat et l'énergie, l'entraide en cas de catastrophe, l'agriculture, l'économie. Pour accompagner la Conférence du Rhin, le Conseil Rhénan a été créé en 1997. Ce dernier est composé d'élus ayant pour objectif d'appuyer ou d'enrichir le travail effectué par la Conférence, à un niveau plus local. Dans la limite de leur compétence, ils cherchent à développer ou approfondir les échanges transfrontaliers existant et à « contribuer au développement harmonieux et cohérent de l'espace du Rhin supérieur » (MOT, 2015).

À un niveau plus local, l'Eurodistrict joue un rôle dans l'initiation de projets transfrontaliers depuis sa création en 2005. C'est à l'occasion de la commémoration du 40ème anniversaire du traité de l'Elysée que l'idée, déjà évoquée dans les années 80, d'un Eurodistrict entre Strasbourg et Kehl a refait surface. À cette occasion, le président français Jacques Chirac et le chancelier allemand Gerhard Schröder affirment leur volonté de développer la coopération. Cette dernière s'était néanmoins déjà développée à l'échelle des municipalités de Strasbourg et de Kehl suite aux accords de Bonn signés en 1975 prévoyant entre autres une information mutuelle sur les projets d'aménagement. Le travail de concertation effectué a permis la réalisation de projets tels que :

- Les centres Infobest : centre d'information et de conseil pour les professionnels exerçant une activité transfrontalière.
- L'Euro-Institut de Strasbourg-Kehl : organisme Franco-allemand de formation continue.
- L'aménagement du Jardin des Deux Rives : Parc de 150 ha s'étendant sur chaque rive du Rhin relié par une passerelle piétonne permettant de traverser la frontière, symbole de l'amitié franco-allemande.

Dans l'esprit du traité de Bonn, à savoir améliorer la connaissance réciproque des territoires et institutions afin de permettre d'approfondir la coopération entre les acteurs, le Livre Blanc transfrontalier Strasbourg - Ortenau a été réalisé par le même syndicat mixte chargé de l'élaboration du SCoTERS (Schéma de Cohérence Territoriale pour la Région de

Strasbourg). Le livre blanc « constitue la base d'un développement cohérent et concerté pour toute la région transfrontalière » (MOT). Les partenaires sollicités dans la démarche étaient :

- L'État Français
- Le Conseil régional d'Alsace
- Le Conseil départemental du Bas-Rhin
- L'Ortenaukreis
- Le Regionalverband Südlicher Oberrhein
- Les Villes d'Offenburg, Kehl, Lahr et Achern
- L'Union européenne (par le biais du programme Interreg).

Cette liste de partenaires souligne l'implication des différentes institutions nationales, régionales, locales et européennes. Dans l'intérêt de l'Union européenne, celui de la France et de l'Allemagne, l'Eurodistrict de Strasbourg - Ortenau doit :

- Devenir au sein de l'Union Européenne un territoire-pilote en matière d'intégration approfondie dans les régions frontalières ;
- Contribuer au développement durable de la région et renforcer son rayonnement européen ;
- Améliorer le quotidien des citoyens.

Au-delà de l'aspect pratique visé à travers la coopération dans la réflexion sur l'aménagement du territoire transfrontalier cet Eurodistrict est pensé comme le symbole de la réussite de la coopération européenne à travers la réussite de l'intégration d'un territoire transfrontalier appartenant à deux puissances majeures de l'Union que sont la France et l'Allemagne, et dont les relations n'ont historiquement pas été sans conflit.

L'Eurodistrict a permis à des projets variés de voir le jour, parmi eux :

- L'agence pour l'emploi franco-allemande : lever les obstacles de l'accession à l'emploi de part et d'autre de la frontière.
- Start'Hop et Kiosk Office : favoriser la coopération des jeunes entrepreneurs de part et d'autre de la frontière.
- Eveil 3D : apprentissage de langues étrangères à l'aide de jeux virtuels en immersion pour favoriser le bilinguisme.
- Maison de la petite enfance transfrontalière : accueil des enfants français ou allemands au bénéfice de l'éducation bilingue.
- L'harmonisation des connexions TGV et ICE.
- L'extension de la ligne D du tramway jusqu'à Kehl.
- Programme universitaire franco-allemand.

La mise en place de tels projets nécessite un apport financier conséquent. L'Eurodistrict, à travers sa position transfrontalière, a pu bénéficier des fonds européens FEDER dans le cadre du programme Interreg. Ces fonds ont été mis en place en 1989 par l'Union Européenne, dans le but de financer des projets transfrontaliers aussi bien sur les frontières intérieures qu'extérieures. Le Rhin Supérieur bénéficie actuellement de la 5ème génération d'appel à projets menés durant la période 2014 – 2020 (INTERREG V). Les précédents

programmes ont permis de financer plus de 300 projets transfrontaliers, portant sur des thématiques variées, soutenant la démarche de coopération et favorisant l'intégration des territoires frontaliers à travers les initiatives transfrontalières. Le renouvellement de l'allocation des fonds témoigne de l'implication de l'UE et des États, dans la coopération transfrontalière.

Pour finir, l'UE s'implique également dans l'acquisition de données pour l'analyse des territoires européens dans un but de développement concerté et d'amélioration de la cohésion territoriale. Par exemple, le programme ESPON (European Observation Network for Territorial Development and Cohesion), cofinancé par les fonds FEDER, « propose une meilleure connaissance des territoires européens aux acteurs de l'aménagement et du développement » sous la forme de données issues de domaines variés (environnement, développement durable, économie verte, développement économique et social, emploi, flux migratoires, réfugiés, exclusion, pauvreté, développement urbain/rural, transports, etc.). Ce programme a notamment servi la réalisation du projet de recherche ULYSSE (Using applied research results from ESPON as a yardstick for cross-border spatial development planning) entre 2010 et 2011. ULYSSE, qui a d'ailleurs été développé en Alsace, permet d'explorer l'impact des politiques de développement territorial, de cohésion ou de compétitivité à l'échelle de la zone frontalière, ou encore les perspectives de développement européen. Le développement de plateformes scientifiques et d'outils concernant des indicateurs permettant l'analyse des territoires a aussi été inclus dans ce projet.

Les différents programmes présentés précédemment font figures d'outils de coopération. Par des co-financements de projets, ils permettent d'initier les collaborations entre acteurs de part et d'autre de la frontière. Les partenaires peuvent être des instituts publics, mais aussi issus du secteur privé. Si les programmes et financement ont pour objectif d'initier la coopération, la pérennisation de la démarche est visée au-delà des périodes de financements des projets.

La présentation des différents programmes de coopération avait pour objectif principal de souligner l'implication des différents acteurs (locaux ou nationaux, français ou allemands) dans la mise en place de projets favorisant les dynamiques transfrontalières. L'accroissement de la porosité de la frontière mentionnée précédemment est donc aussi le résultat de la volonté des acteurs.

Malgré une culture et un fonctionnement institutionnel différent, la présente section a mis en lumière les enjeux communs des pays frontaliers motivant l'appréhension d'une part de l'aménagement du territoire à l'échelle de l'espace frontalier. Pour ce faire, différents programmes et zonages, notamment soutenus par l'UE, officient en tant qu'outils de coopération, effaçant en partie la frontière étatique. Cette section rappelle par ailleurs les éléments historiques particuliers temporisant également l'association de la frontière étatique comme une barrière hermétique, en raison de sa mobilité spatiale. Pour autant, elle reste la limite administrative contemporaine. Ces éléments permettent de questionner l'homogénéité

de la zone d'étude et l'influence de la frontière. C'est au moyen de la modélisation et la simulation du développement urbain qu'est étudiée l'importance de la localisation géographique dans l'un ou l'autre des pays frontaliers. La section suivante présente plus précisément les données libres exploitées pour modéliser la zone transfrontalière et décrire les processus de développement urbain.

2. RÉALITÉS GÉOGRAPHIQUES

Avant de se focaliser sur les données d'occupation du sol et les traitements préalables sous-jacents à une future modélisation, la section s'attarde sur le tracé des limites de la zone franco-allemande et apporte des précisions sur les raisons de leur localisation. S'en suit la présentation des tendances nationales de changements d'occupation du sol observées dans les deux pays constituant la zone transfrontalière de Strasbourg - Kehl. Pour finir, les tendances de changements sur la zone sont étudiées d'après l'analyse des données diachroniques d'occupation du sol.

2.1. Limite

La définition des limites d'une zone d'étude est une étape importante de la modélisation en raison l'impact potentiel que cette définition peut avoir sur la représentation et l'analyse des processus spatiaux. La sous-section précédente présentait plusieurs zonages liés aux programmes de coopération, notamment celui de l'Eurodistrict de Strasbourg-Ortenau. Une approche faisant abstraction d'un découpage administratif a été privilégiée pour une focalisation sur un zonage représentant une réalité spatiale.

La définition de « l'épaisseur » de la zone transfrontalière, au regard de la localisation de son tracé, n'est pas universelle. En cherchant à fixer une distance limite qui la prendrait en compte, les critères peuvent varier d'un pays à l'autre. Par exemple, dans le cas de la convention fiscale Franco-Allemande, il est précisé que la zone frontalière englobe les communes situées dans un périmètre n'excédant pas 30 km du côté allemand alors qu'il n'est que de 20km du côté français. Ce zonage n'est pas symétrique. C'est également le cas des limites de l'Eurodistrict (Figure 3. 10). La zone française est beaucoup moins étendue que la zone allemande. Elle est par contre, proportionnellement, beaucoup plus urbanisée.

Les limites retenues dans ce travail se basent donc sur d'autres critères afin de disposer d'une délimitation plus homogène de part et d'autre de la frontière. Notre zonage est défini d'après l'accessibilité des communes, au centre urbain principale qu'est Strasbourg, qui préserve une certaine symétrie. Nous reviendrons sur la hiérarchie des différents pôles ultérieurement (Section 3.1).

Figure 3. 10 : Limites de l'Eurodistrict (Eurodistrict, 2013)

Par ailleurs, la démarche méthodologique porte sur l'utilisation de données libres, notamment celles produites par le programme européen Corine Land Cover. Le choix de travailler avec une donnée déjà produites impose de prendre en compte la précision de son acquisition initiale. Elle est en effet acquise à l'échelle européenne pour une résolution de 400 m x 400 m, discrétisé à 100 m x 100 m par la suite. Cette résolution peut être considérée comme adaptée à l'échelle européenne, tout en étant insuffisante pour une étude à l'échelle d'une ville. Le choix d'un périmètre régional est donc en partie motivée par l'utilisation des données CLC. Pour finir, nous avons fait état précédemment, d'une dynamique des changements relativement faibles en Europe en générale. Une échelle régionale permet d'observer plus de changement en termes de superficie. Ce point est abordé plus précisément par la suite.

Les critères soutenant la délimitation de la zone d'études sont donc les suivants :

- **Un espace adapté à la capture du phénomène de développement urbain.** Le phénomène étant plus limité en Europe que dans d'autre partie du monde, la prise en compte d'un espace favorise positivement le nombre de transitions urbaines observées.
- **Un espace adapté à la résolution de la donnée d'occupation du sol CLC.** La donnée CLC n'est pas adaptée à l'étude du phénomène de développement urbain à l'échelle d'une ville. Elle est, au contraire, adaptée à l'échelle régionale.

- **Un espace comprenant des points de passages de la frontière.** La présence du Rhin implique un nombre limité de points passages le long de la frontière. Sur la zone franco-allemande au niveau de Strasbourg-Kehl ils sont de différentes natures. Ainsi, le passage est rendu possible par la présence de ponts (routiers, piétons, ferroviaires) ou de bacs.
- **Un espace symétrique en termes de surfaces de part et d'autre de la frontière.** La symétrie permet d'évaluer l'homogénéité de la zone sur la base d'une superficie similaire.

Ces critères se traduisent par la prise en compte des communes présentent dans un périmètre autour du centre urbain de Strasbourg-Kehl n'excédant pas 30 minutes tout en assurant une certaine symétrie spatiale. La limite de 30 min a été choisie d'après le temps de trajet domicile travail, qui en France comme en Allemagne est en moyenne d'une heure par jour (Jenik, 2017; Zilloniz, 2015). Par ailleurs, ce choix permet néanmoins de rester à distance de l'influence du pôle de Karlsruhe au Nord et de Bâle – Mulhouse – Fribourg au Sud. Pour finir, les limites Est et Ouest englobent respectivement la Forêt Noire et le piémont des Vosges. Un gradient topographique s'opère de chaque côté de la frontière.

La délimitation de la zone d'étude est présentée Figure 3. 10 :

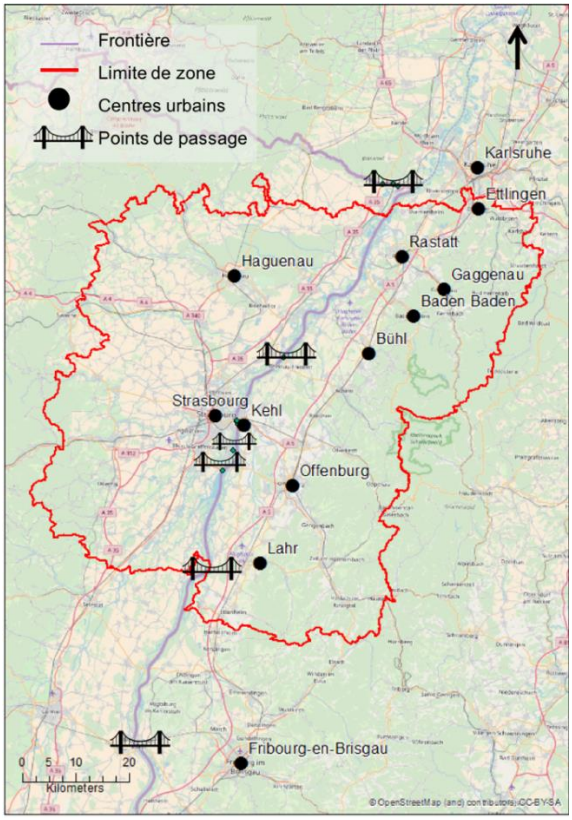


Figure 3. 11 : Limites de la zone d'étude

C'est à l'intérieur de ces limites que le processus de développement urbain est étudié dans le cadre de ce travail. Le choix de la donnée CLC nécessite une reclassification pour réduire le nombre de classe d'occupation du sol permettant la description d'un voisinage favorable à la transition urbaine. Ce point fait l'objet de la sous-section suivante.

2.2. Occupation du sol

La section 1 du chapitre 2 a permis de mettre en lumière l'importance, mais aussi la part de subjectivité liée à la réalisation des classifications. CLC a été élaboré à l'échelle de l'UE. Certaines classes ne sont donc pas pertinentes[30] pour l'étude de la frontière franco-allemande. La multiplication des classes d'occupation du sol ne favorise pas, non plus, l'élaboration et l'interprétation d'un modèle simple. Ensuite, un nombre plus important de classes, induit une plus grande diversité des changements d'occupation observables, au dépend de l'importance de chaque changement en termes de surfaces. Ce point est essentiel et sera abordé plus précisément dans la partie qui concerne la méthodologie (Chapitre 5). Pour obtenir une classification plus générale, tout en permettant d'observer le développement urbain nous présentons une nouvelle classification composée de 8 classes : urbain, agricole, forêt, vignes, zones industrielles et commerciales, zones récréatives urbaines, zones immergées, transport. Le détail de la classification initiale est, ainsi, également mis à profit. Il permet de mettre en avant des occupations du sol particulières, inhérentes au territoire comme les vignes. Cette classe présente un intérêt particulier pour caractériser l'environnement des cellules ou encore différencier les tendances entre la France et l'Allemagne. Les choix des classes fusionnées sont illustrés dans la Table 3. 1.
La classe urbaine regroupe les deux catégories de tissus urbain. La classe forêt est constituée des par la majorité des zones naturelles. Les zones agricoles regroupent l'ensemble des territoires agricoles à l'exception des vignes. Ces dernières ont été conservée comme une classe à part entière en raison de l'importance culturelle et économique liée aux exploitations viticoles dans ces régions. Elle se distingue d'une exploitation agricole plus classique. Elles présentent également une certaine valeur paysagère. Les zones industrielles et commerciales regroupent une partie des territoires artificialisés. Les transports ont été eux aussi isolé en raison de leur spécificité. Ils traduisent également une connectivité plus importante de certains espaces. Les espaces verts urbains et les équipements sportifs et de loisirs, représentent les aménités urbaines. Ils sont donc regroupés dans la classe des zones récréatives urbaines. Pour finir les zones immergées représentent les surfaces en eau.

[30] Citons en exemple les marais maritimes, marais salants, etc.

Table 3. 1 : Classification 8 classes élaborée sur la base des 44 classes Corine Land Cover (Copernicus)

Classification Corine Land Cover - 44 classes		*Re-Classification*
Territoires artificialisés - Zones urbanisées		
1	111 : Tissu urbain continu	1
2	112 : Tissu urbain discontinu	1
Territoires artificialisés - Zones industrielles ou commerciales et réseaux de communication		
3	121 : Zones industrielles et commerciales	2
4	122 : Réseaux routier et ferroviaire et espaces associés	3
5	123 : Zones portuaires	2
6	124 : Aéroports	2
Territoires artificialisés - Mines, décharges et chantiers		
7	131 : Extraction de matériaux	2
8	132 : Décharges	2
9	133 : Chantiers	1
Territoires artificialisés - Espaces verts artificialisés, non agricoles		
10	141 : Espaces verts urbains	4
11	142 : Equipements sportifs et de loisirs	4
Territoires agricoles - Terres arables		
12	211: Terres arables hors périmètres d'irrigation	5
13	212 : Périmètres irrigués en permanence	5
14	213 : Rizières	5
Territoires agricoles - Cultures permanentes		
15	221 : Vignobles	6
16	222 : Vergers et petits fruits	5
17	223 : Oliveraies	5
Territoires agricoles - Prairies		
18	231: Prairies	5
Territoires agricoles - Zones agricoles hétérogènes		
19	241 : Cultures annuelles associées aux cultures permanentes	5
20	242 : Systèmes culturaux et parcellaires complexes	5
21	243 : Surfaces essentiellement agricoles, interrompues par des espaces naturels importants	5
22	244 : Territoires agro-forestiers	5
Forêts et milieux semi-naturels - Forêts		
23	311 : Forêts de feuillus	7
24	312 : Forêts de conifères	7
25	313 : Forêts mélangées	7
Forêts et milieux semi-naturels - Milieux à végétation arbustive et/ou herbacée		
26	321 : Pelouses et pâturages naturels	7
27	322 : Landes et broussailles	7
28	323 : Végétation sclérophylle	7
29	324 : Forêt et végétation arbustive en mutation	7
Forêts et milieux semi-naturels - Espaces ouverts, sans ou avec peu de végétation		
30	331 : Plages, dunes et sable	7
31	332 : Roches nues	7
32	333 : Végétation clairsemée	7
33	334 : Zones incendiées	7
34	335 : Glaciers et neiges éternelles	7
Zones humides - Zones humides intérieures		
35	411 : Marais intérieurs	7
36	412 : Tourbières	7
Zones humides - Zones humides maritimes		
37	421 : Marais maritimes	8
38	422 : Marais salants	8
39	423 : Zones intertidales	8
Surfaces en eau - Eaux continentales		
40	511 : Cours et voies d'eau	8
41	512 : Plans d'eau	8
Surfaces en eau - Eaux maritimes		
42	521 : Lagunes littorales	8
43	522 : Estuaires	8
44	523 : Mers et océans	8

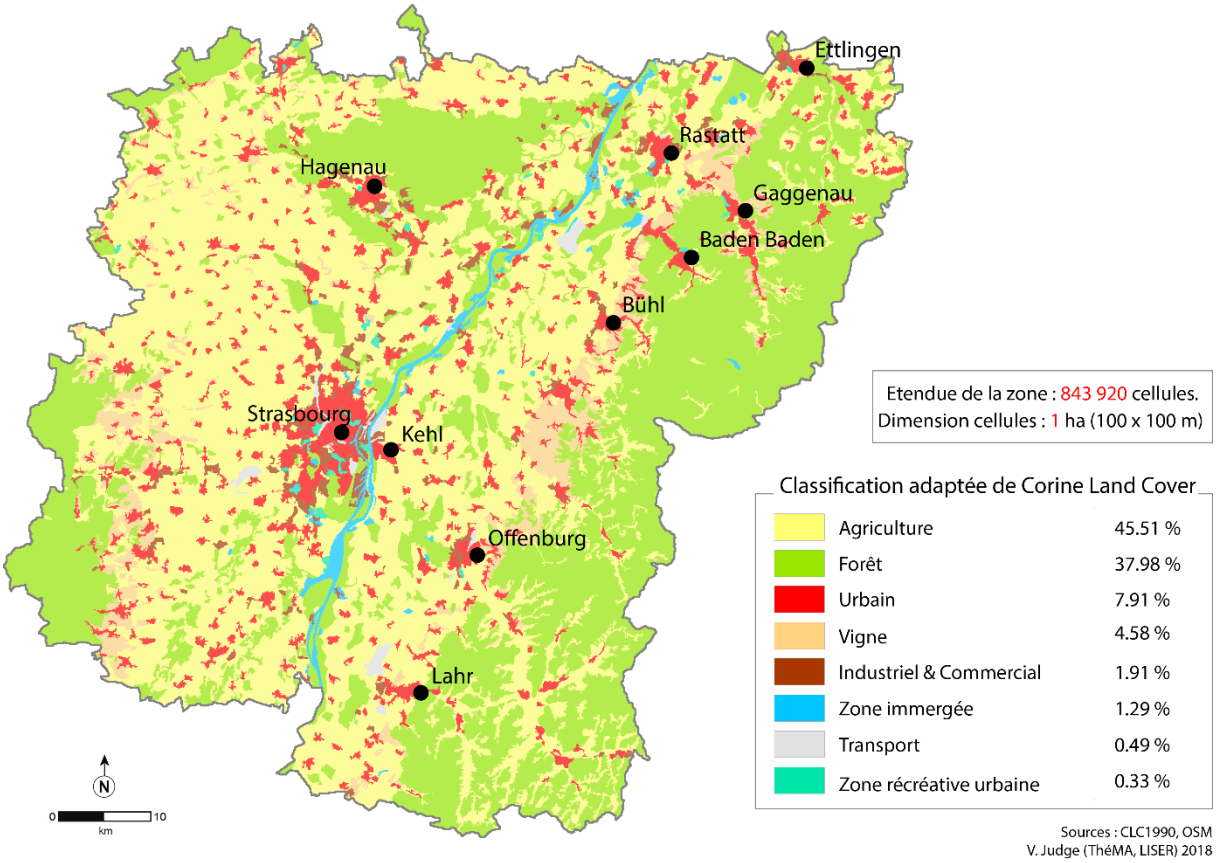


Figure 3. 12 : Occupation du sol en 1990 après fusion des classes

La Figure 3. 12 présente la nouvelle classification accompagnée des proportions de chaque type d'occupation sur la zone d'étude en 1990. Elle permet d'observer une relative homogénéité de la répartition des différents types d'occupation du sol de part et d'autre de la frontière. Une légère différence émerge au niveau de la superficie recouverte par la forêt qui semble être un peu plus conséquente en Allemagne, et agricole qui semble être, elle, plus conséquente en France (Figure 3. 13). C'est ce qu'illustre le graphique ci-dessous présentant plus en détails les proportions d'occupation du sol par pays en 1990. L'urbain, les vignes ainsi que les autres types d'occupation, sont présentes dans les proportions équivalentes.

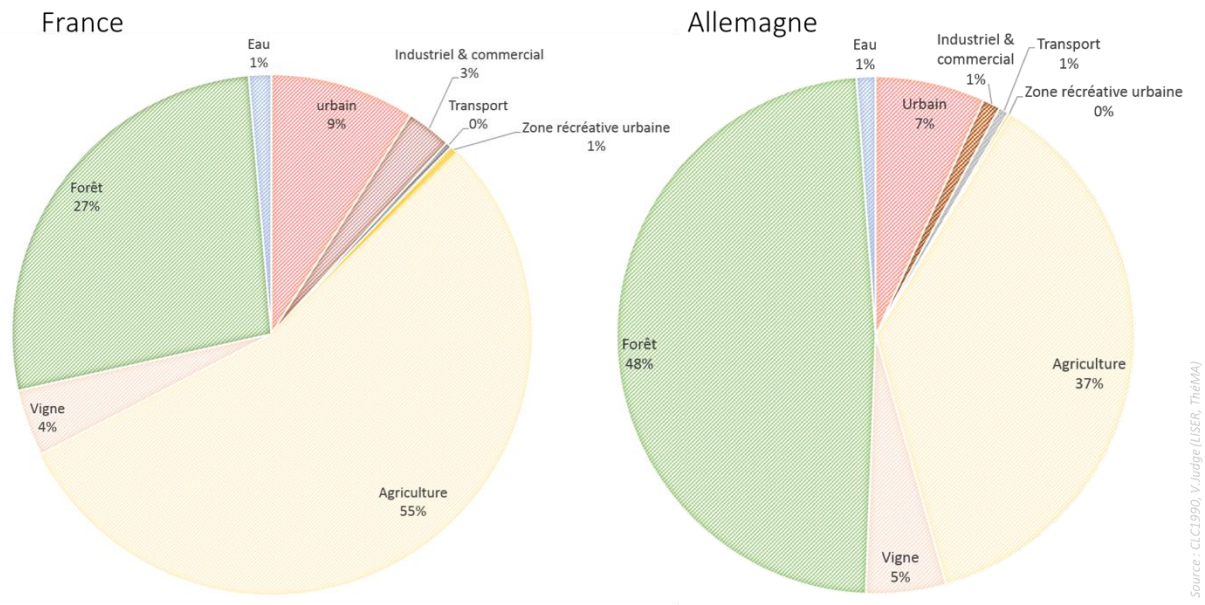


Figure 3. 13 : Répartition des différents types d'occupation du sol par pays

La disponibilité diachronique des données d'occupation du sol est un prérequis indispensable à l'étude des changements d'occupation du sol sur la zone d'étude, donc à celle du développement urbain. La sous-section suivante s'attache à la présentation des tendances de changement observées sur la zone d'étude, d'après les données CLC reclassées.

2.3. Évolution

Le choix de l'unité spatiale a été réalisé en considérant la surface de la zone d'étude et la précision des données d'occupation du sol. L'hectare est l'unité spatiale la plus fine pouvant être sélectionnée sur la plateforme mettant à disposition les données Corine Land Cover. Ce choix permet d'apporter un maximum de précision à l'échelle de la zone d'étude tout en considérant la qualité de l'acquisition ou de la mise à jour de la donnée en elle-même (Chapitre 2 Section 2).

L'acquisition des données Corine a été effectuée pour les années 1990, 2000, 2006 et 2012. Le choix a été fait ici de n'exploiter que les données 1990 et 2006. En effet, comme mentionné précédemment, les changements d'occupation du sol sur le territoire européen ne représentent que 0,7% du territoire en moyenne (Chapitre 1 Section 2). Tout en considérant les transitions urbaines comme transitions majoritaires, elles restent une composante d'un phénomène exceptionnel. Le choix d'étudier conjointement les périodes de 1990-2000 et 2000-2006 permet d'observer un phénomène plus conséquent.

Les données de 2012 ont volontairement été mises à l'écart en raison du changement des protocoles et techniques d'acquisition. L'analyse des changements (Figure 3. 2)met en lumière certains problèmes de cohérence au regard de l'évolution des processus tels qu'ils

sont traduits par la donnée, vis-à-vis de celui existant au vu du contexte socio-économique et démographique de la région étudiée à travers la zone d'étude. Il est en effet, peu probable que dans les faits, les transitions impactant l'agricole entre 2006 et 2012 (période de 6 ans) soient presque 5 fois supérieures à celles observées entre 1990 et 2006 (période de 16 ans), sachant qu'il n'y a pas eu d'évènements particuliers soutenant un tel changement. Les zones forestières sont au contraire en recrudescence. Elles présentaient un solde négatif de 84 ha entre 1990 et 2006, pour présenter un solde positif de 4 243 ha en 2012.

Table 3. 2 : Évolution de l'occupation du sol entre 1990, 2006 et 2012 en hectare (CORINE Land Cover fusionné)

Types d'occupation du sol	1990	2006	2012	Changement 1990 - 2006	Changement 2006 - 2012
Urbain	42143	44612	48181	2469	3569
Industrie et commerce	10163	11977	14527	1814	2550
Transport	2628	2634	1934	6	- 700
Zone récréative urbaine	1747	2078	2615	331	537
Agricole	242397	238826	221073	- 3571	- 17753
Vignes	24410	23556	31115	- 854	7559
Forêt	202321	202237	206480	- 84	4243
Zone immergée	6850	6739	6734	- 111	- 5

Les transitions répertoriées manquent également de cohérence, dans leurs types et/ou dans leur quantité (Table 3. 3). Ainsi 3 701 ha d'urbain effectuent une transition vers un autre état, dont 1 400 ha vers un état agricole. Ce constat nous permet de statuer sur le manque de pertinence de l'utilisation des données de 2012 pour notamment tester et valider le scénario de prolongement de tendance.

Table 3. 3 : Transitions d'occupation du sol 2006-2012 (Source : CORINE Land Cover fusionné)

Type de transitions	Nombres d'hectares
Urbain → Agriculture	1432
Urbain → Industries et Commerces	1239
Urbain → Vignes	413
Urbain → Forêt	368
Urbain → Zone récréatives urbaines	223
Industries et Commerces → Agriculture	369
Industries et Commerces → Zones Immergées	295
Agriculture → Vignes	8992
Agriculture → Forêt	7244

CLC 2012 traduit l'amélioration de la méthode d'acquisition, ainsi que celle de la précision des zonages. Les changements d'occupation du sol représentent donc en partie la correction des biais présents dans les précédentes classifications. Sa mise à disposition aurait cependant pu être exploitée des deux façons suivantes : i – utiliser la donnée en tant que matière d'apprentissage pour l'arbre de décision, ii – utiliser la donnée à des fins de validation.

L'utilisation d'une donnée établie sur un plus long terme peut être intéressante puisque qu'à fortiori le nombre de transitions urbaines entre 1990 et 2012 devrait être en toute logique plus importante que celui entre 1990 et 2006. La validation d'un modèle basé sur l'apprentissage automatique pourrait s'avérer extrêmement riche.

L'occupation du sol en 2006, suivant la classification établie précédemment est présenté Figure 3. 14. Le constat de similitude avec la Figure 3. 12 est flagrant et supporte la caractérisation du phénomène de changement comme « exceptionnel ». Le changement d'occupation du sol entre 1990 et 2006 représente 9 240 ha de la zone d'étude soit 1,74 %. Parmi ces transitions, 0,46 % relève d'une augmentation de la surface urbaine.

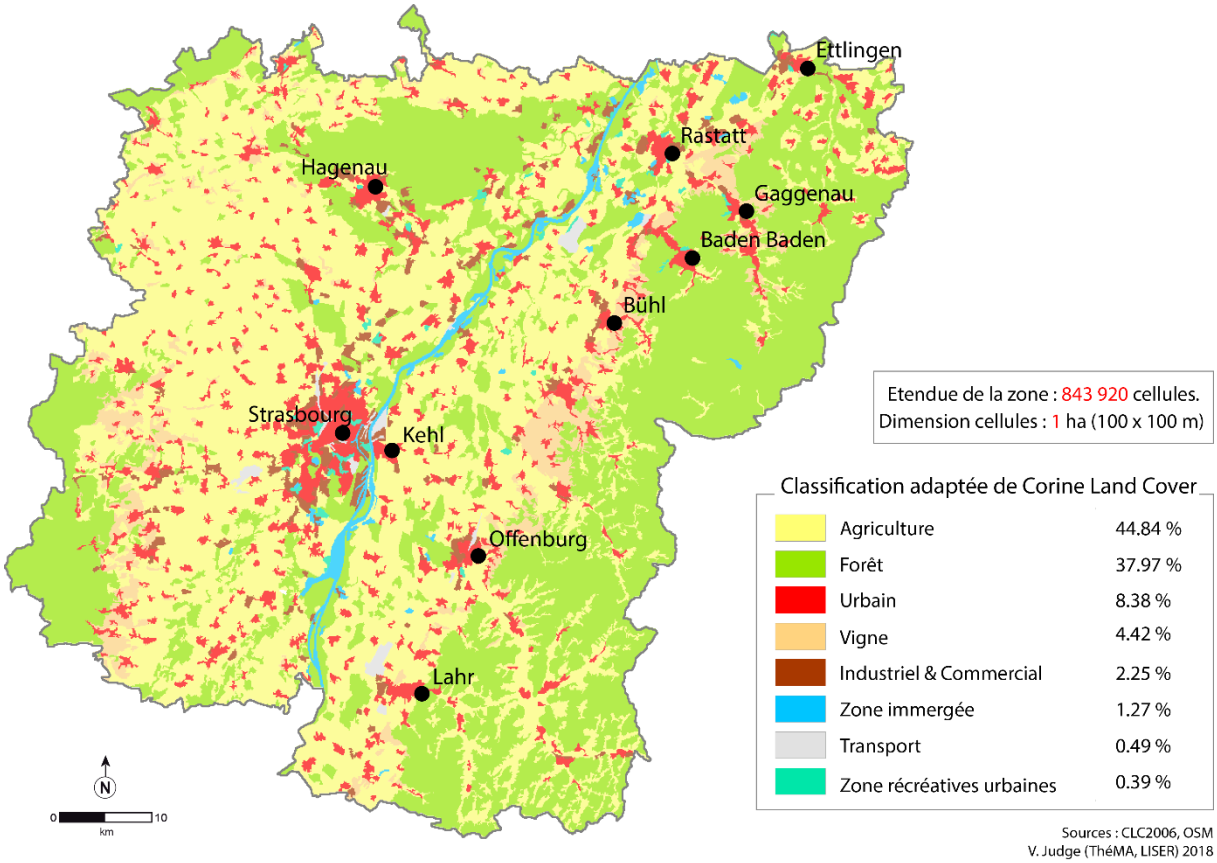


Figure 3. 14 : Occupation du sol en 2006 (CLC fusionné)

La matrice de transition permet de présenter de façon synthétique des changements d'occupation du sol, d'après la classification établie précédemment (Table 3. 4).

Table 3. 4 : Changement d'occupation du sol entre 1990 et 2006 (CLC fusionné)

		2006							
	Types d'occupation du sol	Urbain	Industrie et commerce	Transport	Zone récréative urbaine	Agricole	Vignes	Forêt	Zone immergée
1990	Urbain	41694	265	6	0	162	12	0	4
	Industrie et commerce	95	9694	6	77	122	0	102	67
	Transport	0	8	2587	31	2	0	0	0
	Zone récréative urbaine	13	24	0	1710	0	0	0	0
	Agricole	2528	1672	34	260	237397	110	247	148
	Vignes	258	40	0	0	669	23416	27	0
	Forêt	24	255	0	0	364	18	201559	101
	Zone immergée	0	19	1	0	109	0	302	6419

Sur la zone d'étude, c'est l'occupation du sol agricole qui diminue majoritairement au profit de l'urbain. L'agriculture occupe 45,51% de l'espace en 1990 pour 44,84 % en 2006, soit une perte de 3 571 ha. Cette perte représente plus d'un tiers des changements. La Table 3. 4 permet néanmoins le constat suivant : l'espace viticole bien que présentant un caractère exceptionnel de par son rôle en tant que patrimoine culturel et économique, semble voir sa surface diminuer. La Figure 3. 15 permet de localiser plus précisément les zones d'urbanisation pendant la période 1990 -2006.

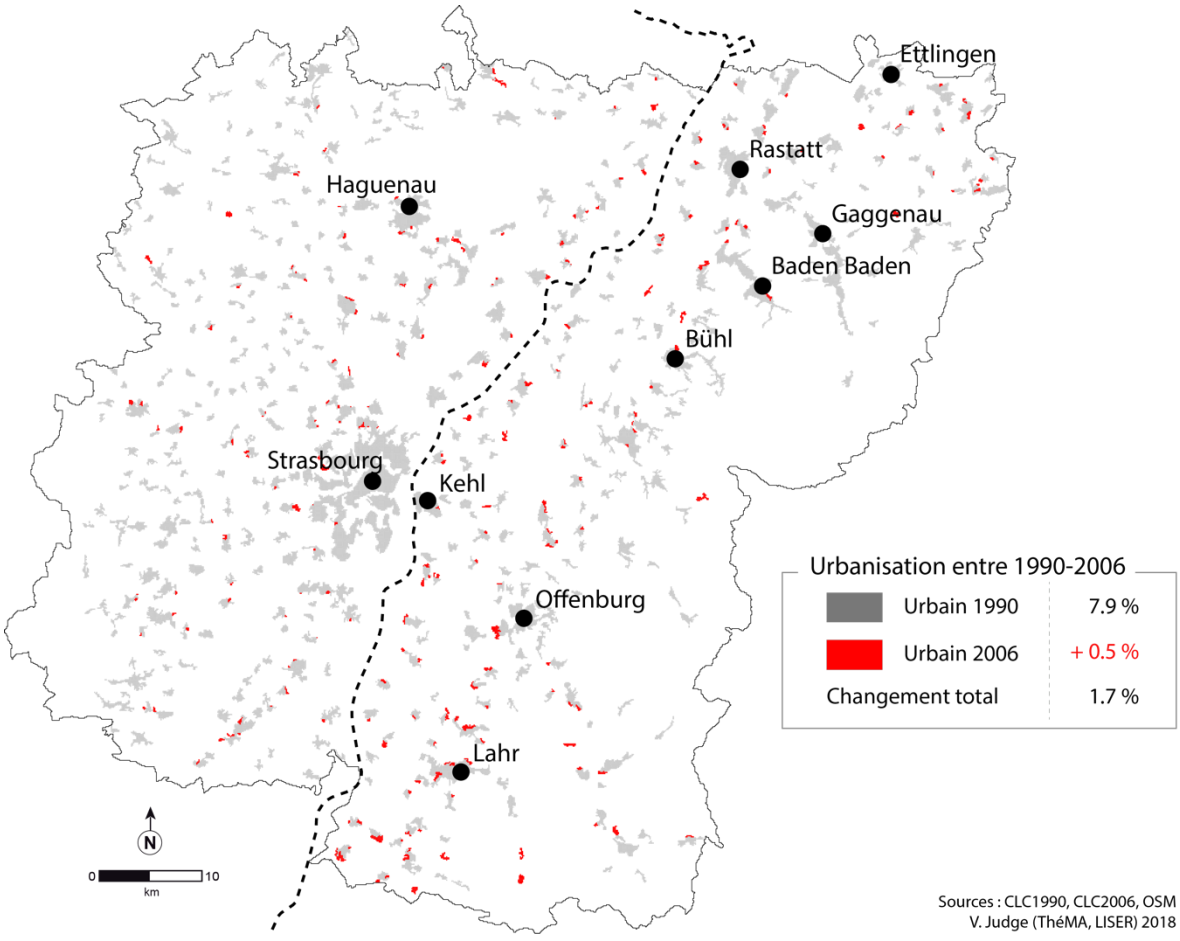


Figure 3. 15 : Urbanisation entre 1990 et 2006 (CLC fusionné)

L'épaisseur des contours des cellules urbanisées met en lumière un phénomène difficilement observable sur les cartes présentant l'occupation du sol en 1990 puis 2006 (Figure 3. 12 & Figure 3. 14). Bien qu'exceptionnel, le phénomène d'urbanisation est d'une ampleur similaire en France et en Allemagne à l'échelle de la zone d'étude. Cette observation souligne le caractère homogène de cette zone, décrit précédemment. En regardant les chiffres de plus près, il est toutefois intéressant de noter que l'Allemagne participe pour 3/5 au phénomène d'urbanisation sur la zone d'étude contre 2/5 pour la France. Au regard des espaces étatiques propres, l'Allemagne se situe au niveau de la tendance générale européenne, puisque l'urbanisation entre 1990 et 2006 a représenté presque 0,7% sur la partie allemande de la zone d'étude. Au contraire, en France, elle est légèrement en dessous de la moyenne avec 0,4%.

En conclusion, les critères délimitent une zone dont les changements sont représentatifs des tendances observées à l'échelle européenne. La reclassification des différentes classes d'occupation du sol facilite l'observation des tendances de changements au sein de la zone d'étude, tout en gardant des spécificités territoriales culturelles telle que la vigne. La sous-section suivante s'attache à exposer les données présentant les caractéristiques intrinsèques utilisées pour définir les cellules de la zone d'étude.

3. DÉTERMINANTS FONCTIONNELS

Le lien entre urbanisation et localisation dans l'espace est une réalité mise en évidence en géographie, notamment par la relation qui existe entre l'attractivité d'un lieu de résidence et son accessibilité. L'utilisation de données libres donne la possibilité de caractériser de plusieurs façons l'accessibilité au sein de la zone d'étude. Elle dépend notamment de la définition des pôles attractifs présents sur le territoire. Dans le cadre de cette recherche, ces pôles sont représentés par les zones urbaines majeures, qui, de par leur importance, sont considérées comme le plus à même d'abriter des zones d'emplois et/ou des aménités diverses. L'importance de ces pôles est ici définie à travers l'application d'une Loi Rang-Taille. C'est donc un critère démographique qui permet de localiser les pôles. Cette loi est une « généralisation empirique selon laquelle la taille des villes est distribuée de manière inverse à leur rang, indépendamment de l'espace et du temps » (Antoni, 2009). La sous-section qui suit permet dans un premier temps d'établir une hiérarchie entre les principaux pôles urbains de notre zone d'étude. Dans un second temps, nous présenterons une cartographie de l'accessibilité de chacune des cellules aux différents niveaux des pôles. Pour finir, nous reviendrons sur les données qui n'ont pas été mobilisées pour décrire la zone d'étude dans ce travail.

3.1. Hiérarchie urbaine

La définition des pôles et la définition de leurs niveaux d'importance (rang) nécessite l'utilisation de données démographiques. Dans un premier temps, la sous-section permet de

présenter les données démographiques et l'application de la loi Rang – Taille pour localiser les pôles urbains majeurs au sein de la zone.

Les données démographiques sont issues des instituts statistiques nationaux, respectivement pour la France et l'Allemagne, l'Institut National de la Statistique et des Études Économiques (INSEE) et Statistisches Bundesamt (Destatis). Dans le cas français, les données de population disponibles à l'échelle communale, ont été agrégées sur la tâche urbaine[31]. Ainsi, la constitution des pôles s'appuie sur une définition morphologique prenant en compte les surfaces urbanisées et non un simple découpage administratif. Pour être en mesure de comparer les données de part et d'autre de la frontière, l'année 2012 a été retenue comme référence commune.

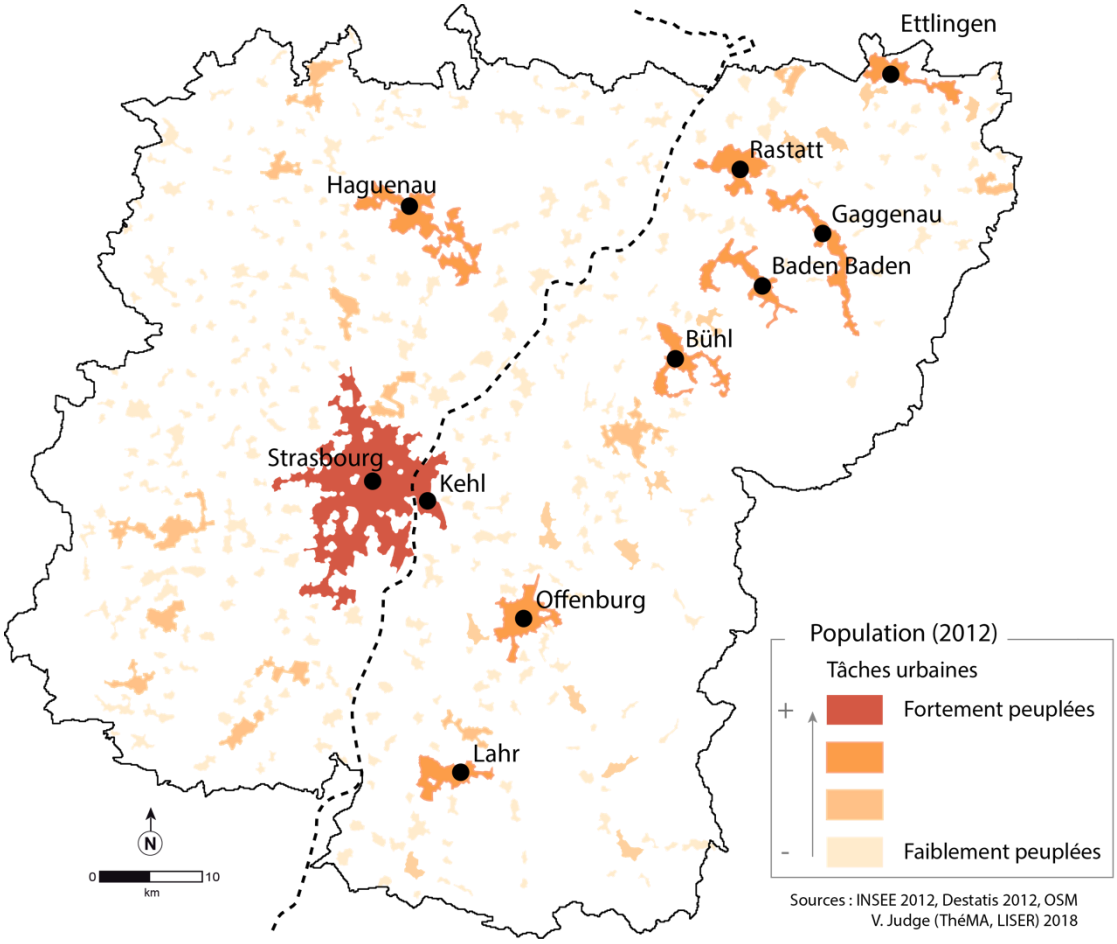


Figure 3. 16 : Répartition démographique (INSEE 2012 ; Destatis 2012)

[31] La tâche urbaine est l'expression du tissu bâti continu. Un périmètre de 200 m est défini dans un premier temps formant une seule et même entité. Cette zone « tampon » est par la suite déduite du précédent zonage, d'après les nouvelles limites. Le zonage permet donc de définir une zone urbanisée uniforme, incluant les vides d'urbanisation entre deux zones urbanisées localisées à moins de 200 m. Cette méthode permet de définir spatialement les unités urbaines d'après la définition de l'INSEE : « Unité urbaine une commune ou un ensemble de communes présentant une zone de bâti continu (pas de coupure de plus de 200 mètres entre deux constructions). Ces seuils, 200 mètres pour la continuité du bâti et 2 000 habitants pour la population des zones bâties, résultent de recommandations adoptées au niveau international » (INSEE, 2016).

Avec 1 677 640 habitants sur la zone d'étude, la zone française est légèrement plus peuplée (903 618 habitants) que la zone allemande (774 022 habitants). Le nombre d'habitants reste pour autant du même ordre de grandeur de part et d'autre de la frontière. Leur répartition diverge cependant légèrement (Figure 3. 16). La donnée de population, agrégée sur l'espace de la tâche urbaine permet déjà de visualiser l'importance de Strasbourg en France. En Allemagne, la répartition de la population est plus homogène sur la totalité du transect frontalier. L'application d'une loi Rang-Taille permet de localiser plus précisément les pôles de différents niveaux.

La loi Rang-Taille, également connue sous le nom de loi de Zipf, a été mobilisée pour appréhender la hiérarchie des pôles urbains dans la caractérisation des processus de développement urbain. Cette dernière a été établie dans les années 1930 par G.K. Zipf à partir de tests empiriques étudiant la fréquence d'apparition des mots dans différents textes (Zipf, 1929). G.K. Zipf étend ses recherches à l'étude des villes, en s'appuyant sur une base de données de 256 villes australiennes (Zipf, 1949). La ville la plus peuplée se voit attribuer le rang 1, la seconde la plus peuplée le rang 2, etc. Il découvre qu'une distribution de Pareto, forme particulière de loi de puissance, s'observe en attribuant un rang à la ville en fonction de sa population (Pareto, 1897). La ville ayant la population maximale, présente deux fois plus d'individus que la ville occupant la seconde place, qui elle-même présente deux fois plus d'individus que la troisième ville la plus peuplée, etc. Ainsi, la population des villes peut être déterminée à partir de la population abritée par la première.

$$f(K_n) = \frac{F}{n}$$

Où : K est la population observée
n est le rang
F est la population de référence (maximale)

V. Pareto à découvert cette distribution en 1897 (Pareto, 1897). Son étude de la répartition de la propriété des terres, en Italie, lui a permis d'émettre le constat suivant : 20% de la population détient 80% des terres. Ce ratio est également retrouvé dans de nombreux contextes. C'est aussi 80% des richesses mondiales qui sont détenues par 20% de la population, ou encore 20% de la population aux États Unis bénéficient de 80% des fonds délayés à la santé en 2006.

Les travaux de G.K Zipf montrent que la loi portant son nom s'applique en géographie pour la superficie des espaces agricoles dans une région, la dimension des territoires dans le monde, et la taille des villes estimée d'après leur population, sont des exemples répondant positivement à l'application de la loi rang-taille. « La hiérarchie urbaine décrite par une loi-rang taille est une forme particulièrement stable et universelle d'organisation du peuplement et des activités dans un territoire » (Pumain, 2008).

La hiérarchie des pôles urbains de la zone d'étude est réalisée à travers l'application d'une loi rang-taille mettant en relation le rang de l'unité morphologique en fonction de sa population. Plus une ville est peuplée, plus son rang est élevé. La projection de ces données

sur un graphique X,Y permet de retrouver une distribution de Pareto (Figure 3. 17). La transformation de l'échelle arithmétique en échelle bi-logarithmique permet d'obtenir une courbe plus ou moins brisée. Les « cassures » mettent en évidence les pôles de différents niveaux (Figure 3. 18).

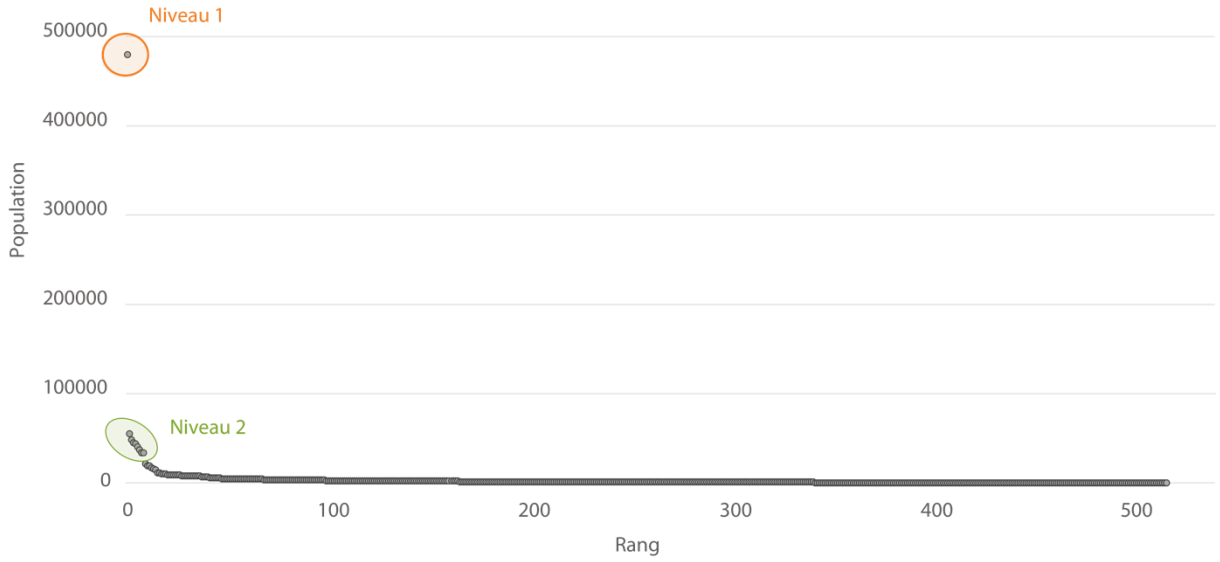


Figure 3. 17 : Rang des unités urbaines en fonction de leur population (Judge, 2017)

La projection sur un graphique Log-Log du rang des villes de la zone d'étude en fonction de leur population permet de définir un pôle de niveau 1. Le pôle Strasbourgeois est le seul pôle de niveau 1 de la zone d'étude. Plusieurs pôles de niveaux 2 peuvent être caractérisés avant la seconde cassure observée sur la courbe. La Figure 3. 19 montre les regroupements des pôles de niveaux 1 et 2 pour l'ensemble des unités de la zone d'étude.

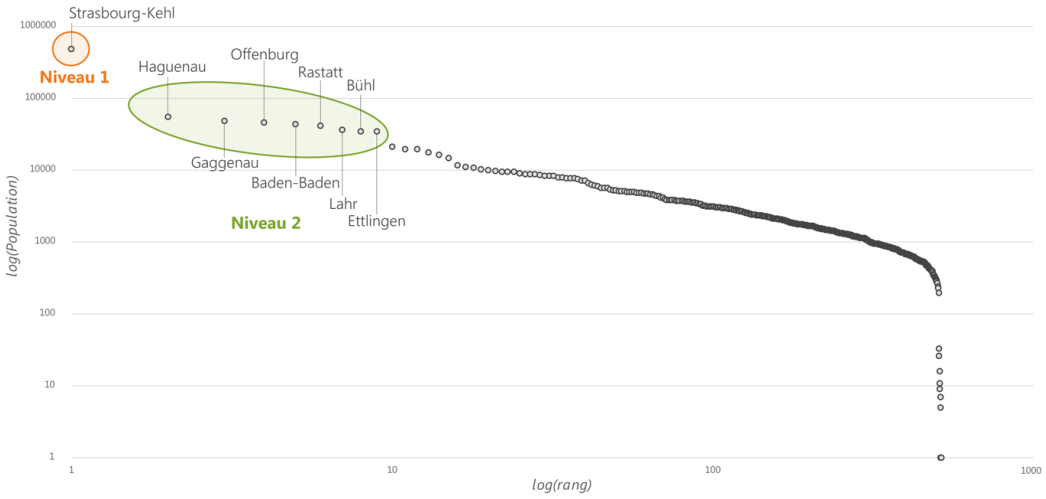


Figure 3. 18 : Loi Rang-Taille - distribution Log-Log sur la zone d'étude (Judge, 2017)

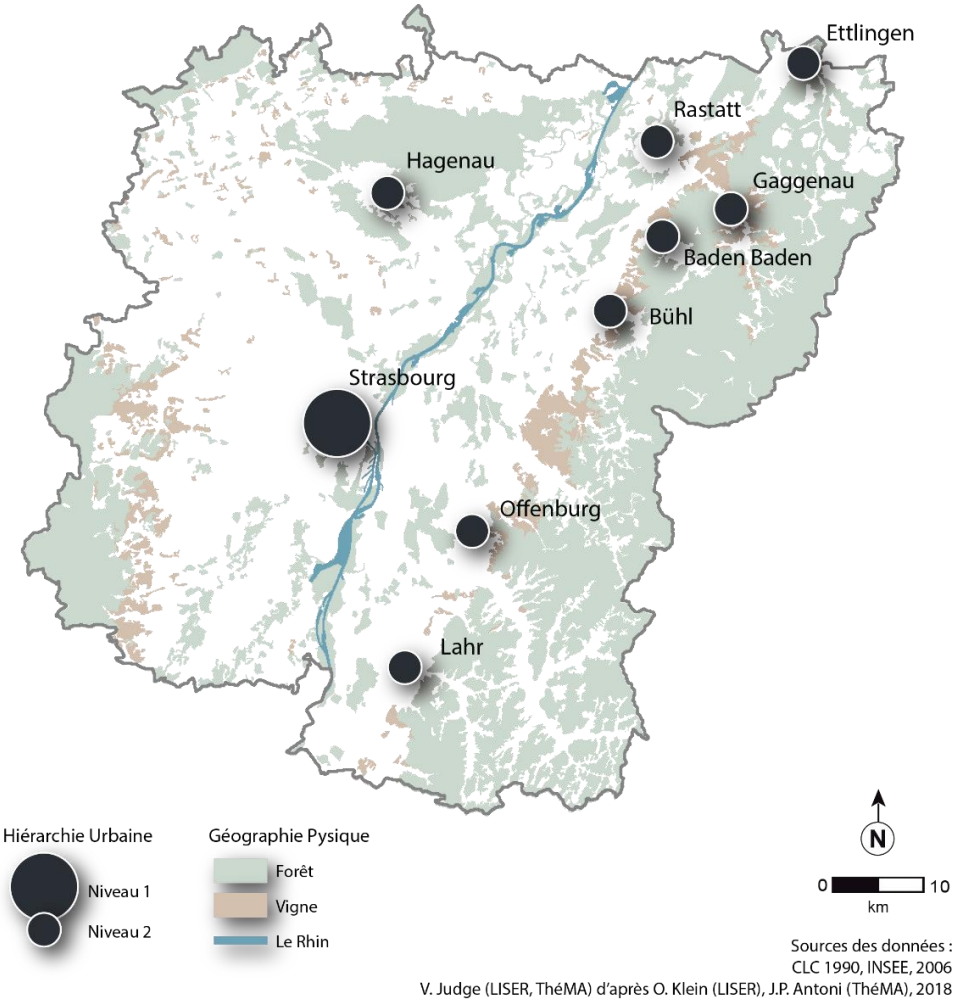


Figure 3. 19 : Localisation et hiérarchisation des pôles de niveau 1 et 2 sur la zone de Strasbourg-Kehl

La répartition des pôles Figure 3. 19 est plutôt monocentrique pour la zone française, bien qu'Haguenau présente un certain poids démographique. Tout au contraire, la zone allemande présente 8 pôles de seconde importance répartis le long de la frontière. C'est à partir de ces pôles, après mobilisation du réseau routiers, que les différents types d'accessibilités sont calculés.

3.2. Localisation et accessibilité

« L'accessibilité d'un lieu mesure le degré de facilité avec lequel ce lieu peut être atteint à partir d'un ou de plusieurs autres lieux, en utilisant tout ou partie des moyens de transport existants » (Bavoux *et al.*, 2005). L'utilisation de données libres permet de mobiliser des données représentant le réseau routier. L'objectif de cette exploitation est la caractérisation de l'accessibilité sur la zone d'étude, notamment dans le cadre d'une mobilité quotidienne influençant la localisation des zones urbaines. En effet, le lien entre accessibilité et les changements d'occupation du sol, notamment le développement urbain, a été mis en évidence dès les années 1960s (Hansen, 1959).

Le calcul d'une accessibilité spatiale généralisée sur la totalité de la zone d'étude permet d'introduire l'influence des pôles urbains principaux, dans la caractérisation du développement urbain 1990 et 2006. L'accessibilité spatiale est définie ici au regard du temps de parcours en minutes, suivant le réseau routier, entre le centroïde d'une cellule et le centre des pôles de différents niveaux localisés précédemment. Dans ce travail, le trajet le plus court entre les centroïdes de chacune des cellules desservies et les pôles urbains de différents niveaux, est estimé en minutes, d'après le réseau routier. Si plusieurs pôles d'un niveau identique sont identifiés, c'est le trajet le plus court qui est retenu. L'hypothèse est, dans ce cas, l'utilisation du pôle le plus proche à service égal. Le calcul d'accessibilité est réalisé à l'aide de GraphHopper.

GraphHopper Direction API (*Application Programming Interface*)
Outils libre disponible en ligne permettant des calculs d'itinéraires basés sur le réseau OpenStreetMap. Développé en 2014 par GraphHopper GmbH à Munich, il offre une richesse et une flexibilité de paramétrages. Pour exemple, différents types de véhicules peuvent être sélectionnés : voiture, camion, vélo, etc. L'élévation peut être prise en compte, ainsi que la congestion à l'aide des données de circulation TomTom© dans le cadre d'une étude plus précise que celle visée dans ce travail. Le calcul réalisé mobilise le paramétrage par défaut (Il ne prend pas en compte la congestion).
https://www.graphhopper.com/

La loi rang-taille permet de déterminer plusieurs pôles dont les niveaux d'importance varient. Sur la base de la localisation des pôles de différents niveaux, plusieurs calculs d'accessibilité ont donc été réalisés pour mettre en relief leurs influences respectives :

- L'accessibilité au pôle de niveau 1 influence l'urbanisation ;
- L'accessibilité au pôle de niveau 2 influence l'urbanisation ;
- Une accessibilité privilégiant le pôle de niveau 1 mais prenant en compte celui de niveau 2 dans une moindre mesure, influence l'urbanisation ;
- Une accessibilité privilégiant le pôle de niveau 2 mais prenant en compte celui de niveau 1 dans une moindre mesure, influence l'urbanisation.

La première accessibilité localise les zones optimisant le temps de trajet vers Strasbourg – Kehl (pôle de niveau 1), la seconde vers les pôles de niveau 2. Les deux autres localisent des zones prenant en compte les pôles de chaque niveau. Ainsi la troisième privilégie l'accessibilité à Strasbourg-Kehl, tout en considérant la localisation d'un pôle de niveau 2. La quatrième privilégie un pôle de niveau 2 mais prend en compte le temps de trajet jusqu'au pôle de niveau 1. Les pondérations utilisées sont 2/3 - 1/3. Ainsi, l'attractivité des zones en

fonction de l'accessibilité aux pôles urbains majeurs pourra être définie lors de l'élaboration des règles (Chapitre 5).

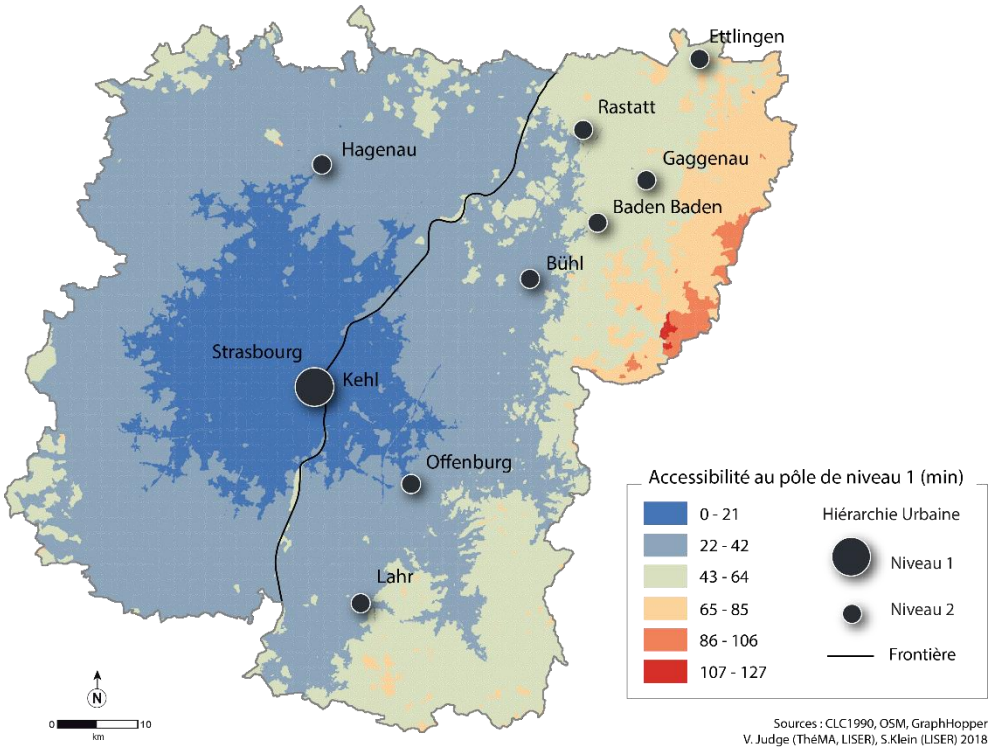


Figure 3. 20 : Accessibilité réseau routier au pôle de niveau 1

Le réseau routier permet d'accéder à l'agglomération de Strasbourg-Kehl dans l'heure, pour la majeure partie de la zone frontalière (Figure 3. 20). Une zone étendue est également localisée à moins de 20 min de ce pôle. Du fait de la structure du réseau autoroutier et routier, cette dernière est plus étendue du côté français.

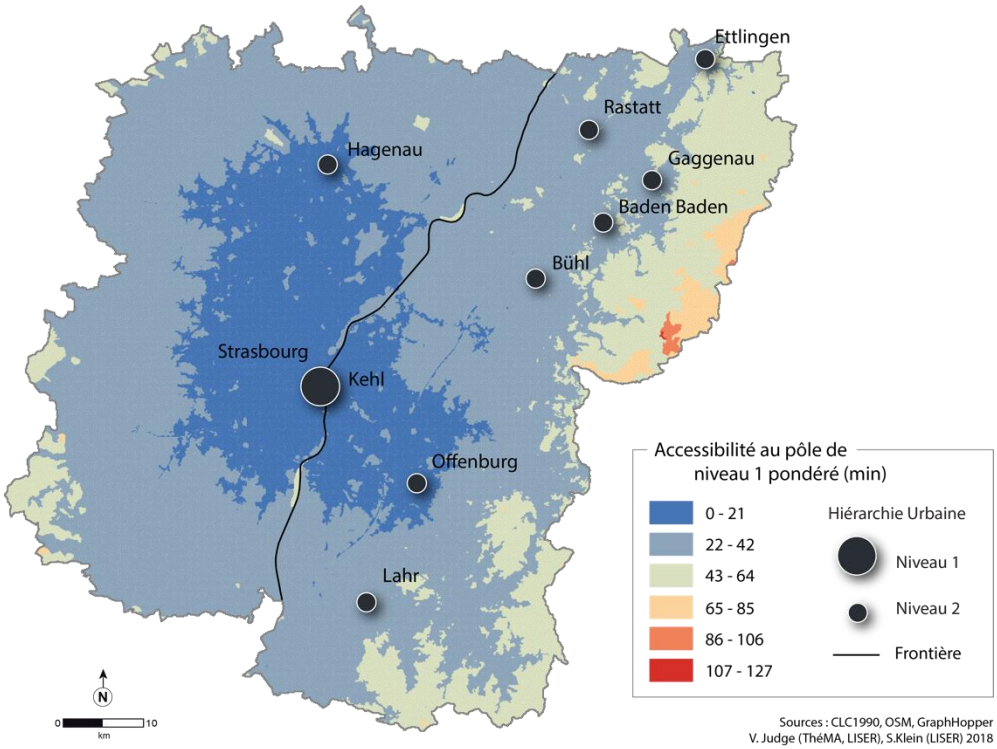


Figure 3. 21 : Accessibilité réseau routier au pôle de niveau 1 pondérée niveau 2

La Figure 3. 21 permet d'observer l'influence de la pondération intégrant la prise en compte d'un niveau 2. En effet, la zone d'accessibilité de moins de 20 minutes s'étend en direction d'Offenburg à l'Est et de Haguenau au Nord lors de.

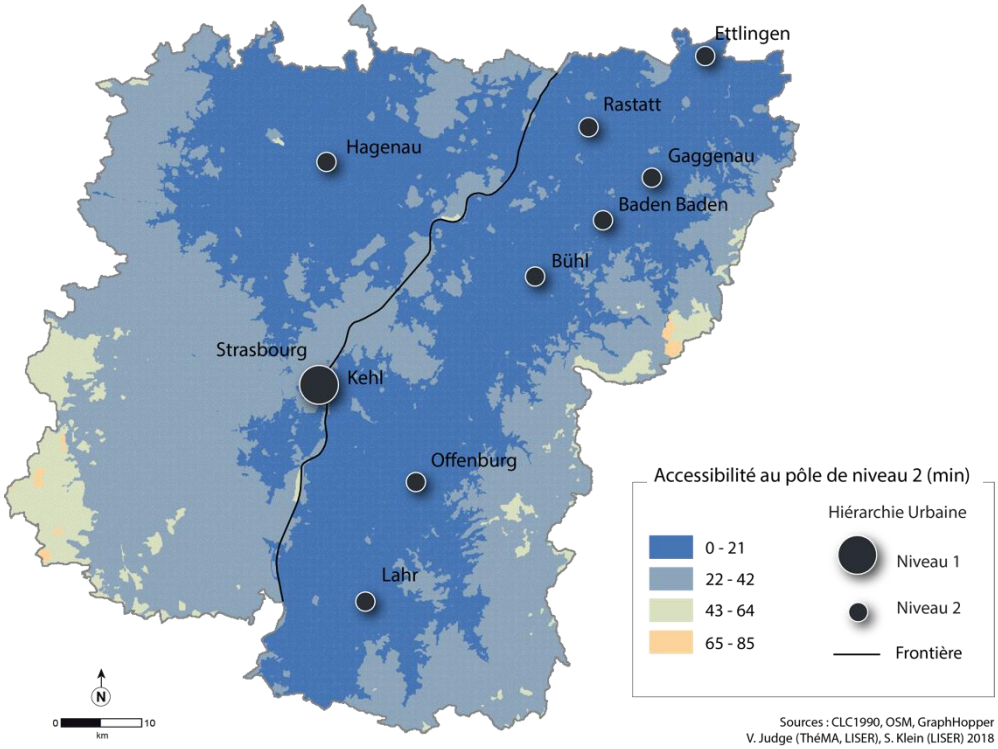


Figure 3. 22 : Accessibilité réseau routier au pôle de niveau 2

Les pôles de niveau 2 sont, eux, accessibles en mois d'une heure depuis la presque totalité de la zone transfrontalière (Figure 3. 22). Pour finir, la pondération le l'accessibilité aux pôles de niveau 2 d'après la localisation du pôle de niveau 1 concentre la zone d'accessibilité de moins de 20 min, notamment du côté allemand (Figure 3. 23).

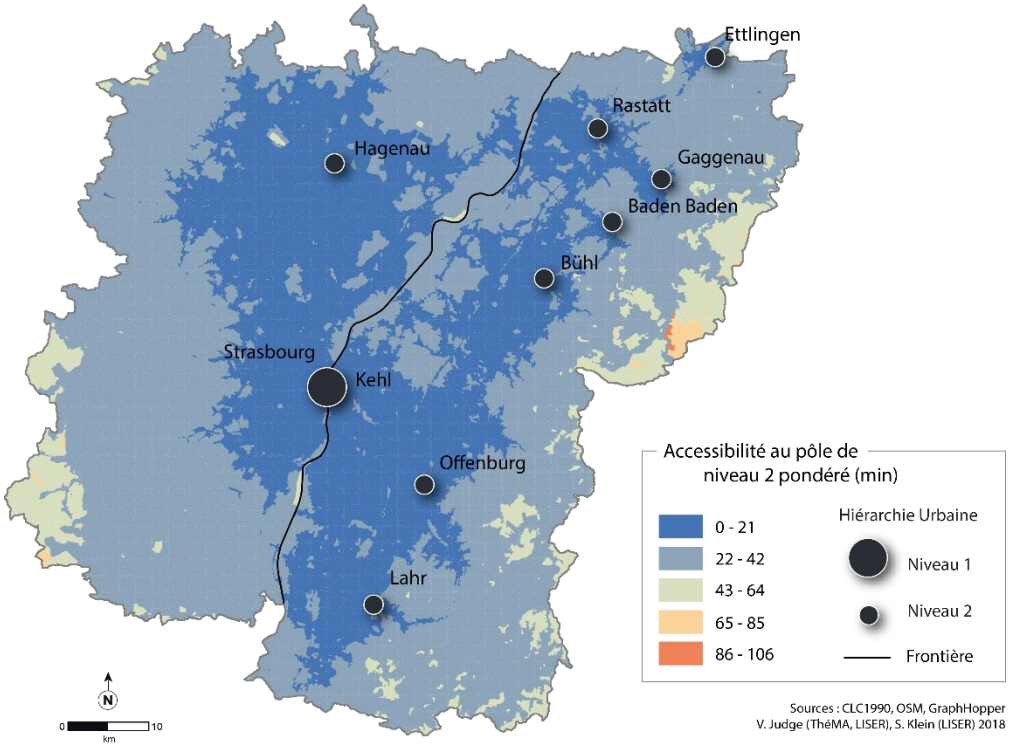


Figure 3. 23 : Accessibilité réseau routier au pôle de niveau 1 pondérée niveau 2

À travers l'utilisation des différents calculs d'accessibilité, l'étude permet de déterminer l'importance de l'accessibilité aux différents pôles dans la localisation des nouvelles zones urbaines. L'objectif à travers cette intégration est d'extraire les temps de trajets qui permettent de rejoindre les différents pôles urbains, ou les intervalles de temps de trajet, qui sont favorable à l'urbanisation. L'intégration de données d'accessibilité dans la démarche méthodologique permet également d'analyser l'apport de l'accessibilité dans l'élaboration du modèle de simulation. L'analyse effectuée à travers l'utilisation des différents calculs d'accessibilité s'affranchit de la frontière étatique, puisque ce n'est pas en fonction de cet objet spatial que l'accessibilité est calculée, mais bien en fonction des pôles urbains de différents niveaux. L'accessibilité est néanmoins contrainte par les points de passages, permettant de joindre les deux rives du Rhin.

Pour finir de caractériser la localisation de la cellule urbanisée, une variable binaire de localisation géographique a été créée dans le cadre de cette recherche pour préciser l'appartenance de la cellule à un espace national particulier (France ou Allemagne). Cet indice permet, par la suite, de définir l'importance de la localisation géographique dans la caractérisation des processus de développement urbain (Chapitre 6).

3.3. Données non mobilisées

Les principaux critères de sélection des données spatiales intégrées dans le processus de modélisation peuvent être résumés comme suit : une donnée libre, harmonisée (même définitions, même temporalité de part et d'autre de la frontière). Ainsi, certaines données n'ont pas été retenues puisqu'elles ne remplissent pas la totalité des critères.

D'autres données sont disponibles mais ne sont pas pertinente dans le contexte de cette recherche. Ainsi, la donnée topographique, et sa variable dérivée la pente, bien que disponible suivant les critères établis précédemment, ne sont pas pertinentes et ne constituent qu'un facteur limitant. Dans la capacité à se développer, elles n'opposent que la limite du budget à investir. Par ailleurs, les données réglementaires ont également été écartées car la littérature met en évidence une inadéquation entre les plans d'aménagement et les résultats observés à l'échelle de plusieurs mandats politiques, notamment sur des périodes concernant plusieurs mandats politiques. Des études ont par ailleurs montré l'écart entre le cadre fixé par la réglementation ou les plans d'aménagement, et la réalité du développement et des changements d'occupation du sol (Alfasi *et al.*, 2012). L'actualité des catastrophes naturelles en France peut nous diriger vers des conclusions similaires, par exemple lors de la tempête Xynthia survenue en hiver 2010, au regard de son impact sur des constructions situées dans des zones pourtant définies comme inondables (Chauveau *et al.*, 2011). Notre durée d'analyse du changement étant fixé à 16 ans dans cette recherche, il convient de ne pas inclure de données réglementaires. Le prix du foncier est, quant à lui, une donnée peu stable au court du temps et extrêmement hétérogène dans l'espace. Elle est transformée la plupart du temps en index permettant d'établir un différentiel spatial à plus grande échelle. La situation économique entre la France et l'Allemagne ne présente pas de différentiel majeur, vis-à-vis des situations observables aux frontières Franco-Suisse ou Franco-Luxembourgeoise. Elle n'est par ailleurs par disponible sous une forme harmonisée à l'échelle de la zone d'étude.

Au-delà de la disponibilité d'une donnée harmonisée, ce travail vise à étudier l'homogénéité du processus de développement urbain sur la zone transfrontalière. L'objectif est de laisser l'opportunité à un phénomène différencié d'émerger, sans intégrer au préalable d'éléments discriminants en entrée. Ce choix incite à épurer le modèle dès l'étape de sélection des données retenues pour modéliser la zone frontalière.

Conclusion du chapitre 3

Ce chapitre a permis de présenter plusieurs arguments soutenant la zone frontalière franco-allemande, comme terrain pertinent pour l’étude l’analyse de l’influence culturelle dans la caractérisation du processus de développement urbain. En effet, cet espace est, d’un certain point de vue, relativement homogène puisqu’il présente des caractéristiques physiques et environnementale similaire. La frontière n’est pas non plus marquée par une fort différentiel économique, ni par une différence monétaire. Elle est également le symbole d’une coopération européenne grandissante à travers le rapprochement de deux des puissances majeures au sein de l’UE. Si tous ces éléments tendent à atténuer l’influence de la frontière en tant que barrière, il n’en reste pas moins que les deux pays voisins que sont la France et l’Allemagne, possèdent une langue, un fonctionnement institutionnel, une politique et des pratiques qui leurs sont propres. En d’autres termes, la différence culturelle persiste. C’est à travers la modélisation de l’espace frontalier et du développement urbain que nous étudierons l’influence culturelle sur le caractère homogène de la zone frontalière. Pour ce faire, le présent chapitre a permis de définir les limites de notre zone d’étude et de présenter les données qui seront mobilisées dans le protocole méthodologique. Nous nous baserons principalement sur deux types de données, les d’occupation Corine Land Cover et les données d’accessibilité. Elles présentent toutes deux l’avantage d’être disponible en tant que données libres et harmonisées à l’échelle de la zone frontalières franco-Allemande.

Conclusion de la première partie

Cette première partie a permis de présenter le cadre théorique et les objectifs associés à la modélisation du développement urbain. Nous sommes, dans un premier temps, revenu sur l'enjeu majeur que représente la compréhension des processus de changements d'occupation du sol en général. Le sol constitue, en effet, une ressource pour l'être humain et présente un caractère fini. Il est, dès lors, nécessaire de comprendre les processus conduisant à son changement. De plus, du fait de l'impact rétroactif sur l'Homme de ces changements, son étude est primordiale afin notamment d'évaluer les conséquences environnementales. L'analyse du processus de développement urbain, en particulier, a permis de montrer qu'il s'exprime sous différentes formes ou organisations, suivant sa localisation, en dépit d'un objectif unique quel que soit le lieu : la construction de bâtiments pour accueillir des habitations ou des activités humaines. Nous avons montré que cette diversité de forme peut être en partie expliquée par la différence dans la composition culturelle locale des sociétés humaines. Ce point met donc en évidence un lien entre la forme du développement et ses acteurs. Dès lors notre étude de la spécificité locale ou, au contraire, de l'existence d'une part généralisable d'un processus de développement urbain dont la fonction reste similaire indépendamment de la localisation s'est orientée sur l'analyse d'un espace frontalier.

Nous avons ensuite présenté notre approche de l'espace frontalier, à travers la modélisation du développement urbain. Notre protocole a pour objectif d'être simple et transparent pour favoriser la compréhension et l'interprétation thématique à travers l'étude de données et variables explicites, mais également soutenir son caractère transposable. Cette approche et la composante frontalière de notre zone d'étude induisent une réflexion sur l'utilisation des données. Après avoir délimité spatialement notre zone d'étude, la question de la disponibilité de données libres et harmonisées a été abordée. Définir les limites de la zone avait pour objectif de refléter une cohérence spatiale basée sur une certaine symétrie au regard de la frontière. Ce dernier point a influencé notre choix de nous affranchir des limites administratives existantes, pour leur préférer une limite basée sur l'accessibilité au centre urbain principal, c'est-à-dire Strasbourg-Kehl. Le choix de recourir à des données libres permet, non seulement, de travailler dans un contexte binational, mais aussi, d'optimiser l'interopérabilité de la méthodologie développée. Trois types de données libres et harmonisées ont été retenus à l'échelle de notre zone d'étude : l'occupation du sol, le réseau routier et la démographie.

Pour finir, les spécificités de la zone frontalière franco-allemande ont été présentées. Son histoire est particulière notamment en raison de la mobilité spatiale du tracé de la frontière au cours des siècles. Cette mobilité a été accompagnée de politiques locales visant à influencer le sentiment d'appartenance des populations. Leurs traces persistent notamment spatialement à travers l'occupation du sol et la structure urbaine. Nous questionnons le rôle de la frontière en tant que marqueur spatial de différence culturelle entre les deux pays. En effet, chacun d'entre eux possède une politique, une hiérarchie institutionnelle, une langue

et bien d'autres caractéristiques spécifiques, pour autant l'histoire et la politique de cohésion menée par l'Union Européenne et soutenue par les pays membres notamment la France et l'Allemagne tendent à augmenter la porosité de la frontière.

C'est à travers l'élaboration d'un modèle de simulation du développement urbain que ce travail de recherche analyse l'influence de la frontière sur la zone d'étude. La partie suivante permet de présenter la méthodologie élaborée pour répondre aux questions suivantes : La frontière apparaît-elle dans la caractérisation du développement urbain ? En d'autres termes, les variables de localisation étatiques sont-elles mobilisées dans l'élaboration automatique du modèle de simulation ?

Deuxième partie

Modéliser le développement urbain en zone frontalière

Introduction de la deuxième partie

La modélisation des processus de développement urbain, à travers des règles qui représentent ses dynamiques, est une étape préalable fondamentale à l'étude de ses caractéristiques au regard de sa localisation au sein de l'espace frontalier. Nous avons précédemment établi les enjeux liés à l'étude des changements d'occupation du sol, et l'intérêt de privilégier celle des changements urbains pour analyser l'influence de la culture locale, sur la forme prise par le changement. Dans ce travail doctoral, nous proposons une approche modélisatrice qui s'appuie sur des données spatiales libres et harmonisées à l'échelle de la zone frontalière franco-allemande. Il s'agira, dans ce contexte, de présenter l'élaboration du protocole de modélisation.

Cette deuxième partie consiste à présenter les outils de modélisation mobilisés et à expliquer les choix méthodologiques effectués dans cette recherche. L'approche modélisatrice est, en effet, une approche classique en géographie théorique et quantitative. L'étude et l'analyse des interactions spatiales par le biais de modèles qui permettent une simulation dynamique des processus spatiaux se développe encore davantage grâce à l'évolution des performances informatiques. La quantité et le volume des données disponibles, associés à l'augmentation des capacités de calculs informatiques, inspirent le développement d'une multitude de méthodes et de modèles, dont la complexité varie. Ainsi, nous discuterons dans un premier temps du choix de l'automate cellulaire, comme outil de modélisation dans ce travail.

L'automates cellulaires ont suscité l'intérêt de nombreuses disciplines. Ils permettent, sur la base de règles simples, d'observer l'émergence de processus complexes. Cet outil a donc été mobilisé à de nombreuses reprises dans les études portant sur la modélisation des changements d'occupation du sol. Pour autant, la question de la définition des règles qui régissent la dynamique du processus modélisé perdure encore aujourd'hui, notamment dans l'étude des changements d'occupation du sol. Et pour cause, comme le rappel P. Torrens (2011) : « dans l'étude des systèmes urbains, nous n'avons, le plus souvent, pas la moindre idée de ce que sont les ″bonnes règles″ ». L'élaboration et le calibrage des règles qui permettent de modéliser les changements d'occupation du sol, et le développement urbain, restent encore aujourd'hui un enjeu majeur. C'est en effet, une composante importante de la démarche de modélisation, puisqu'elle influence les résultats produits. Ainsi, nous discuterons dans un deuxième temps du choix de l'arbre de décision, comme outil permettant de générer automatiquement les règles d'urbanisation nécessaire à l'automate cellulaire.

L'étape de calibrage qui, dans ce travail passe par le paramétrage de l'arbre de décision, est indissociable de l'étape de validation des résultats produits. La pluralité des objectifs de modélisation se reflète également à travers des approches variées de l'étape de validation des résultats. Ainsi, nous discuterons dans un troisième temps des choix concernant la validation du modèle dans ce travail.

Chapitre 4

Une base : les automates cellulaires

La modélisation des changements d'occupation du sol a pris une dimension particulière en raison de l'intérêt croissant pour la compréhension des processus dynamiques qui influencent certes le quotidien des individus, mais présentent également une influence sur les cycles naturels à l'échelle globale (Chapitre 1). Le développement des instruments d'acquisition et de traitement des données qui présentent une composante spatiale a également favorisé cet engouement (Chapitre 2). Le travail présenté ici s'intéresse particulièrement à la zone frontalière franco-allemande de Strasbourg-Kehl. L'objectif est d'étudier l'homogénéité de la zone d'étude par le prisme de la modélisation du développement urbain qui reflète notamment une culture locale. Il s'agit donc d'étudier les réminiscences d'une composante nationale, sur une zone dont la frontière présente une porosité croissante (Chapitre 3). Pour ce faire, notre démarche s'oriente vers l'utilisation de données libres en raison du caractère frontalier de l'espace étudié, et sur le développement d'une méthodologie simple et transposable, qui peut s'adapter efficacement à d'autres contextes. C'est donc à travers l'élaboration d'une méthode nécessitant l'utilisation d'un modèle urbain que ces points sont abordés dans notre travail.

L'automate cellulaire, un outil initialement développé par des mathématiciens, rencontre un succès incontestable dans la modélisation des systèmes urbains et des changements d'occupation du sol en géographie. La première section s'attache donc à retracer succinctement l'histoire de cet outil en présentant les caractéristiques pour lesquelles il est principalement apprécié en géographie. Le développement des modèles permettant de simuler les changements d'occupation du sol et le développement urbain en géographie s'est inspiré de théories et d'outils issus de divers champs de recherches. Le contexte entourant l'émergence de la modélisation des systèmes urbains est présenté dans la seconde section. En effet, plusieurs modèles ont vu le jour au cours des dernières décennies, néanmoins, leurs caractéristiques ne permettent pas leur mobilisation dans le présent travail en raison de nos objectifs de modélisation et des caractéristiques frontalières de notre zone d'étude. Ainsi, la troisième section revient elle sur la formalisation des données nécessaires à leur exploitation dans un automate cellulaire, puis présente l'outil de modélisation, LucSim, spécifiquement utilisé et adapté dans le cadre de cette recherche.

"The art of modelling lies in understanding
how much to include but not more."
R.White 2015

1. UN OUTIL SPÉCIFIQUE

Les automates cellulaires (AC) ont constitué une approche privilégiée de modélisation des changements d'occupation du sol en géographie pendant plusieurs décennies (Kocabas & Dragicevic, 2006). L'idée d'une discrétisation de l'espace en cellules, dont l'état est issu d'une liste finie d'états et dont l'évolution serait dépendante de son propre état mais aussi de celui des cellules composant son voisinage, a été utilisée à maintes reprises dans les années qui suivirent sa première exploitation en géographie par W. Tobler en 1979 (Couclelis, 1985, 1988; Phipps, 1989). L'application au milieu urbain a ensuite été démocratisée, notamment par R. White et G. Engelen (1993), M. Batty et Y. Xie (1997; 1999), Clarke et al. (1997) pour ne citer que quelques exemples.

Cette section a pour objet principal de présenter l'outil de modélisation qui a été retenu d'un point de vue théorique et appliqué. La sous-section 1 permet de revenir sur l'historique de son apparition et de son utilisation majeure dans le cadre de la modélisation de processus spatiaux, notamment celui du changement d'occupation du sol en géographie et des divers processus urbains étudiés dans ce domaine. La sous-section 2 revient sur l'utilisation des modèles en géographie et présente les caractéristiques fondamentalement intéressantes qui retiennent l'attention des utilisateurs d'automates cellulaires. La sous-section 3 permet ensuite de préciser les caractéristiques nécessaires aux données mobilisées dans un automate cellulaire. Elle permet également de présenter l'outil qui sert de support à l'implémentation de la méthodologie développée dans ce travail de recherche. Et finalement de faire un état des lieux du processus de calibrage des modèles de simulation des changements d'occupation du sol.

1.1. Historique

L'étude des automates cellulaires a été initiée par le mathématicien S. Ulam, dans les années 1940s à Los Alamos (Nouveau Mexique, USA). Ce dernier s'intéressait, entre autres, à l'étude des formes géométriques, ainsi qu'au processus de cristallisation des cristaux de roches. Son attrait pour la géométrie le poussa à explorer l'évolution des formes en fonction de règles simples. Il eut pour la première fois l'idée de disposer sur une grille représentant un espace cellulaire, des cellules pouvant adopter deux états : allumé ou éteint. Partant d'une configuration initiale, il observa l'évolution à l'étape suivante, en fonction d'une règle simple : *Une cellule s'allume si elle est en contact avec deux cellules allumées, sinon elle s'éteint.*

Cette démarche lui a permis de constater qu'une règle simple peut faire émerger des structures « complexes et esthétiques » qui parfois se répliquent dans le temps (Renard,

2000). J. Von Neumann qui travaillait au même moment sur les machines auto-réplicatrices d'après les travaux d'Alan Turing, suivit les conseils d'Ulam lui suggérant d'utiliser un espace cellulaire pour sa machine auto-réplicatrice. Il s'affranchissait ainsi de toutes contraintes physiques limitant l'exploration théorique (Von Neumann & Burks, 1966). Dans ce cadre, il élabora, à Princeton (USA), un AC de 200 000 cellules, contenant un copieur universel. Ce même mécanisme est observé dans l'étude de l'ADN en biologie. Princeton devint un lieu de passage symbolique puisque cette université accueillit par la suite J. H. Conway, plus connu pour l'élaboration du Jeu de la Vie.

Le Jeu de la Vie a participé au rayonnement des AC dans la recherche scientifique. Ce modèle, développé par J.H. Conway (Gardner, 1970), s'applique sur une grille cellulaire dont les cellules suivent 3 règles extrêmement simples. Le cadre d'application du modèle est illustré Figure 4. 1-a. Il s'applique dans un environnement à deux dimensions, considérant un voisinage de Moore. Après avoir défini une situation initiale qui présente une répartition aléatoire des cellules caractérisées par l'un ou l'autre des états (Figure 4. 1-b), le modèle produit plusieurs types de structures obtenus après l'application des règles sur plusieurs pas de temps (Figure 4. 1-c).

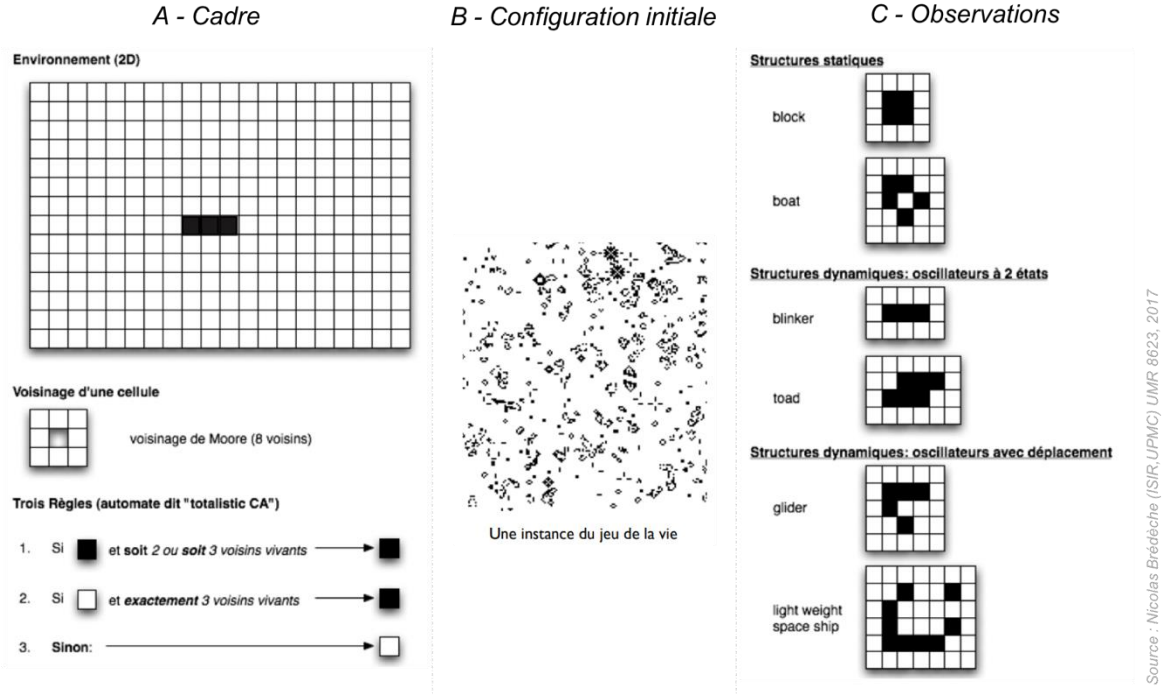


Figure 4. 1 : Le jeu de la Vie de Conway (Bredèche, 2017)

À travers ce modèle, Conway met en évidence la création parallèle d'états stationnaires, d'états cycliques et de mouvements irréguliers chaotiques, sur la base de la simulation de trois règles extrêmement simples. Il reste donc célèbre pour avoir démontré qu'un phénomène dynamique présentant une apparente complexité pouvait être décrit à travers des règles simples.

S. Ulam, J. Von Neumann et J. H. Conway appuient tous les trois leurs recherches sur un automate cellulaire dans un espace à 2 dimensions alors que S. Wolfram décide lui d'explorer un espace à 1 dimension (Wolfram, 1982). Intéressé par l'étude des systèmes complexes et l'exploration théorique, il testa l'ensemble des combinaisons possibles dans le cadre d'un automate cellulaire à 1 dimension (Figure 4. 2), présentant deux états (noir/blanc) et dont l'évolution est régie par des règles basées sur un voisinage adjacent (cellule à droite et cellule à gauche de la cellule interrogée). Ces caractéristiques limitent la combinatoire à 2^8 règles.

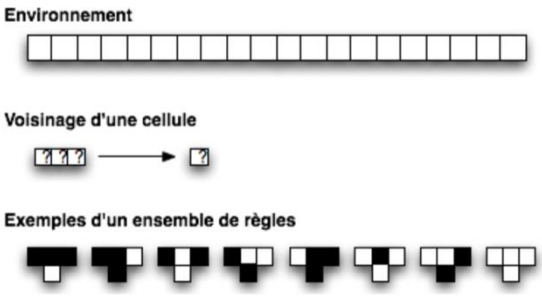


Figure 4. 2 : Automate cellulaire à 1 dimension selon S. Wolfram (Bredèche, 2017)

L'exploration de l'espace des combinaisons possibles permit à Wolfram de mettre en évidence des similitudes dans les configurations produites par les règles (Wolfram, 1984). Ces dernières définissent 4 classes dans la classification de Wolfram (Figure 4. 3).

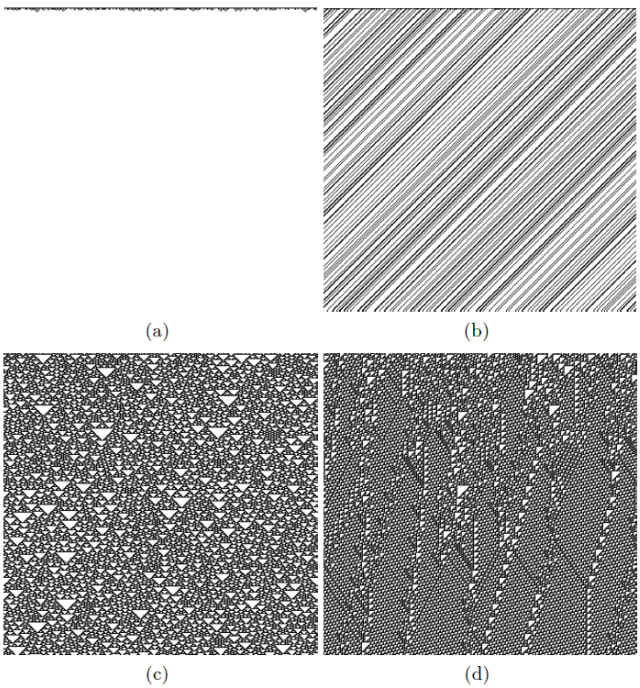


Figure 4. 3 : Les 4 classes d'automate d'après Wolfram : a) Classe I règle 32 - Structure uniforme ; b) Classe II règle 10 – Structure périodique simple ; c) Classe III règle 126 – Structure apériodique, imprédictible, chaotique ; d) Classe IV règle 110 – Structure complexe et imprédictible (Martinez et al.*, 2013)*

Le point remarquable de ce travail est bien la mise en évidence de différentes configurations, de l'émergence de structures uniformes, périodiques simples, apériodiques chaotiques imprédictibles, et complexes imprédictibles, basées sur l'utilisation de règles simples et de même calibre initial. Il est donc impossible de savoir en amont, quelle configuration sera associée à la règle.

Ces divers travaux soulèvent alors des questions fondamentales : La complexité n'est-elle qu'apparente ? Est-elle le fruit de lois fondamentales simples ? Si oui quelles règles régissent le phénomène ? Comment déterminer ces règles ?

C'est la simplicité du paradigme, sa capacité à faire émerger des processus complexes à partir de règles simples basées sur une relation de voisinage, qui attira l'attention de chercheurs dans d'autres domaines, notamment en géographie. L'intégration entre les AC et les Systèmes d'Information Géographiques (SIG) a aussi eu une influence notable sur la démocratisation des AC en géographie (Couclelis, 1997). Dès lors, nous nous intéressons plus spécifiquement à la définition de l'automate cellulaire en géographie et ses propriétés associées.

L'automate cellulaire en géographie

Les systèmes urbains et l'occupation du sol, qui résultent de l'interaction d'une multitude de facteurs, montrent des caractéristiques d'auto-organisation et d'émergence (Benenson & Torrens, 2004). Ils peuvent alors être modélisés comme des systèmes complexes (White *et al.*, 2015). L'automate cellulaire est, ainsi, un outil approprié pour explorer la dynamique de ces systèmes, notamment en raison de son potentiel en termes de calculs (exploration de multiples paramètres à moindre coût). L'intérêt des géographes à son égard est en partie issu du caractère intrinsèquement spatial de l'AC. De plus, l'utilisation d'une machine pour réaliser les calculs liés à l'application de règles ou algorithmes « simples » rend la tâche réalisable plus facilement.

W. Tobler (1970) a été le premier à définir l'AC comme un modèle géographique (Kocabas & Dragicevic, 2006). Néanmoins T. Schelling comme T. Hagerstränd les avaient déjà mobilisés dans leurs travaux respectifs sur la ségrégation (Schelling, 1971) ou sur la diffusion spatiale (Hagerstränd, 1967). H. Couclelis s'appuie à son tour sur un « monde cellulaire » pour modéliser des processus dynamiques dans les années 1980 (Couclelis, 1985). R. White et G. Engelen développent, eux, un automate cellulaire contraint (White & Engelen, 1993) dont les règles de transitions traduisent une probabilité de transition d'un état d'occupation du sol à un autre, dès le début des années 1990. K. Clarke et L. Gaydos se focalisèrent, à leur tour, fin des années 1990, sur l'étude rétrospective du développement de la ville de San Francisco et des régions de Baltimore et Washington à travers le développement du modèle SLEUTH, basé sur un AC (Clarke & Gaydos, 1998). X. Li et A. Yeh intégrèrent une contrainte spatiale dans l'automate cellulaire, définissant différentes zones plus ou moins intéressantes en amont (Li & Yeh, 2000). Pour R. White (2015),

l'important est que l'AC soit capable de capturer le phénomène modélisé et de générer un fonctionnement suffisamment riche.

Depuis ses premières utilisations en géographie, un regain évident d'intérêt pour les AC a émergé en raison du développement des SIG. L'environnement, en partie raster, des SIG permet une intégration directe des résultats produits par l'automate, facilitant d'autant plus leurs analyses d'un point de vue spatial. Ainsi, les AC ont commencé à être « torturés » (White *et al.*, 2015). En s'écartant de la définition fondamentale, des composantes ont été ajoutées ou modifiées (Batty, 1998; Wu & Webster, 1998). Torrens (2001) s'est intéressé au couplage AC et modèles multi-agents, alors que Yeh et Li (2003) explorèrent le couplage avec des réseaux neuronaux dans le cadre de la définition des paramètres.

Ici apparaît la notion de modélisation et simulation « basée » sur les automates cellulaires (*eng. Cellular Automata based)*. Effectivement, les différentes recherches et utilisations de cet outil ont fait évoluer le paradigme de base. Ce point est plus largement abordé dans la sous-section suivante, définissant plus précisément l'Automate en lui-même.

1.2. Concept

L'automate cellulaire est un outil de modélisation, utilisé en géographie pour apporter des éléments de compréhension lors de l'étude de processus complexes, tel que celui du changement d'occupation du sol. Il est particulièrement apprécié pour la simplicité de son paradigme. Dans sa forme primaire, l'AC en géographie peut être décrit comme une représentation discrète de l'espace à travers des cellules uniformes (taille et forme) généralement carrés. Ces dernières sont caractérisées par un état, appartenant à une liste finie. L'évolution d'un état initial à un instant t vers un état à un instant t+1 est contrainte par des règles de transitions basées sur l'état initial de la cellule et des cellules présentes dans son voisinage. Les AC possèdent plusieurs propriétés intéressantes pour l'étude des systèmes urbains. Nous retiendrons ici l'émergence et l'autoréplication.

Émergence

En 1875, G. H. Lewes suggère que « des entités émergentes peuvent être le résultat de l'action d'entités plus fondamentales et peuvent être parfaitement nouvelles ou irréductibles par rapport à ces dernières » (Lewes, 1875). Un phénomène nouveau peut alors découler d'une combinaison de propriétés, sans pour autant en revêtir aucune caractéristique fondamentale (Juignet, 2015; O'Connor & Wong, 2015). La conscience humaine est un exemple couramment utilisé pour illustrer l'émergence. C.Langton (1986) a constaté lors de son exploration des AC « l'émergence de comportements complexes et extrêmement structurés, à la frontière entre un ordre simple et le chaos » (White *et al.*, 2015). S. Wolfram a, lui, observé l'émergence de structures fractales à travers l'élaboration de règles simples dans un AC à une dimension seulement (Wolfram, 1984). La classe IV de la classification portant son nom en est l'illustration (Figure 4. 3). R. White effectue le constat suivant lors

de ses recherches : les villes dont la planification est absente présentent tout de même une structure pouvant être décrite. « Elles sont en réalité extrêmement structurées mais dans une autre forme que la forme euclidienne classique » (White *et al.*, 2015). L'observation de structures urbaines résultant d'un schéma d'aménagement parallèlement à celles n'ayant pas de contraintes, permet d'observer la différence entre la planification et les structures dont l'organisation émerge par un processus d'auto-organisation. L'automate cellulaire, dans l'étude des changements d'occupation du sol, laisse l'opportunité d'observer l'émergence de structures ou formes spatiales, à travers l'étude du développement urbain.

Auto-réplication

L'objectif d'une machine auto-réplicatrice est d'être « capable de produire n'importe quelle machine décrite dans son programme, y compris une copie d'elle-même, à partir de matériaux trouvés dans l'environnement » (Hutzler, 2015). C'est J. Von Neumann qui pour la première fois met en lumière le potentiel auto-réplicateur de l'automate cellulaire (Von Neumann & Burks, 1966). La machine qu'il a mise au point, le Kinématon, moyennant une quantité de données suffisante en entrée, a la capacité de se dupliquer. Von Neumann cherchait à reproduire le mécanisme observé dans la production de l'ADN. Cette propriété est également illustrée par l'automate de Fredkin (Fredkin & Toffoli, 1982). Ce dernier reproduit la configuration initiale en neuf exemplaires (Figure 4. 4). L'automate de Langton (1984) permet lui aussi d'illustrer ce phénomène (Figure 4. 5).

Figure 4. 4 : Simulation de Fredkin (Calicode, 2018)

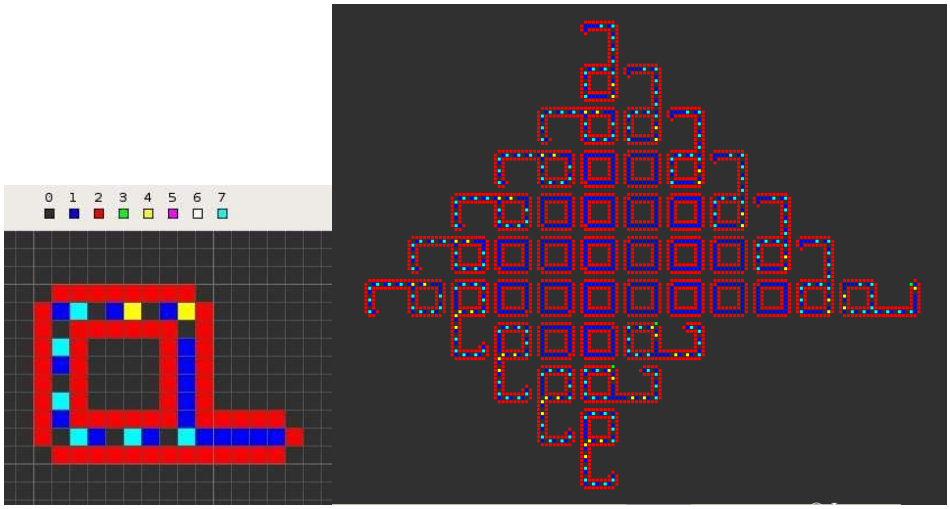


Figure 4. 5 : Boucle de Langton (Calicode, 2018)

L'illustration du phénomène apparaissant dans la simulation de Fredkin peut être mise en relation avec une composante du processus de développement urbain, notamment celui du phénomène de mitage. Celui apparaissant dans l'automate de Langton est quant à lui intéressant dans la perspective établissant que 29 règles d'évolutions appliquées à 7 états permettent de créer un développement extrêmement structuré. Réciproquement, la définition des règles de développement urbain en fonction d'autres états d'occupation du sol, dans un automate cellulaire, conduirait à la caractérisation d'un processus lui-même présentant une logique spatiale émergente. L'intérêt de s'appuyer sur un environnement d'automate cellulaire est donc une fois de plus mis en lumière.

Les propriétés de l'automate cellulaire sont particulièrement intéressantes pour l'étude du développement urbain puisqu'il permet d'appréhender un phénomène complexe à travers des règles de transitions simples, avec l'émergence de structures spatiales particulières par le biais de règles reproduisant des structures observées. La sous-section suivante présente plus précisément les différentes structures d'AC.

1.3. Définition et structure

J.P. Renard fait la distinction entre les règles structurelles et fonctionnelles qui peuvent intervenir dans la construction des AC (Renard, 2000). « Les premières définissent la structure spatiale du réseau d'automates, soit son nombre de dimensions, la forme des cellules (carrées, hexagonales, etc.) et le mode de détermination du voisinage. Les secondes déterminent le nombre d'états et les règles de transition. Le choix de ces deux types de règles permet de construire un univers adapté à l'objectif recherché » (Renard, 2000). Nous revenons dans cette sous-section sur l'aspect structurel de l'AC. L'aspect fonctionnel étant abordé dans la section suivante (1.3).

Dans notre démarche, l'objectif est d'analyser des données spatialisées, sans en perdre le sens et d'en extraire les règles qui régissent le processus d'urbanisation. Or, un moyen de

décrire l'espace est sa segmentation en entités présentant les attributs nécessaires à son analyse. C'est donc d'une analyse spatiale dont il s'agit. « L'analyse spatiale apparait comme une sorte de statistique spatiale appliquée à l'espace. [… Elle] consiste globalement à mettre en relation les données et leurs dispositions géographiques, à travers une démarche cartographique intégrant des techniques et des algorithmes d'interpolation et de lissage, voire d'intelligence artificielle » (Zaninetti, 2005 dans Antoni 2016). L'analyse spatiale « tente in fine d'établir des lois et des règles permettant d'expliquer des phénomènes » (Antoni, 2016) à travers la description de l'espace. C'est souvent une grille composée de cellules homogènes qui fait office de support de référence pour cette description, rappelant directement l'AC.

Les propriétés des AC ont été amenées à évoluer au cours du temps, dans certain cas par nécessité, parfois à des fins exploratoires. H. Couclelis résume dans sa publication de 1997 leurs principales caractéristiques, et permet de faire la distinction entre l'AC classique tout en répertoriant les évolutions développées (Couclelis, 1997). Bien que cette forme d'automate ne soit que rarement mobilisée actuellement, dans la modélisation du changement d'occupation du sol, nous nous attardons dans un premier temps sur la structure classique de l'AC pour revenir aux origines.

Automate cellulaire classique

Le paradigme de l'AC a particulièrement séduit par sa simplicité. Il rend la modélisation abordable dans une période où les ordinateurs n'offrent pas une puissance de calculs suffisante et nécessaire à l'utilisation de modèles qui présentent une structure plus complexe. Par ailleurs, un autre point attire l'attention des chercheurs : la possibilité de visualiser des phénomènes émergeants, des formes stables stationnaires ou mobiles. Ceci est rendu possible lors de l'étude de processus complexes, par l'exécution de règles simples. Le Jeu de la Vie en est un exemple.

Un AC classique est donc défini par les 5 critères de base (White *et al.*, 2015) :

- un espace cellulaire,
- un temps,
- des états,
- un voisinage,
- des règles de transitions.

Ces critères peuvent être formalisés à l'instar de la définition formelle de P. Torrens (2009) :

$$A \sim (S, R, N);\ S = \{S^1, S^2, \dots, S^k\};\ R\colon (S_t, I_t) \rightarrow\ S_{t+1}$$

Où *A* représente l'automate caractérisé par une liste finie d'états *S* ainsi que par la ou les règles de transitions *R*. *N* représente le voisinage contigu. La fonction de règles de transition permet l'évolution d'un état *S* d'un instant *t* à un instant *t+1*, cela étant donné l'information

spécifique des états des cellules du voisinage *N* à ce même instant *t*. Ces critères seront abordés chacun dans le détail par la suite.

Automate cellulaire et la modélisation des changements d'occupation du sol

La définition classique représente la formulation la plus simple et basique de ce que peut être un AC en géographie. Elle peut tout à fait être mobilisée pour simuler les changements d'occupation du sol. Néanmoins, la recherche a exploré d'autres alternatives dans le but de rendre plus précis les modèles développés pour l'étude des changements d'occupation du sol. Ils cherchent, dans la plupart des cas, à obtenir une représentation initiale et un modèle plus proche de la réalité. Les critères de bases ont donc évolué dans cette perspective, à mesure des avancées de la recherche, concomitantes à celle de l'amélioration de la capacité de calculs en informatique. La définition originelle, perçue comme « limitante », a été modulée suivant les besoins ou intentions en matière d'exploration de l'outil. L'AC peut alors être utilisé en s'affranchissant, par exemple, de l'uniformité des cellules composant la grille de base (Pinto & Antunes, 2010). Ainsi, la « grille » peut accueillir un maillage non rectangulaire. L'automate peut être contraint, en limitant par exemple le nombre de transitions ou en interdisant la transition dans une zone particulière. Il peut aussi être multi-niveaux, la cellule est alors par exemple définie à la fois par un état et une densité (White *et al.*, 2012). La majeure partie des adaptations du paradigme sont résumées par H.Couclelis (Figure 4. 6).

Quand les études présentent des AC qui s'éloignent de leur définition originelle, elles ont, par exemple, recours à la mention « basé sur » un automate cellulaire (*eng. CA-based modelling*). Les modèles présentés section 1.3 en sont d'ailleurs des exemples concrets.

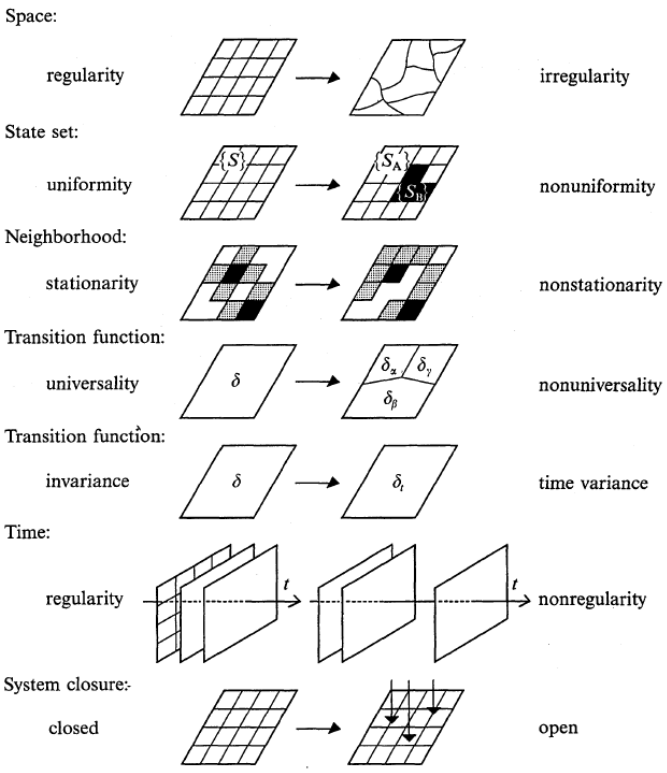


Figure 4. 6 : Généralisation commune des automates cellulaires (Couclelis, 1997)

Les composantes d'un AC

Carroyage

Le carroyage est une technique de quadrillage dont le but est de rassembler et de traiter des données en vue d'une exploitation cartographique ou statistique. Cette discrétisation de l'espace en cellules de formes carrées permet la représentation de divers phénomènes géographiques. Elle peuvent, en effet, incarner de nombreux « items » géographiques : des voitures (Nagel & Schreckenberg, 1992), une occupation du sol (Engelen *et al.*, 1995), un éventail d'écosystèmes (De Angelis & Gross, 1992) ou encore des microorganismes (Deutsch & Dormann, 2004). La cellule peut être régulière, comme dans la définition classique, ou irrégulière. Sa forme et sa surface peuvent aussi varier. Souvent carrée, elle peut tout aussi bien être triangulaire ou hexagonale (Figure 4. 7). Les cartes cadastrales, par exemple, sont composées de cellules dont la surface et la forme présentent une très grande variabilité puisque chaque cellule représente une parcelle spécifique. L'utilisation de cellules anisotropes permet d'augmenter la cohérence de la discrétisation vis-à-vis de la réalité. Pour autant, si un découpage non-isotrope peut s'avérer plus réaliste, la prise en compte de ce facteur augmente, dans la majeure partie des cas, le temps de calcul et complexifie la définition du voisinage (White *et al.*, 2015). De surcroit, ce cas s'avère limitant pour la réalisation de tests systématiques visant à évaluer la performance du modèle.

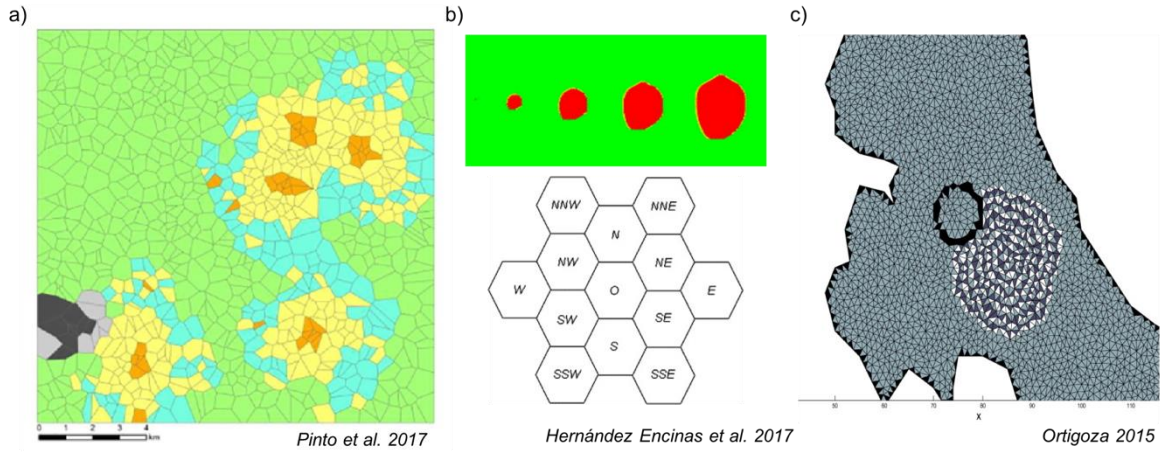


Figure 4. 7 : Échantillon de carroyage utilisé dans le cadre de simulations d'automates cellulaires. a) grille polygonale anisotrope basée sur les limites des différentes occupations du sol (Pinto et al., 2017) ; b) Carroyage hexagonal isotrope utilisé pour la simulation de la diffusion des feux de forêt (Hernández Encinas et al., 2007) ; c) Carroyage triangulaire anisotrope pour simuler la diffusion d'un phénomène géographique (Ortigoza, 2015)

Dans le cadre de cette recherche, le carroyage utilisé est une grille classique composée de cellules carrées (1ha). Cette forme est privilégiée par l'exploitation des données Corine Land Cover.

États

Un état se définit comme un ensemble d'informations qui caractérisent un objet à un moment donné. Appliqué à une cellule, il permet de spécifier la caractéristique principale de l'élément qu'elle représente. L'état des cellules est représenté visuellement au pas initial de la simulation et pourra être amené à changer au cours du temps. C'est aussi l'état, final cette fois-ci, qui sera présenté comme le résultat de la simulation. La définition classique de l'AC fait donc de l'état un point essentiel puisqu'en tant que caractéristique intrinsèque ou en tant que caractéristique de voisinage, il oriente la réponse (positive ou négative) d'une cellule à une règle de transition.

Dans l'étude de l'urbanisation, ou des changements d'occupation du sol, le nombre d'états considérés varie. Certains modèles comme MUP-City (Section 2.3) s'appuient, par exemple, sur une définition binaire de l'occupation du sol caractérisée par deux états : urbanisé ou non-urbanisé. D'autres modèles ou plateformes utilisent une caractérisation plus riche et plus nuancée des différents types d'occupation en distinguant généralement urbain, agricole, forêt, eau. Certaines classes peuvent être parfois précisées. Ainsi l'urbain peut être différencié en dense ou diffus lors d'études s'intéressant à la densité de bâti. De même, les surfaces artificialisées peuvent être précisées comme étant commerciales ou industrielles tout comme les types de végétations qui peuvent être précisés pour l'étude des dynamiques spatiales de certaines espèces. La présente étude s'appuie sur 8 états (Figure 4. 8) pour décrire l'occupation du sol sur la zone d'étude (Chapitre 3 section 2.3).

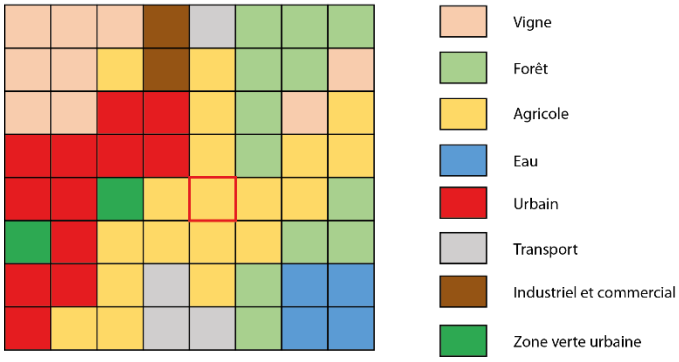


Figure 4. 8 : Automate cellulaire à 8 états (Judge, 2017)

L'état d'une cellule sera amené à changer au cours du temps suivant une règle de transition prenant en compte l'état initial ainsi que l'état des cellules présentent dans son voisinage.

Voisinage

Dès l'origine, la notion de voisinage dans un automate cellulaire a fait l'objet de nombreuses interrogations. P. Torrens (2011) définit le voisinage comme la zone entourant un automate, formée par des cellules spatialement interconnectées. Les voisinages de Moore et Von Neumann font partie des voisinages classiques. Ces derniers peuvent être combinés et/ou étendus. L'utilisation d'un voisinage ne comprenant pas uniquement les cellules adjacentes s'éloigne de la définition classique.

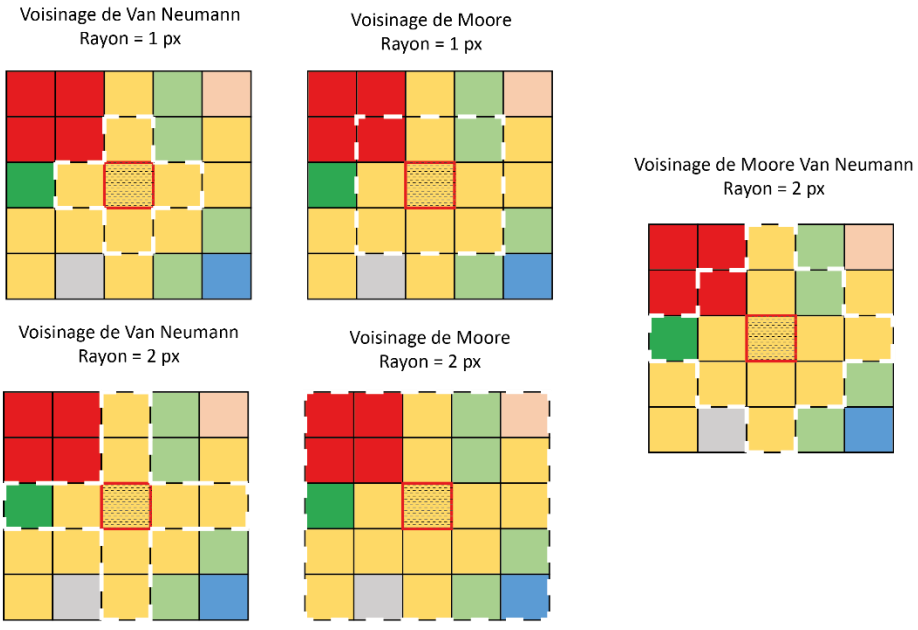


Figure 4. 9 : Les voisinages classiques et leurs extensions (Judge, 2017)

Les voisinages peuvent parfois également être définis par le bais de métriques représentant une distance d'interaction à partir de la cellule centrale. Ils ne sont donc plus définis par le

biais d'une cellule. Dans la définition classique de l'AC, le voisinage reste stable au cours de la simulation. Il peut néanmoins être également non-constant (Figure 4. 6).

Le voisinage est un élément indispensable dans l'écriture des règles de transitions régissant la dynamique du modèle. Cependant, sa forme ou son rayon influence les résultats des simulations (Kocabas & Dragicevic, 2006). Le type de voisinage retenu dans ce travail est celui de Moore étendu. Ce dernier maximise l'information sur l'occupation du sol avoisinante. En effet, l'un des objectifs de ce travail est d'étudier la définition du voisinage à travers son rayon et son occupation du sol sur la zone Franco-Allemande de Strasbourg-Kehl. Il convient alors de s'interroger sur sa taille optimale pour caractériser définir les caractéristiques d'un voisinage propice à la transition urbain au regard de l'occupation du sol.

Le temps

Le temps fait partie des composants essentiels de l'AC. Ce dernier, discret, introduit la dynamique et permet d'implémenter le changement en suivant les règles de transition. L'intervalle de temps, représenté par un pas de simulation, est généralement constant quand bien même H. Couclelis mentionne une possibilité de non régularité (Figure 4. 6). Ce changement est introduit de manière synchrone quand après examen de l'état de chaque cellule et l'estimation des transitions, tous les changements sont réalisés en même temps. Il est au contraire introduit de manière asynchrone quand les transitions sont réalisées les unes après les autres. Une cellule sélectionnée aléatoirement, ou d'après un choix particulier, est examinée puis effectue ou non une transition. L'action est répétée à tour de rôle pour chacune d'entre-elle. Ainsi, lors d'un même pas de simulation, une nouvelle transition aura pris en compte les transitions réalisées précédemment. Les résultats d'un même jeu de règles, suivant une simulation synchrone ou asynchrone seront donc probablement différents. Si le choix synchrone tend à s'éloigner de la réalité puisque tous les changements sont effectués en bloc, le paramétrage asynchrone pose également question. En effet, il nécessite d'estimer l'impact de la sélection aléatoire de la première cellule examinée.

Règles de transition

Les règles de transition représentent le « carburant » de l'AC. Il est possible de considérer les AC comme des machines universelles, à l'image de la machine de Turing. En effet, les règles peuvent prendre une infinité de formes (fonction, opérateur, expression, …), qui décriront comment l'automate/la cellule doit réagir considérant la donnée d'entrée (Torrens, 2009). L'état initial est toujours connu du modélisateur, qu'il représente une situation réelle telle que la répartition de l'occupation du sol, ou une situation fictive distribuée aléatoirement ou non, à des fins exploratoires. L'état final est, lui aussi, parfois connu pour devenir l'objectif à atteindre lors de la définition et de la calibration des règles à travers une approche rétrospective. Les règles de transition peuvent être définies manuellement, à travers l'utilisation d'outils statistiques (régressions, etc.), ou encore grâce à la mobilisation

d'algorithmes issus de l'intelligence artificielle. Il est nécessaire de dégager les règles qui permettent de décrire le processus dynamique étudié pour pouvoir l'analyser à travers la simulation. L'étape de définition et de calibrage des règles représente donc un des enjeux principaux dans la modélisation par AC, puisque ces mêmes règles permettent d'obtenir les résultats issus de la simulation ; résultats interprétés et analysés par la suite.

P. Torrens (2011) évoque le concept de « bonnes règles » qui, replacé dans la vision énoncée par W. Tobler (1970), seraient « simples » tout en permettant de simuler un processus complexe. Par conséquent, dans notre étude, une bonne règle est encadrée par le respect de deux contraintes conjointes :

(1) une règle doit pouvoir être interprétée thématiquement ;
(2) elle doit permettre de reproduire des phénomènes observés.

Or, « lorsque les systèmes urbains sont évoqués, nous n'avons souvent pas la moindre idée de ce que sont les "bonnes" règles » (Torrens, 2011). De ce fait, la définition des règles ainsi que leur calibration représentent une étape centrale de ce travail de modélisation – simulation. J. Van Vliet (2016), dans un état de l'art portant sur les méthodes actuelles de calibration et validation couramment utilisées pour les modèles de changement d'occupation du sol, constate qu'il n'existe pas, à ce jour, de méthode standardisée.

Dans son appréhension classique, la règle, ou le jeu de règles, s'applique indifféremment sur l'ensemble de la zone d'étude. Dans les évolutions du paradigme, leur application peut être restreinte à une zone en particulier. L'AC dans ce cadre devient multidimensionnel, d'un certain point de vue. H. Couclelis (1997) présente le système comme ouvert (Figure 4. 6). La dernière évolution répertoriée considérant la définition classique est la persistance du jeu de règles au cours de la simulation. Ce dernier peut être amené à changer à partir d'un certain seuil atteint (une certaine surface urbanisée, une date où interviendrait un changement politique ou un changement d'occupation du sol lié à l'élaboration d'un projet d'aménagement, etc.). C'est la variance temporelle du jeu de règles dont fait mention H. Couclelis (Figure 4. 6).

L'enjeu autour de la création et du calibrage des règles de transition est manifeste. Il s'agit de s'assurer de la performance des modèles créés pour simuler les processus étudiés. En effet, la modélisation des processus dynamiques spatiaux sous-tend généralement plusieurs objectifs : localiser, quantifier et parfois même comprendre et décrire le changement. Pour ce faire, différents modèles de simulation ont succédé aux premières théories mises au point en recherche sur la localisation et l'organisation des entités dans l'espace. La plupart d'entre eux sont, par ailleurs, basés sur un automate cellulaire. La sous-section suivante permet de revenir succinctement sur l'émergence des théories et sur l'émergence des premières approches modélisatrices des systèmes urbains, avant de présenter les caractéristiques de quelques modèles, parmi les plus connus, développés en géographie pour simuler les changements d'occupation du sol.

2. UN OBJECTIF GÉNÉRAL : LA MODÉLISATION

Modéliser l'occupation du sol et ses dynamiques a pour objectif, dans la majeure partie des cas, de quantifier et de localiser les processus à l'œuvre sur des espaces délimités. Cette démarche peut être étendue à la compréhension des processus, et à leur description de manière explicite. Pour ce faire, différentes méthodes de modélisation peuvent être mobilisées. Après un retour sur l'historique de leur développement et de leur utilisation, cette section présente les plus communes en géographie. Certains modèles populaires sont ensuite présentés de manière plus détaillée afin supporter les choix méthodologiques effectués dans cette recherche.

2.1. Historique

L'étude des changements d'occupation du sol et des systèmes urbains prend un nouvel essor avec le développement de l'informatique. La modélisation assistée par ordinateur est la clé de voûte qui permet l'élaboration de telles approches, notamment au regard de nouvelles possibilités d'acquisition en matière de données spatialisées. Auparavant principalement descriptive, l'étude des changements d'occupation du sol devient progressivement dynamique (White *et al.*, 2015).

C'est, en effet, par une approche moins dynamique que l'histoire de la modélisation urbaine débute au 19ème siècle avec J. Von Thünen (Von Thünen, 1826). Ce dernier élabora une théorie qui lie la composante agricole de l'occupation du sol à la valeur de son potentiel de production. Sa théorie prend également en compte les prix du marché de la denrée produite sur l'espace cultivable. Dans cette modélisation, l'occupation du sol *in fine* dépend à la fois de la distance à un lieu d'intérêt, ici celle du marché, du transport (temps et moyen) et de la productivité de la parcelle, donc de son potentiel. Ainsi, nous pouvons extrapoler que, dès les premiers modèles, l'occupation du sol dépend de caractéristiques intrinsèques de la zone considérée, mais aussi de son environnement.

Différentes approches et théories ont par la suite été élaborées sur le thème de l'étude des systèmes urbains, plus ou moins directement en phase avec la thématique de l'occupation du sol. Ainsi, W. Christaller (1933) et A. Lösch (1940) développaient dans les années 1930, chacun de leur côté, la théorie des lieux centraux présentée précédemment (Chapitre 2 section 2). A. Weber se concentre, quant à lui, sur les structures spatiales liée à la situation économique (Weber & Friedrich, 1929). Il publie pour cela une théorie sur la localisation optimale des industries manufacturières. Pour R. White (White *et al.*, 2015), ces quatre chercheurs sont à l'origine des théories de la localisation, de la géographie quantitative et des sciences régionales, posant ainsi des fondements que d'autres utiliseront pour élaborer leurs propres théories, telles que les approches d'Alonso (1964) et Muth (1969), qui elles-mêmes servirent de base aux travaux de Fujita (1986). Quelque que soit le champ de recherche (économie, géographie), la description des processus est basée sur l'élaboration de théorie qui permettent de décrire des interactions spatiales entre des éléments

géographiques. C'est ainsi, entre autres, à travers les liens spatiaux, que ces chercheurs tentent d'expliquer la localisation des entités géographiques. Ce constat renforce notre intérêt pour l'AC, puisque la dynamique des processus qu'il permet de simuler est exclusivement basé, dans sa définition classique, sur la relation spatiale d'une entité géographique avec celles localisées dans son voisinage.

Pour autant, ces divers travaux restent, en grande partie, théoriques. La formulation des hypothèses nécessaires à leur application rend délicat le transfert vers le test empirique. Par exemple, l'application du modèle d'Alonso est basée sur l'hypothèse d'une ville monocentrique implantée sur une plaine isotrope, dont le centre attire les emplois et la périphérie les résidences, soit des hypothèses fortes. L'application empirique nécessite une représentation plus explicite du cadre de modélisation. Cette étape peut induire la mise en place d'hypothèses restrictives rendant laborieuse la généralisation des modèles créés. Ils ne présentent alors un intérêt fondamental pour certains, qui préfèrent se focaliser sur l'exploration théorique plutôt que sur la réflexion qui légitime un test empirique. Les applications réelles sont donc presque inexistantes, excepté pour la théorie des lieux centraux. Néanmoins, l'objectif reste inchangé : apporter des éléments de compréhension sur l'allocation spatiale d'éléments géographiques, que ce soit celles des villes, des industries, ou des parcelles agricoles, à travers leurs interactions.

Outre leur orientation principalement théorique, les premières approches présentent également un point commun : elles n'utilisent pas encore la modélisation combinée à la simulation. Les premiers balbutiements de l'informatique, qui n'est alors pas très accessible, ni performant, peuvent en partie expliquer cette situation. Seul T. Hagerstränd fait office de précurseur à cet égard lors du développement de la *Time geography* (Hägerstrand, 1985). Il peut parallèlement être considéré comme un des pionniers de la modélisation individu-centrée (Hägerstrand, 1970). La géographie théorique et quantitative émerge alors. C'est notamment grâce au développement de la télédétection et des systèmes d'information géographique que le développement de modèles pour l'exploration des processus spatiaux dynamiques s'est démocratisé. Ces modélisations sont le plus souvent basées sur un environnement d'AC à travers l'application de technique issues des statistiques pour étudier les changements d'occupation du sol (White *et al.*, 2015).

Les différentes théories et approches présentées précédemment montrent l'intérêt des chercheurs pour la localisation des entités composant les systèmes urbains, ou plus généralement pour la description du monde et des processus notamment liés au développement anthropique. Ils cherchent à comprendre et organiser les facteurs dirigeant leur évolution. Mesurer, localiser, comprendre le phénomène étudié devient un enjeu majeur en géographie. Les chercheurs élaborent ainsi des théories dans un premier temps, puis des modèles empiriques permettant parfois même d'extrapoler des règles plus ou moins générales à partir du paramétrage des modèles et de l'observations des résultats. Pour ce faire, plusieurs types d'outils de modélisation peuvent être mobilisés ; ceux-ci sont développés au sein de la section suivante.

2.2. Outils et approches

L'intérêt pour l'étude de l'évolution dans le temps des systèmes urbains, parallèlement au progrès des technologies, a incité les chercheurs à mobiliser ou à développer des outils pour analyser spatialement les processus dynamiques. Deux principales familles d'outils retiennent notre attention : les automates cellulaires et les systèmes multi-agents (SMA). L'AC sert en effet de base dans la plupart des modèles qui permettent d'étudier les systèmes urbains. Pour autant, les SMA se sont eux développés à mesure de l'adoption d'approches plus fines, notamment basées sur des individus mobiles dans l'espace. Ces deux types d'outils sont en effet, les plus mobilisés dans la modélisation des processus spatiaux dynamiques en géographie. L'utilisation de l'un ou l'autre des outils peut être conjointe : des individus mobiles qui réagissent en fonction des changements de leur environnement simulé par AC. Pour autant chacun d'entre eux présente des caractéristiques qui lui sont propre.

Comme mentionné précédemment, plusieurs définitions s'appliquent à décrire les Automates Cellulaires dans la littérature(Benenson & Torrens, 2004; Engelen *et al.*, 1997; Torrens, 2011; White *et al.*, 2015; White & Engelen, 1993; Wolfram, 1984). En général, les AC sont considérés comme des outils composés d'automates permettant de reproduire des processus dont les caractéristiques changent dans le temps suivant leurs états en fonction de règles qui régissent l'aspect dynamique. Wolfram les décrits plus en détails sur la base de quatre caractéristiques : i – une représentation discrète de l'espace ; ii – chaque cellule est caractérisée par un état au sein d'une liste finie d'états ; iii – les règles de transitions régissant la dynamique dépendent de l'état initial de la cellule et des états des cellules présentes dans le voisinage ; iv – les états changent en fonction des mêmes règles de transitions sur l'espace considéré (Wolfram, 1984). L'avantage que présentent les AC est d'être intrinsèquement spatial et rapide (Benenson & Torrens, 2004; Kocabas & Dragicevic, 2006; Pinto & Antunes, 2010; White *et al.*, 2015). Les phénomènes spatiaux peuvent être directement cartographiés à travers un espace cellulaire. R. White (2015) envisage même les AC à deux dimensions comme des Systèmes d'Information Géographique (SIG). Il avance l'argument qu'un raster exploité dans un SIG n'est pas différentiable d'un AC à un certain instant t. Pour lui, l'AC permet d'ajouter très simplement un processus dynamique au raster SIG en lui indiquant l'évolution à suivre à travers l'écriture de règles de transitions. La structure du raster associée à des règles simples, et un voisinage fixe, font de l'AC une alternative efficace en termes de calculs informatiques.

Les modèles basés sur des systèmes multi-agents (SMA) présentent également une alternative reconnue pour l'étude des interactions du couple humain-environnement, tel que l'occupation du sol. Les différences majeures entre le SMA et l'AC sont : la modélisation explicite des processus de décision, les interaction des acteurs à l'échelle individuelle (Sun *et al.*, 2016), et le caractère mobile dans l'espace des individus modélisés. Les SMA peuvent donc être adaptés pour modéliser des systèmes urbains car ils permettent une description plus précise de l'environnement de modélisation comme dans le cas des modèles de trafics.

Ils sont la plupart du temps couplés à un AC (White *et al.*, 2015). Cette approche de la modélisation peut, néanmoins, s'avérer limitante suivant l'échelle d'étude du système urbain. Le fait de modéliser chaque individu induit une phase de paramétrage et de calibrage plus conséquente que la normale, toujours en termes de ressources informatiques. Comme l'individu ne peut pas être étudié sans intégrer des choix dépendant de probabilités, les modèles doivent être lancés un certain nombre de fois dans le but de cartographier une convergence de résultats. Le temps de simulation dépendra donc du nombre d'individus considérés et du nombre d'interactions étudiées. Cet aspect peut notamment engendrer des temps de calculs importants. Par exemple, une modélisation individus centrés sur notre zone d'étude nécessiterait une mobilisation de ressources informatiques très conséquente. C'est également l'agrégation des résultats qui peut poser problème puisque l'objectif n'est pas d'étudier un individu en particulier mais bien les résultats de la combinaison des comportements d'un ou plusieurs groupes d'individus, à l'échelle méso ou macro-spatiale. L'échelle d'étude peut donc s'avérer déterminante quant au choix de l'outil de modélisation utilisé. La dernière limite selon White (2015) est la difficulté finale d'associer l'émergence d'un phénomène à une des composantes du SMA, or l'objectif ici serait d'apporter une information explicite lors de l'analyse des résultats.

C'est bien l'objectif poursuivit dans cette recherche à travers la modélisation du processus de développement urbain qui oriente principalement notre choix vers l'AC : l'élaboration d'une méthode simple, rapide et explicite permet de différencier une composante locale d'une composante généralisable à la totalité de la zone d'étude, lors de la caractérisation du développement urbain. Par ailleurs, l'étendue de notre zone frontalière, semble également orienter logiquement notre choix vers l'AC.

Différents modèles ont été élaborés, différents outils ont été mobilisés en géographie dans ces perspectives. La sous-section suivante propose de s'attarder sur les modèles de systèmes urbains les plus populaires et les plus communément utilisés.

2.3. Modèles

Depuis plusieurs décennies, la recherche en géographie s'intéresse au développement de modèles à des fins exploratoires pouvant ou non servir de support d'aide à la décision. Ainsi SLEUTH (Clarke & Gaydos, 1998), Metronamica (van Delden *et al.*, 2010), SpaCelle (Dubos-Paillard & Langlois, 2018), Land Change Modeler (Eastman, 2009), MUP-City (Frankhauser *et al.*, 2010; Tannier *et al.*, 2010) en sont les modèles les plus connus. Nous nous attachons dans la suite de cette section à présenter Metronamica, SLEUTH, LCM, MUP-City car ces modèles sont, pour chacun d'entre eux, représentatifs d'une approche particulière de la modélisation, tant par la méthode mobilisée que par les hypothèses qu'ils adoptent. En effet, plusieurs approches de la modélisation sont adoptées en géographie : une approche rétrospective cherchant à visualiser et comprendre les processus dans le passé (t-1 à t), une analyse de la situation actuelle (t), ou encore une approche prospective (t à t+1). Cette dernière peut être effectuée à travers l'application de scénarios, suivant un modèle issu

d'une analyse rétrospective, agrémenté ou non de modifications liées à la connaissance préalable d'un futur plan d'aménagement (continuation de tendance, exploration de scénarios envisagés), ou appliqué directement à l'instant t en s'appuyant sur l'exploration de logiques spatiales comme celles du développement fractal.

Quelle que soit l'approche, la majeure partie de ces outils évolue dans le cadre élargi d'un AC. Ils mobilisent des méthodes de modélisation variées, s'appuyant sur un jeu de données plus ou moins fourni. Les étapes de calibrage et de validation sont, elles aussi, variées. Si la majorité des modèles s'appuient sur une analyse rétrospective des évolutions, ce n'est pas l'unique possibilité. Il est également possible d'élaborer des modèles sur la base de postulats (dire d'expert, hypothèse de développement fractal, etc.) puis de les utiliser dans le cadre de scénarios d'urbanisation dont les objectifs sont explicites. Une présentation succincte de ces modèles et de leurs caractéristiques principales permet de considérer un échantillon des possibles en termes d'approches et de modélisation.

Metronamica

Metronamica a été développé par le RIKS Institut au Pays-Bas (Delden & Hurkens, 2011; RIKS, 2009). Cet outil de modélisation, à portée prospective, est orienté sur l'analyse de la mise en place de politiques d'aménagement. Il permet d'établir une évolution spatiale prévisionnelle des effets de l'application de réglementations en matière d'aménagement urbain, à l'échelle locale ou régionale. Il intègre ainsi des informations sur les facteurs qui influencent les changements d'occupation du sol (augmentation de la population, changement économique, etc.) tout en spatialisant la réglementation en vigueur ou en projet (politique de densification, zones contraintes ou indisponibles au développement d'après les textes réglementaires, etc.). Les décideurs peuvent alors tester le résultat de scénarios d'aménagement, ou la mise en place de politiques locales ou régionales.

Metronamica se base sur un environnement AC. Le processus d'allocation des changements d'occupation du sol est dirigé par un principe de « compétition pour l'espace » entretenu par les différents types d'occupation du sol. Le modèle adopte une approche imbriquant plusieurs échelles. Il inclut, à l'échelle globale, les changements démographiques et le potentiel en matière d'emplois dans un modèle macro-économique. L'échelle régionale est caractérisée par les changements socio-économiques influencés par l'attractivité des différentes parties de la zone étudiée. L'attractivité est liée à l'accessibilité sur la zone définie par un modèle de transport. Ainsi, à l'échelle locale, les changements d'occupation du sol sont influencés par les facteurs définis aux échelles supérieures, induisant la caractérisation de zones comme ayant un potentiel de développement supérieur à d'autres de par leur attractivité (localisation avantageuse, topographie, réglementation, etc.). Les différentes échelles sont interdépendantes. Leurs facteurs respectifs sont donc liés par des boucles de rétroaction.

Le modèle de changement d'occupation du sol est basé sur celui développé par R. White et G. Engelen (1993) au sein duquel quatre facteurs sont pris en compte lors de l'évaluation du potentiel de changement :

- la pertinence physique, précisant si la localisation peut accueillir l'activité ;
- le zonage réglementaire, décrivant la possibilité ou l'interdiction en matière de développement ;
- l'accessibilité, énonçant si le changement peut être effectué au vu de l'accessibilité ;
- le comportement humain, indiquant aux différents acteurs le comportement à adopter à travers un jeu de règles dépendant de la localisation la plus attractive.

Chacun des trois premiers facteurs cartographiés est démultiplié en fonction du type de sol concerné. De ce fait il est possible d'analyser parallèlement plusieurs types de changements d'occupation du sol. L'interface utilisateur du logiciel intégrant les différents aspects de Metronamica permet d'extraire des indicateurs socio-économiques et environnementaux. Le calibrage et le paramétrage nécessitent un certain nombre d'informations, notamment la disponibilité de données diachroniques qui permettent l'analyse des tendances de changements à travers le temps. Des métriques déjà intégrées peuvent être mobilisées dans l'analyse des données d'entrée. Le calibrage consiste à élaborer un « modèle neutre » de référence (Hagen-Zanker & Lajoie, 2008) et permet d'analyser les effets des différents paramétrages à travers la comparaison de leurs résultats vis-à-vis de ceux du ou des modèles neutres. D'autres mesures, notamment l'évaluation visuelle des résultats, complémente l'analyse de la performance du modèle au regard de la phase de calibrage, notamment grâce au *Map Comparison Kit* (Visser & de Nijs, 2006). Metronamica s'appuie également sur la précision de la prédiction réalisée lors de la simulation pour évaluer la performance du modèle, ainsi que sa capacité à reproduire un processus cohérent avec les processus réels observés. Pour finir, une série d'indicateurs peut être utilisée pour évaluer la performance du calibrage, ainsi le « kappa flou » (van Vliet *et al.*, 2013a), un indicateur de concordance dans la structure du paysage (« clumpiness index »), l'évaluation de la dimension fractale des structures urbaines (Frankhauser *et al.*, 2010), la distribution Rang-Taille sur la zone simulée, l'élaboration du facteur d'enrichissement (van Vliet *et al.*, 2013b), ou encore une vérification visuelle de la cohérence des résultats. L'étape de calibrage est suivie comme dans la majeure partie des cas, par une étape de validation. Celle-ci, dans Metronamica, s'appuie sur les mêmes indicateurs que ceux utilisés dans la phase de calibrage. Pour cause, les étapes de calibrage et de validation sont difficilement dissociables puisque l'évaluation des résultats issue d'une calibration particulière mènera à réitérer l'exploration en fonction de leur adéquation avec la performance nécessaire à l'utilisation du modèle.

SLEUTH (Slope, Land Cover, Excluded, Urban, Transportation, Hillshade)

SLEUTH a été développé aux États-Unis (Clarke & Gaydos, 1998). Comme Metronamica, il s'appuie sur une base AC pour simuler explicitement l'expansion urbaine. Des règles qui représentent l'évolution souhaitée sont élaborées à l'aide de données d'entrée. Son nom est par ailleurs tiré des données d'entrée nécessaires à son utilisation : la pente, l'occupation du

sol, les zones exclues de l'urbanisation, les zones urbanisées, le réseau de transport, et l'ombrage. SLEUTH ne permet pas l'utilisation de données d'entrées aussi variées que Metronamica, néanmoins, ce point en particulier favorise sa notoriété (Triantakonstantis & Mountrakis, 2012). Les prérequis minimums en matière de données d'entrée sont : une donnée SIG diachronique de la pente, des zones exclues de l'urbanisation et de l'ombrage ; au moins deux couches SIG présentant les réseaux de transports et l'occupation du sol ; au minimum quatre couches SIG présentant les zones d'extension urbaines. Le modèle permet de mobiliser quatre types de jeux de règles associés à un type de développement particulier : l'étalement diffus, la diffusion de nouveaux centres éparpillés, l'étalement naturel et l'étalement influencé par la localisation des routes (Clarke & Gaydos, 1998). Ces comportements sont déterminés par 5 coefficients dont la valeur varie entre 0 et 100. La performance de la calibration des coefficients est évaluée à travers 13 indicateurs moyennés sur une série d'itérations de Monte Carlo (Clarke, 2018). Auparavant paramétré pour utiliser la « force brute »[32] pour définir la meilleure combinaison de paramètres, la dernière version de SLEUTH s'appuie sur l'utilisation d'un algorithme génétique (Clarke-Lauer & Clarke, 2011). En effet, le recours à l'algorithme génétique permet de réduire considérablement le temps de calcul nécessaire à la calibration en comparaison avec celui indispensable à l'utilisation de la « force brute ». La validation est également une étape réalisée dans les études mobilisant SLEUTH. Ainsi K.C. Clarke (2018) décrit le modèle comme étant un des plus validés. Il précise que ce dernier est évalué sur la précision des prédictions, qui peut atteindre plus de 80% lors de la phase de calibration.

MUP-City

MUP-City est un modèle développé au sein du laboratoire ThéMA, pour l'étude de la croissance résidentielle uniquement (Tannier *et al.*, 2010). L'état des cellules ne représente que cette caractéristique binaire, sans distinction aucune concernant les autres types de sols. Les données de bases nécessaires à MUP-City sont : le réseau routier, le bâti existant, les zones non-constructibles. Elles permettent de définir un référentiel spatial et de caractériser l'accessibilité sur la zone étudiée. Des données plus fines sont également importantes. Ainsi les commerces et services, les espaces verts et de loisirs, sont intégrés. De plus, trois niveaux de recours potentiels aux aménités sont définis en amont, à savoir : quotidien, hebdomadaire, mensuel ou plus rare (Frémond, 2015). La surface urbanisée est indiquée par l'utilisateur. La création des règles régissant l'évolution de l'urbanisation d'un instant t à un instant t+1 est le fruit d'une appréhension fractale de l'évolution de la morphologie urbaine (Frankhauser *et al.*, 2010). L'aspect multi-échelles est pris en compte, inhérent à l'application d'une logique fractale. Dans MUP-city les mêmes règles sont appliquées quelle que soit l'échelle.

[32] Test de l'ensemble des possibles concernant les combinaisons de paramètres (Dietzel & Clarke, 2007)

Land Change Modeler (LCM)

LCM est quant à lui disponible à travers le logiciel de télédétection IDRISI©. La création d'un modèle sous LCM est également basée sur une analyse rétrospective des changements. Deux familles de méthodes sont mobilisées pour la création du modèle, celles issues des statistiques (régression logistique, etc.) et celles issues de l'intelligence artificielle (réseaux de neurones). Comme Metronamica, les données mobilisées s'étendent aux données démographiques, au réseau routier, à la proximité des services et aménités urbaines et à l'emploi, entre autres (Aguejdad *et al.*, 2016). La localisation des pixels ayant changé d'occupation du sol est étudiée à travers les variables explicatives mentionnées précédemment. L'utilisateur peut établir une relation mathématique en utilisant une des trois méthodes disponibles :

- Régression logistique (Kleinbaum, 1994),
- Méthode SimWeight (Sangermano *et al.*, 2010),
- Perceptron multicouches (famille des réseaux de neurones).

Dans le cadre de scénarios exploratoires, des projections markoviennes peuvent être déterminées sur la base de l'estimation des probabilités de transition d'après l'analyse diachronique. La quantité de changements peut tout aussi bien être directement définie par l'utilisateur, comme dans MUP-City. LCM ne s'appuie donc pas sur un environnement d'automate cellulaire mais présente des caractéristiques communes à celles des autres modèles, pour un même objectif général, à savoir la modélisation d'un processus spatial dynamique en l'occurrence le changement d'occupation du sol.

Tout en ayant un objectif commun, celui de la modélisation de processus spatiaux pour une meilleure compréhension et adaptation aux changements, les modèles présentés dans cette sous-section proposent tous une approche particulière. Les données d'entrée indispensables et optionnelles ne sont pas identiques. Les modèles prennent parfois en compte des données réglementaires ou s'en affranchissent, nécessitent des données diachroniques ou non, etc. MUP-City à la différence de Metronamica ou SLEUTH, ne dépend pas d'une analyse rétrospective. Comme Metronamica, l'aspect multi-échelle est pris en compte, inhérent à l'application d'une logique fractale. Néanmoins, l'approche diffère dans l'application des règles. Dans Metronamica, différentes règles sont appliquées aux différentes échelles étudiées, alors que pour MUP-City les mêmes règles sont appliquées quelle que soit l'échelle. Certains modèles, comme LCM, sont également directement implémentés dans des logiciels propriétaires. Ce point peut entrer en considération lors de la sélection de l'outil, puisque l'utilisation de ce modèle nécessite donc une licence et impose un cadre de travail spécifique (données, processus de calibrage, etc.).

La description du fonctionnement et des méthodes mobilisées par les différents modèles permet d'illustrer certaines différences fondamentales dans le développement ainsi que dans les hypothèses sous-jacentes (développements fractals, analyses rétrospectives et poursuite de tendances, projections basées sur des chaînes de Markov, etc.). Ces exemples indiquent

la diversité des approches qui reposent sur un environnement de type automate cellulaire composé d'un espace discret (grille), d'une définition des états de cellules, et parfois de règles de voisinage. Pour autant, peu de modèles mobilisent, de nos jours, les AC dans leur forme fondamentale. Leur développement a été adapté aux phénomènes étudiés dans les différents champs de recherches se focalisant sur l'exploration de processus dynamique spatiaux.

La sélection de l'AC comme outils de modélisation dans cette recherche repose sur son caractère intrinsèquement spatial et sur le lien direct pouvant être mis en place avec les Systèmes d'Information Géographique (White *et al.*, 2015). L'AC permet à la fois d'étudier les changements en quantité, moyennant la mobilisation de données adéquates, mais aussi de les localiser. Cet outil, par son approche simple à appréhender, est également en totale adéquation avec la démarche initiée dans ce travail, à savoir le développement d'une méthodologie simple pour pouvoir faciliter la communication des résultats à un public, qu'il soit expert ou non. Nous privilégierons donc ici une approche plus simple de la modélisation basée sur les AC, que celles liées aux modèles présentés précédemment.

Si l'approche initiée dans ce travail de recherche se veut plus basique que celles des modèles présentés dans cette sous-section, l'enjeu autour de la création et le calibrage des règles de transition est manifeste. Ce dernier est principalement lié aux données et à l'outil mobilisé. Ces points font l'objet de la section suivante.

3. UNE IMPLÉMENTATION ADAPTÉE : LUCSIM

R. White pose l'approche algorithmique comme étant « la clé pour modéliser les villes comme un système auto-organisé » (White *et al.*, 2015). Nous étendons ici ces propos à la modélisation de l'occupation du sol et plus particulièrement au développement urbain. Ce dernier doit être considéré comme un processus et non comme un artefact (White *et al.*, 2015). L'algorithme est le langage informatique permettant de modéliser des processus. C'est, par définition, une suite finie et non ambiguë d'instructions permettant de résoudre un problème. Appliqué au développement urbain, il permet à l'aide d'une suite d'instruction de modéliser des processus d'urbanisation. Il incorpore par ailleurs une dimension temporelle dans la modélisation en raison du caractère dynamique des processus dont il est question de représenter l'évolution spatiale. Les algorithmes sont donc utilisés comme une méthode de modélisation des processus dynamiques. Leur utilisation nécessite une phase de calibrage. En effet, toute création de modèle s'effectue par la réalisation des étapes suivantes :

- sélection des données,
- sélection d'un outil,
- sélection d'une méthode,
- création et calibrage du modèle,
- validation du modèle.

A fortiori, les choix effectués à chaque étape dépendent du contexte et des objectifs de la modélisation. Les choix relatifs à cette recherche ont été argumentés dans des sections précédentes : la sélection des données dans le chapitre 2, la sélection de l'outil dans le présent chapitre.

L'étape de création et de calibrage du modèle, qui se matérialise par la création et le calibrage de règles pour un automate cellulaire, peut prendre différentes formes et présente un enjeu majeur. Cette étape doit être en adéquation avec les objectifs de recherche. Pour se faire, cette section revient sur les pratiques généralement utilisées dans la définition et le calibrage des règles de transition pour l'AC. La première sous-section se focalise notamment sur la manière de créer ces règles et de les calibrer. La deuxième sous-section introduit l'automate cellulaire LucSim développé au sein du laboratoire ThéMA, utilisé et adapté dans le cadre de cette thèse, afin répondre aux besoins spécifiques de la méthodologie développée dans cette recherche. La troisième sous-section fait état des méthodes de création et de calibrage des modèles permettant la simulation des changements d'occupation du sol.

3.1. Données

L'élaboration du jeu de données est étroitement liée à la donnée utilisée ainsi qu'à l'outil de modélisation sélectionné. Notre recherche s'est orientée vers l'exploration des données libres disponibles sur la zone d'étude et harmonisées à l'échelle transfrontalière (Chapitre 2).

Notre jeu de donnée est dès lors composé de trois types de données pour chaque cellule :

- Occupation du sol

Il est nécessaire, dans un premier temps, d'incorporer l'état initial de la cellule, ici l'occupation du sol en 1990, ainsi que son état final, ici l'occupation du sol en 2006. Ces deux variables permettent de localiser les transitions urbaines effectives entre les deux dates.

- Voisinage

Le voisinage est défini d'après le type d'occupation du sol des cellules localisées autour de la cellule décrite. Une opération de comptage est effectuée pour répertorier la quantité en hectare de chaque type d'occupation du sol tous les 100 mètres sur une étendue de 3 000 mètres autour de chaque cellule. Le choix de décrire le voisinage à l'échelle du pixel (tous les 100 m) permet d'obtenir une description précise. La méthode a permis de fixer la taille du voisinage et cette question est abordée plus en détails dans le chapitre 6.

- Localisation géographique

Il est communément admis que l'accessibilité aux pôles d'activité joue un rôle dans le développement urbain (Chapitre 2). Chaque cellule est donc définie par son accessibilité aux pôles de niveau 1 et 2 de la zone d'étude (Chapitre 3 section 3). Dans le cadre de ce travail, quatre façons d'appréhender le paramétrage de l'accessibilité d'une cellule sont étudiés en

parallèle. Chaque cellule est donc caractérisée par quatre variables d'accessibilité. La localisation géographique dans l'un ou l'autre des États est aussi retenu, ainsi que celle qui place la cellule dans la partie Nord ou dans la partie Sud de la zone d'étude.

Chaque cellule de la zone d'étude est donc caractérisée par plusieurs variables qu'il sera possible de mobiliser pour définir les règles de transition appropriées à la simulation du processus de développement urbain (Figure 4. 10).

Occupation du sol			Voisinage											Localisation			
			1 pixel (0-100 m)								2 pixels (100-200 m)						
ID	CLC 90	CLC 06	Urbain	Agri.	Forêt	Vignes	Indus.	ZRU	Eau	Transp.	Urbain	Agri.	{...}	IDP	Access1	Access2	{...}
0	4	4	2	2	0	0	2	1	0	1	4	4	{...}	0	12	23	{...}
1	4	4	0	4	2	0	1	1	0	0	4	5	{...}	1	34	76	{...}
2	4	8	1	2	4	0	1	0	0	0	5	9	{...}	0	23	12	{...}
3	4	2	5	0	2	0	0	0	1	0	9	0	{...}	1	98	32	{...}
4	5	3	0	3	2	1	0	0	0	2	4	2	{...}	0	40	25	{...}
5	9	6	1	0	7	0	0	0	0	0	3	0	{...}	0	23	17	{...}
6	4	8	0	3	2	0	0	0	1	2	0	0	{...}	0	65	95	{...}
7	3	0	0	2	4	0	0	1	1	0	0	3	{...}	1	13	65	{...}
{...}	{...}	{...}	{...}	{...}	{...}	{...}	{...}	{...}	{...}	{...}	{...}	{...}	{...}	{...}	{...}	{...}	{...}

Agri. : Agricole
ZRU : Zone récréative urbaine
Indus. : Industrie et commerce
Transp. : Transport
IDP : Indice de pays
Access 1 : Temps d'accès aux pôles de niveau 1
Access 2 : Temps d'accès aux pôles de niveau 2

Figure 4. 10 : Structure du jeu de données

Nous pouvons, par exemple, observer que la cellule identifiée par l'ID 0 n'a pas changé d'occupation du sol entre 1990 et 2006, puisque son état d'occupation est 4 dans les deux cas. Elle présente dans son voisinage à 100m, 2 ha de type urbain, 2 ha de type agricole, 2 ha de type industriel ou commercial, 1 ha de type zones récréatives urbaines et 1 ha de type transport. Cette description correspond donc à celle des occupations du sol dans un voisinage de 100m. La même information est présente pour définir le voisinage à 200m, ce tous les 100m jusqu'à 3 km. Parallèlement, la cellule dont l'identifiant est 0, est également définie par sa localisation sur la zone d'étude. Dans un premier temps, grâce à l'Indice de Pays, nous savons que cette cellule est localisée en Allemagne (IDP égal à 0), mais aussi à 12 min du pôle de niveau 1 et 23 min d'un pôle de niveau 2. La même information est disponible pour localiser la cellule au Nord ou au Sud de la zone, et relativement aux accessibilités pondérées aux pôles de niveaux 1 et 2.

L'extraction des données nécessaires à l'étude de notre zone, sous la forme présentée Figure 4. 10, est effectuée à l'aide du logiciel LucSim, qui permet notamment d'automatiser le comptage de la surface des types d'occupation du sol présent dans le voisinage de chaque cellule. Ce logiciel est présenté dans la sous-section suivante.

3.2. Logiciel

LucSim est un automate cellulaire dédié à l'analyse des dynamiques de changement d'occupation du sol par le biais d'une interface utilisateur adaptée pour un large public (expert ou non). Développé au sein du laboratoire ThéMA, sur la base d'un logiciel préexistant (CWS/Camdeus) ce dernier intègre la méthodologie développée par J.P Antoni pour l'exploration des changements d'occupation du sol sur le Territoire de Belfort (Antoni, 2003). LucSim permet de développer une approche purement experte dans l'écriture des règles de transitions. Il permet également de définir une matrice de transition entre les différentes occupations du sol, sur la base d'une analyse diachronique. Plusieurs modules peuvent ensuite être mobilisés, à des fins exploratoires ou dans le cadre de la définition de scénarios de développement. Un module intégrant les chaines de Markov permet de contraindre l'AC pour une approche prospective de l'application. Un second module permet de mettre en place un modèle de potentiel basé principalement sur le principe d'interaction spatiale. Pour finir, LucSim intègre un module spécifique pour la création automatique de jeux de règles de transitions. Ce module additionnel est directement issu des travaux réalisés dans cette recherche doctorale et a été spécifiquement développé pour mener à bien l'étude cas (Chapitre 5).

Après avoir construit le jeu de données, la question consiste à définir correctement les variables à mobiliser dans l'écriture des règles de transition. Cette étape est non seulement fondamentale pour l'élaboration des règles permettant la simulation du développement urbain sur la zone d'étude, mais rend également possible l'analyse d'une différence entre le développement localisé en France et celui localisé en Allemagne. En effet, l'implication de la variable de localisation géographique au sein d'un État en particulier, lors de la définition automatique des règles, induirait une particularité nationale donc une distinction liée à la frontière. En résumé, l'étape de définition et de calibrage des règles de transition permet déjà d'apporter une réponse au questionnement soulevé dans ce travail de recherche, à savoir l'influence de la frontière en tant que limite administrative, sur la définition des processus d'urbanisation.

L'étape de création et de calibrage reste encore aujourd'hui l'un des enjeux principaux lors du développement d'un modèle de simulation urbaine puisqu'elle influence en grande partie la performance du modèle et la qualité des résultats interprétés (Torrens, 2011; van Vliet *et al.*, 2016). Un certain nombre de méthodes de calibration ont été développées depuis l'utilisation des AC dans la modélisation des processus urbains. Ces dernières font l'objet de la sous-section suivante.

3.3. Calibrage

La création des règles régissant la dynamique des processus urbains explorée par le biais des AC peut prendre différentes formes. L'étape de création est suivie par une étape de calibrage. Cette dernière comprend l'ajustement des paramètres utilisés dans la structure du

modèle, permettant *in fine*, d'obtenir un résultat « optimal ». La notion d'optimal reste relative aux objectifs de modélisation. Le calibrage est un processus itératif puisqu'il sera réitéré jusqu'à obtenir un modèle dont les résultats sont considérés comme satisfaisants. Il est ainsi étroitement lié à l'étape de validation qui permet, elle, d'évaluer la performance du modèle (Figure 4. 11).

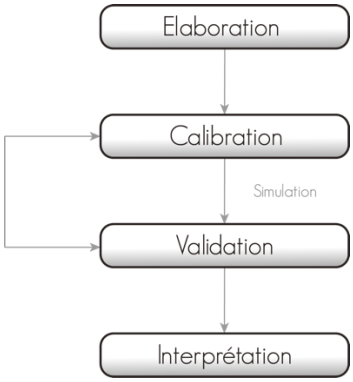


Figure 4. 11 La calibration - un processus itératif (Judge)

Les sources d'information utilisées pour la modélisation du phénomène spatial orientent la démarche de calibrage. J.M. Mas (2018) identifie deux approches majeures : le dire d'expert et l'approche par la donnée. La première permet d'obtenir l'information auprès d'experts ou individus susceptibles d'apporter des éléments de compréhension sur la base de leurs connaissances au regard du processus étudié. La seconde impose l'analyse des données pour en extraire l'information pertinente à l'étude de ce même processus. Le calibrage, à dire d'expert, est donc entièrement basé sur la connaissance de l'expert du champ de recherche exploré à travers la modélisation (Antoni *et al.*, 2018). Il peut dépendre de l'expertise du modélisateur ou encore de celle d'institutions. L'utilisation d'une approche à dire d'expert rend la méthodologie difficilement reproductible et transposable puisque les informations utilisées sont extrêmement spécifiques à la zone étudiée. Au regard de l'étude de la zone frontalière, cette démarche induirait la prise en compte en amont de la différence existant entre la France et l'Allemagne, ou la prise en compte d'experts français et allemands qui ne partageront peut-être pas la même vision. Cette démarche n'est donc pas la plus efficace pour observer ou non l'émergence de cette différence, et ne sera pas mobilisée dans ce travail.

J. van Vliet (2016) identifie plus précisément six approches différentes, dont celle à dire d'expert, pour effectuer l'étape de calibrage des modèles de changement d'occupation du sol (Table 4. 1).

Table 4. 1 Calibrations et modèles de changement d'occupation du sol (van Vliet et al., 2016)

Table 1
Calibration approaches used in the reviewed model applications (N = 114). Some applications are calibrated using a combination of approaches, therefore, the values do not add up to 100%.

Calibration approaches	Applications
Based on expert knowledge	25%
Manual calibration	8%
Automated procedures	36%
Statistical analysis	44%
Based on another application	15%
Calibration not reported	18%

La revue de littérature menée par J. van Vliet et ses collaborateurs porte sur 114 articles publiés entre 2010 et 2014, sur les modèles de changement d'occupation du sol appliqués à des cas d'études réels. Elle montre une prédominance des procédures de calibrages statistiques et automatisées. Le calibrage à dire d'expert reste cependant important. Les publications antérieures n'ont pas été analysées, car les auteurs considèrent qu'elles pouvaient présenter un biais puisque « la validation des modèles de changement d'occupation du sol n'est arrivée à maturité que récemment » (van Vliet *et al.*, 2016).

L'importance des procédures de calibrages statistiques et automatiques est liée à l'évolution de la puissance de calculs des ordinateurs corrélée à celle de l'utilisation d'un nombre, ou d'un volume, toujours plus important de données. L'exploitation d'un volume important de données nécessite le recours aux statistiques pour extraire la part significative d'informations contenues dans l'ensemble. Cette étape ne peut plus être effectuée manuellement. L'utilisation de procédures automatisées est également favorisée par le développement d'outils d'apprentissages automatiques dans le domaine des sciences informatiques et dans celui de l'intelligence artificielle. La communication des résultats, notamment par le biais des revues scientifiques, permet l'émergence d'applications extrêmement diverses d'une même méthode d'analyse ou d'apprentissage et favorise d'autant plus leur utilisation. Les procédures d'apprentissage automatisé ou d'analyses statistiques apparaissent également à plusieurs reprises dans les travaux de synthèses de P. Ghosh (Figure 4. 12). Ce type de procédures est donc particulièrement intéressant pour extraire l'information significative contenue dans un jeu de donnée.

Cette information sert ensuite de base à la rédaction des règles de transition pour l'AC. Les règles peuvent également être extraites directement sous forme de jeu à travers le processus d'apprentissage automatique. C'est par exemple le cas lors de l'utilisation d'algorithmes utilisant les réseaux de neurones ou les arbres de décisions.

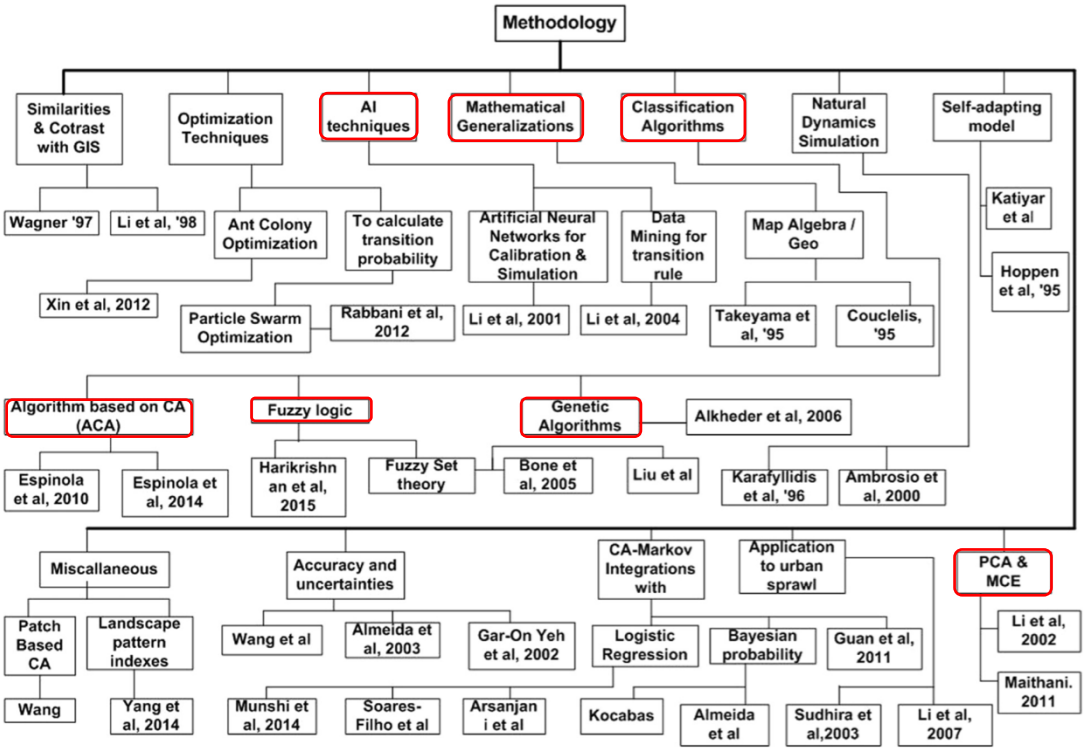


Fig. 1. Methodological developments in CA-Markov Chain in geospatial environmental modeling.

Figure 4. 12 : : Développements méthodologiques appliqués aux AC contraints par les chaînes de Markov, dans la modélisation de l'environnement géospatial (Ghosh et al., 2017)

Pour autant, l'influence du modélisateur est toujours omniprésente puisqu'il décide des données initialement à disposition lors de l'apprentissage ou de l'analyse statistique. Néanmoins, au contraire du dire d'expert, le modélisateur crée de l'information par l'utilisation de ces techniques. Il n'intègre pas directement l'information dont on observe le résultat lors de la simulation. Les résultats spatialisés issus du modèle créé par l'apprentissage automatique ou par l'analyse statistique sont également observés par la suite. Leur cohérence est établie à travers l'étape de validation mentionnée précédemment.

L'apprentissage permet donc d'effectuer les deux tâches automatiquement : création et calibrage. L'objectivité est maximisée lors de l'étape de création puisqu'elle dépend de l'analyse de données spatiales, non d'enquêtes ou de dire d'experts. Il en est de même pour l'étape de calibrage. Le calibrage automatique permet par ailleurs un apport de nouvelles connaissances liées à la réalisation du calibrage. En d'autres termes, au lieu d'indiquer en entrée les informations nécessaires pour obtenir le modèle attendu et d'analyser « uniquement » les résultats de la simulation, nous avons déjà l'opportunité d'analyser le processus de création et de calibrage en lui-même. Cette démarche permet, par exemple, une première analyse de la zone d'étude à travers la sélection automatique des variables utilisées. Cette étape peut notamment nous permettre de déterminer l'intérêt de la localisation géographique dans la caractérisation du développement urbain.

C’est dans l’esprit de la création de règles répondant à notre définition de « bonnes règles » et celui du développement d’une méthodologie simple et transposable qui puisse s’adapter au contexte transfrontalier, que notre choix s’est porté sur l’utilisation d’un algorithme en particulier. C’est l’apprentissage supervisé par Arbres de Décision (*Decision Tree*), issu du domaine de l’intelligence artificielle qui retient notre attention dans la suite de ce travail (Chapitre 5). Cette méthode permet de définir et de calibrer automatiquement des règles de transition déterministes à partir de l’analyse d’un jeu de données. Elle permet également de conserver le caractère explicite des règles, indispensable à la réalisation de notre analyse. Ces règles sont ensuite utilisées pour simuler le processus de développement urbain dans LucSim.

Conclusion du chapitre 4

Ce chapitre a permis de souligner l'adéquation entre les caractéristiques de l'AC et nos objectifs de recherche : un outil simple qui permet toutefois de modéliser et simuler rapidement un processus complexe tel que le développement urbain sur la zone frontalière de Strasbourg-Kehl.

L'automate cellulaire a été choisi ici pour son potentiel concernant l'observation et la caractérisation de phénomènes émergents mais également pour sa simplicité d'utilisation et de compréhension. Comme le souligne W. Tobler « ce n'est pas parce qu'un processus paraît compliqué qu'il y a cependant une raison de supposer qu'il est le résultat d'une règle compliquée » (Tobler, 1970). C'est en s'appuyant sur cette idée que l'utilisation d'un automate cellulaire pour la simulation du développement urbain prend tout son sens.

L'utilisation d'un AC nécessite la création de règles de transition pour simuler le développement urbain en question. Les jeux de règles reposent au minimum sur l'état initial des cellules et l'état de celles présentes dans leurs voisinages respectifs (Benenson & Torrens, 2004). Pour construire un jeu de règles basé sur les caractéristiques du voisinage, l'étape primaire est la définition dudit voisinage notamment sa taille et sa description. Les questions de la taille optimale du voisinage à prendre en compte lors de l'élaboration du jeu de règles, ainsi que celles de la définition des caractéristiques favorables à la transition urbaine doivent alors être abordées. Il est nécessaire d'être en capacité de déterminer l'importance des variables d'occupation du sol décrivant le voisinage, mais également celle des différentes variables de localisation géographique. Il sera ainsi possible d'aborder la question de la localisation nationale dans la caractérisation du processus de développement urbain sur la zone d'étude.

La notion de distance est extrêmement importante en géographie. W. Tobler, en 1970 énonce la loi connue comme étant la première loi de la géographie : « Toute chose est reliée à toute autre chose, seulement, les choses les plus proches présentent des liens plus forts que celles éloignées », basée sur la loi de la gravitation de Newton. Pour autant, elle est également liée au contexte géographique de la zone d'étude. Quelque chose de « loin » ne va pas être à la même distance pour un habitant insulaire, ou un citoyen russe ou américain. Bien que la première étape soit de définir si l'utilisation de l'accessibilité rend le modèle de développement urbain plus performant, c'est l'importance de cette variable et notamment sur celle du temps d'accessibilité aux pôles suivant la localisation nationale qui nous intéresse particulièrement. En résumé, il est question d'établir si l'accessibilité des zones urbaines aux pôles de différents niveaux est différente entre la France et l'Allemagne, soulignant alors l'importance de la frontière comme marqueur spatial de la différence.

Pour répondre à ces questions, l'objectif est de développer une méthodologie s'affranchissant au maximum d'un calibrage manuel, qu'il soit à dire d'experts ou déterminé

après un processus d'enquête. L'intention affichée est de limiter, tant que faire se peut, la subjectivité, le manque d'exhaustivité, et la mise en place de démarches chronophages. Plusieurs processus d'apprentissage assisté par ordinateur, dit apprentissage automatique *(Machine Learning)*, existent actuellement. Il est proposé d'utiliser un algorithme d'apprentissage par Arbre de Décision (AD), dans un environnement basé sur celui des Automates Cellulaires, pour définir et calibrer automatiquement les règles de transition nécessaires à la simulation du développement urbain. Ce point fait l'objet du chapitre suivant.

Chapitre 5

Un couplage : les arbres de décision

Le Chapitre précédent a mis en lumière les atouts que présente l'utilisation d'un automate cellulaire pour simuler le développement urbain. Ses propriétés en font un outil particulièrement adapté à l'approche modélisatrice adoptée dans ce travail (Chapitre 2). Les automates cellulaires sont appréciés pour leur simplicité, notamment à travers l'utilisation de règles basées sur l'appréhension des caractéristiques intrinsèques d'une cellule et de celles présentes dans son voisinage. Des règles simples peuvent favoriser l'émergence de dynamiques complexes, de phénomènes d'auto-organisations, de réplications, etc. comme l'ont montré les travaux menés, entre autres, par S. Ulam (Metropolis & Ulam, 1949) dans les années 1950 à Princeton puis ceux de S. Wolfram (1982). Il est donc envisageable de penser qu'un processus spatial, comme celui du développement urbain, peut être décrit à travers des règles simples et simulées dans un automate cellulaire. Toute la question reste alors de déterminer ces règles.

Ce chapitre permet de proposer une approche méthodologique de création et de calibrage automatique des règles pour un automate cellulaire, à travers l'utilisation des arbres de décision. Dans un premier temps, le chapitre reviendra sur l'émergence de l'utilisation des algorithmes d'apprentissage par ordinateur *(Machine Learning)* qui débute dans les années 1990 (Sous-section 1). Après avoir explicité les choix poussant à l'utilisation des arbres de décision comme support d'apprentissage, ces derniers sont présentés plus en détail (Sous-section 2). La sous-section 3 s'attarde sur l'élaboration des jeux de règles ainsi que sur la sélection du jeu considéré comme le plus adéquat pour répondre à la problématique posée dans ce travail de recherche.

"La connaissance mène à la sagesse et à la compréhension" (Williams, 2009).

1. PRINCIPE DU *MACHINE LEARNING*

En passant d'une géographie descriptive à une géographique quantitative, les géographes ont dû s'approprier de nouvelles méthodes issues des statistiques ou des sciences informatiques. La science des données est justement un domaine qui allie statistique et informatique, et qui permet d'analyser des processus spatiaux pour lesquels il existe une quantité croissante de données (Tufféry, 2017). Au sein de ce domaine, l'apprentissage automatique est l'une des méthodes d'analyse utilisable.

L'apprentissage automatique se base sur « l'étude de méthodes de calculs pour l'amélioration des performances [d'apprentissage] à travers la mécanisation de l'acquisition de connaissances/informations d'après l'expérience » (Langley & Simon, 1995). L'objectif de la démarche est l'analyse automatique de larges jeux de données. L'importance de ce champ de recherche, issu de l'intelligence artificielle[33], s'est accrue à partir des années 1950 (Quinlan, 1986). A. Turing illustre pour la première fois le concept d'intelligence artificiel grâce au développement de sa « machine universelle » (Turing, 1936), suivi par W. McCulloch et W. Pitts qui développèrent, eux, le premier réseau de neurones artificiels en 1943 (McCulloch & Pitts, 1943). Les machines « intelligentes » sont pilotées par des algorithmes. L'apprentissage automatique permet d'analyser des bases de données empiriques pour en extraire des logiques ou tendances générales. Il s'agit de ce que l'on nomme l'exploration de données ou *Data Mining*[34]. L'extraction de connaissances par l'analyse algorithmique peut dès lors être effectuée sur des jeux de données spatiales ou non, qualitatives comme quantitatives.

D'après S. Tufféry, les méthodes de *Data* Mining peuvent être regroupées en deux familles majeures : les méthodes descriptives et les méthodes prédictives. Les premières servent à réduire, résumer, synthétiser les données, sans variable à expliquer. Les secondes ont pour but de comprendre les données, l'évolution d'un comportement, non plus seulement en fonction de variables explicatives mais aussi selon une variable à expliquer, plus spécifiquement examinée dans l'évolution de son influence (Tufféry, 2017). Les principales méthodes de *Data Mining* utilisées actuellement en sciences sociales sont présentées dans la suite de cette sous-section.

[33] Le terme « intelligence artificielle » ou l'abréviation IA (eng. *Artificial Intelligence, AI*) a été utilisé pour la première fois lors de la conférence de Dartmouth (1956). Il est défini pat M. Minsky lors de cette conférence comme « la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique »(Minsky, 1965).

[34] Processus consistant à trouver des tendances générales utiles à l'analyse ou l'interprétation d'un large volume de données (Aghimien *et al.*, 2019)

1.1. Principes généraux

Plusieurs techniques d'apprentissages ont été développées pour analyser un jeu de données et en extraire une information utile à un questionnement spécifique. Quelle que soit la technique sélectionnée, certaines caractéristiques principales similaires et grandes étapes communes permettent d'achever un apprentissage. Nous présentons ici ces caractéristiques en nous appuyant sur notre cas d'étude.

Le présent apprentissage est réalisé sur le jeu de données élaboré précédemment (Chapitre 3). Ce dernier est composé d'individus caractérisés par des variables. Ici, les individus sont les cellules de notre AC. Les variables qui les caractérisent sont l'état initial d'occupation du sol en 1990, puis en 2006, l'occupation du sol des cellules composant le voisinage de l'individu, et la localisation géographique (Figure 4. 10). La question de la taille du voisinage sera traitée ultérieurement (Chapitre 6).

Le jeu de données peut être composé de variables qualitatives (indice de pays, occupation du sol en 1990, etc.), quantitatives discrètes (proportion des types de sols du voisinage, etc.) ou quantitatives continues (accessibilité). Si le jeu de données présente des lacunes (données manquantes), des méthodes permettant de combler ces lacunes sont généralement appliquées (interpolations, etc.). La méthode d'apprentissage utilisée dépendra donc des types de données composant le jeu initial (qualitatives, quantitatives, discrètes ou continues).

Une variable dite « à expliquer » (ou encore « cible », « réponse », « dépendante », « endogène ») doit être définie dans les paramètres de l'apprentissage. Son rôle est d'indiquer à l'algorithme la cible de l'analyse. Par exemple, la variable cible de cette recherche est celle qui informe de l'existence ou de l'absence d'une transition urbaine, l'objectif étant de définir les critères favorables à cette transition. La prédiction s'effectue généralement dans le but de caractériser l'appartenance à deux classes distinctes, répondant à la question de l'analyse par « oui » ou « non ». Ici, la prédiction permet de caractériser l'appartenance à la classe « ayant été urbanisée entre 1990 et 2006 » ou non.

Sur la base de cette phase d'apprentissage, le modèle créé permettra d'établir, en fonction des variables explicatives (ou encore « indépendante », « de contrôle », « exogène ») qui auront été observées durant ladite phase, l'appartenance d'un individu (cellule) à la classe analysée : urbaine ou non urbaine. Grâce à cette phase d'apprentissage, le modèle peut ainsi aider à prédire des transitions, et les vérifier d'après les données observées (cette phase de « test » est abordée ci-dessous). Le modèle issu de l'apprentissage pourra donc fournir un résultat sous forme de probabilités ou de résultats absolus.

En résumé, le choix de la méthode d'apprentissage se base, entre autres, sur la variable à expliquer puis sur les variables explicatives de cette dernière.

Les différentes phases dans l'élaboration du modèle

L'élaboration du modèle s'effectue en trois phases :

1- L'apprentissage
2- Le test
3- L'application

L'apprentissage est réalisé sur les individus dont le comportement est connu au regard de la question posée. Dans le cas d'espèce, l'apprentissage est réalisé sur une variable qualitative booléenne (binaire, « oui » / « non ») spécifiant si la cellule a effectué une transition urbaine entre 1990 et 2006. Dans un premier temps, le jeu de données initiales est partitionné en deux sous-ensembles :

- « Entraînement » : Il permet à l'algorithme d'effectuer l'apprentissage et de créer les règles appropriées
- « Test » : il permet de tester le jeu ne comportant que des données « vierges », non incluses dans la phase d'apprentissage (*unsee data*).

Le pourcentage que représente chaque sous-ensemble est laissé à la discrétion du modélisateur. Les valeurs retrouvées dans la littérature présentent souvent un ratio entraînement/test de l'ordre de 70/30. 70% du jeu de données est utilisé pour l'apprentissage, les 30% restant pour tester le modèle créé. Chaque sous-ensemble est échantillonné aléatoirement, tout en conservant un caractère représentatif du jeu total (proportion de transitions, de type d'occupation du sol, etc.). La partition est effectuée pour permettre la validation du modèle. En effet, son efficacité ne peut pas être objectivement testée sur les données ayant permis son élaboration. Le « biais d'optimisme » est ainsi évité (Tufféry, 2017). Le rôle du jeu test est de permettre l'évaluation de l'erreur du modèle, non biaisée, en utilisant seulement les données isolées au départ (non utilisée à l'élaboration du modèle). Ainsi, le modèle retenu est le plus performant pour la prédiction des transitions urbaines dans un jeu de données n'ayant pas servi à la phase d'apprentissage.

Critères d'évaluation du modèle

Plusieurs modèles peuvent être obtenus à partir d'un même jeu de données en fonction du paramétrage de l'algorithme d'apprentissage. L'évaluation de sa performance dépend des objectifs de modélisation. Ils peuvent, par exemple, nécessiter l'élaboration d'un modèle dont la prédiction est extrêmement précise, aux dépens d'un caractère éventuellement moins complexe, ou inversement. Pour ce faire, le modélisateur peut jouer sur le paramétrage. S. Tufféry (2017) identifie ainsi 7 critères pour évaluer la performance d'un modèle :

1- La précision

Le nombre d'individus mal classés par l'algorithme au regard de la question posée, permet de définir un taux d'erreur. Par exemple, une cellule urbanisée d'après les résultats issus de la simulation du modèle mais non urbanisée d'après les données observées constitue une erreur. Dans l'idéal, ce taux doit être minimum. Différents indicateurs permettent de mesurer

la précision, suivant la méthode d'apprentissage appliquée. Par exemple, lors de l'utilisation d'une méthode de classement, la courbe ROC[35] ou l'indice de Gini[36] peuvent être envisagés. La précision d'une régression peut, elle, être évaluée grâce au R^2 ou à l'erreur quadratique moyenne.

2- La robustesse

La robustesse traduit le détachement existant entre la réussite du modèle créé et l'échantillon utilisé. En d'autres termes, le modèle ne doit pas être sensible à la fluctuation aléatoire d'une variable. Il doit être capable d'atteindre une certaine performance quand l'application concerne des variables ayant évolué dans le temps. Cette stabilité pourra être testée lors de la construction du modèle, si le modélisateur dispose de données diachroniques par exemple. Le modèle sera d'autant plus performant s'il s'appuie sur des variables suffisamment stables et présentes d'une période à l'autre.

3- La concision

La parcimonie peut être un critère privilégié. Des règles simples et peu nombreuses seront d'autant plus faciles à interpréter, et plus enclines à la généralisation si l'utilisation du modèle est étendue à une autre population d'individus. La concision peut être un facteur de robustesse quand elle n'est pas privilégiée aux dépens de la précision.

4- Des résultats explicites

Les règles sont considérées comme explicites quand elles sont accessibles et compréhensibles. La compréhension directe est un facteur important au regard de la communication autour des résultats. Elle facilitera aussi l'intégration des règles dans un logiciel si tel est l'objectif.

5- La diversité des types de données prises en charge

Ce point peut favoriser l'utilisation d'une méthode. Certaines méthodes sont exclusivement dédiées à un type particulier de données (exclusivement quantitatives continues par exemple), quand d'autres au contraire permettent un traitement de données mixtes. La prise en charge des données concerne également celle des données manquantes ou fortement corrélées, ce qui peut revêtir un caractère discriminant quant au choix de la méthode d'apprentissage.

6- La rapidité du calcul

Le temps d'apprentissage peut être limitant notamment dans le cadre de l'élaboration de scénarios participatifs en temps réel. Dans ce cadre, le modèle doit être modulable rapidement pour prendre en compte de nouvelles caractéristiques ou encore adapter un scénario en cours de réflexion. Ce point impose la création du modèle et l'exportation des résultats de simulation dans un laps de temps réduit (durée d'une réunion, journée d'étude). Or, les résultats de certains modèles sont obtenus après plusieurs jours de simulation.

7- Les possibilités de paramétrage

---

[35] La courbe ROC (*Receiver Operating System*) est utilisé « pour décrire un test qui prédit la présence ou l'absence d'un trait binaire » (Wang *et al.*, 2019), pour une présentation approfondie voir (Krzanowski *et al.*, 2009; Pepe, 2003; Zhou *et al.*, 2009)

[36] L'indice de GINI permet de mesurer l'impureté d'une classe au regard de l'objectif de classification. L'indice de Gini peut également être considéré comme « le taux d'erreur attendu si le libellé de la classe est choisi de manière aléatoire dans la distribution de la classe au niveau du nœud » (Brown & Myles, 2009).

La richesse du paramétrage permet au modèle d'être flexible et de pouvoir s'adapter à divers contextes. Lors de l'application sur la zone d'étude plusieurs pistes auraient pu être envisagées : la production d'un modèle possédant un nombre très limité de règles simples, ou au contraire la production d'un modèle plus performant en termes de prédictions mais qui nécessite des règles plus longues et plus précises. C'est à travers le paramétrage d'apprentissage que ces deux objectifs peuvent être atteints.

L'importance des critères de performance variera suivant l'objectif de modélisation. L'équilibre de leur ajustement confère à l'exercice sa délicatesse, puisque l'augmentation de la performance de l'un d'entre eux risque de se faire aux dépens d'un autre critère de la liste. La section 1.3 revient plus précisément sur la configuration sélectionnée dans le cadre de ce travail après la présentation de la méthode (section 1.2).

Le réglage des divers paramètres doit parfois prendre en compte le phénomène de sur-apprentissage pouvant affecter la performance de certains types d'algorithmes. Ce phénomène est d'autant plus important lors de l'étude d'un processus peu représenté dans les données d'entrée. C'est notamment le cas du processus d'urbanisation de la zone d'étude du présent travail. Ce point fait l'objet de la sous-section suivante.

1.2. Phénomène exceptionnel et « sur-apprentissage »

Deux points majeurs doivent retenir l'attention du modélisateur lors de l'utilisation d'algorithmes d'apprentissage automatique. Le premier concerne un apprentissage basé sur un jeu de données déséquilibré et dont les prédictions intéressantes ne portent que sur une classe faiblement représentée au regard du nombre total d'individus. Le second concerne le sur-apprentissage, lequel affecte la propension à la généralisation du modèle en lui conférant un caractère descriptif ponctuel excessivement précis. Cet écueil dévalue alors la performance du modèle en tant que processus dégageant des tendances de transition.

Phénomène exceptionnel

La capacité d'une machine à apprendre d'un phénomène en fonction de données fournies dépend en partie de la composition du jeu de données, et notamment de l'équilibre existant dans la proportion des classes à prédire[37]. Le Chapitre 3 soulignait le caractère exceptionnel du phénomène d'urbanisation, lequel comprend un nombre restreint d'individus appartenant à la classe « urbanisé entre 1990 et 2006 ».

L'efficacité d'un modèle peut être évaluée dans sa globalité de façon binaire, par sa capacité à prédire correctement l'appartenance d'un individu : ce dernier est correctement classé ou ne l'est pas. Dans le cas de l'urbanisation, une cellule urbanisée (individu) dans les données

[37] La recherche d'équilibre peut amener une correction visant à tendre vers une proportion de 50/50 entre la classe à prédire et l'autre classe.

observées l'est également dans les données simulées. Elle est alors correctement classée. Dans le cas contraire, la prédiction est fausse.

À ce titre, l'apprentissage d'un processus exceptionnel est plus délicat à paramétrer (Haixiang *et al.*, 2017; López *et al.*, 2013). En effet, sans spécifier l'important de la prédiction d'une classe en particulier, l'algorithme se satisfera de sa capacité de prédiction globale. Ainsi, dans le cas des classes déséquilibrées, le modèle pourra être performant sur prédiction de la classe la plus importante et pauvre dans la prédiction de la classe la moins importante, tout en étant performant dans la globalité des prédictions. Dans le cadre de ce travail, la classe des transitions urbaines est faiblement représentée, au regard du non-changement (0,05% de la zone d'étude). L'absence de prise en compte du déséquilibre a pour conséquence directe la création d'un modèle capable de prédire le phénomène de non-changement, puisqu'il est de toute façon majoritaire, au lieu du phénomène d'urbanisation. Pour illustrer nos propos, considérons les proportions de chaque classe sur notre zone d'étude. Les cellules présentant un changement d'état entre 1990 et 2006 représentent 0,55 % des individus par conséquent les cellules ne présentant aucune évolution, représentent, elles, 99,45%. L'hypothèse menant à la construction d'un modèle nul suggèrerait de ne rien faire évoluer sur la zone d'étude. Ce modèle nul serait néanmoins efficace à 99,45% dans la justesse de ses prédictions, puisque l'erreur ne concernerait que les cellules ayant changé d'état, soit 0,55%. Le modèle serait étonnamment extrêmement efficace dans la globalité de ses prédictions, mais incapable de prédire le phénomène étudié dans ce projet.

D. Cieslak (2008) précise que le problème de l'apprentissage sur des jeux de données dont les classes sont déséquilibrées est un problème persistant lors de la mobilisation d'outils d'apprentissage automatique, notamment dans le cas des arbres de décision. Trois alternatives, entres autres, peuvent être envisagées pour pallier le déséquilibre (Haibo & Garcia, 2009).

- Un ré-échantillonnage différent

Cette méthode consiste à modifier l'échantillonnage d'apprentissage pour favoriser la classe auparavant sous-estimée.

- Pénaliser le modèle

Lorsque l'erreur de prédiction concerne la classe sous-représentée, elle sera artificiellement considérée dans les paramètres du modèle comme plus importante.

- Utiliser d'autres indicateurs de performance

La performance d'un modèle peut être évaluée suivant différentes modalités. Un seuil à atteindre peut-être prédéfini par exemple dans le cadre d'une analyse ROC.

L'apprentissage automatique du processus d'urbanisation sur la zone d'étude nécessite la prise en compte du caractère exceptionnel de ce phénomène. Cependant, la mise en place des mesures nécessaires ne doit pas s'effectuer aux dépens de la qualité de l'apprentissage, laquelle peut être dégradée par le « sur-apprentissage ».

Sur-apprentissage

Certains algorithmes d'apprentissage automatique sont sensibles au phénomène de sur-apprentissage. Ce phénomène est lié à la difficulté de généralisation rencontrée lors de l'élaboration de modèles présentant une certaine complexité. La Figure 5. 1 illustre le phénomène en question :

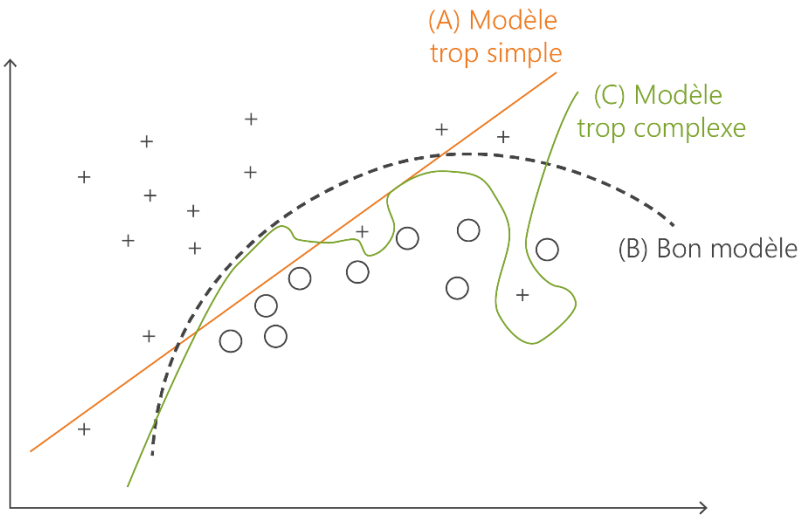


Figure 5. 1 : Sous- et sur-apprentissage d'un modèle de classement (d'après Bousquet dans Tufféry 2017)

Le modèle A, excessivement simple, ne sera pas en capacité de saisir les subtilités de la distribution, et présentera un fort biais de prédiction. Le modèle C, trop complexe, ne sera au contraire pas capable de généraliser le phénomène observé. Le modèle B, ou le « bon modèle », pour S.Tufféry, se trouve dans un juste milieu, dont la localisation est toujours dépendante des besoins en termes de modélisation. La Figure 5. 2 propose une image du « juste équilibre » en question, à travers la projection du taux d'erreur lié au modèle créé. À mesure de l'augmentation de la complexité du modèle, le taux d'erreur diminue (repère A). Les prédictions du modèle sur les données d'apprentissage et celles obtenues sur les données test convergent vers une diminution du taux d'erreur induisant une « bonne généralisation » (repère B). L'augmentation de la complexité au-delà de ce repère permet d'obtenir un modèle d'autant plus précis sur les prédictions lors de l'apprentissage mais dont le taux d'erreur, qui se manifeste sur les données test croît en raison de l'augmentation de la spécificité du modèle créé. Le taux d'erreur alors important induit une mauvaise généralisation (repère C).

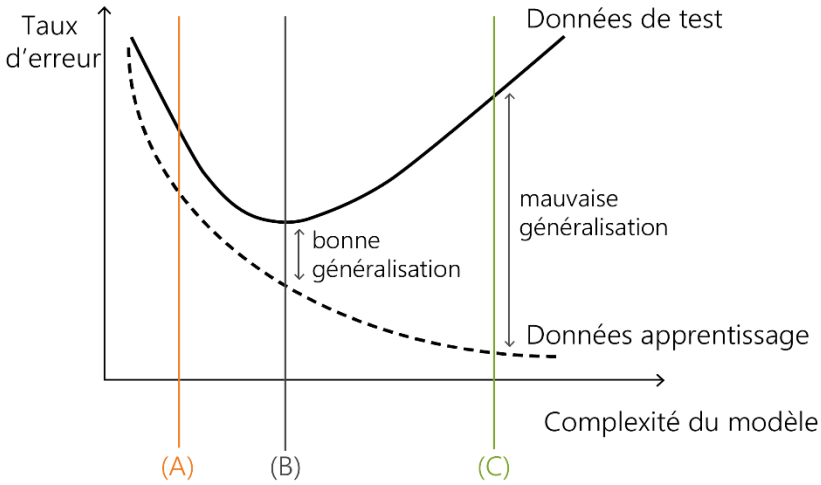


Figure 5. 2 Taux d'erreur en fonction de la complexité du modèle (Tufféry, 2017)

Le problème du sur-apprentissage peut apparaître lors de l'exploitation d'un échantillon trop minime au regard du nombre de paramètres contenus dans le modèle. D'autres sources de sur-apprentissage sont aussi à considérer : la présence d'une variable explicative fondamentalement corrélée dans sa nature à celle dont on cherche l'explication, ou l'attribution d'un poids excessif lors du recours à une pondération des variables explicatives. Le dernier cas à mentionner, relevant plus d'un biais général que celui du sur-apprentissage tel qu'on l'imagine intuitivement, est celui de l'apprentissage réalisé sur un jeu de données erronées ou biaisées à l'acquisition. Dans ce cas, l'apprentissage ne sera pas correct non plus.

Si plusieurs cas de figure peuvent mener à la réalisation d'un modèle sujet au « sur apprentissage », des solutions permettent de l'éviter. Outre la vérification ou l'ajustement des différents paramètres au regard des sources de « sur-apprentissage » (nombre de variables, diversité de l'échantillon, etc.), le test du modèle sur un échantillon vierge, c'est-à-dire qui n'as pas servi à la création du modèle lors de l'apprentissage, est la méthode de vérification principale attestant de l'efficacité du modèle sur un « nouveau » jeu de données.

Ceci permet d'éclairer la marche à suivre lors de l'utilisation d'algorithmes d'apprentissage pour créer des règles issues de l'analyse d'un jeu de données. Les règles issues de l'analyse traduisent les comportements relatifs à la question d'intérêt. Nous pouvons voir ici, comme dans la démarche de modélisation initiée par une sélection arbitraire des données, le caractère subjectif, intrinsèquement lié à la problématique recherchée, de la forme du modèle et de ses résultats. Ainsi, les choix d'ajustement des critères privilégiant l'un ou l'autre des paramètres influenceront les résultats et la vision exprimée. L'ajustement des paramètres variera aussi en fonction de la méthode et de l'algorithme utilisé.

Les principales méthodes décrites dans la littérature et liées à l'élaboration de modèle traitant notamment de l'occupation du sol sont présentées dans la sous-section suivante.

1.3. Méthodes d'apprentissage automatique

De nos jours, les statistiques sont couramment mobilisées en sciences humaines. Comme mentionné précédemment, S. Trufféry (2017) distingue deux familles de méthodes statistiques, descriptive et prédictive, dont la différence principale réside dans l'absence de variable à expliquer concernant l'approche descriptive. Ainsi son objet est de résumer et synthétiser les données. Notre démarche s'inscrit quant à elle, dans l'approche prédictive de la modélisation statistique qui vise à expliquer une variable. Nous cherchons à expliquer la variable « urbanisation entre 1990 – 2006 ». La suite de cette sous-section présente succinctement les différents modèles mentionnés dans la Table 5. 1 et liés à l'approche prédictive. Nous nous limiterons à la description des méthodes pouvant être mobilisées selon les conditions initiales imposées par les types de variables de notre jeu de données (Table 5. 2). Pour rappel, notre variable cible est binaire, nos variables explicatives sont qualitatives, quantitatives discrètes et continues. Nous nous concentrerons également sur les solutions dont l'exploitation est libre de droits[38].

Table 5. 1 : Deux grandes familles de méthodes statistiques en sciences sociales (Trufféry, 2017)

Type	Famille	Sous-famille	Algorithme
Méthodes descriptives	Modèles géométriques	Analyse factorielle (projection et visualisation dans un espace de dimension inférieure)	Analyse en composantes principales ACP (variables quantitatives)
			Analyse factorielle des correspondances AFC (variables qualitatives et binaires), Analyse des correspondances multiples ACM (variables qualitatives et binaires)
		Analyse typologique (regroupement dans tout l'espace en classes homogènes)	Méthode de partitionnement (centres mobiles,k-means, nuées dynamiquesk-medoids, DBSCAN, ...)
			Méthodes hiérarchiques (ascendantes et descendantes)
		Analyse typologique + réduction de dimension	Classification neuronales (cartes de Kohonen)
	Modèles combinatoires		Classification par agrégation des similarités (variables qualitatives)
	Modèles à base de règles logiques	Détection de liens	Recherche de règles d'associations Recherche de séquences similaires
Méthodes prédictives	Modèles à base de règles logiques	Arbres de décision	Arbres de décision (variable à expliquer quantitative ou qualitative)
	Modèle à base de fonctions mathématiques	Réseaux de neurones	Réseaux à apprentissage supervisé (perceptron, réseau à fonction de base radiale, ...)
		Modèles paramétriques ou semi-paramétriques	Régression linéaire, ANOVA, MANOVA, ANCOVA, MANCOVA, modèle linéaire général, régression PLS, ridge ou lasso, SVR (variable à expliquer continue)
			Analyse discriminante de Fisher, régression logistique PLS, ridge ou lasso, SVM (variable à expliquer continue)
			Modèle log-linéaire (variable à expliquer = comptage = nombre d'individus prenant une combinaison donnée de modalités de variables qualitatives)
			Modèle linéaire généralisé, modèle additif généralisé (variable à expliquer continue, discrète, comptage ou qualitative)
	Prédiction sans modèle	Analyse probabiliste	K-plus proches voisins (k-NN)

[38] Certaines méthodes étant implémentées uniquement dans des logiciels nécessitant l'acquisition d'une licence.

Table 5. 2 : Méthodes prédictives mobilisables en fonction des types des variables explicatives ou à expliquer (Tufféry, 2017)

Explicatives / A expliquer	**Quantitative** *(covariable)*	**n quantitatives** *(covariables)*	**1 qualitative** *(facteur)*	**n qualitatives** *(facteurs)*	**Mélange**
1 quantitative	régression linéaire simple, régression spline, régression robuste, arbres de décision, MARS, SVR (support vector regression), k-plus proches voisins	régression linéaire multiples, régression spline, régression robuste, régression PLS, arbres de décision, MARS, SVR, k-plus proches voisins	ANOVA, arbres de décision, MARS, SVR, k-plus proche voisins	ANOVA, arbres de décision, MARS, réseau de neurones, SVR, k-plus proche voisins	ANOVA, GLM univariée, arbres de décision, MARS, réseau de neurones, SVR, k-plus proche voisins
n quantitatives *(représenant des quantités différentes)*	régression multivariée, régression PLS2	régression multivariée, régression PLS2, réseaux de neurones	MANOVA	MANOVA, réseaux de neurones	MANCOVA, GLM multivariée, réseaux de neurones
1 qualitative nominale ou binaire	analyse discriminante de Fisher, régression logistique, modèles linéaires généralisés pénalisés, arbres de décision, MARS, SVM, classifieur bayésien naïf, k-plus proche voisins	analyse discriminante de Fisher, régression logistique, régression logistique PLS, modèles linéaires généralisés, modèles pénalisés, arbres de décision, MARS, réseau de neurones, SVM, classifieur bayésien naïf, k-plus proche voisins	régression logistique, analyse discriminante DISQUAL, modèles linéaires généralisés, modèles pénalisés, arbres de décision, MARS, SVM, classifieur bayésien naïf, k-plus proche voisins	régression logistique, analyse discriminante DISQUAL, modèles linéaires généralisés, modèles pénalisés, arbres de décision, MARS, réseau de neurones, SVM, classifieur bayésien naïf, k-plus proche voisins	régression logistique, modèles linéaires généralisés, modèles pénalisés, arbres de décision, MARS, réseau de neurones, SVM, classifieur bayésien naïf, k-plus proche voisins
n qualitatives nominales ou binaires *(représentant des quantités différentes)*	arbres de décision, modèle linéaire généralisé vectoriel, modèle additif généralisé vectoriel	arbres de décision, modèle linéaire généralisé vectoriel, modèle additif généralisé vectoriel, réseaux de neurones	arbres de décision, modèle linéaire généralisé vectoriel, modèle additif généralisé vectoriel	arbres de décision, modèle linéaire généralisé vectoriel, modèle additif généralisé vectoriel, réseaux de neurones	arbres de décision, modèle linéaire généralisé vectoriel, modèle additif généralisé vectoriel, réseaux de neurones
1 quantitative asymétrique	régressions gamma et log-normale				

La Table 5. 1 permet de regrouper les différents types de modèles statistiques en fonction de leur famille de référence. L'utilisation d'un type de modèle en particulier dépendra principalement du type des variables analysées (qualitatives/quantitatives, continues/discrètes) ainsi que du volume de données (nombre de variables). La Table 5. 2 synthétise les différentes combinaisons possibles en fonction des types de variables.

Comme notre démarche s'inscrit dans une approche prédictive, la suite de cette sous-section présente succinctement les différents modèles mentionnés dans la Table 5. 2 liés à cette approche, et qui présentent un caractère favorable au regard des types de nos données d'entrées. L'exploitation des Tableaux 1 et 2 permet de restreindre la liste des méthodes envisageables à celles de la régression logistique, des réseaux de neurones et des arbres de décision. En effet, les régressions linéaires, l'analyse discriminante, l'analyse discriminante DISQUAL ne permettent pas d'appréhender la configuration de nos types de données.

La régression logistique

La régression logistique est une méthode courante en statistiques. Elle permet de mesurer l'association entre l'occurrence d'un événement, décrit par une variable qualitative, et les facteurs susceptibles de l'influencer, décrits par des variables qualitatives ou quantitatives. Le choix des variables explicatives est basé sur une connaissance antérieure de l'évènement étudié. Si cette méthode permet l'élaboration d'un modèle précis, elle est plus généralement appliquée à un échantillon de petite taille et modélise directement une probabilité. Le jeu de données étudié dans ce travail de recherche est d'une taille conséquente puisqu'il contient 532 659 individus pouvant être caractérisés par 80 à 126 variables explicatives. L'objectif du travail est d'utiliser un algorithme capable de déterminer l'importance ou l'absence d'importance des variables explicatives fournies. En raison du volume de notre échantillon la régression logistique ne correspond donc pas exactement aux objectifs fixés.

Les réseaux de neurones

Parmi les modèles à base de fonctions mathématiques, les réseaux de neurones (*Neural Network – NN*) sont incontournables. Ils peuvent tout aussi bien s'appliquer dans un cadre de modélisation descriptive ou prédictive. Cette approche s'avère puissante notamment dans l'appréhension de phénomènes complexes, de formes irrégulières et de données ne suivant pas de lois probabilistes particulières. Leur utilisation peut néanmoins présenter certaines limites :

1- L'aspect boîte noire (poids des variables, poids des liens, fonction de combinaison, fonction de transfert, structure avec ou sans couches cachées, etc.)
2- Des réglages délicats
3- La puissance informatique nécessaire
4- Le sur-apprentissage.

L'aspect boîte noire et le manque de transparence des règles sont les facteurs limitant le plus l'utilisation des réseaux de neurones dans ce travail puisque l'objectif est d'obtenir des règles explicites permettant une caractérisation thématique du processus de développement urbain.

Modèles à bases de règles logiques : Arbres de décision.

Les arbres de décision font partie des méthodes les plus intuitives à appliquer pour étudier un jeu de données. Ils présentent l'avantage de pouvoir s'appliquer directement « sans avoir à mettre en forme, à homogénéiser, (…) » la donnée (Tufféry, 2017). S. Tufféry définit ce cas de figure comme « *off the shell* ». Leur pouvoir prédictif est parmi les plus performants et les règles fournies sont totalement explicites. Ce point soutient majoritairement notre préférence pour cette méthode. Les arbres de décision se situent à la limite entre la description et la prédiction, positionnement extrêmement avantageux dans l'étude du développement urbain sur la zone transfrontalière de Strasbourg-Kehl.

Le résumé des points forts et faibles des méthodes est présenté ci-dessus (Table 3).

Table 5. 3 Synthèse des forces et faiblesses des méthodes statistiques mobilisables

	Précision	Robustesse	Concision	Lisibilité	Peu de données	Variables corrélées	Rapidité de calcul	Off the shell
Régression logistique	+	+	+	+	+	-	=	-
Réseaux de neurones	+	-	-	-	-	=	-	-
Arbre de décision	=	-	=	+	-	+	+	+

L'analyse comparative des points forts et faibles de ces trois méthodes nous oriente vers la sélection de la méthode d'apprentissage par arbres de décision. Si la régression logistique peut s'avérer plus robuste et concise, elle ne permet pas la création d'un modèle performant dans l'étude de variables potentiellement corrélées notamment dans le cas d'un jeu de données important. Les réseaux de neurones quant à eux présentent une limite concernant la lisibilité des règles, un point indispensable à l'analyse des différences ou similitudes entre les pays de la zone d'étude. Les arbres de décision ne sont pas affectés par les limites soulignées dans les deux autres méthodes. Ils permettent de traiter rapidement des variables de divers types, en fournissant des résultats explicites, quand bien même les variables explicatives présenteraient une corrélation. Le point de la robustesse est néanmoins à prendre en compte et sera discuté plus en détail dans la section suivante, qui se focalise plus précisément sur la méthode d'apprentissage par arbres de décision.

2. PRINCIPE DES ARBRES DES DÉCISION

2.1. Historique

Le concept des arbres de décision (AD) a été développé dans le champ de recherche de l'apprentissage automatique (*Machine Learning*) et parallèlement au développement de l'exploration de données (*Data Mining*). L'arbre de décision appartient au groupe des outils statistiques de classification supervisée. Son objectif est de classer un nouvel élément, au regard des caractéristiques des éléments déjà classés. Pour se faire, des règles permettant d'attribuer l'élément à une classe en particulier sont définies à partir de variables qualitatives ou quantitatives caractérisant ces objets.

Comme pour les AC, c'est dans le champ de recherche des mathématiques, par J. N. Morgan et J. A. Sonquist, que les premières méthodes d'analyse à l'aide d'arbres ont été développées (Sonquist & Morgan, 1964). L'apprentissage par arbre de décision s'étend, par la suite, très largement à divers champs de recherche, allant du domaine médical au domaine économique, avec par exemple l'évaluation des risques financiers (Kotsiantis, 2011; Mitchell, 1997; Safavian & Landgrebe, 1991). Les chercheurs y recourent donc, de manière générale, pour extraire des règles de classification à travers l'analyse d'un jeu de données. Wang (2014) estime que la précision des classifications issues de l'utilisation d'algorithmes

d'arbres de décision est égale, voire supérieure, à celles des autres modèles élaborés pour un cadre identique. Le développement des algorithmes AD a pour but premier de résumer l'information contenue dans un jeu de données. Ce résumé est orienté sur l'étude d'une caractéristique ou particularité concernant les individus étudiés (dans le cas de l'occupation des sols, un type de transition spécifique par exemple) et permet d'extraire l'information mettant en perspective leurs autres caractéristiques communes. Pendant les dernières dizaines d'années, plusieurs algorithmes d'arbres de décision ont été développés. Les plus connus d'entre eux sont ID3 (Quinlan, 1986), CART (Breiman *et al.*, 1984), C4.5 (Quinlan, 1993), SPRINT (Shafer *et al.*, 1996), SLIQ (Mehta *et al.*, 1996). Une description plus détaillée de ces différents algorithmes, ainsi que d'autres moins connus, et leur comparaison, peut être trouvée dans Lim *et al.* (2000) ou Murthy (1998). L'utilisation de ces algorithmes d'apprentissage est souvent justifiée par plusieurs points. Ils présentent généralement une différence dans les critères de séparation ou de stop utilisés.

Pal (2003) entre autres auteurs (Kotsiantis, 2011; Therneau *et al.*, 1997) liste les avantages attenants à l'utilisation des arbres de décision. L'un des principaux se trouve être, tout comme l'automate cellulaire en modélisation, la simplicité de la démarche, associée à la rapidité d'exécution de l'algorithme. Cependant, celui qui retient particulièrement notre attention dans le cadre de ce travail de recherche, et soutient fondamentalement ce choix méthodologique, est l'opportunité de pouvoir interpréter les règles créées ; ce n'est pas une « boîte noire », comme peuvent être décrits les réseaux de neurones par des utilisateurs non experts (Kotsiantis, 2011; Pal & Mather, 2003). Ils sont « simples » à manipuler et à comprendre au regard des autres méthodes présentées. Le dernier point fort que nous mentionnerons correspond à leur capacité à traiter des données de tous types. Or, notre jeu de données initial est composé de données aussi bien qualitatives que quantitatives. L'arbre de décision présente donc les caractéristiques nécessaires à l'élaboration de notre modèle. La section suivante s'attache donc à en souligner les principes et concepts de base.

2.2. Concept et définition

Principe d'apprentissage

L'arbre de décision est une méthode connue pour l'extraction de règles de classification de données à travers une analyse globale d'un jeu de données. Leur dénomination fait analogiquement référence à la structure d'un arbre en biologie : branche, feuilles, etc. Un arbre de décision « pousse suivant une approche descendante (*top down*), quand nous partitionnons successivement la partie entraînement du jeu de données (*training set*) en sous-jeux contenant des individus plus similaires, ou ayant le même label » (Wang, 2014). Comme dans la plupart des processus impliquant de l'apprentissage automatique, la première étape consiste, en effet, à réaliser la partition du jeu original en deux sous-ensembles :

- Le sous-ensemble « entraînement » permet à l'algorithme de créer les règles appropriées ;

- Le sous-ensemble « test » permet de tester le jeu ne comportant que des données « vierges » non-incluses lors de l'apprentissage initial (Section 1.1.1).

Une fois le jeu de données divisé en parties « entraînement » et « test », l'objectif de l'algorithme est de développer, « faire pousser », l'arbre de décision à partir de l'échantillon « entraînement » (discrétisation récursive). C'est en effet, la méthode sur laquelle l'algorithme s'appuie. Il est successivement divisé pour classer les individus dans l'un ou l'autre des groupes qui répondent à la question posée. Par exemple, cet individu a effectué une transition urbaine (goupe 1) celui-là non (groupe 2). L'objectif est de créer des classes de plus en plus homogènes, à mesure des divisions successives, au regard du critère étudié. Par exemple, l'objectif de ce travail de recherche et de classer l'individu en fonction de son urbanisation entre 1990 et 2006. La division de l'échantillon d'apprentissage s'effectue sur la base du seuil de la variable explicative permettant de maximiser l'homogénéité des classes obtenues après division. Ainsi, le seuil de la variable sélectionnée permet de répartir les individus en sous-populations à l'homogénéité croissante et d'obtenir les critères permettant de caractériser au mieux la transition urbaine. L'objectif est, *in fine*, d'obtenir les classes les plus homogènes possibles, présentant d'un côté les individus urbanisés et de l'autre les non-urbanisés (Table 5. 3). L'action est donc répétée jusqu'à ce que « la séparation des individus ne soit plus possible ou plus souhaitable » (Tufféry, 2017). Dans le cas d'un seuil « souhaitable » c'est l'utilisateur qui règle ce paramètre.

La division est basée sur un critère de division (*splitting criteria*) et prend fin quand le critère d'arrêt (*stopping criteria*) est atteint. Ces critères sont présentés plus précisément dans la sous-section suivante (1.2.3). Chaque division du sous-ensemble en fonction d'une variable spécifique permet de constituer progressivement les « branches » de l'arbre. L'ensemble obtenu après la dernière division est considéré comme une « feuille ».

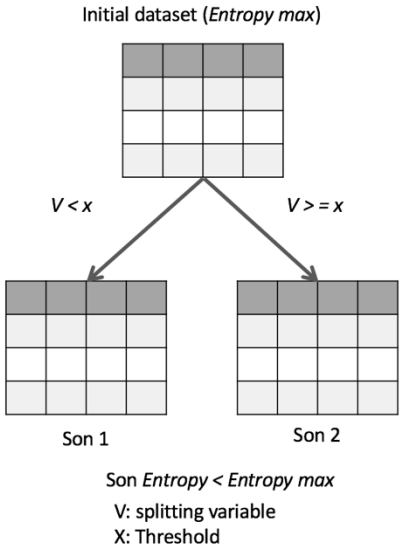


Figure 5. 3 Division récursive de l'échantillon apprentissage

D'après L. Rokach (2010) dans la plupart des cas, les fonctions discrètes de divisions sont univariées. Univarié est, ici, entendu comme une division basée sur la valeur d'un seul et unique attribut devant être défini. Plusieurs critères de division existent et peuvent être classés d'après le type de mesure qu'ils représentent (*information theory*, dépendance, distance, etc.) (Rokach & Maimon, 2010). La division récursive des sous-ensembles continue donc jusqu'à ce que le gain d'information apporté par la division ne soit plus considéré comme « suffisant », ou encore jusqu'à ce que le critère d'arrêt soit atteint.
Sur le même plan que les critères de division, différents critères d'arrêt peuvent être envisagés. Le choix de ces critères est toujours intrinsèquement lié aux objectifs de l'étude. Un des critères peut être l'obtention d'une branche dite « pure ». « Si seulement une seule classe est représentée dans le sous-ensemble, cette classe est considérée comme pure, dans le cas contraire, comme impure » (Kotsiantis, 2011). La longueur maximale de la branche peut aussi être contrainte et entrainer l'arrêt du processus visant à « faire pousser » l'arbre. L'apprentissage peut également être stoppé quand un nombre minimum d'individus compose un sous-ensemble final. Le processus peut finalement être interrompu si la meilleure variable de division ne remplit plus les prérequis initiaux (par exemple si la division ne permet pas d'atteindre un certain niveau de gain d'information, etc.).

Le critère de division ou le critère d'arrêt peuvent varier en fonction de l'algorithme utilisé ou de l'objectif de la classification. Les critères ne seront probablement pas les mêmes si la finalité est d'obtenir un jeu de règles extrêmement précis, induisant la création d'un arbre composé d'une multitude de longues branches, ou bien si l'objectif est de rendre le jeu de règles plus général en imposant par exemple un nombre de niveaux de divisions maximum. Une liste détaillée des critères pouvant diriger l'étape de division peut être trouvée dans Safavian & Landgrebe (1991).

Quand la division de l'échantillon est finalisée, des règles qui permettent de classer les individus au sein d'un groupe spécifique sur la base de leur similarité, peuvent être extraites suivant les « branches » qui mènent aux « feuilles » issues des divisions successives. « Une fois le processus de construction de l'arbre achevé, les règles de décisions générées peuvent être utilisées pour prédire les classes des individus restants du jeu de données » (Padua *et al.*, 2014), soit ceux de la partie « test » du jeu de données original. L'efficacité des règles peut ainsi être vérifiée.

La description du processus d'élaboration de l'arbre de décision permet de faire le lien entre les objectifs du travail et le potentiel offert par l'utilisation de cette méthode d'apprentissage automatique : une sélection des variables explicatives et de leurs seuils, permettant d'obtenir des règles de classification des individus ayant été urbanisés. Les règles fournissent en effet une description de l'environnement et des critères intrinsèques principaux des cellules urbanisées.

L'apprentissage sur la zone d'étude

Pour faciliter la réalisation des différentes étapes par l'utilisateur, la méthodologie et l'outil développés ont été pensés de manière à être flexibles et ouverts. La finalité du travail est de fournir un support capable de remplir les objectifs de notre travail de recherche, mais encore de penser le caractère adaptable de ce support à d'autres zones d'études, d'autres contextes de recherches, et d'autres problématiques. En résumé, un outil de modélisation transposable.

L'apprentissage par arbre de décision est utilisé pour définir et calibrer automatiquement les règles de transitions pour automate cellulaire dans le cadre de l'étude des processus de développement urbain. Comme le rappelle Paul Torrens (2011) « dans le cadre de l'étude des systèmes urbains, nous n'avons le plus souvent aucune idée de ce que sont les bonnes règles ». (Chap. 4 section 1.3)

L'objectif est donc ici de créer automatiquement les règles en considérant les variables utilisées par l'arbre de décision comme représentatives de l'environnement et des caractéristiques intrinsèques des cellules effectuant une transition urbaine. L'arbre de décision permet de répondre aux questions : où ? quoi ? combien ? nécessaires à l'écriture des règles pour l'automate cellulaire.

L'image utilisée pour définir ce type d'algorithme est particulièrement adaptée et explique l'utilité et notamment la flexibilité d'un tel outil pour les objectifs de modélisation recherchés. Au fur et à mesure de l'apprentissage, l'acquisition d'information grâce à l'algorithme permet de faire croître un arbre, en démultipliant ou en allongeant ses branches. Pour autant, à l'image d'un jardinier qui taillerait ses arbres, l'utilisateur peut contraindre ou diriger ce développement de branches dans le but d'offrir à l'arbre une structure adaptée. Il peut aussi laisser l'arbre se développer totalement, lui garantissant le maximum de branches et des réponses plus précises vis-à-vis de la question posée. Si la démultiplication des branches (des réponses) devient un inconvénient plutôt qu'un avantage, l'utilisateur peut choisir de restreindre son nombre. Une structure moins riche ou précise en informations, mais néanmoins plus claire, se dégagera alors.

Comme le jardinier, le modélisateur doit avoir défini, avant de faire pousser l'arbre, les objectifs de son travail. Faut-il un arbre bien fourni en branches, pour un résultat se rapprochant au maximum de celui que fournirait la nature dans tous ses détails ? Au contraire faut-il pouvoir apprécier la structure des branches principales en taillant celles de moindre importance, par souci esthétique ou pratique ? C'est à ces mêmes questions que le modélisateur doit répondre. Veut-il un arbre maximisant l'information et la précision des résultats au risque de réduire l'aspect généralisant du modèle ? Au contraire, préfèrera-t-il moins d'informations ou une information plus générale dans le but de pouvoir mieux saisir les points clés du processus étudié, comme c'est souvent le cas en modélisation ?

Cette flexibilité de l'outil permet également d'effectuer une démarche itérative, afin de tester les différentes données et leur potentiel de modélisation. Il est alors possible de mobiliser seulement certaines données, par exemple seulement l'occupation du sol, pour tester l'efficacité du modèle, et de réitérer le processus en ajoutant progressivement de nouvelles variables. Cet aspect de l'outil laisse tout autant la possibilité d'enrichir le modèle ou de le contraindre suivant l'évolution des objectifs après réflexion, ou au cours d'un projet.

L'essentiel de ces considérations démontre que l'utilisation des données d'entrées autant que le paramétrage de l'arbre de décision peut être adaptés suivant les besoins et les objectifs définis. L'outil n'est donc pas dédié. L'apprentissage par arbre de décision est donc retenu pour sa flexibilité au regard des données d'entrées, et du paramétrage.

Cette étape de paramétrage de l'apprentissage, tout comme l'étape de calibrage d'un modèle en général, est une étape importante puisqu'elle impacte les résultats analysés. L'importance du paramétrage ainsi qu'une illustration des différentes possibilités en la matière, sont présentées plus en détail dans la sous-section suivante.

2.3. Paramétrage

Plusieurs paramètres comme le critère de division ou le critère d'arrêt, influencent la construction de l'arbre. La division récursive continuera tant qu'aucun seuil d'arrêt n'est atteint. Ces critères font l'objet de la présente sous-section.

Critères de division

La première étape de construction de l'arbre de décision dépend du critère de division spécifié dans l'algorithme. Ce critère permet de séparer, classer les individus dans l'un ou l'autre des sous-ensembles, par division du jeu de données initiales. Ce dernier est divisé en sous-ensemble par le biais de la sélection d'une variable explicative à un seuil spécifique qui permet d'augmenter l'homogénéité des classes obtenues. Par exemple, l'objectif de l'apprentissage sur notre jeu de données est de classer les individus en fonction de leur urbanisation entre 1990 et 2006. Une division du jeu de donnée pourra par exemple être basée sur la présence de plus de 4 ha d'urbain dans un rayon de 200m, cette variable s'avérant la plus discriminante dans la séparation des cellules urbanisées des cellules non-urbanisée. Elle crée par conséquent des sous-ensembles les plus homogènes.

Les algorithmes peuvent être paramétrés de diverses façons pour tester l'homogénéité des sous-groupes obtenus après division. L'un de ces paramétrages revient à tester les différentes valeurs de chaque variable explicative, puis de sélectionner la variable présentant la valeur la plus intéressante pour diviser le jeu de données. Une autre méthode peut être de sélectionner la première variable et la valeur qui lui est associée, concordant avec une performance de classification minimale définie en amont par le modélisateur. Cette alternative pourra réduire le temps de traitement, mais ne testera pas toutes les possibilités

avant de sélectionner la variable et sa valeur pour la division. Quand l'homogénéité, autrement nommée pureté, des sous-ensembles est évaluée pour sélectionner la variable la plus performante, plusieurs critères peuvent être utilisés. Nous nous focaliserons ici sur la présentation des critères permettant de traiter tous types de variables :

- Le critère de Gini (CART) ;
- L'entropie de Shannon ou *Information Gain* (CART, ID3, C4.5, C5.0).

L'entropie quantifie, pour chaque division potentielle, le gain d'information suivant sa propre diminution dans chacun des sous-ensembles créés. Cet indice est préféré à celui de Gini pour son caractère explicite, quand bien même L. Breiman indique que le choix entre l'un ou l'autre des indice influe peu sur la construction de l'arbre (Breiman *et al.*, 1984).

La division des sous-ensembles est réitérée jusqu'à avoir atteint un critère d'arrêt spécifié, ou encore lorsqu'il ne reste plus qu'un individu dans chaque feuille, auquel cas la division devient impossible. L'étape de division est dite binaire quand elle se réduit à la création de deux sous-ensembles. Certains algorithmes ne se limitent pas à une division binaire.
D'un point de vue pratique, chaque sous-ensemble peut également être caractérisé par un poids relatif au nombre d'individus qu'il contient vis-à-vis du nombre total d'individus. Ce poids peut devenir un seuil relatif à un critère de séparation comme à celui d'un critère d'arrêt. Il permet, si besoin, de ne pas créer de classes trop peu fournies.

Critères d'arrêt

L'arrêt du processus de séparation peut dépendre d'un ou plusieurs critères. Dans la majeure partie des cas, plusieurs critères sont programmés. Lorsque le premier critère est atteint, cela signifie, de fait, la fin du processus de croissance. Les critères d'arrêts sont résumés dans la liste suivante :

- La profondeur de l'arbre ;
- Le nombre de feuilles ;
- Un effectif minimum d'individus dans chaque sous-ensemble - la division ultérieure risquerait de dépasser le seuil minimum d'individus dans un sous-ensemble ;
- La qualité globale de l'arbre est suffisante - la prochaine division n'améliorerait pas suffisamment la qualité de l'arbre.

Les arbres diffèrent majoritairement sur les deux derniers critères.

Les processus de divisions et d'arrêts sont donc principalement dirigés par la notion de pureté du sous-ensemble, aussi appelé nœud, obtenus par les divisions successives. La pureté peut être mesurée à l'aide de trois fonctions : l'indice de Gini, l'entropie, la fonction d'impureté élémentaire. Elle est maximale quand la totalité des individus d'un sous-ensemble appartient à la même classe et au contraire minimale lorsque les individus composant un sous-ensemble appartiennent à des classes différentes, dans les mêmes proportions. La pureté permet à la fois l'initiation du processus de division par la sélection de la variable et du seuil optimal permettant d'obtenir les meilleurs sous-ensembles, et l'arrêt

de ce même processus quand cette limite est atteinte (pureté maximale ou pureté acceptable signifiée dans les paramètres). Le processus d'élaboration de l'arbre peut s'arrêter à cette étape.

Des arbres plus sophistiqués intègrent une étape d'élagage.

L'élagage

L'algorithme construit en premier lieu un arbre de taille maximale pour ensuite élaguer les branches dont l'intérêt est limité. La complexité de l'arbre est définie par son nombre de feuilles. Les branches qui influencent le plus l'aspect complexe de l'arbre, et réduisent donc son pouvoir de généralisation, sont coupées. L'élagage permet ainsi d'éviter le phénomène de sur-apprentissage en réduisant la longueur des branches.

La construction de l'arbre ainsi que la complexité de sa structure dépendront du paramétrage. Un changement dans le paramétrage mène dans la majeure partie des cas à une évolution dans la structure de l'arbre, donc dans celle du jeu de règles issu du processus de classification. La sous-section suivante permet de faire le lien entre le potentiel issu de l'apprentissage automatique par arbre de décision, et les besoins de l'automate cellulaire dans l'élaboration d'un modèle de simulation performant.
La section suivante explore l'influence des différents paramètres utilisables pour la sélection d'un modèle (jeu de règles) remplissant les objectifs de simulation de l'urbanisation de notre zone d'étude.

3. PRINCIPE DU COUPLAGE AC - AD

Parmi les différents algorithmes pouvant être mobilisés, *J48 tree* présente l'avantage d'être disponible librement dans la suite de logiciels d'apprentissage automatique développée à l'Université Waikato en Nouvelle Zélande. Cette implémentation libre Java reprend l'algorithme C4.5 originalement développé par R. Quinlan (1993). C4.5 s'appuie à l'origine sur le critère d'entropie pour réaliser la division récursive. L'utilisation de cet algorithme est parmi l'une des plus courantes dans la littérature.

Pour rappel, l'objectif est la création d'un modèle qui permette de localiser la totalité des transitions de l'échantillon d'entraînement, et fournisse un résultat convenable[39] sur l'échantillon test. Concrètement, le modèle doit être performant sur des données vierges. L'analyse systématique des paramètres et de l'influence des variables dans l'évaluation de l'efficacité du modèle permet d'ores et déjà de répondre à la première partie du questionnement posé dans ce travail de recherche, à savoir d'un point de vue méthodologique : l'utilisation de la seule donnée d'occupation du sol permet-elle la création

[39] Plusieurs études montrent qu'une précision de 60 % à 80% est considérée comme exceptionnelle (Hema Laboratory, 2006; Mishra *et al.*, 2010; Newman *et al.*, 2016)

d'un modèle satisfaisant ? La prise en compte des données de localisation géographique améliore-t-elle la modélisation ?

3.1. Performance du modèle

L'efficacité d'un modèle peut être évaluée par le biais d'indicateurs qui rendent compte de sa capacité à reproduire une situation réelle. Dans ce travail de recherche, l'important est d'évaluer la capacité du modèle à reproduire le phénomène d'urbanisation modélisé à travers la transition d'une cellule dont l'état est « non urbanisé » en 1990 d'après les données CORINE Land Cover, vers un état « urbain » en 2006, tout en minimisant le nombre de règles nécessaires. La première étape de l'estimation des performances généralement utilisée dans l'évaluation d'un modèle créé à base d'observations rétrospectives est l'évaluation de sa capacité à reproduire des données observées à t+1, à partir des données observées à l'instant t. Ici, une simulation lancée à partir de données de 1990, a pour objectif de retrouver l'occupation du sol observée en 2006.

La comparaison entre la donnée d'occupation du sol observée et la donnée simulée peut, dans un premier temps, être réalisée par le biais d'une matrice de confusion classique (Table 5. 4).

Table 5. 4 Matrice de confusion

		Simulé	
		0	1
Observé	0	VN	FP
	1	FN	VP

Vrai Négatif (VN) : pas de transition observée, ni simulée -> Prédiction correcte
Vrai Positif (VP) : transition observée, et simulée -> Prédiction correcte
Faux Négatif (FN) : transition observée, mais non simulée -> Prédiction incorrecte
Faux Positif (FP) : transition simulée, mais non observée -> Prédiction incorrecte

La matrice de confusion, aussi appelée matrice d'erreur, est fréquemment utilisée pour évaluer la précision des simulations de changement d'occupation du sol (Stehman, 1997). De manière plus générale, c'est un outil qui sert à mesurer la qualité d'un système de classification. Différents indicateurs de performances peuvent être définis à travers son utilisation, dans le but d'évaluer la performance du modèle suivant différents angles (Fawcett, 2006; Stehman, 1997). Nous retiendrons particulièrement la précision et la sensibilité pour une évaluation de la performance dans la modélisation de l'urbanisation, ainsi que l'efficacité pour une évaluation de la performance globale du modèle, tant dans la modélisation des cellules évoluant, que dans celles présentant une absence d'évolution entre 1990 et 2006.

La précision

Cet indicateur rend compte de l'efficacité du modèle dans la prédiction des Vrais Positifs (VP), donc des transitions observées et simulées. Cette information est mise en perspective avec l'erreur produite en termes de transitions simulées mais non observées (Faux Positif). En résumé, ce paramètre permet d'évaluer la performance en prenant en compte les cellules « rajoutées » par le modèle.

$$Précision = \frac{VP}{VP + FP}$$

La sensibilité

Au contraire de la précision, ce paramètre évalue l'erreur en incluant les cellules « oubliées ».

$$Sensibilité = \frac{VP}{VP + FN}$$

La précision comme la sensibilité rendent compte de la capacité du modèle à reproduire le phénomène d'urbanisation. Cependant, d'un point de vue hiérarchique, la sensibilité sera considérée comme plus importante. En effet, il importe ici davantage de minimiser le nombre de cellules dont la transition observée n'est pas simulée (Faux Négatifs) plutôt que le nombre des celles dont la transition, simulée, n'a pas été observée (Faux Positifs) (Chapitre 4).

L'efficacité

Cet indicateur rend compte de la performance globale du modèle, en considérant aussi bien les transitions correctement simulées, que les absences de transition.

$$Efficacité = \frac{VN + VP}{VN + FN + VP + FP}$$

Soit la somme des prédictions correctes aussi bien dans les transitions que dans l'absence de transition, au regard du total de cellules composant la zone d'étude.

Coefficient de Corrélation de Matthews

Le Coefficient de Corrélation de Matthews (CCM) est mobilisable pour évaluer un modèle créé par arbre de décision (Matthews, 1975). Comme la matrice de confusion, il s'appuie sur la capacité du modèle à effectuer correctement les prédictions, c'est-à-dire, de prédire correctement l'évolution des états des individus entre deux jeux de données. Ce coefficient

diffère d'une matrice de confusion classique, car il permet d'évaluer la performance globale en prenant en compte un éventuel déséquilibre dans les classes à prédire (Boughorbel *et al.*, 2017). C'est pourquoi le CCM a été préféré à la matrice de confusion classique. La formule originale utilisée par B. W. Matthews, prend la forme suivante :

$$MCC = \frac{\frac{VP}{N} - S \times P}{\sqrt{PS\,(1-S)(1-P)}}$$

Où : $N = \text{VN} + \text{VP} + \text{FN} + \text{FP}$

$$S = \frac{VP + FN}{N}$$

$$P = \frac{VP + FP}{N}$$

L'efficacité globale mesurer d'après le CCM est basée sur l'expression d'une matrice de confusion classique et prendra la forme suivante :

$$MCC = \frac{VP \times VN - FP \times FN}{[(VP + FP) \times (FN + VN) \times (FP + VN) \times (VP + FN)]^{\frac{1}{2}}}$$

Les résultats du calcul du CCM sont compris entre -1 et 1. Une valeur de CCM égale à 1 indique un accord parfait entre l'observation et la prédiction ; une valeur de CCM égale à 0 indique que la prédiction n'est pas plus performante qu'une prédiction aléatoire ; une valeur de CCM égale à -1 indique un désaccord total entre l'observation et la prédiction.

Pourquoi ne pas se contenter de la mesure d'efficacité ?

Quand elle est confrontée à l'étude d'un phénomène exceptionnel, la matrice de confusion brute peut « biaiser » l'interprétation des résultats suivant l'objectif. S'il s'agit de rendre compte de l'efficacité du modèle en général (qu'il y ait transition ou non), la matrice de confusion classique est adaptée. Il faut cependant garder à l'esprit que si l'objectif est d'étudier l'efficacité du modèle dans l'apprentissage du phénomène exceptionnel, la matrice de confusion peut présenter une efficacité générale du modèle très satisfaisante, tout en ne permettant absolument pas de caractériser le phénomène exceptionnel. Ce point a été illustré dans la Section 1.2 à travers la création d'un modèle nul sur la zone d'étude.

Après avoir défini les indicateurs nous permettant d'évaluer la performance du modèle, la prochaine sous-section s'attache à évaluer la performance du processus selon les différents paramètres pouvant influencer la construction de l'arbre. Des tests ont, de ce fait, été réalisés afin d'évaluer la sensibilité du processus au changement de paramètres, tout en donnant la possibilité de sélectionner le paramétrage le plus à même de satisfaire les objectifs de modélisation visés dans ce travail de recherche.

3.2. Sélection du jeu de règles

Comme mentionné précédemment, les paramètres de l'algorithme doivent être spécifiés. Leur réglage influence le résultat produit, soit la structure et le contenu de l'arbre. Il est alors nécessaire de réaliser l'espace des paramètres pour évaluer la sensibilité de l'arbre au regard de leur modification. Le test systématique sur l'espace des paramètres est effectué sur les paramètres modulables dans notre arbre, à savoir : la taille du voisinage, la pondération, l'intervalle de confiance, le nombre minimum d'instances par feuilles. Ces derniers sont présentés en détails dans la suite de cette section.

La démarche de test se base sur la variation d'un seul paramètre, tout en fixant les autres. Elle offre la possibilité de tester l'ensemble des possibles dans les intervalles définis pour chaque paramètre. Cette étape permet de définir le paramétrage le plus intéressant en accord avec nos objectifs. Plus concrètement, le but est d'optimiser l'efficacité du modèle, en même temps que d'établir l'influence des paramètres et de leur modulation. Il est ainsi possible de répondre à une des questions fondamentales émergeant lors de l'utilisation d'un AC : celle de la taille du voisinage approprié pour caractériser au mieux le voisinage des cellules urbanisées. L'influence des variables de localisations géographiques est, à son tour, analysée en utilisant le paramétrage optimal de l'algorithme défini précédemment. Ces tests systématiques relèvent donc de deux étapes décrites par la Figure 5. 4.

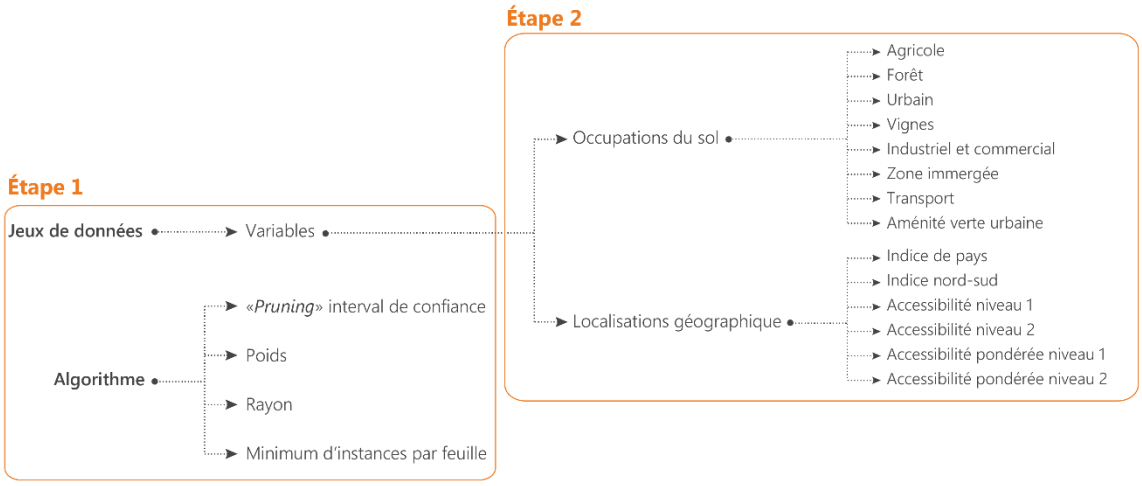


Figure 5. 4 : Démarche effectuée lors des tests systématiques

Dans un premier temps, l'étape 1 permet de trouver la meilleure combinaison de paramètres à sélectionner pour l'arbre de décision. Dans un second temps, l'étape 2 permet d'évaluer l'apport des variables de localisation géographique, en évaluant la performance du modèle obtenu en prenant uniquement en compte les variables d'occupation du sol.

Description des paramètres

Les paramètres peuvent être classés en différents types. En effet, certains paramètres testés interviennent sur la construction de l'arbre en amont du processus (l'intervalle de confiance, le nombre d'instances par feuilles, partition de l'apprentissage). Un autre type de paramètres est spécifique à l'utilisation de l'algorithme pour l'étude d'un phénomène exceptionnel (le poids). D'autres encore concernent l'évaluation de l'influence de la distance d'un type de sol, dans le voisinage des cellules urbanisées entre 1990 et 2006 sur notre zone d'étude (taille du voisinage). L'étape 1 de la Figure 5. 4 renvoie au test de ces paramètres.

L'impact de l'utilisation des variables de localisation géographique est, ensuite, plus finement étudié d'après le paramétrage de l'apprentissage sélectionné à l'étape 1. Cette deuxième étape est également schématiquement représentée sur la Figure 5. 4. Ainsi, l'influence de la localisation géographique dans l'un ou l'autre des pays peut être évaluée à travers le processus de construction de l'arbre : la variable représentant cette localisation intervient-elle dans la construction ? Si oui, quelle est son importance relative au sein de l'arbre ? L'utilisation de cette variable améliore-t-elle la performance générale de l'apprentissage en termes de nombre de règles, ou d'après les indicateurs mobilisés précédemment ?
L'ensemble des tests, et le paramétrage de l'arbre qui en résulte, permet d'apporter une réponse à ces questions. L'intervalle de test des différents paramètres est présenté dans le tableau (Table 5. 5) suivi de la description littérale de chaque paramètre.

Table 5. 5 : Liste des paramètres et intervalle test

Paramètres	***Intervalles***
Voisinage	10 px – 15 px – 20 px – 25 px – 30 px
Poids	20 – 40 – 60 – 80 – 100 – 120 – 140 – 160 – 180 – 200
Nombre d'instances par feuilles	1 – 2 – 3 – 4 – 5 – 15
Intervalle de confiance de l'élagage	0,5 – 0,25 – 0,1 – 0,01

La taille du voisinage

L'apprentissage est réalisé d'après la construction du jeu de donnée à partir de différents rayons de voisinage. En effet, la composition du voisinage en termes d'occupation du sol est extraite tous les 100 m jusqu'à 3 000 m. Le test permet de répondre à la question de la distance nécessaire à la définition du voisinage approprié dans l'automate cellulaire, pour obtenir un modèle performant.

Le poids

L'utilisation du paramètre « poids » a pour objectif de pondérer les individus que l'on cherche à étudier, soit les cellules urbanisées entre 1990 et 2006. La mise en place de cette pondération est rendue nécessaire en raison du déséquilibre des classes composant notre jeu de donnée. Elle fait référence au problème de l'apprentissage d'un phénomène exceptionnel

présent dans la littérature et abordé précédemment (Section 1.2). Comme mentionné dans le Chapitre 2, le développement urbain est une tendance majeure de changement d'occupation du sol. Néanmoins le ratio des changements concernant l'urbanisation dépasse rarement 0,5 % de la superficie des espaces européens, selon l'étude régionale menée en Europe (Chapitre 3). Pour cette raison, une pondération est utilisée entre la création du jeu de données et sa segmentation en partie entraînement et test. Elle permet de donner de l'importance aux individus de la classe à prédire puisque c'est celle qui représente les transitions urbaines. L'intervalle de pondération étudié est borné par le seuil qui permet un rééquilibrage total de l'échantillon, c'est-à-dire celui permettant à l'algorithme de détecter la classe à prédire comme étant aussi importante que la classe représentant le non-changement. Le ratio entre les deux classes est donc de 1/1. Cette valeur seuil est ensuite divisée pour faire varier l'importance de la classe à prédire en la réduisant. Comme la pondération engendre la création d'un artefact, l'objectif est de la minimiser tout en cherchant le modèle le plus efficace. L'expression du poids dans notre étude s'effectue d'après la division du ratio 1/1. En d'autres termes, un poids de 80 divise par 80 le nombre permettant d'obtenir l'équilibre parfait entre les classes. Un poids de 120 le divisera par 120. Un poids de 80 engendre donc un rééquilibrage plus important qu'un poids de 120.

Le partitionnement

Ce test permet d'établir la performance maximale de l'algorithme lors de l'utilisation d'un apprentissage sur l'ensemble des données. Cet apprentissage est réalisé une première fois sur un jeu de données construit uniquement à partir des variables d'occupation du sol, puis sur un jeu élaboré à partir des variables d'occupation du sol auxquelles s'ajoutent les variables de localisation géographique. La mise en relation entre la performance du modèle et la mobilisation des variables peut ainsi être réalisée, tout en évaluant l'influence des autres paramètres. En effet, un apprentissage réalisé à partir d'un échantillon créé par une sélection représentative, mais aléatoire au sein du jeu de données total, ainsi qu'un apprentissage sur 75% du jeu et 25% de test, nécessite la mise en place d'une évaluation de la sensibilité liée à cet échantillonnage. Ainsi, l'élaboration du modèle doit être testée sur la base de plusieurs tirages aléatoires avant d'être sélectionnée. L'utilisation d'un apprentissage sur la totalité du jeu de données dans cette étape permet de s'en affranchir. Par ailleurs, cette étape permet également de comparer la performance du modèle lors d'un apprentissage sur la totalité du jeu ou lors de la mobilisation de 75%.

Le nombre d'individus par feuille et l'intervalle de confiance de l'élagage

Ces paramètres influencent la longueur des branches ainsi que la pureté des feuilles obtenues, soit, globalement, la complexité de l'arbre créé. Comme l'objectif est de créer un modèle relativement précis, nous ne ferons pas intervenir ces paramètres dans leur caractère limitatif. Nous permettrons à l'arbre de se développer de façon précise, sans imposer un nombre d'individus par feuille et sans l'élaguer drastiquement. Ces paramètres peuvent néanmoins être modulés et mobilisés dans le cadre d'autres applications.

3.3. Paramétrage optimal

Comme mentionné précédemment, la définition du paramétrage optimal qui favorise la création d'un modèle efficace, dépend de la notion d'efficacité relative au contexte de travail et à l'objet d'étude. On s'attachera donc ici à analyser les résultats du test systématique d'après :

- Le coefficient de corrélation de Matthews
- L'efficacité de prédiction des transitions
- Le nombre de règles nécessaires
- La surestimation du nombre de transitions

Le coefficient de corrélation de Matthews de l'échantillon d'apprentissage devra tendre vers 1, tandis que celui de l'échantillon test devra être maximisé. L'objectif est de retrouver un maximum de transitions observées, cela aux dépens de l'augmentation éventuelle des transitions simulées mais non observées (FP). En effet, nous privilégions ici les vrais positifs (VP) tout en gardant l'opportunité d'évaluer les FP en termes de quantité ou de localisation. Ces derniers pourront être utilisés dans l'amélioration du modèle a postériori. Le nombre de règles doit être minimisé pour augmenter la portée générale du modèle et faciliter l'interprétation thématique du jeu créé.

Cette démarche permet de répondre aux interrogations suivantes :

- Un voisinage plus étendu induit-il systématiquement une performance supérieure du modèle ?
- Comment la pondération de la classe représentant un phénomène exceptionnel influence-t-elle les résultats de l'apprentissage ?
- Quel est l'intervalle de confiance optimal ?

La sélection du paramétrage optimal passe par une étape de validation « numérique » du modèle. La validation numérique contrairement à la validation spatiale du modèle, évoquée brièvement dans le précédent chapitre, a pour objectif de contrôler la structure du jeu de règles (nombre de règles, nombre de variables) parallèlement aux performances du modèle en matière de prédiction. L'étape de validation est présentée dans le chapitre suivant après la définition de paramétrage de l'algorithme d'apprentissage.

Conclusion du chapitre 5

La création automatique de règles de transition pour automate cellulaire devient d'autant plus intéressante quand les arbres de décisions sont mobilisés à cet effet. Ce chapitre a permis de développer les avantages inhérents à l'utilisation d'algorithmes d'apprentissage automatique et de souligner l'intérêt particulier que représentent les arbres de décision. Ces derniers sont extrêmement explicites et simples d'utilisation. Ils ne nécessitent aucun prétraitement des données qui peuvent aussi bien être qualitatives que quantitatives, discrètes ou continues. Suivant l'algorithme utilisé, la variable à expliquer représentant le phénomène modélisé, peut être catégorielle binaire ou non. Ce point est intéressant dans le cadre de ce travail de recherche puisqu'il permet d'étudier et de créer les règles illustrant plusieurs types de transition vers l'urbain.

Si l'arbre de décision semble approprié au questionnement développé dans ce travail, le modèle issu de l'apprentissage doit néanmoins passer par une phase de validation. La validation peut être divisée en deux étapes principales : celle effectuée en amont de l'interprétation et de la simulation du jeu de règles, qui évalue la performance et la sensibilité du modèle aux changement des paramètres présentés dans ce chapitre, suivie d'une validation spatiale pour observer la cohérence du développement au-delà de l'analyse purement numérique du modèle. La démarche de validation dans sa globalité est exposée dans le chapitre 6.

Chapitre 6

Une difficulté : le calibrage

Le chapitre 4 a associé succinctement la démarche de calibrage à celle de validation dans le cadre du questionnement sur l'utilisation des modèles et notamment des automates cellulaires. Le présent chapitre porte plus précisément de la phase de calibrage. En effet, comme le mentionne J. M. Epstein, la seconde question « abordée par curiosité », après celle du « pourquoi modéliser ? » est bien celle de la validation (Epstein, 2008).

Si certaines applications récentes ne présentent pas d'étape de validation des résultats produits (van Vliet *et al.*, 2016), l'évaluation de leur qualité reste un élément important qu'il apparaît délicat de négliger (Paegelow *et al.*, 2018). En effet, la validation permet de communiquer les résultats dans un contexte précis et transparent. Elle peut aussi s'avérer rassurante lors du développement d'outils destinés à un public plus large, notamment non-issu du monde académique, qui chercherait dans le modèle, des possibilités d'applications concrètes.

L'étape de validation permet par ailleurs de cerner plus objectivement les performances d'un modèle, en soulignant notamment ses limites. En ce sens, elle met en lumière l'incertitude liée à l'utilisation du modèle, la « perfection relative » des résultats obtenus. Cette étape est extrêmement intéressante, à la fois pour parfaire la démarche de modélisation, mais également pour adapter l'utilisation du modèle à ses performances, lesquelles s'illustrent dans la validité des résultats. Comme le souligne J. M. Epstein, un modèle peut s'avérer précieux, sans pour autant être totalement « juste ». Cette considération est également illustrée par G. Box : « Tous les modèles sont faux, mais certains sont utiles[40] » (Box & Draper, 1987).

Les chapitres 4 et 5 ont présentés respectivement les raisons et objectifs associés à la sélection, d'une part, de l'outil de simulation, l'automate cellulaire, et d'autre part, de la méthode menant à la création et au calibrage des règles de transition, à savoir l'apprentissage automatique par arbres de décision. Si la création et le calibrage des règles pour l'AC sont automatiques, le calibrage des paramètres de l'algorithme AD ne l'est pas. Or cet élément influe le jeu de règles, et permet *in fine* d'en apprécier la validité, raison pour laquelle la

[40] *All models are wrong, but some are useful* (Box and Draper, 1987).

réflexion sur les paramètres de l'AD est utile au choix du modèle comme à l'appréciation de sa performance.

Le présent chapitre permet donc, dans un premier temps, de revenir sur la notion de validation dans les modèles de changement d'occupation du sol. Le modèle utilisé dans ce travail est ensuite sélectionné après une première phase de validation, intrinsèquement liée à la phase de calibrage du paramétrage de l'arbre de décision. Cette phase permet également d'établir l'influence des variables d'occupation du sol et de localisation géographique sur la performance du modèle. Enfin, un paramétrage est sélectionné pour présenter l'enjeu de la validation spatiale dans l'évaluation de la performance du modèle ainsi qu'en tant que support de visualisation des résultats.

1. VALIDATION ET CALIBRAGE

L'étape de validation, comme celle du calibrage, reste sujette à discussion encore aujourd'hui. Aucune méthode particulière ne fait actuellement consensus pour valider les modèles de changement d'occupation du sol. Nous retenons deux limites majeures rencontrées lors de cette étape (van Vliet *et al.*, 2016) :
L'impossibilité de déterminer l'exactitude d'une proposition lors de l'étude d'un système non clos, comme celui des changements d'occupation du sol. En effet, les modèles s'appuient sur des données d'entrées qui ne pourront jamais être connues dans leur totalité (les processus résultent de la combinaison d'une infinité de variables).
Les modèles de changement d'occupation du sol ne permettent pas une validation absolue en raison de la part d'incertitude existante au sein même de la donnée observée, mais également en raison du caractère non stationnaire[41] du processus.
L'objectif recherché à travers cette étape est donc principalement une évaluation de la performance du modèle au regard de l'objet d'étude, et non une recherche de la perfection (Houet *et al.*, 2016).

Le besoin d'évaluer la pertinence d'un résultat est inhérent au travail de recherche, d'autant plus lors d'une démarche d'aide à la décision. Les résultats obtenus pendant l'étape de validation ne présentent donc pas un caractère binaire (juste ou faux), mais permettent de faire évoluer le modèle en fonction de l'évaluation de sa performance. C'est pourquoi, l'étape de validation reste « un défi ouvert sans fin clairement définie » (Aumann, 2007).

La validation d'un modèle en géographie peut être abordée suivant deux angles, parfois conjugués. Le premier est porté sur l'analyse de la performance à travers les résultats numériques produits lors de l'élaboration du modèle. Ici, nous considérons, par exemple, le coefficient de corrélation de Matthews (CCM), l'efficacité de prédiction des transitions urbaines (Vrais Positifs), le nombre de règles nécessaires, la surestimation des transitions (Faux Positifs) et l'erreur de prédiction (Faux Négatifs). Ces données sont issues d'une matrice de confusion classique présentée dans le chapitre 5, parallèlement aux différents indicateurs de performance (Chapitre 5 section 3.1). Les résultats peuvent donc être analysés indépendamment de leur spatialisation, puisque l'analyse des transitions est effectuée de manière binaire. Elle est donc identique aux données observées (juste – VP et VN) ou non-identique (fausse – FP et FN). Le second angle est basé sur la spatialisation des résultats, qui, par l'appréhension d'une cohérence spatiale, apparaît moins manichéenne que l'approche binaire.

Cette section permet dans un premier temps, de présenter différentes formes de validations numériques ou spatiales. L'étape de validation est ensuite discutée d'après la structure du modèle développé (AC contraint ou non). Pour finir, la combinaison des méthodes retenues est argumentée.

41 Processus dont la structure évolue dans le temps.

1.1. Des contraintes à identifier

La difficulté associée à l'étape de validation est variable. Elle pourra dépendre de plusieurs paramètres, notamment des contraintes intégrées au préalable lors du développement du modèle ou celle intégrées lors de la simulation. Le chapitre 4 mentionnait l'évolution des AC, amenés à s'éloigner progressivement de leur définition classique (Couclelis, 1997). Nous retiendrons ici deux types de contraintes majeures pouvant intervenir dans l'élaboration du modèle et impactant le processus de validation : l'intégration d'une contrainte quantitative et l'intégration d'une contrainte spatiale.

Contraintes quantitatives

L'intégration d'une contrainte quantitative est une démarche qui permet de limiter les transitions proposées par le modèle à un nombre prédéfini. Le seuil limite de transition est ainsi imposé en amont de la simulation et prémunit le modèle d'une surestimation des changements. L'intégration de cette contrainte dans le processus de modélisation est courante dans le développement de modèles AC étudiant les changements d'occupation du sol (Ghosh *et al.*, 2017).

Dans le cas d'une approche rétrospective, la contrainte est déterminée d'après les changements observés des cellules existantes. Par exemple, nous pouvons établir un seuil de transition en comptant le nombre de transitions urbaines observées entre 1990 et 2006 pour contraindre la simulation AC. Dans un cadre prospectif, une contrainte quantitative peut être imposée, selon différentes modalités. Lors d'une approche statistique, la contrainte quantitative peut par exemple être déterminée d'après l'utilisation des chaînes de Markov[42]. À défaut, elle peut répondre à un objectif politique (limitation de l'étalement urbain, urbanisation de zones spécifiques, etc.), dont il sera question d'étudier les conséquences à travers l'utilisation de modèles de simulation (Frémond, 2015). Ainsi, à partir de la quantité de transitions prédéfinie, statistiquement ou à dire d'expert politiquement, différents scénarios d'aménagement peuvent être testés et comparés par le biais de simulations. L'utilisation du modèle est, dès lors, prémunie de toute surestimation des changements.

À l'issue de la détermination de la contrainte, qu'elle résulte d'un cadre rétrospectif ou prospectif, deux cas de figure sont susceptibles de se présenter, selon les caractéristiques du modèle et sa zone d'application. Dans une première hypothèse, le modèle devra avoir recours à plusieurs pas de simulation pour atteindre le seuil établi. Dans la seconde, un seul pas de simulation suffira au modèle pour atteindre le seuil de transitions, voire le surpasser. Dans ce cas, un choix aléatoire au sein des transitions proposées devra être effectué. Ce choix peut être dépendant d'une probabilité de transition variable ou encore hiérarchisé d'après des critères exogènes (zone attractive, zone réglementée, etc.). Au regard de cette sélection aléatoire, la sensibilité du modèle devra néanmoins être évaluée, notamment d'un point de vue spatial, puisque chaque essai pourra engendrer une image différente des changements en matière de localisation (Chapitre 5 – Section 2.3).

[42] Processus stochastique permettant de définir un changement d'état dans le futur à partir d'une donnée dans le présent.

Contraintes spatiales

Les contraintes spatiales sont définies ici comme les zonages préexistants interdisant une transition, ou augmentant sa probabilité de façon exogène. Elles peuvent se matérialiser par l'intégration de plans réglementaires ou d'objectifs d'aménagements (densification, préservation de certains espaces, etc.). Le recours aux zonages réglementaires est courant dans l'élaboration des modèles[43]. Leur implémentation peut notamment être utilisée pour évaluer leur impact sur les changements d'occupation du sol, ou encore l'adéquation des changements vis-à-vis des contraintes réglementaires. Pour autant des études ont montré l'écart entre le cadre fixé par la réglementation ou les plans d'aménagement, et la réalité du développement et des changements d'occupation du sol (Alfasi *et al.*, 2012). Ces deux points soutiennent notre choix de ne pas intégrer les plans réglementaires pour limiter et délimiter en amont les zones pouvant accueillir le développement urbain.

L'interdiction ou la limitation exogène, par l'utilisateur, des changements possibles peut s'avérer favorable à la performance générale du modèle. En effet, dans le cadre d'un AC, des cellules répondant positivement à une règle mais localisées dans un périmètre proscrit, n'effectueront pas la transition. Les faux positifs (FP) peuvent donc être limités en amont. Ils n'intégreront pas les cellules répondant positivement aux règles mais localisées dans un périmètre proscrit, ni celles induisant un dépassement du seuil fixé de transitions.

Or, l'objectif de ce travail n'est pas d'augmenter la performance du modèle en intégrant des zonages restrictifs, ou des seuils dans le nombre de transitions. Il s'agit de parvenir à une certaine performance à travers le modèle lui-même, par l'apprentissage basé sur la description du voisinage et la localisation de la cellule. De plus, la différence entre le modèle et les données observées présente en elle-même un intérêt, puisque comme le souligne P. H. Verburg « la différence entre la réponse donnée par le modèle et les changements d'occupation du sol observés pourrait indiquer que lors d'une des étapes du développement du modèle, le processus réel est capturé de manière inadéquate. Tracer l'origine de cette différence pourrait révéler des informations concernant le fonctionnement du processus réel, et fournir des suggestions concernant l'amélioration du modèle à cet égard » (Verburg *et al.*, 2016). Ainsi, le modèle présenté ici n'intègre ni contrainte quantitative, ni contrainte spatiale.

1.2. Des possibilités à adapter

Notre démarche de modélisation induit une simulation spatialement explicite[44]. L'échantillon des techniques de validation présentées dans cette sous-section concerne donc celles relevées dans un tel cadre de simulation. L'objectif est de donner un aperçu global des techniques mobilisées pour la validation. Les modèles de changement d'occupation du sol peuvent être validés :

- d'après la précision de leurs résultats (localisation ou forme urbaine des changements) ;
- par l'analyse de leur sensibilité lors de la modification des variables ou des seuils des paramètres ;

[43] Metronamica, parmi d'autres modèles, permet par exemple l'intégration de plan réglementaires.

[44] « Approche basée sur une carte initiale d'occupation du sol, typiquement représentée par une grille raster composée de cellules, sur laquelle des changements vont apparaître au cours du temps » (van Vliet et al., 2016).

- par l'incertitude du modèle concernant la variation des résultats, à paramétrage et variables identiques.

La validation peut être effectuée indépendamment des données utilisées pour construire le modèle, ou d'après ces mêmes données. L'utilisation de données externes (données issues d'un espace ou d'une acquisition temporelle différente, ou encore un échantillon de données isolées lors de la création du modèle) permet une validation plus objective et une estimation plus réaliste de la performance du modèle notamment lors de l'utilisation d'algorithmes d'apprentissage automatique. Nous nous intégrons dans cette démarche en isolant 25% du jeu de données lors de l'apprentissage par arbre de décision. Quand une portion du jeu de données est isolée pour le test (Chapitre 5) la sélection des cellules mises à l'écart est aléatoire. Elle préserve néanmoins les caractéristiques globales du jeu complet, notamment la proportion des cellules urbanisées et celles qui ne présente pas de changement entre 1990 et 2006. Dans ce cas, la sensibilité du modèle vis-à-vis de la composition de l'échantillon isolé doit être testée. Ce point est développé plus précisément dans la suite du chapitre (section 2.3).

Dans la continuité, l'étape de validation peut être réalisée suivant une approche quantitative ou spatiale, stricte ou progressive. Les différents indicateurs quantitatifs tels que la précision, la sensibilité, l'efficacité et le Coefficient de Corrélation de Matthews (CCM), font partie de ceux qui permettent d'évaluer la précision de la localisation dans un objectif de réponse binaire (juste ou fausse). Cette approche est la plus stricte et ne laisse pas de flexibilité dans l'appréciation de la performance du modèle. Elle est utilisée dans la plupart des études de cas récentes (van Vliet *et al.*, 2016). Il est néanmoins légitime de se poser la question du caractère adéquat de cette méthode de validation lors de l'utilisation de modèles de changement d'occupation du sol, d'autant plus lorsque l'on souhaite conserver une précision relativement importante à des échelles néanmoins larges. Ainsi, le modèle est-il faux si la transition modélisée n'est pas localisée à l'endroit exact de la transition observée, mais dans un certain périmètre ? Par exemple, une analyse « *near hit* » (Hagen, 2003; van Vliet *et al.*, 2013a) permet d'intégrer la logique floue dans l'évaluation de la précision de localisation de la transition ou celle du changement de classe (Figure 6. 1).

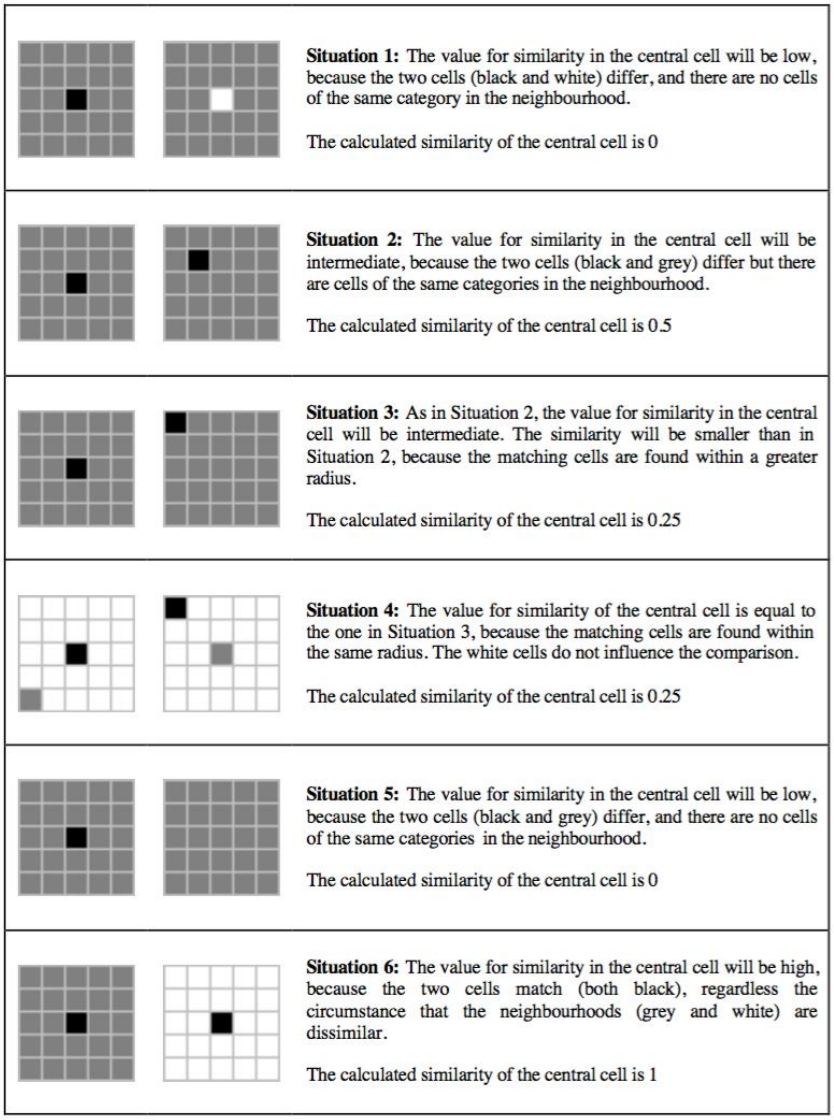


Figure 6. 1 Six situation ou les cartes de gauche et droite sont comparées, en considérant l'aspect flou de la localisation (Hagen, 2003):

Ainsi, l'analyse des résultats n'est plus binaire mais présente un seuil de tolérance plus flexible dans la modélisation d'une transition considérée comme juste. Cet élément fait l'objet de développements lors de la validation spatiale (Section 3).

1.3. Un choix à justifier

L'étape de validation est particulièrement enrichissante pour l'étude du processus de développement urbain sur la zone frontalière franco-allemande. Elle permet la sélection d'un meilleur modèle en fonction du paramétrage dont est issu l'arbre de décision. Pour rappel, la définition du meilleur modèle est basée sur les critères suivant spécifiés chapitre 5 section 3.3 :

- Le coefficient de corrélation de Matthews ;
- L'efficacité de prédiction des transitions ;
- Le nombre de règles nécessaires ;
- La surestimation du nombre de transition.

L'objectif de l'étape de validation dans ce travail est double. Elle permet dans un premier temps de sélectionner le meilleur modèle en étudiant la sensibilité des résultats obtenus en fonction du paramétrage. Dans un second temps, l'influence des variables de localisation géographique peut être analysée, inférant notamment l'impact de la frontière dans la caractérisation du développement urbain. En effet, puisque l'algorithme intègre la sélection des variables influentes dans la description du voisinage ou dans la caractérisation intrinsèque de la cellule, il sera possible d'indiquer si la localisation géographique dans l'un ou l'autre des pays se révèle être un critère retenu.

Il s'agit de nous appuyer, dans un premier temps, sur une validation stricte cellule par cellule, permettant de sélectionner le meilleur modèle. Bien que basée sur une vision binaire, elle présente l'avantage d'être simple et permet de faciliter la communication autour de l'outil à tout type d'utilisateur (Sous-section 3.2). L'utilisation d'une méthode plus flexible incluant par exemple de la logique floue améliorerait la caractérisation de la performance du modèle mais ne rendrait pas le modèle en lui-même, plus performant. Elle n'est donc pas privilégiée à ce stade. Une méthode est cependant utilisée dans un second temps, afin de réaliser une approche spatiale de l'évaluation des résultats, laquelle permet d'analyser l'homogénéité des performances de modélisation au sein de la zone d'étude (Sous-section 3.3).

Les étapes de paramétrages et de validation sont étroitement liées puisque les tests de paramétrage sont réitérés jusqu'à leur validation. Leur combinaison permet *in fine* la sélection du paramétrage optimal, donc du meilleur modèle.

La première étape menant à la sélection du meilleur modèle sur notre zone d'étude est celle de la définition du paramétrage de l'arbre de décision, incluant la taille du voisinage et la pondération servant à équilibrer l'effectif de la classe à prédire (transition urbaine). L'objectif est donc d'évaluer la sensibilité du modèle à la variation de ces deux paramètres. L'exploration de leur influence permet de valider le choix des seuils qui mènent à la sélection du meilleur modèle, mais apporte également une réponse quant à la taille du voisinage à considérer dans la simulation AC. Cette étape relève donc du calibrage de l'algorithme de l'arbre de décision.

Pour ce faire, nous maximisons, dans un premier temps, l'information contenue dans le jeu de données, dans le but de maximiser les performances de l'apprentissage. Ainsi, toutes les variables sont incluses (occupation du sol et localisation géographique) et l'apprentissage a été effectué sur l'intégralité du jeu de données des données. Cette étape permet de s'abstraire du biais lié à l'échantillonnage aléatoire et de la nécessité d'évaluer la sensibilité du modèle comme dans le cas de l'utilisation d'une partie entraînement et test. Ces éléments sont explorés par la suite (Section 2.3). Ainsi, pour commencer, les seuls points considérés sont :

- La performance du modèle lié au paramétrage de l'AD ;
- L'apport des variables de localisation géographique.

C'est dans un second temps qu'un échantillonnage est effectué pour dissocier une partie entraînement et une partie test du jeu de données, et qu'une analyse de la sensibilité du modèle liée à l'échantillonnage est réalisée. Chaque simulation est alors réalisée 10 fois pour évaluer le lien entre l'échantillonnage et la performance des modèles produits. Le meilleur paramétrage est ensuite choisi d'après les valeurs médianes des indicateurs de performances : médianes des CCM (entraînement et test), mise en relation avec la valeur médiane du nombre de règles nécessaire, et de la valeur médiane de la capacité de prédiction

des transitions. Une fois le choix du modèle arrêté, une analyse plus spécifique de l'apport des variables de localisation géographique est considérée.

L'étape de validation est donc répétée autant que nécessaire à l'obtention du paramétrage du modèle optimal pour la simulation du développement urbain sur la zone d'étude. La section suivante présente les tests réalisés à cet effet.

2. TEST DES PARAMÈTRES

Comme expliqué précédemment (Sous-section 1.3) cette première étape de calibrage explore la performance du modèle, en s'appuyant uniquement sur l'occupation du sol comme facteur explicatif de la transition urbaine. Cette simplification permet de focaliser exclusivement l'analyse sur le lien entre la taille du jeu de règles (nombre de règles, nombre de variables mobilisées) et la performance du modèle.

Le travail de calibrage présente également un intérêt thématique dans notre travail, puisqu'à l'issue, il devient possible de mesurer le bien-fondé de l'utilisation des variables qui spécifient la localisation géographique, et notamment, la localisation dans l'un ou l'autre des pays de la zone d'étude ou encore l'accessibilité des pôles de différents niveaux (Chapitre 7).

1.1.1 Test systématique : influence de la distance et de la pondération

Le test systématique est réalisé dans le cadre d'un apprentissage effectué sur la totalité du jeu de données (sans échantillon test) exclusivement sur les données d'occupation du sol. L'objectif est, ici, d'évaluer la performance optimale de l'apprentissage, et plus précisément d'interroger le rayon de voisinage adéquat. Ainsi, quel rayon de voisinage doit être pris en compte lors de l'étude des transitions urbaines ? Un voisinage plus grand permet-il systématiquement d'obtenir un modèle plus performant ?

Pour répondre à ces questions, l'objectif est dans un premier temps d'étudier l'influence de l'intervalle de confiance et du nombre minimum d'instances par feuille, à travers la réalisation de tests systématiques. Pour se faire, la performance du modèle produit en fonction de la valeur d'un des deux paramètres est mis en relation avec le CCM obtenu.

Pour rappel, les tests suivants ont été effectués sur le jeu de données (chapitre 5) :

Table 6. 1 : Rappel de la liste des paramètres et intervalle test

Paramètres	***Intervalles***
Voisinage	10 px – 15 px – 20 px – 25 px – 30 px
Poids	20 – 40 – 60 – 80 – 100 – 120 – 140 – 160 – 180 – 200
Nombre d'instances par feuilles	1 – 2 – 3 – 4 – 5 – 15
Intervalle de confiance de l'élagage	0,5 – 0,25 – 0,1 – 0,01

Dans la réalisation de l'espace des paramètres ci-dessus, 1200 combinaisons sont donc analysées.

Auparavant, il faut néanmoins spécifier le type de voisinage étudié. Les différentes formes de voisinage ont été présenté précédemment (chapitre 4). Nous avons opté pour un voisinage de Moore étendu pour inclure un maximum de cellules périphériques. La taille du voisinage évolue également en fonction du rayon considéré (Figure 6. 4). Elle peut être appréhendée en nombre de cellules selon deux approches. Le voisinage propre fait référence au voisinage restreint à la seule distance étudiée (Figure 6. 2). Le voisinage cumulé représente quant à lui la totalité des cellules entourant la cellule d'intérêt, et ce jusqu'à une distance spécifiée en pixel (Figure 6. 3).

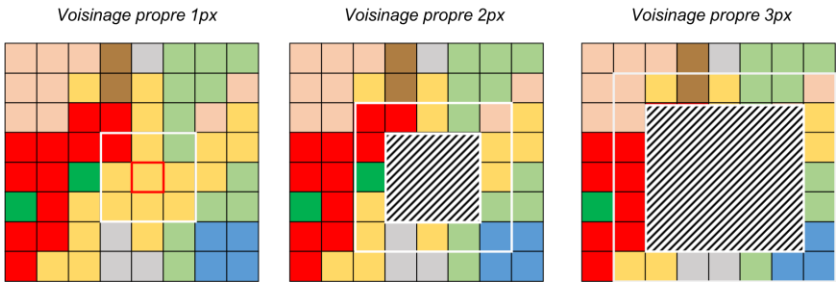


Figure 6. 2 :Voisinage propre

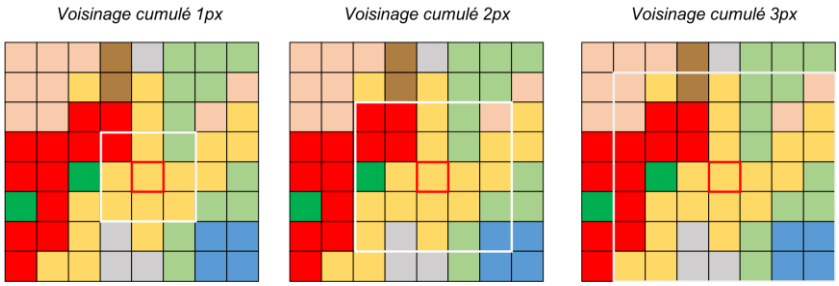


Figure 6. 3 :Voisinage cumulé

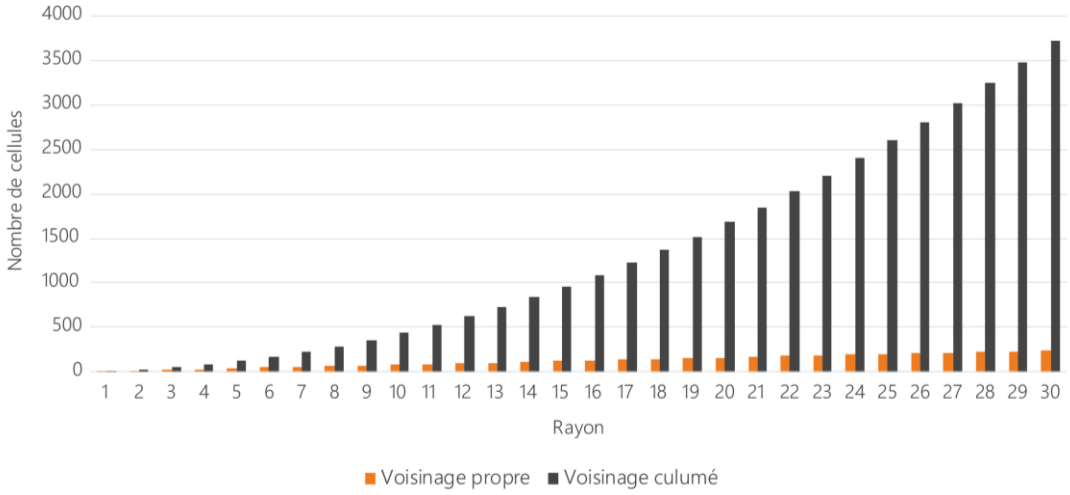


Figure 6. 4 : Évolution du nombre de cellules dans le voisinage en fonction du rayon en pixel

Le voisinage cumulé ne restreint pas l'information d'occupation du sol aux seules cellules de la distance spécifiée. Il présente une plus grande richesse en termes de description, puisqu'il inclut cette information pour l'ensemble des cellules se situant entre ces dernières et la cellule d'intérêt. C'est pourquoi cette approche du voisinage est préférée à celle du voisinage propre.

Les tests systématiques sont donc réalisés sur une donnée d'occupation du sol extraite d'un voisinage cumulé. Leur objectif dans un premier temps est d'évaluer l'influence de l'intervalle de confiance et celle du nombre d'instances par feuille, d'après le CCM. La figure 4 présente conjointement la variation du CCM et celle de l'intervalle de confiance en fonction des différentes combinaisons de paramètres testées. L'intervalle de confiance le plus stable qui permet d'obtenir les CCM les plus performants est celui de 0,25 (Figure 6. 5).

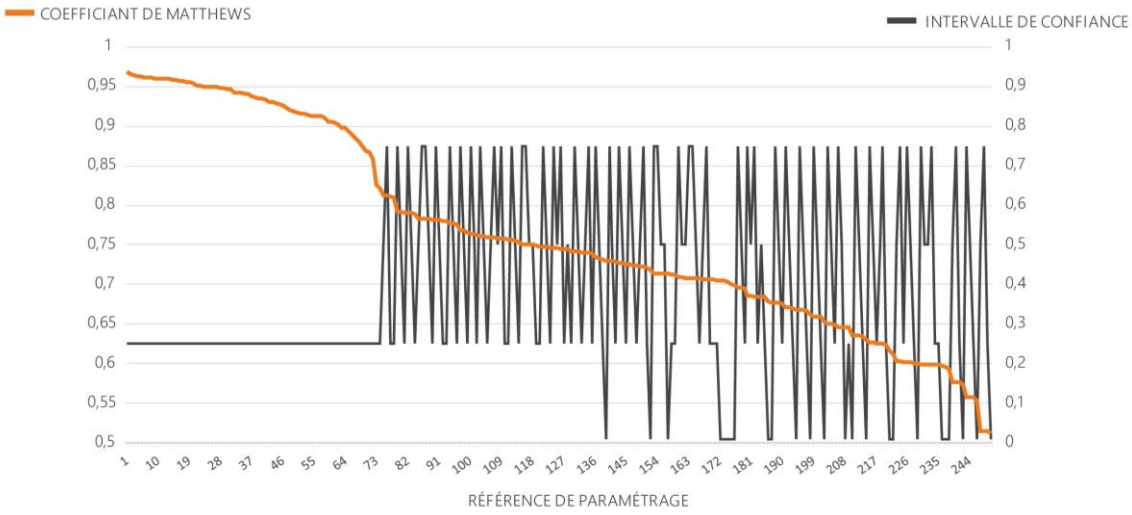
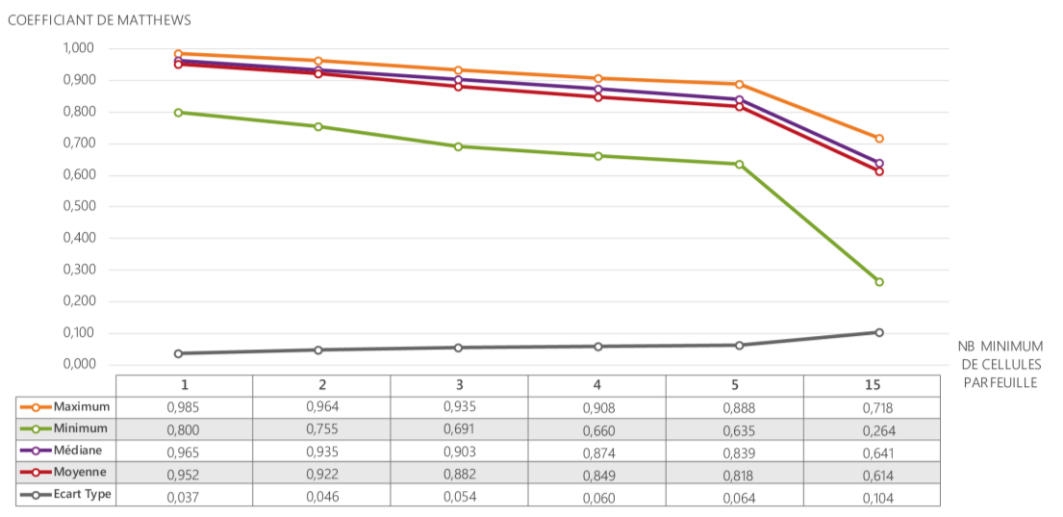


Figure 6. 5 : Influence de l'intervalle de confiance sur le CCM

La figure 5 présente elle une projection des valeurs maximales, minimales, moyennes et médiane des CCM obtenus suivant le paramétrage du nombre minimum d'instances par feuilles. La valeur la plus basse de ce paramètre est retenue puisqu'il permet d'obtenir les meilleurs CCM moyen, médian, maximum ou minimum, ainsi que le plus faible écart type, sur l'ensemble des combinaisons. Ce paramètre est fixé à 1, dans le but de garantir la plus grande précision possible (Figure 6. 6). Concrètement, ce paramétrage induit la possibilité d'obtenir des règles qui ne concernent qu'un nombre limité de cellules (jusqu'à une seule cellule). L'intérêt de ce choix est discuté plus en détails dans la partie suivante (Chapitre 7 section 1).



	1	2	3	4	5	15
Maximum	0,985	0,964	0,935	0,908	0,888	0,718
Minimum	0,800	0,755	0,691	0,660	0,635	0,264
Médiane	0,965	0,935	0,903	0,874	0,839	0,641
Moyenne	0,952	0,922	0,882	0,849	0,818	0,614
Ecart Type	0,037	0,046	0,054	0,060	0,064	0,104

Figure 6. 6 : Influence du nombre minimum d'individus par feuille sur le Coefficient de corrélation de Matthews (apprentissage occupation du sol seulement)

Les modèles les plus performant, d'après le CCM, au regard des paramètres précédents, soit limitant le nombre minimum de cellules par feuille à 1 et fixant l'intervalle de confiance à 0,25, sont retenus pour analyser l'influence du rayon du voisinage et celle de la pondération des cellules urbanisées. Ainsi, le nombre de combinaisons de paramètres testé est, dès lors, réduit à 50. L'intervalle des CCM des modèles sélectionnés d'après les paramètres fixés précédemment pour étudier l'évolution l'influence du rayon et de la pondération est compris entre 0,7999 et 0,984.

La performance du modèle est donc explorée à travers le CCM, en fonction de la taille du voisinage, donc de son rayon, et de la pondération des cellules urbanisées. L'objectif est de déterminer si un modèle créé à partir d'un voisinage plus large ou d'une pondération plus importante est automatiquement plus performant.

La Figure 6. 7 permet de constater qu'un rayon supérieur, contenant à *fortiori* plus d'informations, ne permet pas toujours d'obtenir le modèle le plus performant. Ainsi, un modèle créé sur la base d'un voisinage de 15px, soit 1 500m, peut permettre d'atteindre un CCM proche de 0,98. Néanmoins, un rayon compris entre 25 et 30 px, soit compris entre 2 500m et 3 000m, semble plus adapté à la réalisation d'un apprentissage performant. Les rayons au-delà de 30 px, qui n'apportent pas d'amélioration en termes d'apprentissage au regard du CCM, n'apparaissent pas dans l'analyse présentée ici.

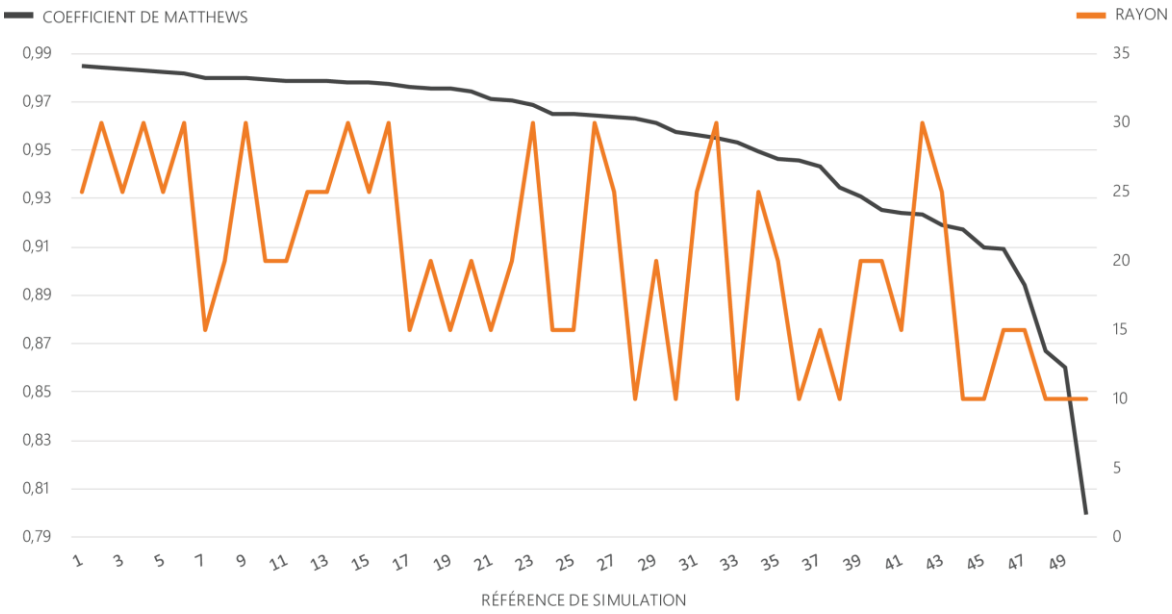


Figure 6. 7 : Évolution du coefficient de corrélation de Matthews en fonction rayon du voisinage

La Figure 6. 8 présente la projection du poids, qui permet de rééquilibrer les classes artificiellement pour l'apprentissage d'un phénomène exceptionnel, et du CCM. Cette figure montre là aussi que ce n'est pas la pondération la plus importante qui permet d'obtenir un apprentissage plus performant. C'est dans un intervalle de pondération compris entre 80 et 120 que le CCM est maximal (au-dessus de 0,98).

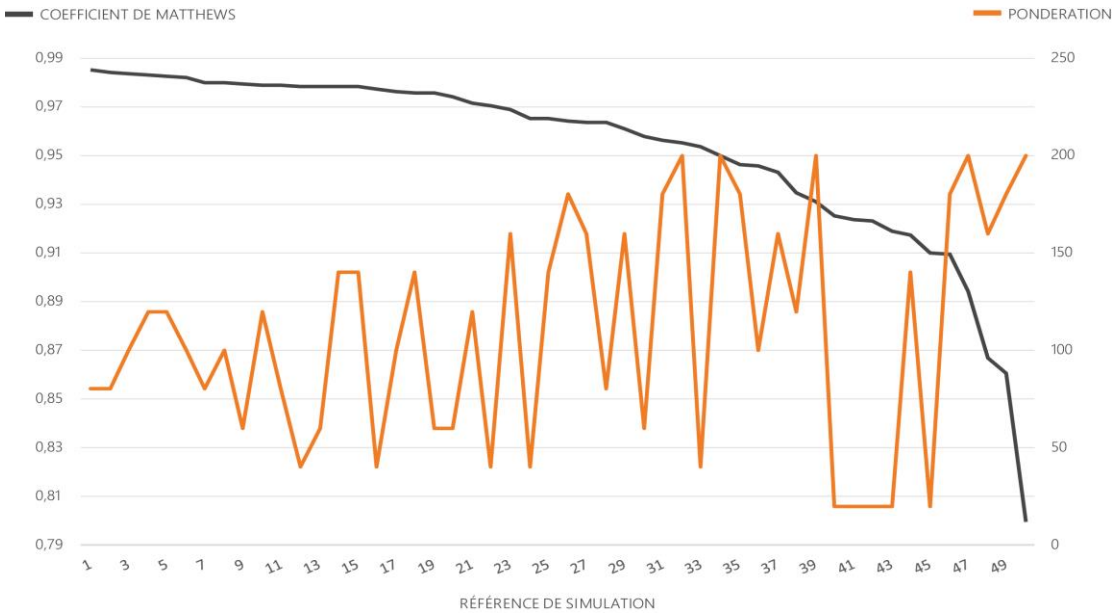


Figure 6. 8 : Évolution du coefficient de corrélation de Matthews en fonction du poids attribué aux transitions exceptionnelles

La performance du modèle est, ensuite, explorée en fonction du nombre de règles définies, et toujours à travers le coefficient de corrélation de Matthews. L'objectif est d'observer si un modèle contenant une plus grande quantité de règles est automatiquement plus performant. La Figure 6. 9 montre qu'un arbre plus grand (composé de plus de règles) ne présente pas une performance supérieure au regard du CCM. Ce point enrichit la réflexion

quant au sur-apprentissage. Ici, un arbre plus précis, donc un apprentissage plus poussé, ne donne pas le meilleur résultat. Le même constat peut être effectué grâce à l'analyse du nombre de variables mobilisées dans la construction (Figure 6. 10).

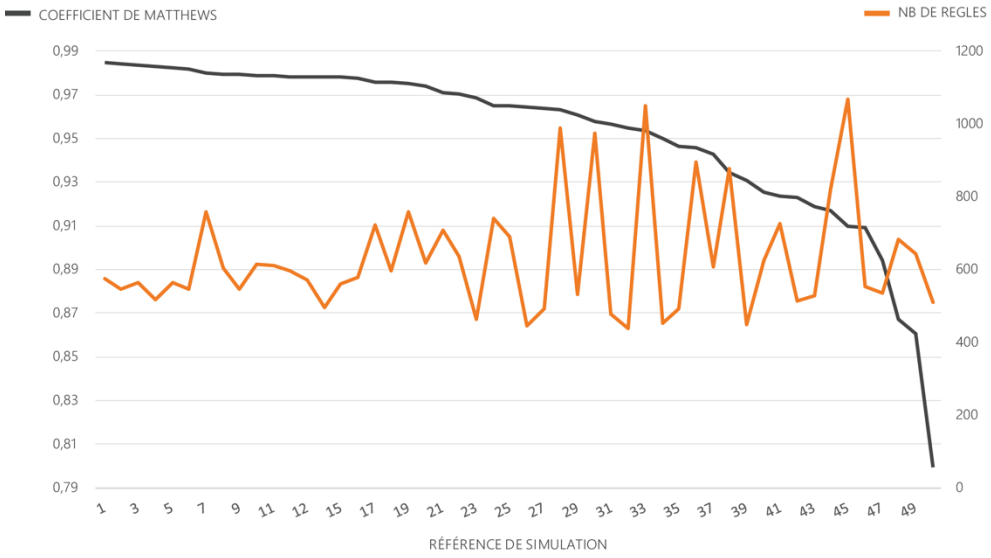


Figure 6. 9 : Évolution du coefficient de corrélation de Matthews en fonction du nombre de règles présentes dans le modèle

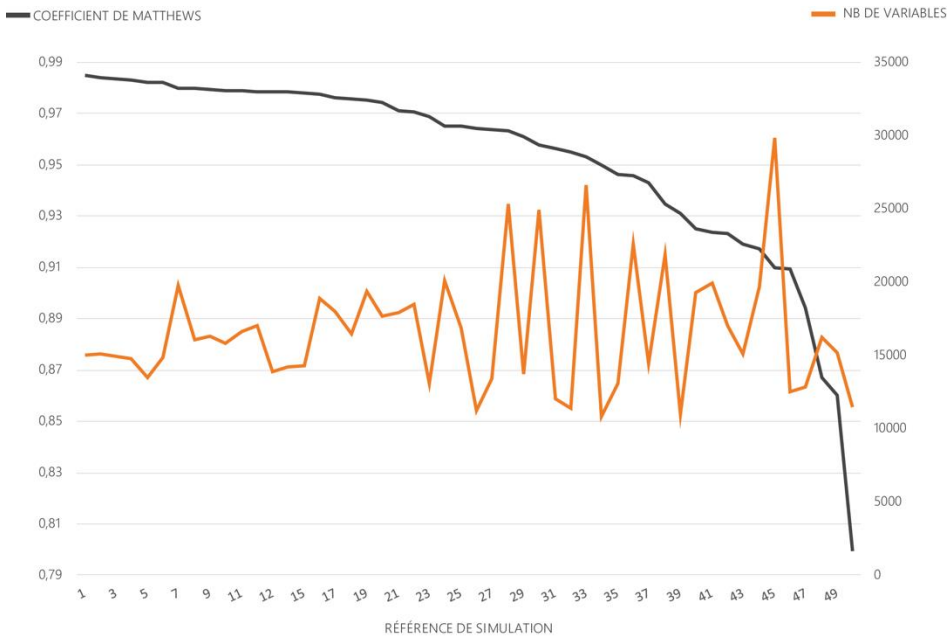


Figure 6. 10 : Évolution du coefficient de corrélation de Matthews en fonction du nombre de variables

Le processus d'apprentissage pourrait laisser penser que l'augmentation du nombre de règles, serait corrélée à celle de la performance du modèle, puisqu'elle engendre une description plus précise et spécifique du voisinage et des caractéristiques des cellules urbanisées. En d'autres termes, la maximisation du nombre de règles augmenterait les performances numériques du modèle. L'analyse contredit l'idée reçue, montrant que des jeux de règles moins importants permettent un meilleur apprentissage d'après le CCM (Figure 6. 9). Ce constat est également vérifié dans la section suivante, lors de

l'apprentissage sur un échantillon d'entraînement du jeu de données, puis lors du test du modèle produit sur des données neutres isolées au départ.

Les tests systématiques permettent également une analyse des valeurs de paramètres dont la combinaison engendre les modèles les plus performant. Le modèle le plus performant d'après un coefficient de Matthews présente un CCM égal à 0,984978. Sur les 50 combinaisons de paramètres testées, sept combinaisons présentent un CCM supérieur ou égal à 0,98. Il est donc intéressant de revenir sur la valeur des paramètres utilisés dans ces modèles. Le choix d'en faire la présentation se base sur la volonté d'étudier les différences de paramétrage qui mènent à un résultat compris dans cet intervalle. La Table 6. 2 explicite lesdits paramétrages et permet d'observer plus précisément la relation existante entre les différents paramètres.

Table 6. 2 : Paramétrages et performances de l'apprentissage sur les données d'occupation du sol uniquement

Simulation	Rayon	Poids	CCM	Nb de règles	Nb de variables	Vrai Positif	Faux Positifs	Vrai Négatif	Faux Négatifs
1	25	80	0,984978	574	15020	2915	86	529655	3
2	30	80	0,984143	547	15084	2914	90	529651	4
3	25	100	0,983365	566	14954	2902	82	529659	16
4	30	120	0,982918	516	14808	2889	71	529670	29
5	25	120	0,982251	566	13510	2889	75	529666	29
6	30	100	0,982012	547	14843	2899	87	529654	19
7	15	80	0,98	759	19765	2910	111	529630	8

Le poids le plus important donne les CCM les plus performants puisqu'une pondération de 80 donne le meilleur CCM (paramétrage 1). Il semble néanmoins qu'un poids moins important permettre l'obtention de jeux de règles moins fournis (paramétrages 4 et 5) tout en restant performant puisque le CCM est toujours au-dessus de 0,98. Les paramétrages limitant les faux positifs, maximisent les faux négatifs (paramétrages 4 et 5). S'ils sont plus précis sur la prédiction des transitions observées en limitant l'erreur à ce niveau, ils ne permettent pas de retrouver autant de transitions que les autres paramétrages. Les jeux de règles moins fournis tant en nombre de règles qu'en nombre de variables sont moins performants pour retrouver les transitions observées, mais présentent un caractère plus général, et synthétique, appréciable pour l'analyse thématique. Le nombre de règles semble, lui, diminuer avec l'augmentation du rayon du voisinage. La Figure 6. 11 confirme cette présomption. On constate en effet une forte diminution du nombre de règles entre 10 et 15 px, suivie d'une diminution moins importante mais continue jusqu'à 30px soit 3 000m.

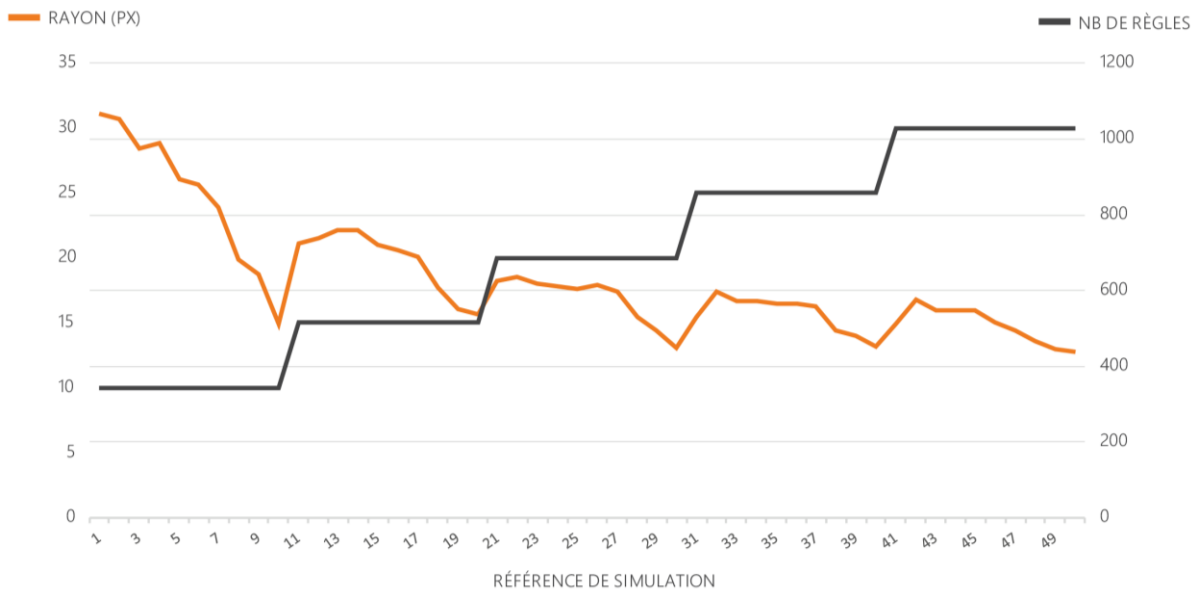
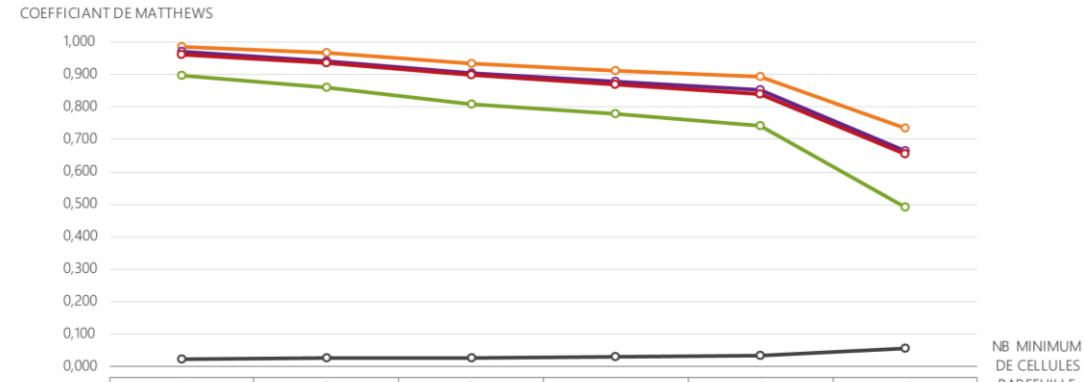


Figure 6. 11 : Évolution du nombre de règles en fonction du rayon du voisinage utilisé

Si ces simulations montrent un CCM extrêmement performant, la table 1 dévoile cependant que quel que soit le paramétrage, le nombre de règles reste important. Ainsi, le paramétrage 4, possédant le jeu le moins fourni présente plus de 500 règles. Ce constat est à mettre en parallèle avec l'incapacité de ces modèles à retrouver toutes les transitions urbaines : la colonne Faux Négatif, donc celle des transitions observées mais non simulées par le modèle, n'est jamais nulle. La question qui se pose alors, au vu de ces performances générales, est de savoir si l'ajout des variables de localisation géographique améliore la performance du modèle en termes de prédiction des transitions, tout en diminuant le nombre de règles contenu dans le jeu. En d'autres termes, il s'agit d'explorer si la localisation géographique présente une pertinence explicative et thématique. Pour ce faire, les tests sont réalisés sur l'ensemble des variables et présentés dans la sous-section suivante.

2.1. Sensibilité à la taille du voisinage

Le même dispositif est mobilisé dans le cadre du test de l'espace des paramètres sur le jeu de données incluant toutes les variables. L'objectif est d'évaluer si l'ajout des variables de localisation géographique améliore la performance du modèle. Comme dans la sous-section précédente, la performance des modèles produits est également évaluée à travers le coefficient de corrélation de Matthews mais aussi à travers les structures du jeu de règles (nombre de règles, faux négatifs ou erreurs, faux positifs ou surestimations).



	1	2	3	4	5	15
Maximum	0,986	0,969	0,935	0,912	0,893	0,735
Minimum	0,895	0,858	0,809	0,781	0,743	0,492
Médiane	0,970	0,942	0,904	0,877	0,852	0,665
Moyenne	0,962	0,936	0,901	0,871	0,842	0,656
Ecart Type	0,023	0,025	0,026	0,029	0,033	0,056

Figure 6. 12 : Influence du nombre minimum d'individus par feuille sur le Coefficient de Corrélation de Matthews (apprentissage sur totalité des variables)

L'apprentissage effectué sur le jeu de données incluant la totalité des variables confirme le constat effectué d'après l'apprentissage basé sur les données d'occupation du sol uniquement au regard de l'influence du nombre minimum d'individus par feuilles. La performance des modèles, exprimée par le CCM, décroît quand la limite du nombre d'individus par feuille augmente. Les deux figures (5 et 11) sont relativement similaires, la projection des deux jeux de courbes (AO : apprentissage occupation du sol seulement ; AT : apprentissage totalité des variables) permet néanmoins de mettre en lumière la contribution des variables de localisation géographiques. En effet, pour une même valeur du paramètre limitant le nombre d'instance par feuilles, le CCM est plus performant et souligne, d'ores et déjà, l'existence d'un apport des variables de localisation lors de l'apprentissage (Figure 6. 13).

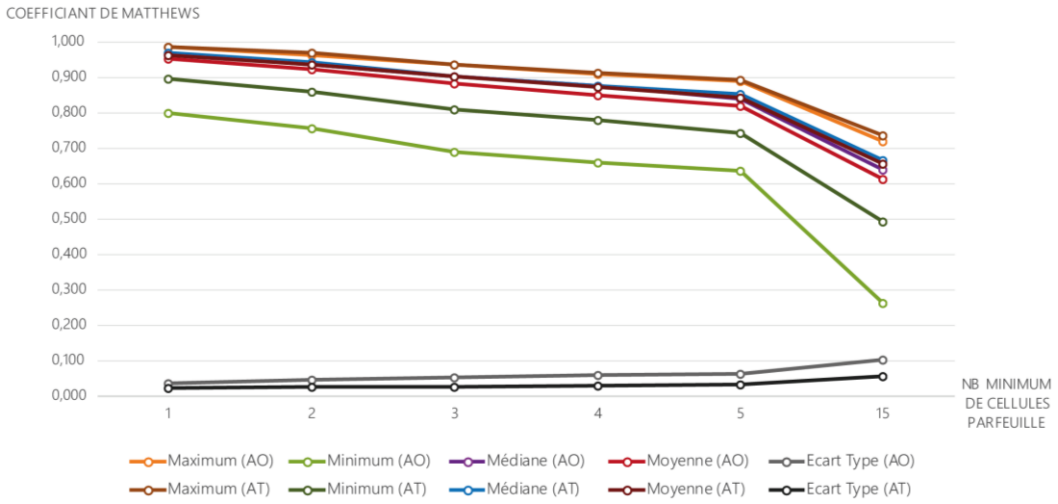


Figure 6. 13 : Comparaison des performances des modèles obtenus après apprentissage sur le jeu de données partiel ou total

Comme mentionné précédemment, l'évaluation de la performance d'un modèle n'est pas réalisée uniquement sur la base du CCM. Les valeurs des paramètres et leur contribution à la structure des modèles produits sont donc mises en parallèle (Table 6. 3).

Table 6. 3 : Paramétrages et performances de l'apprentissage sur la totalité des variables

Simulation	Rayon	Poids	CCM	Nb de règles	Nb de variables	Vrai Positif	Faux Positifs	Vrai Négatif	Faux Négatifs
1	25	120	0,985963	534	12527	2893	57	529684	25
2	30	80	0,985327	514	14004	2917	86	529655	1
3	25	80	0,985327	568	14852	2917	86	529655	1
4	25	100	0,985185	542	13564	2901	70	529671	17
5	20	80	0,984318	587	15684	2915	90	529651	3
6	30	100	0,984065	512	13508	2906	82	529659	12
7	30	120	0,983985	506	14051	2897	73	529668	21
8	20	100	0,983199	571	15115	2902	83	529658	16
9	15	80	0,982489	674	17142	2913	99	529642	5
10	30	140	0,980128	478	12268	2875	73	529668	43
11	30	40	0,980102	564	16485	2918	119	529622	0

Le modèle le plus performant d'après un coefficient de Matthews présente, cette fois, un CCM égal à 0,985963, soit légèrement plus performant que celui obtenu par apprentissage sur la seule occupation du sol (paramétrage 1). Sur les 50 combinaisons de paramètres testées, onze combinaisons présentent un CCM supérieur ou égal à 0,98, soit 4 de plus que lors de l'apprentissage sur l'occupation du sol uniquement. L'un des paramétrages permet de trouver la totalité des transitions observées moyennant une surestimation (faux positifs) et la mobilisation d'un nombre de variables légèrement plus important (paramétrage 11). L'ajout des variables de localisation géographique permet d'obtenir des modèles présentant moins de règles, et constitués d'un nombre inférieur de variables. Ce point est particulièrement intéressant au regard des objectifs de la modélisation réalisée sur notre zone d'étude.

Nous privilégierons ici la sélection d'un paramétrage limitant la présence de faux négatifs tout en maximisant le CCM. En effet, l'objectif principal est d'obtenir un modèle capable de retrouver la majorité des transitions observées. Pour autant, sa performance globale évaluée à travers le CCM reste importante en tant qu'indicateur générique. Dès lors, le second paramétrage pourrait être envisagé. Ce dernier permet de limiter le nombre de règles, en retrouvant la presque totalité des transitions observées, tout en gardant un CCM extrêmement performant. Néanmoins, l'évaluation de la performance des modèles au regard de leur paramétrage a été réalisé jusqu'ici sur des apprentissages mobilisant la totalité du jeu de données. Cette démarche a permis de mettre en lumière l'influence de chaque paramètre, ce sans introduire de biais lié à l'échantillonnage nécessaire à la composition d'un partitionnement apprentissage et test. Elle a permis d'établir l'apport explicatif des variables de localisation géographique. Pour autant, le chapitre précédent souligne également l'importance du test sur des données vierges pour évaluer la performance d'un modèle issue d'un apprentissage automatique. La prochaine sous-section s'attache donc à évaluer la sensibilité des modèles lors d'un échantillonnage aléatoire nécessaire au partitionnement apprentissage/test.

2.2. Sensibilité à la localisation géographique.

La séparation du jeu de données en deux parties, entrainement et test, présente un intérêt majeur eu égard au problème de sur-apprentissage lié à l'utilisation d'algorithmes d'apprentissage automatique (Chapitre 5). Une partie du jeu de données est donc isolée lors de la phase d'apprentissage, pour avoir l'opportunité de tester la performance du modèle sur des données vierges. La sélection des données isolées est réalisée aléatoirement tout en conservant une cohérence dans la représentativité des classes (cellules urbanisées ou non) dans chaque échantillon (entrainement et test). Cependant, la sélection aléatoire peut entrainer une variation du modèle issu d'un même paramétrage de l'arbre puisque chaque sélection aléatoire ne répartit pas exactement les mêmes individus dans les deux échantillons (entrainement et test). La conséquence directe peut être une variation des indicateurs de performance ou de la structure des modèles. Il est dès lors nécessaire d'évaluer leur sensibilité, au regard de l'échantillonnage effectué pour isoler la partie entrainement de la partie test du jeu.

Pour cela, la présente sous-section évalue la sensibilité de l'échantillonnage aléatoire à travers l'analyse des résultats fournis par 10 répétitions d'un apprentissage au paramétrage identique. La précédente sous-section ayant mis en lumière l'apport des variables de localisation géographique lors de l'apprentissage, la totalité des variables est donc mobilisée dans ce cas.

L'échantillonnage test représente 25% du jeu de données. Celui de l'entraînement contient donc les 75% restants. Ainsi le jeu de règles (modèle) est élaboré à partir de la partie entraiment, puis testé sur la partie test. L'ensemble des valeurs des indicateurs de performance du modèle sont alors obtenues pour la partie entrainement (MCC, VP, FP, VN, FN). Dans un second temps, le modèle est appliqué sur la partie test. Cette application permet d'obtenir ces mêmes indicateurs de performance mais issus de données non mobilisées lors de l'apprentissage, donc de données vierges n'ayant pas influencé l'apprentissage initial. La totalité du processus est répété à dix reprises, pour chaque paramétrage. L'enchaînement de ces deux étapes donne la possibilité d'évaluer à la fois la sensibilité du modèle vis-à-vis de l'échantillonnage ; mais aussi la performance du modèle sur des données vierges.

L'analyse de performance est orientée dans un premier temps sur les CCM entrainement et test. Puisqu'une valeur de chaque critère de performance est obtenue par répétition, la performance générale d'un paramétrage, qui permet de sélectionner le meilleur modèle, est estimée d'après leur médiane.

Avant d'évaluer plus précisément la sensibilité des critères de performance à l'échantillonnage aléatoire, une première analyse globale des valeurs médianes permet d'observer que la qualité du CCM test apparaît corrélée à celle du CCM entrainement (Figure 6. 14).

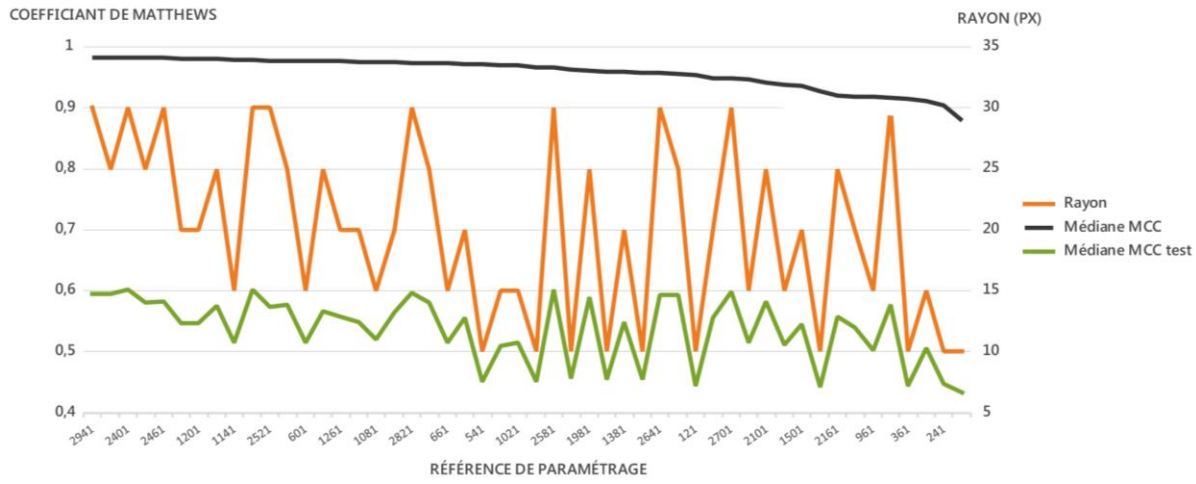


Figure 6. 14 : Relation entre les MCC entraînement et test lors de l'apprentissage 75%

Le CCM test peut néanmoins varier pour un même intervalle du CCM entrainement (Table 6. 4). Ainsi, pour un CCM entrainement de 0,98, le CCM test varie de 0,547 à 0,602. Cette variation semble, par ailleurs, être liée au rayon du voisinage pris en compte. Le CCM test de 0,6 est obtenu sur un rayon de 30 px, alors que le plus faible est, quant à lui, obtenu sur un rayon de 20 px. Cette variation s'observe également sur la Figure 6. 14.

Table 6. 4 : Variation du CCM test pour un CCM apprentissage de 0,98

Référence	Rayon	Pondération	Médiane CCM	Médiane CCM test	Nb de règles	Nb variables
2941	30	80	0,982578	0,595761	421	10652
2341	25	80	0,9823235	0,594148	444	10913
2401	30	100	0,982262	0,601527	439	12191
1801	25	100	0,98189	0,580513	448	10634
2461	30	120	0,981423	0,5831525	449	11909
1741	20	80	0,9806675	0,547727	476	12675
1201	20	100	0,980289	0,5472035	478	11678

Pour restreindre le nombre de paramétrages étudier en détails lors de l'évaluation de la sensibilité liée à l'échantillonnage, et facilité par la suite la sélection du meilleur modèle, seulement certains paramétrages sont étudiés en détails. Ils sont sélectionnés d'après les valeurs médianes de leurs critères de performances mais aussi en fonction de la structure des arbres produits. Les paramétrages analysés en détails présentent donc au moins l'une des caractéristiques suivantes :

- CCM entrainement médian performant (supérieur à 0,98) ;
- CCM test médian performant (supérieur à 0,6) ;
- Nombre de règles minimum ;
- Nombre de variables minimum ;
- Faux négatif proche ou égale à zéro.

Ces critères illustrent notre définition d'un modèle performant : un modèle composé d'un nombre minimum de règles et de variables, tout en étant performant d'après les MCC, avec la capacité de prédire le maximum de transitions observées. La sensibilité des modèles est donc étudiée pour chaque paramétrage répondant à l'un de ces critères. Les valeurs maximales, minimales, médianes et l'écart type sont calculées. La Table 6. 5 contient les valeurs des paramètres des modèles issus de cette sélection. Les paramètres, coefficient de

partition et minimum d'individus par feuille, restent fixés respectivement à 0,25 et 1, comme argumenté précédemment.

Table 6. 5 : Évaluation de la sensibilité du modèle suivant l'échantillonnage entraînement - test

ID	Rayon	Pondération	Mediane MCC	Médiane MCC test	Mediane Rules	Médiane Variables	Médiane VP	Médiane FP	Médiane VN	Médiane FN	Médiane VP test	Médiane FP test	Médiane VN test	Médiane FN test
2101	25	200	0,9409405	0,5825715	354,5	7800,5	1970	33,5	397270	216,5	398	235	132202,5	339
2401	30	100	0,982262	0,601527	435,5	11309	2160	64,5	397251,5	13	451	282,5	132127,5	298,5
2581	30	160	0,965759	0,6018415	383	8983,5	2077,5	37	397266,5	102	416,5	239,5	132197	312,5
2701	30	200	0,948424	0,597666	352	8035,5	2005	36	397265	186	402	226	132220,5	323,5
2881	30	60	0,978618	0,6015805	443	12239,5	2202,5	97	397201	0	435	284,5	132152	289
2941	30	80	0,982578	0,595761	440	11530	2183,5	74,5	397236	2	439,5	292,5	132131,5	294,5

ID	Rayon	Pondération	Ecart type MCC	Ecart type MCC test	Ecart type Rules	Ecart type Variables	Ecart type VP	Ecart type FP	Ecart type VN	Ecart type FN	Ecart type VP test	Ecart type FP test	Ecart type VN test	Ecart type FN test
2101	25	200	0,007966468	0,01729497	10,54356044	308,1457267	40,43320967	6,828046736	28,07450405	31,90541577	15,21001132	15,64360004	19,04643449	25,695006
2401	30	100	0,002544403	0,015339045	14,57394936	790,2471414	32,20144924	9,362454569	37,6657817	3,984692934	23,87839563	14,74260643	35,0580471	18,2455474
2581	30	160	0,005689733	0,028312772	12,38682095	539,2460683	25,93067681	9,016035098	21,30310358	19,73716179	30,44284262	14,09846327	19,10817161	17,25913607
2701	30	200	0,00439664	0,013740911	17,86492529	690,9779221	43,60122322	7,133644853	38,42684247	18,52325865	12,49044079	24,49512423	32,54893751	27,89842688
2881	30	60	0,002662353	0,013861185	18,70709907	624,7770802	19,00409313	12,32882801	22,37160899	0	11,71134303	14,7576722	15,45926547	22,43410103
2941	30	80	0,001279818	0,011164935	14,66704545	743,0407571	22,92233263	5,35931173	21,39080799	0,994428926	11,34117376	17,58187198	24,11292876	20,90188721

ID	Rayon	Pondération	Max MCC	Max MCC test	Max Rules	Max Variables	Max VP	Max FP	Max VN	Max FN	Max VP test	Max FP test	Max VN test	Max FN test
2101	25	200	0,952557	0,602712	366	8247	2028	44	397335	255	414	251	132213	376
2401	30	100	0,988177	0,622954	449	12326	2205	74	397311	20	477	309	132192	337
2581	30	160	0,974249	0,634351	405	9957	2140	63	397304	136	460	263	132236	358
2701	30	200	0,952493	0,611046	369	9029	2053	43	397346	217	416	252	132260	379
2881	30	60	0,982986	0,614557	489	13266	2221	111	397234	0	450	316	132177	317
2941	30	80	0,985402	0,61296	465	13014	2209	82	397272	4	460	326	132183	331

ID	Rayon	Pondération	Min MCC	Min MCC test	Min rules	Min Variables	Min VP	Min FP	Min VN	Min FN	Min VP test	Min FP test	Min VN test	Min FN test
2101	25	200	0,932214	0,545331	335	7245	1889	26	397243	169	361	198	132155	302
2401	30	100	0,979256	0,575418	402	9690	2105	43	397209	8	404	259	132082	268
2581	30	160	0,955324	0,535369	358	8076	2056	33	397230	78	356	223	132175	304
2701	30	200	0,940355	0,56802	318	6839	1936	20	397230	163	375	171	132143	284
2881	30	60	0,975659	0,57193	428	11541	2166	77	397173	0	414	272	132137	247
2941	30	80	0,980787	0,578824	421	10652	2138	63	397207	1	424	274	132086	271

Les performances du modèle présentent une certaine sensibilité suivant l'échantillonnage séparant les parties entraînement et test de l'apprentissage (Table 6. 5). La variation du nombre de règles ou du nombre de variables, illustrée par leur écart-type, est faible au regard des valeurs médianes de ces mêmes critères. Par exemple, la variation de l'écart-type concernant le nombre de règles créées pour un même paramétrage est comprise entre 10 et 20, pour des jeux contenants entre 318 et 489 règles. La sensibilité générale des VP, FP, VN et FN entraînements est faible en comparaison avec celle des tests. Les variations des VP et VN entraînements peuvent être considérées comme négligeables puisqu'elles représentent entre 19 et 43 cellules pour un total minimum d'environ 1 900 cellules VP et de 397 000 cellules VN. La variation est plus importante mais reste faible pour le FP et FN entraînement. S'agissant du FP entraînement, elle peut atteindre 12 cellules mais sur un total de 111 soit 10%, ou encore, pour le FN entraînement, 31 cellules sur un total de 255, soit 12%. La sensibilité pour la matrice test se situe également autour de 10% maximum en termes de nombres de cellules. La sensibilité issue de l'échantillonnage aléatoire est moins importante dans le cas du CCM entraînement que dans celui du CCM test, dont l'écart-type atteint par exemple 0,028 (paramétrage 2581). Si ce paramétrage (2 581) présente la plus grande variation d'écart-type, c'est également celui dont le CCM test, de 0,63, est le plus performant lors du test sur les données n'ayant pas servi en phase d'apprentissage. Ce point particulier

retient notre attention puisque la capacité de ce modèle à performer sur des données vierges est particulièrement intéressante. Néanmoins, avant de procéder à tout choix, il reste important d'analyser sa performance vis-à-vis des autres critères suivant chaque répétition de l'échantillonnage (Table 6. 6).

Table 6. 6 : Les 10 répétitions du modèle de voisinage 30 et de pondération 160

ID	Rayon	Pondération	Répétition	CCM	CCM test	Nb règles	Nb de variables	Vrai Positif	Faux Positif	Vrai Négatif	Faux Négatif	Vrai Positif test	Faux Positif test	Vrai Négatif test	Faux Négatif test
2599	30	160	8	0,96958	0,634351	387	8966	2061	36	397304	93	460	224	132177	304
2600	30	160	9	0,974249	0,603609	384	8076	2107	33	397276	78	427	251	132181	306
2582	30	160	10	0,966778	0,615876	372	8921	2073	45	397279	97	442	242	132175	306
2597	30	160	6	0,971751	0,583684	379	9001	2140	37	397230	87	384	238	132236	307
2585	30	160	4	0,955324	0,608766	358	8559	2059	63	397243	129	423	234	132201	307
2584	30	160	3	0,96875	0,608241	382	9336	2082	37	397278	97	421	223	132203	318
2583	30	160	2	0,962349	0,600074	381	8420	2056	34	397277	127	412	225	132205	323
2586	30	160	5	0,959816	0,580366	405	9239	2064	37	397257	136	393	241	132206	325
2581	30	160	1	0,96474	0,565194	390	9957	2085	45	397257	107	391	263	132176	335
2598	30	160	7	0,964375	0,535369	391	9376	2085	35	397255	119	356	258	132193	358
			Ecart type	0,0057	0,0283	12	539	26	9	21	20	30	14	19	17

Comme mentionné précédemment, le CCM test de ce paramétrage est le meilleur obtenu sur la totalité des tests de paramétrages effectués avec un apprentissage sur 75% du jeu de données. Il présente néanmoins la plus grande sensibilité à l'échantillonnage, et souligne donc l'importance de la multiplication des tests lors d'une démarche incluant une sélection aléatoire. Pour autant, cette variation reste faible en valeur absolue (0,1).

Le second point d'intérêt majeur est le nombre de règles composant le modèle. Ce nombre doit être associé à celui des variables utilisées dans une optique d'équilibre. En effet, en fonction de la proportion entre le nombre de règles et le nombre de variables, l'interprétation du résultat peut être plus ou moins délicate. Un jeu de règles comprenant peu de règles mais mobilisant un plus grand nombre de variables créera des règles plus longues. Celles-ci seront potentiellement difficiles à interpréter, par opposition à un jeu contenant plus de règles courtes et précises. Le paramétrage en question présente 387 règles. Nous pouvons d'ores et déjà souligner une diminution importante du nombre de règles par rapport à celui de l'apprentissage sur la seule occupation du sol : pour un CCM de 0,98, il n'était pas inférieur à 516 règles. Quand les variables de localisation géographique sont ajoutées, le nombre baisse à 478 règles. L'apprentissage à 75% permet donc de réduire encore ce nombre à performance égale du CCM : 421 règles. Nous pouvons donc observer une amélioration des performances pour le critère du nombre de règles.

La diminution du nombre de règles est à considérer parallèlement au nombre de variables utilisées. Le même constat est ici réalisé s'agissant d'une diminution du nombre de variables. D'environ 13 500 variables minimum en apprentissage occupation du sol, ce nombre chute à près de 12 200 en apprentissage sur la totalité des variables. Il est d'approximativement 10 500 sur la partition à 75%, en considérant un CCM de 0,98. En conséquence, l'interprétation des règles sera donc plus simple que précédemment. Le modèle analysé ici mobilise un peu moins de 9 000 variables. Il faut néanmoins souligner que la restriction du nombre de règles ou celui du nombre de variables, tout en permettant une interprétation plus simple, s'effectue généralement aux dépens d'une augmentation de l'erreur (Faux Négatifs) ou de la surestimation (Faux Positifs).

Le critère de la restriction maximale des Faux Négatifs, que ce soit dans l'évaluation de l'entraînement ou du test, est revu de façon plus souple pour préférer un jeu de règles moins fourni et un nombre moins important de variables mobilisées. Le paramétrage 2 881 qui permet de retrouver toutes les transitions observées dans l'échantillon apprentissage, ne permet d'améliorer les FN que d'une cinquantaine de cellules pour une augmentation du nombre de règles équivalente (Table 6. 2). Il n'apparaît donc pas pertinent.

Dans cette perspective, et d'après l'utilisation de la Table 6. 6, il convient d'évoquer le paramétrage 2 599. Sa performance globale représentée par les CCM est maximisée, parallèlement à celle du potentiel en termes d'analyse thématique puisque le jeu de règles et le nombre de variables mobilisées sont restreints. Le modèle utilisé pour l'étude du développement urbain sur la zone transfrontalière de Strasbourg-Kehl est donc celui qui est issu de la 8ème répétition du paramétrage considérant 30px de voisinage et 160 de pondération. Sa matrice de confusion permet de synthétiser les résultats obtenus en phase entrainement et test (Table 6. 7).

Table 6. 7 : Matrice de confusion du paramétrage 2 599

		Simulé	
		0	1
Observé	0	529481	260
	1	397	2521

Cette section a permis de sélectionner le meilleur modèle au regard des critères de performances établis d'après les objectifs du travail de recherche (maximiser le CCM test tout en minimisant le nombre de règles ou de variables du modèle). L'étape de calibrage du paramétrage de l'arbre de décision a permis d'évaluer la performance du modèle dans différents cas de figure (apprentissage d'après l'occupation du sol exclusivement, sur la totalité des variables et suivant une partition entraînement et test). Les performances de l'apprentissage par arbre de décision ont ainsi été appréciées (CCM entraînement supérieur à 0,96 ; CCM test supérieur à 0,63), en conservant la notion de sensibilité de l'algorithme à l'échantillonnage de partitionnement entraînement – test, du jeu de données. La réponse au problème du sur-apprentissage a été abordée lors de l'étude de l'influence du rayon de voisinage et de la pondération. Trois éléments doivent ainsi être mis en exergue :

- Le rayon maximum ne permet pas automatiquement d'obtenir des modèles plus performants.
- La pondération maximale visant l'équilibre parfait (50/50) des classes à prédire (transition urbaine) et sans changement entre 1990 et 2006 n'est pas non plus optimale. Ainsi, conférer à la classe à prédire une importance exacerbée, n'est pas nécessaire à l'obtention du meilleur modèle.
- La réalisation d'un apprentissage sur une partie seulement des données contenues dans le jeu total permet d'évaluer plus objectivement la performance du modèle sur des données vierges. Le modèle créé permet de retrouver plus de 60% des transitions sur des données vierges, ce qui est considéré comme performant dans la modélisation des changements d'occupation du sol (Hema Laboratory, 2006).

Pour conclure, la phase de calibrage de l'arbre de décision menant à la validation quantitative du modèle sélectionné a permis d'établir l'apport des variables de localisation géographique, et notamment de localisation d'une cellule dans l'un ou l'autre des pays. Ce point peut être considéré comme un premier résultat. Le questionnement sous-jacent reste néanmoins la part d'influence de la localisation au regard de la frontière vis-à-vis des autres variables appartenant à la liste des variables de localisation géographique. Ce point est abordé dans le chapitre suivant (Chapitre 7), à l'issue de la section suivante, qui s'intéresse à l'aspect spatial de la validation.

3. SENSIBILITÉ À LA PARTITION DU JEU DE DONNÉES

La validation spatiale a pour objectif d'établir la cohérence du développement urbain simulé, ce au regard des données observées, ainsi que d'évaluer la performance du modèle d'un point de vue spatial. Cette étape permet certes de valider le modèle mais également de localiser ses manquements. Il s'agit donc d'analyser l'erreur d'un point de vue spatial en complément de l'aspect quantitatif développé dans la section précédente. Les questions suivantes peuvent ainsi être adressées :

- *Le modèle permet-il de reproduire un processus cohérent avec celui observé dans la réalité ?*
- *Permet-il de capter différentes formes de développement urbain ?*
- *Les erreurs ou bonnes prédictions sont-elles spécifiques à certains espaces ?*
- *Résident-t-elles dans l'expression d'une forme de développement ou sur une zone particulière ?*

Dans cette section, ces questions sont illustrées à travers le jeu de règles sélectionné précédemment pour simuler le développement urbain entre 1990 et 2006. Les résultats sont mis en relief avec les données observées de 2006 issues directement de Corine Land Cover.

3.1. Contraintes : résolution, étendue, emprise spatiale du phénomène.

La validation spatiale est, dans le cadre de ce travail, réalisée à travers la visualisation de l'urbanisation observée et de celle simulée par le modèle. La résolution de la grille (100 m^2) a été choisie pour maximiser la précision de l'étude du processus d'urbanisation. Elle correspond également à la résolution la plus fine dans les modalités d'export Corine Land Cover. Cependant, quand mise en relation avec l'étendue de la zone d'étude et le caractère exceptionnel du phénomène, l'observation des changements peut s'avérer délicate (Figure 6. 15). La résolution des cellules permet difficilement d'établir distinctement la différence entre les trois classes représentées pour 2006 : urbain observé (Faux Négatifs), urbain simulé (Vrai Positif, Faux Positif). Les Faux Positifs ne sont pas présents en grande quantité, et sont donc difficilement identifiables visuellement. Si les Vrais Positifs et les Faux Négatifs le sont plus facilement, l'appréciation de leur proportion locale est également délicate. À titre d'exemple, la carte ci-dessous ne permet pas d'établir si la performance du modèle est homogène ou non, et donc d'en déduire si elle dépend de la localisation étatique, ou du type de processus de développement (mitage, étalement urbain, etc.).

Il est alors nécessaire de mettre en œuvre d'autres approches, pour permettre d'apprécier plus aisément la performance du modèle en fonction de la localisation dans l'espace de la zone d'étude. Cette question est abordée à travers deux approches faisant l'objet des sous-sections suivantes : l'approche locale, l'approche agrégée.

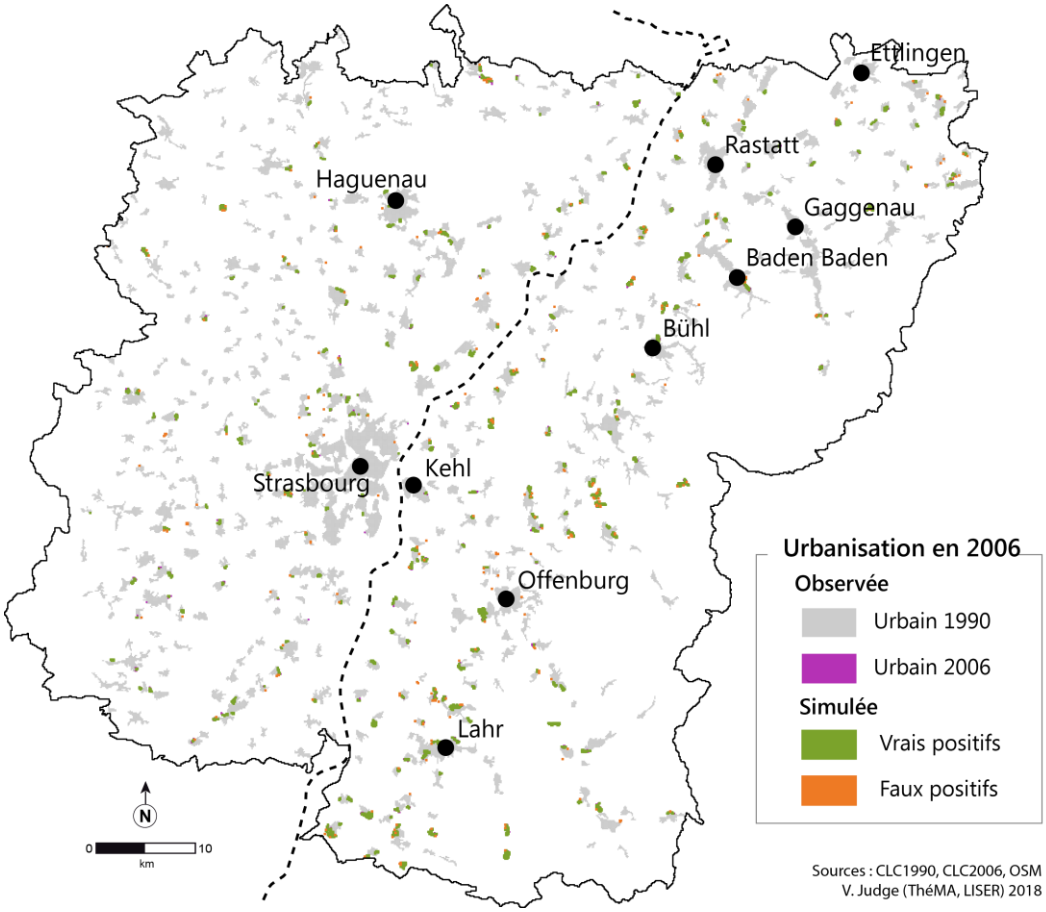


Figure 6. 15 : Données observées et simulées du développement urbain entre 1990 et 2006

3.2. Validation : Approche locale

Une observation plus locale peut être mobilisée ponctuellement (Figure 6. 16). Le développement urbain simulé est spatialement cohérent avec le processus observé, notamment au vu de la quantité importante de transitions urbaines correctement simulées (vrais positifs). Ces résultats sont relatés par la matrice de confusion issue de l'analyse de performance numérique du modèle (Table 6. 7).

Le développement simulé est caractéristique des processus d'étalement urbain, de l'urbanisation des dents creuses, ou encore du mitage dans une proportion plus faible. Un exemple de mitage est mis en relief par la Figure 16 (Zoom 3). Une certaine cohérence est également observée dans la simulation des Faux Positifs : s'ils ne représentent pas de transitions urbaines observées, leurs localisations reflètent les caractéristiques d'un phénomène cohérent de potentiel étalement urbain (Zoom 1) ou d'une urbanisation des dents creuses, par exemple (Zoom 2). Le développement urbain simulé, mais non observé est donc cohérent avec un développement réel. Il apparaît parfois très proche d'un développement observé mais non simulé (Zoom 2). Ce point nous pousse à nous interroger sur l'évaluation de cette erreur, dans la partie résultats (Chapitre 7).

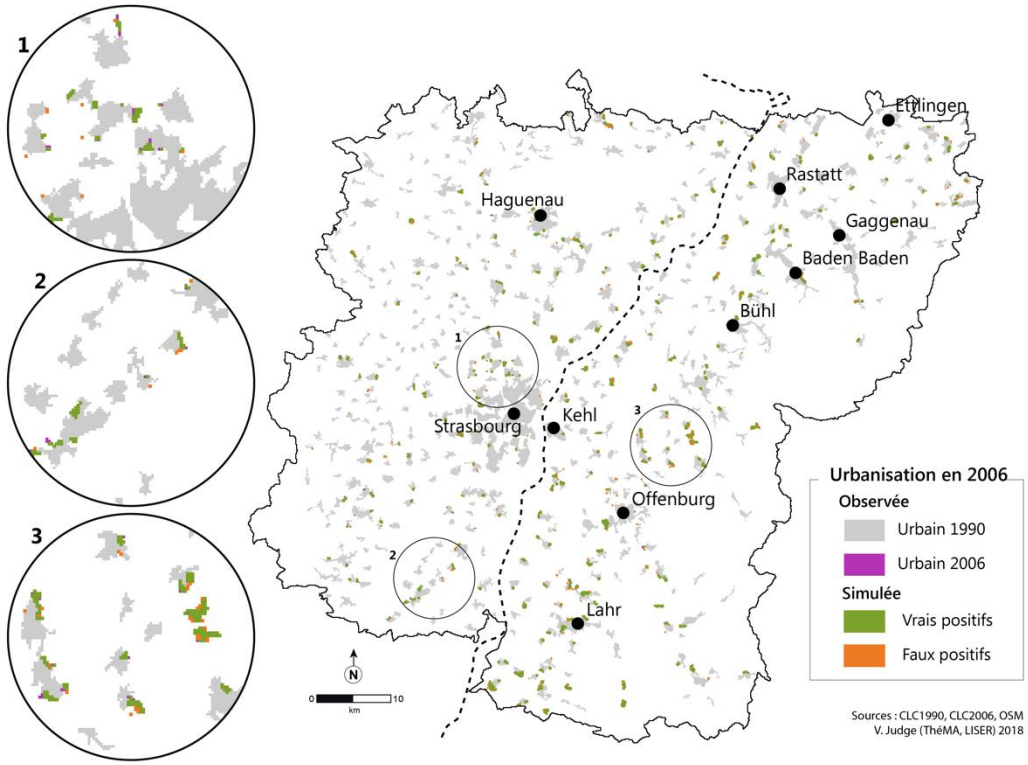


Figure 6. 16 : Une observation plus localisée

L'approche locale n'est cependant pas optimale pour une appréciation plus globale de la performance du modèle à l'échelle de la zone d'étude. La seconde approche de visualisation est développée pour rendre compte de l'efficacité plus générale du modèle à l'échelle de cette zone.

3.3. Validation à l'échelle agrégée

La méthode visant à agréger les données a été mise en place pour remplir deux objectifs :

- Améliorer la visualisation des résultats à l'échelle de la zone d'étude ;
- Ajouter de la flexibilité dans la validation à travers la visualisation.

Comme mentionné précédemment le choix d'une validation stricte (binaire juste/faux) est issu d'une démarche orientée vers une méthodologie simple et transposable, facilitant la communication autour de l'outil. D'autre part, cette approche plus restrictive évalue la performance de modélisation de la manière la plus stricte. Inclure de la flexibilité ne pourra pas engendrer d'influence négative sur la performance du modèle, bien au contraire. Pour autant, même en ayant la perspective stricte et simple de la validation binaire, nous pouvons nous interroger sur la cohérence de cette démarche dans l'analyse d'un phénomène spatial tel que le développement urbain à l'échelle d'une zone aussi large que celle étudiée sur la frontière franco-allemande. En effet, comme le montrait la figure précédente (Figure 6. 16), le développement simulé à tort au regard d'une appréciation binaire de l'erreur, se trouve localisé parfois dans un périmètre extrêmement proche d'une cellule urbanisée dans les données observées. Dans ce contexte, il est opportun de s'interroger sur la portée de l'erreur, à savoir si une prédiction de transition à 100, 200 ou encore à 500 mètres près doit vraiment être considérée comme totalement fausse dans une appréciation plus globale de la

performance d'un modèle de développement urbain, notamment lors d'un travail à une échelle régionale.

Ainsi, la recherche s'est interrogée sur les méthodes de comparaison pertinentes pour évaluer la donnée simulée au regard de la donnée observée. Ces méthodes de comparaison sont notamment étudiées et développées par le RIKS dans son « map comparison kit » de Metronamica (Visser & de Nijs, 2006). La ressemblance entre les données est évaluée à travers des indicateurs comme le *fuzzy inference system*, *fuzzy set*, *fuzziness* of location. Sans en exposer tous les détails, ces méthodes permettent d'intégrer de la flexibilité dans l'évaluation de la similitude entre deux cartes, en recourant à la logique floue pour d'évaluer, par exemple, la localisation au vu des transitions observées, ou intégrer un degré de flexibilité sur le type d'occupation du sol prédis et observés, etc. Nous préférons ici recourir à une méthode d'agrégation plus basique dont le mécanisme est aussi simple à expliquer que celui consistant à s'appuyer sur une validation cellule à cellule. Le nombre de transitions simulées est agrégé sur une grille composée de cellules plus vastes : 2 km x 2 km et 4 km x 4 km. L'agrégation permet d'inclure de la flexibilité dans l'évaluation de la performance du modèle en estimant correctes les prédictions dans un certain périmètre. Ainsi, des prédictions fausses dans une évaluation cellule à cellule deviennent acceptables. Ce phénomène peut être illustré par la figure 6.15, dans le cadre du périmètre rouge. L'étalement urbain a été simulé du côté opposé à celui observé. Une agrégation des transitions observées à l'échelle du zoom 2 de la Figure 6. 16 considèrera, par exemple, que la précision spatiale est acceptable en dépit du fait qu'elle ne soit pas exacte.

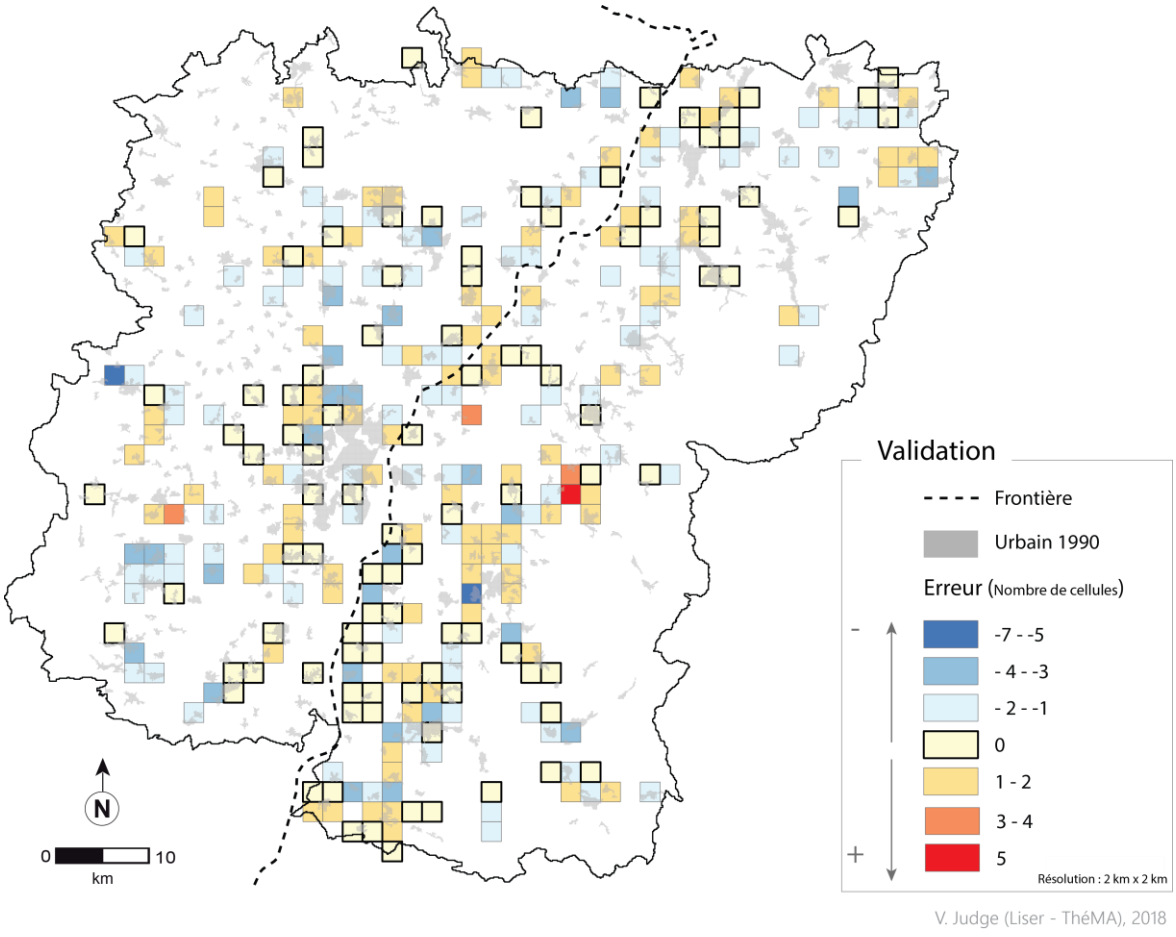


Figure 6. 17 : Transitions agrégées sur une grille de résolution 2 km x 2 km

La représentation utilisée sur les figures 6.16 et 6.17 permet par ailleurs de faire la distinction entre les zones agrégées dont la prédiction est correcte en termes de changements et celles dont la prédiction est correcte en termes de non-changements. Ce travail est concentré sur la prédiction du changement. Ces figures font donc apparaître uniquement les zones de changements. La surface blanche de la zone d'étude, non spécifié suivant la légende présente des zones donc le non-changement est correctement estimé. Ainsi, le modèle est également juste. Pour autant, il ne paraît pas pertinent de faire apparaître spécifiquement les zones où l'absence de transitions est correctement estimée. Au contraire, il est nécessaire de souligner la performance du modèle concernant la prédiction du changement. Le reste de la classification permet d'apprécier visuellement la performance du modèle, à savoir la surestimation ou sous-estimation du nombre de transitions urbaines simulées au regard de celles observées.

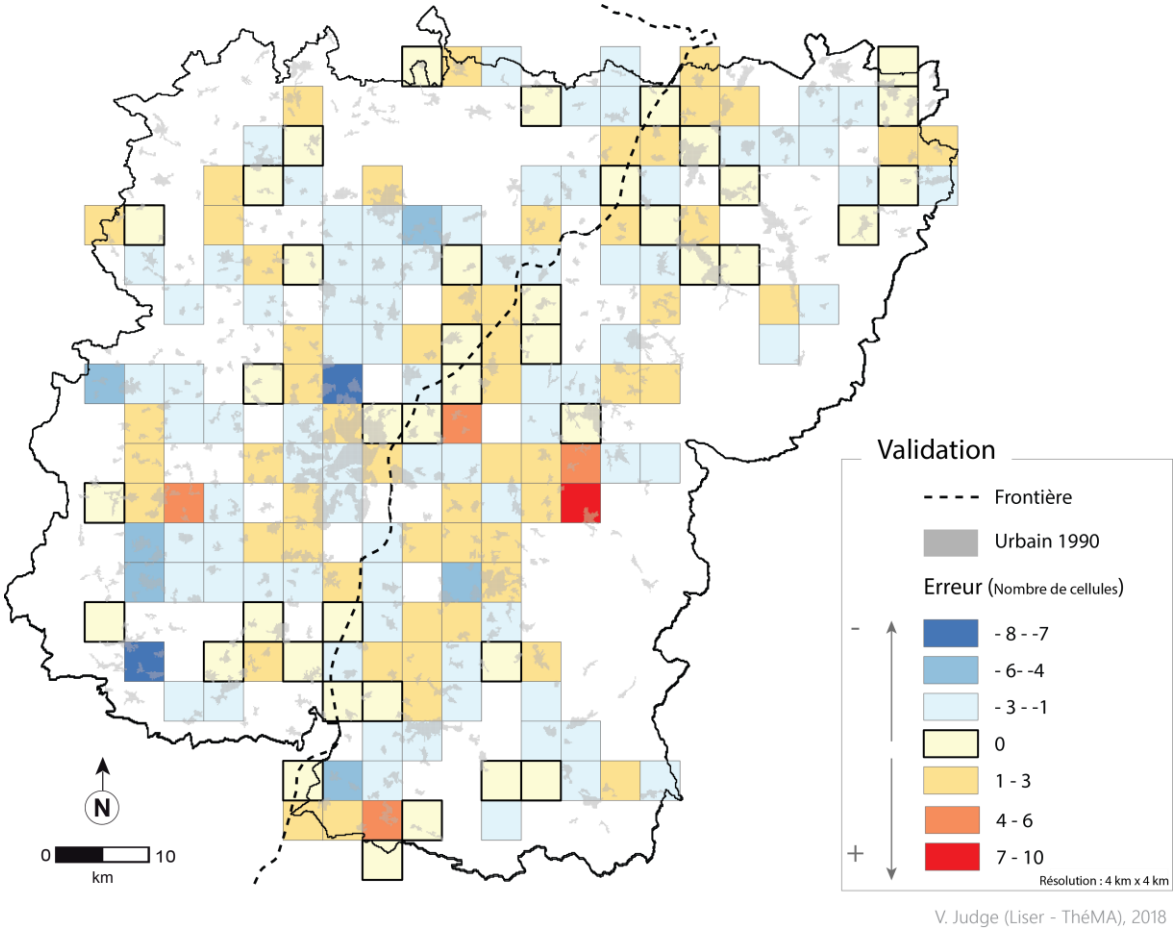


Figure 6. 18 : Transitions agrégées sur une grille de résolution 4 km x 4 km

À l'échelle de leurs résolutions respectives, l'agrégation facilite une d'observation simple des résultats du modèle concernant la spatialisation du développement urbain. Cette observation simplifiée est rendue possible par la flexibilité introduite par la présente méthode. Elle est d'autant plus enrichissante, qu'elle apporte des éléments de réponse, exploités ultérieurement, quant à la thématique frontière présente dans notre travail. Dans le cadre des exemples ci-dessus, il semblerait en effet, que le modèle soit plus performant sur la zone française, la catégorie extrême n'étant présente que sur la partie allemande. La Figure 6. 18 souligne également une différence de performance entre la France et l'Allemagne dans les espaces qui présentent une prédiction correcte du changement. En France 27 cellules

représentent les zones où les transitions urbaines ont été correctement simulées, contre 15 cellules en Allemagne. La différence est également notable à l'échelle supérieure (Figure 6. 18) même si moins marquées (9 pour la France contre 6 pour l'Allemagne) puisqu'elle tend à être lissée avec l'augmentation de la taille du carroyage.

L'agrégation des résultats à l'échelle des grilles de 2 km x 2 km et de 4 km x 4 km permet une estimation plus flexible des performances du modèle en incluant une tolérance dans la précision de la prédiction d'un point de vue spatial. La flexibilité peut également être introduite dans la quantité de prédictions correctes sur une zone. La Figure 6. 19 introduit un seuil de tolérance dans la quantité de cellules, qui s'ajoute au seuil de tolérance spatial induit par l'agrégation sur la grille. Ainsi, une erreur inférieure ou égale à deux cellules est considérée comme acceptable. La visualisation des performances du modèle en est encore légèrement modifiée (Figure 6. 19).

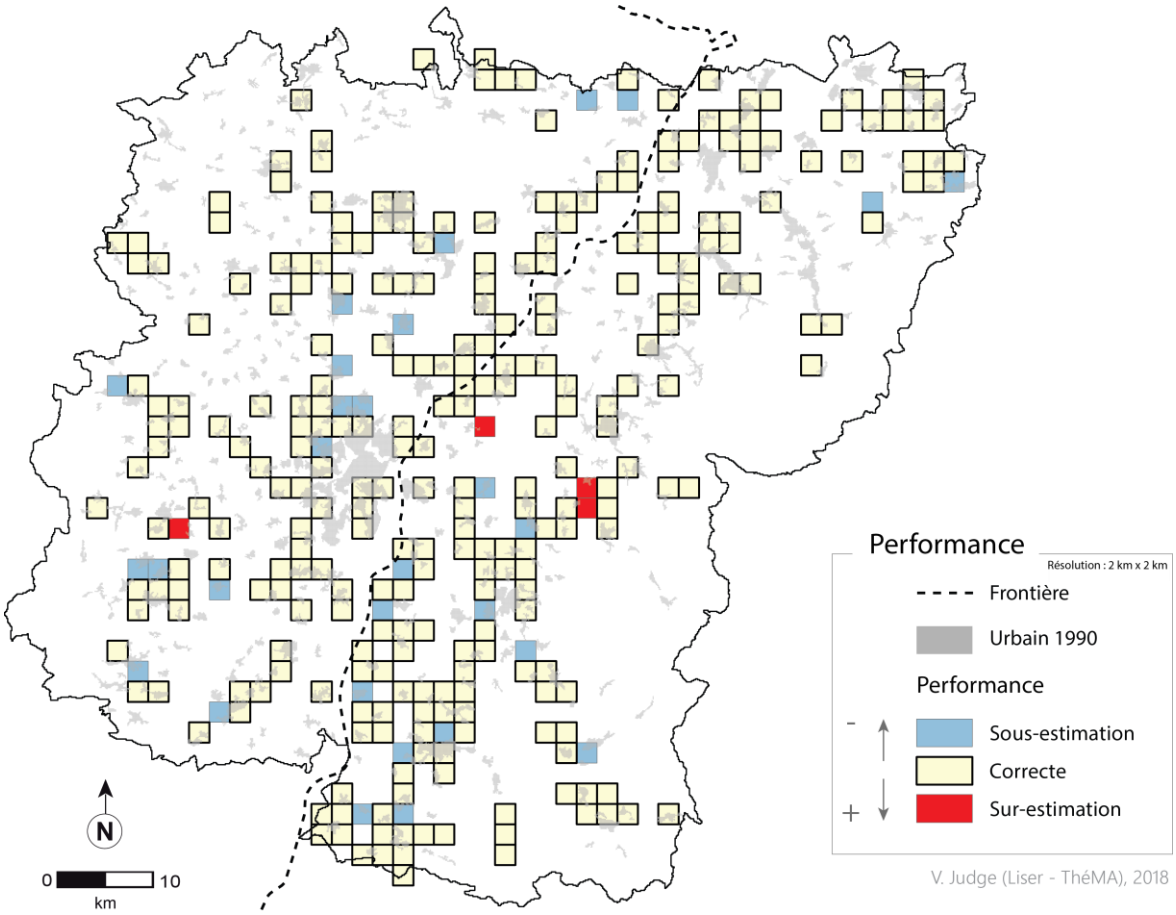


Figure 6. 19 : Évaluation spatiale des performances du modèles. Introduction d'un seuil de tolérance plus flexible

Les résultats ne seront pas discutés plus avant dans cette section, puisque son seul objectif, méthodologique, est de montrer les procédés mobilisés pour l'évaluation de la performance, et donc la validation spatiale du modèle. Les résultats issus de la simulation du modèle sont discutés dans la partie suivante.

Conclusion du chapitre 6

La pluralité de méthodes de validation des modèles de changements d'occupation du sol n'a d'égale que celle des modèles développés à cet effet. Si cette étape présente un enjeu majeur notamment en permettant la diffusion de résultats sur la base d'analyses de performances, elle n'est actuellement pas retrouvée dans tous les travaux concernant les modèles de changement d'occupation du sol (van Vliet *et al.*, 2016). Une explication réside éventuellement dans les propos de Peagelow, qui, prenant du recul sur cette étape, déclare qu'elle « ne peut en aucun cas être complète » (Paegelow *et al.*, 2018), tout comme un modèle ne peut être totalement juste. En tout état de cause, cette étape, en s'appuyant sur la vision de G. Box, même incomplète, s'avère utile (Box & Draper, 1987). La partie 2 de notre travail illustre par ailleurs ces propos puisque, de la validation du calibrage des paramètres de l'arbre de décision, nous pouvons d'ores et déjà établir l'importance de la composante de localisation géographique dans la description du voisinage et des caractéristiques des cellules ayant été urbanisées entre 1990 et 2006. Cette étape permet également de caractériser certaines limites du modèle notamment dans la performance de l'apprentissage ou dans le nombre de règles mobilisées. Le chapitre 7 permet d'analyser plus en détails le jeu de règles issu de l'apprentissage et d'en définir la contribution à une analyse thématique.

Conclusion de la deuxième partie

La deuxième partie a permis de développer le protocole méthodologique mis en place pour simuler le développement urbain sur la zone frontalière franco-allemande. Sa simulation résulte du couplage d'un automate cellulaire et d'un processus d'apprentissage par arbre de décision. Notre protocole propose donc d'associer deux outils généralement privilégiés pour leur caractère simple et transparent : (i) l'arbre de décision qui permet de créer des règles interprétables d'un point de vue thématique, (ii) l'automate cellulaire qui permet de les mettre en œuvre.

En nous appuyant sur une analyse de sensibilité, nous avons paramétré un algorithme permettant de construire l'arbre de décision et démontré qu'il est pertinent de recourir aux variables de localisation pour augmenter la performance du modèle final. Ce point est particulièrement intéressant puisqu'il met notamment en lumière l'influence de l'Indice de Pays, qui localise le processus de développement urbain dans l'un ou l'autre des Pays de la zone d'étude. L'analyse de sensibilité a également révélé que les modèles composés d'un nombre plus important de règles, ou s'appuyant sur une quantité plus importante de variables, ne sont pas toujours les modèles les plus performants. Ce point apporte des réponses positives à la question du sur-apprentissage des algorithmes basés sur les arbres de décisions. De même pour l'évaluation de l'incidence de la pondération nécessaire à l'apprentissage d'un phénomène de changement présentant un caractère exceptionnel. L'analyse montre que ce n'est pas non plus la pondération la plus importante qui permet d'obtenir les modèles les plus performants.

Finalement, le modèle retenu dans la suite de ce travail résulte de choix méthodologiques, liés aux objectifs de modélisation et au contexte de l'étude. Il a ainsi été favorisé pour optimiser la performance de prédiction tout en minimisant le nombre de règles. Ce point offre la possibilité d'interpréter les règles issues de l'apprentissage, et fera l'objet de la troisième partie.

Troisième partie

Interpréter le développement urbain en zone frontalière

Introduction de la troisième partie

Dans la deuxième partie nous avons présenté les différents outils formant la base de notre méthode pour évaluer l'influence culturelle sur la caractérisation du développement urbain. Le choix du modèle s'est porté sur l'automate cellulaire (AC) du fait de sa simplicité de prise en main pour un utilisateur final et sa rapidité de calcul. L'AC a été associé à un arbre de décision (AD) qui permet d'automatiser l'élaboration des règles de transitions et présente de manière transparente leur élaboration. Dès lors, le couplage des deux outils permet la création automatisée d'un jeu de règles transparent, qui permet de modéliser le développement urbain sur la base des données libres et harmonisées retenue dans notre travail.

Le jeu de règle issu de cet apprentissage automatique par arbre de décision sert de support à l'évaluation du caractère favorable du voisinage ou de la localisation d'une cellule, pour l'urbanisation. Si une première évaluation de l'influence des différents types de variables (occupation du sol ou localisation) a déjà été réalisée à travers les tests de sensibilités effectués pour paramétrer l'algorithme, il convient dans la troisième partie de ce travail doctoral, de préciser par une mesure plus fine, l'influence de chaque variable sélectionnée par l'arbre. Nous présentons dans un premier temps une analyse du jeu de règle qui met en lumière l'influence de la localisation géographique, sur la performance d'apprentissage. La mesure de l'influence de chaque variable met en avant un rôle important de la localisation dans l'un ou l'autre des pays frontaliers. Pour autant, une partie du jeu de règles s'affranchit de son influence. Il s'agit ensuite, dans un second temps, d'analyser les spécificités des tendances d'urbanisation nationales et transfrontalières. Pour finir, les résultats obtenus sont mis en perspective et associés à des pistes de développements futurs.

Chapitre 7

L'occupation du sol : Analyse à travers le jeu de règles

La phase de calibrage de l'arbre de décision a permis de mettre en évidence l'influence de la localisation géographique dans la définition automatique des règles de développement urbain. L'élaboration d'un modèle plus performant est rendue possible par la mobilisation des variables localisant la cellule dans l'un ou l'autre de États (France ou Allemagne), au Nord ou Sud de la zone, ou encore au regard des agglomérations de différents niveaux présents sur la zone d'étude. Si l'importance de la localisation géographique est soulignée, il reste cependant une interrogation sur l'importance de son rôle au sein du groupe de variables mobilisé lors du processus d'apprentissage. En d'autres termes, il est question de définir l'importance que revêt la localisation dans l'un ou l'autre des États frontaliers dans la création de règles d'urbanisation. Existe-t-il une part transfrontalière dans le processus de développement urbain ? Ce point est abordé dans la première section. Le caractère explicite des règles créées par le processus de décision issue de l'algorithme d'apprentissage est un des points remarquables de notre travail et soutient le choix de cette méthode. Le présent chapitre, dans les sections 2 et 3, s'attache à en exploiter l'essence thématique dans une analyse détaillée des règles. L'objectif est de définir les particularités spatiales qui permettent de caractériser le développement urbain sur notre zone d'étude, tant du point de vue du type d'occupation du sol présente dans le voisinage (section 2), que de celui de l'accessibilité aux principales agglomérations de la zone d'étude (section 3).

1. L'INFLUENCE DES VARIABLES

Deux analyses sont mobilisées pour établir puis quantifier l'importance des variables caractérisant la cellule urbanisée entre 1990 et 2006. La première est similaire à la démarche mise en place pour établir l'apport des variables de localisation géographique dans l'élaboration d'un modèle plus performant. Elle reproduit la méthode de test établissant l'apport de l'ajout des variables en fonction des critères de performance établis précédemment. La sous-section 1 présente les résultats de cette analyse. La deuxième analyse donne l'opportunité de quantifier l'influence des variables dans la construction de l'arbre dont est extrait le jeu de règles simulé. Le positionnement hiérarchique de chaque variable au regard des autres variables est quant à lui présenté dans la sous-section 2.

1.1. Une influence notable : la localisation

Une démarche itérative est à son tour développée pour rendre compte de l'importance de chaque variable de localisation géographique. La performance des modèles obtenus pour chacune des combinaisons possibles entre les six variables de localisation est alors évaluée. La hiérarchie de leurs performances est établie selon la même démarche que celle effectuée dans le chapitre précédent : évaluation de la performance des modèles obtenus sur la seule base des variables décrivant l'occupation du sol dans le voisinage, au regard de celle obtenue en ajoutant les variables de localisation géographique. Les indicateurs de performances obtenus lors des tests sont donc analysés pour hiérarchiser l'importance des variables.

Le paramétrage de l'arbre de décision retenu ici est celui basé sur les analyses de performance effectuées dans le chapitre précédent. En effet, ce paramétrage apporte la réponse la plus satisfaisante en termes de prédiction mais aussi de structure du jeu de règles. Pour rappel, les paramètres sont les suivant : intervalle de confiance 0,25, minium d'instances par feuille 1, voisinage 30 px, pondération 160. La démarche itérative consiste donc à tester chaque combinaison de variables de localisation géographique : Indice de Pays, Nord – Sud, accessibilité pôle de niveau 1, accessibilité pôle de niveau 2, accessibilité pôle de niveau 1 pondérée, accessibilité pôle de niveau 2 pondérée. L'ensemble des 63 combinaisons de ces six variables sont testées dans le cadre d'un apprentissage à 75% induisant de nouveau 10 répétitions du même paramétrage. La performance de chaque modèle obtenu est évaluée, comme précédemment, à travers la médiane des CCM test dans un premier temps, puis d'après leur nombre de règles et l'erreur où la surestimation. La réduction de l'écart entre les critères de performance des différents paramétrages nous pousse, également, à mobiliser la médiane de l'erreur cumulée des prédictions du modèle sur les échantillons « entrainement » et « test » (FP – FN entrainement et test), en plus des autres critères.

Les tests des différentes combinaisons de variables mettent en évidence l'influence de la variable de localisation dans l'un ou l'autre des pays (Indice de Pays – IDP). Sur la totalité

des 179 combinaisons (répétitions incluses) dont le CCM test est supérieur à 0,6, plus de 60% s'appuient sur cette variable.

Table 7. 1 : Paramétrages les plus performants d'après la somme des erreurs de prédictions

ID	Paramétrages	CCM	CCM test	Nb règles	Nb de variables	Vrai Positif	Faux Positif	Vrai Négatif	Faux Négatif	Vrai Positif test	Faux Positif test	Vrai Négatif test	Faux Négatif test	Somme FN	Somme FP	*Somme Erreur*
1	**IDP, NS, PT1, PT2, T1, T2**	**0,9686745**	**0,6020295**	**390,5**	**9568**	**2105**	**37,5**	**397258**	**99**	**417,5**	**249,5**	**132198,5**	**302**	**401**	**287**	***688***
2	IDP, PT1, T1, T2	0,9678185	0,591695	387	9283,5	2092	37,5	397253,5	103,5	408,5	244,5	132195,5	304	407,5	282	***689,5***
3	IDP, PT2, T2	0,966127	0,5907415	382	9021,5	2091	41,5	397252,5	107	405	239,5	132201,5	304	411	281	***692***
5	T2	0,968127	0,6044955	380	9458,5	2093	37,5	397264	98,5	407,5	240,5	132189	318,5	417	278	***695***
4	IDP, T1, T2	0,967818	0,594532	386,5	9285,5	2096	42,5	397251,5	94	413,5	252,5	132171,5	306	400	295	***695***
7	NS, T1, T2	0,9666145	0,5986635	381	9457	2080,5	42	397270	102,5	423,5	246	132181,5	308	410,5	288	***698,5***
6	PT1, T2	0,9647365	0,592553	384	9342	2096,5	41,5	397256,5	107,5	406,5	250	132199,5	299,5	407	291,5	***698,5***
8	IDP, PT1, PT2, T1	0,9672765	0,595442	389,5	9403	2079	39	397263	104	416,5	247	132185,5	310,5	414,5	286	***700,5***
9	NS, PT2, T1, T2	0,965863	0,595112	383,5	9498	2077,5	36,5	397268,5	105	412	248	132177	311,5	416,5	284,5	***701***
10	NS, PT1, T1, T2	0,9677245	0,5956085	387,5	9476	2075	38	397262	101	411	242	132192,5	320,5	421,5	280	***701,5***
11	IDP, PT2, T1, T2	0,96595	0,6000815	379	8973,5	2069,5	40	397269	102,5	413,5	248,5	132179,5	312	414,5	288,5	***703***
13	IDP, PT2	0,9657975	0,5993005	386	8832,5	2073,5	39	397268	101,5	421	243	132189	320,5	422	282	***704***
12	IDP, NS, PT1, T1, T2	0,9669255	0,596386	389	9275,5	2087	41,5	397268	101,5	416,5	248,5	132182,5	312,5	414	290	***704***
14	IDP, NS, PT2, T2	0,964761	0,598091	385	9317,5	2088	40,5	397257,5	114	412	239	132200,5	312	426	279,5	***705,5***
15	IDP, NS	0,9660575	0,595134	383	9121	2107	40,5	397252,5	108	401,5	253,5	132203,5	305	413	294	***707***

Les quinze paramétrages les plus performant suivant le critère rassemblant la somme des erreurs de prédiction (faux positifs et faux négatifs) sont extraits pour en analyser l'occurrence des variables de localisation géographique. Cette liste présente dix paramétrages s'appuyant sur l'indice de pays, dont les trois premiers, c'est-à-dire les plus performant (Table 7. 1).

Table 7. 2 : Occurrence des variables de localisation géographique dans les 15 meilleurs paramétrages.

Variables de localisation	Occurrence
Accessibilité niveau 2	12
Indice de pays	10
Accessibilité niveau 1	9
Nord - Sud	7
Accessibilité pondérée niveau 1	6
Accessibilité pondérée niveau 2	6

C'est par ailleurs l'accessibilité au pôle de niveau 2 qui apparaît le plus fréquemment dans la liste restreinte (Table 7. 2). Ce point met en exergue le rôle de la localisation au regard des agglomérations, notamment de celles de niveau 2. Un des paramétrages s'appuyant uniquement sur l'accessibilité aux pôles de niveau 2 ressort dans les cinq paramétrages les plus performants. Ce dernier s'avère cependant moins performant dans la localisation des transitions observées.

D'après la table 1, le jeu de données incluant la totalité des variables de localisation géographique s'avère plus performant. Ce dernier est donc retenu pour étudier l'influence des différentes variables dans la définition des règles de développement urbain sur la zone frontalière franco-allemande.

La sélection du meilleur modèle est, là encore, basée sur les calculs de valeurs médianes des critères de performance. L'analyse est complétée par celle de l'amplitude de leur variation lors des différentes répétitions (Table 7. 3).

Table 7. 3 : Répétition du modèle contenant toutes les variables de localisation géographique

Répétitions	CCM	CCM test	Nb règles	Nb de variables	Vrai Positif	Faux Positif	Vrai Négatif	Faux Négatif	Vrai Positif test	Faux Positif test	Vrai Négatif test	Faux Négatif test	Somme FN	Somme FP	***Somme Erreur***
9	0,974511	0,622621	395	10370	2112	38	397272	72	439	234	132197	295	367	272	639
3	0,968642	0,61435	393	9937	2105	36	397253	100	426	244	132208	287	387	280	667
8	0,968416	0,592375	397	9648	2120	36	397236	102	386	220	132249	310	412	256	668
4	0,969935	0,603057	387	9852	2140	52	397222	80	411	250	132217	287	367	302	669
6	0,970859	0,601002	388	9047	2105	36	397263	90	421	253	132189	302	392	289	681
10	0,968809	0,617129	353	7814	2071	38	397290	95	447	246	132167	305	400	284	684
5	0,968707	0,570801	410	9479	2140	35	397216	103	373	255	132235	302	405	290	695
1	0,964322	0,604796	375	8476	2104	57	397235	98	420	249	132200	296	394	306	700
7	0,964804	0,584408	388	9488	2058	37	397286	113	415	255	132163	332	445	292	737
2	0,96604	0,568951	394	9872	2068	43	397281	102	405	267	132150	343	445	310	755

La répétition 9 présente les meilleurs CCM (entrainement et test). Ce point souligne son efficacité à prédire les transitions observées, tant sur le jeu de données ayant servi à l'apprentissage que sur celui isolé pour le test. Cette caractéristique est extrêmement intéressante et est privilégiée au dépend du nombre de variables mobilisées, qui lui n'est pas le plus faible des dix répétitions. La répétition 10 présente en effet un jeu de règles moins conséquent, mobilisant moins de variables. Néanmoins, son erreur de localisation des transitions est supérieure (+45) notamment concernant les faux négatifs (+33). Le choix privilégié ici est celui du modèle limitant l'erreur de prédiction, qui omet la localisation de cellules urbanisées dans les données observées. Il s'agit donc de limiter les faux négatifs. Le modèle sélectionné, qui recours à plus de variables, contient par ailleurs *a fortiori* plus d'information décrivant le voisinage ou les caractéristiques des cellules urbanisées. Les informations contenues dans le jeu de règles serviront ultérieurement de base à la description thématique de l'urbanisation sur la zone d'étude.

Pour rappel, l'influence globale des variables de localisation géographique, mais également celle de la taille du voisinage qui indique le périmètre d'observation de l'occupation du sol permettant de décrire l'environnement des cellules urbanisées et de créer le modèle de développement urbain, ont été établies par les analyses effectuées dans le chapitre précédent.

Cette section quant à elle présente une première hiérarchie de l'influence des variables de localisation géographique d'après la performance des modèles qui les sollicitent. Cette approche montre que l'accessibilité aux pôles de niveau 2 semble être la variable de localisation géographique la plus influente au regard de son occurrence dans les différents modèles, suivit de l'indice de pays, puis de nouveau de l'accessibilité, mais au pôle de niveau 1 cette fois.

Si dans le cadre du modèle retenu, la totalité des variables de localisation géographique sont incorporées dans le jeu de données, nous n'avons cependant pas d'information concernant l'importance du rôle respectifs de chacune d'entre-elle dans la construction de l'arbre. Cette

information est primordiale car, si nous sommes dès lors en mesure d'affirmer que l'accessibilité au pôle de niveau 2 ou que l'indice de pays joue un rôle dans la performance du modèle, nous ne sommes pas en mesure de déterminer l'importance de ce rôle. Il est alors question de hiérarchiser l'influence des variables dans la construction de l'arbre. Ce point est abordé dans la sous-section suivante.

1.2. Des contributions à préciser

La structure du jeu de règles (nombre et longueur des règles le composant) est dépendante :

- du contexte (taille de la zone d'étude, résolution, nombre de classe de LUC, etc.),
- de la question et des objectifs de recherche,
- du paramétrage de l'algorithme.

Les jeux peuvent atteindre plusieurs centaines de règles, composées chacune de plusieurs dizaines de variables. Dans ces conditions, il est alors question d'interpréter l'influence respective de chaque variable, au sein du jeu de règles. Dès lors, la présente section introduit la création d'un index qui permet de rendre compte de l'importance relative des variables mises à contribution dans la création de l'arbre. *In fine*, cet index rend donc compte de leur importance dans la description du voisinage ou dans celle des caractéristiques de localisation géographique d'une cellule urbanisée. L'objectif est donc d'expliciter la pondération établie pour cette analyse.

La définition de l'influence des variables dans la construction de l'arbre de décision dont est issu le modèle servira de base à l'analyse thématique du jeu de règles. À cette fin, l'importance de chaque variable est déterminée et hiérarchisée grâce à l'attribution d'un score relatif à son occurrence dans le jeu de règles. La construction de l'arbre de décision est un processus itératif (Chapitre 5). Ainsi, la première variable sélectionnée est considérée par l'algorithme comme la plus discriminante et à même de définir les classes les plus homogènes après partition (tri entre transitions urbaines et non-changement). Le postulat relevant de l'importance des variables se base donc sur les hypothèses suivantes :

- L'importance de la variable est directement liée au niveau d'avancement du processus de construction de l'arbre, lors de sa sélection : les variables sélectionnées en amont sont plus à même d'être la base d'une description plus générale du processus ; les variables sélectionnées en fin de processus apportent de la précision dans cette description.
- L'importance des variables est relative à la taille de la règle dont elles sont issues : une variable issue d'une règle plus longue aura moins d'importance propre qu'une variable issue d'une règle plus courte.
- L'importance de la variable dépend également de son occurrence de mobilisation au cours de la construction de l'arbre : une variable sélectionnée à plusieurs niveaux du processus d'apprentissage, est plus influente.

Un score est attribué aux variables pour exprimer simultanément les trois aspects listés ci-dessus. Afin de rendre compte de la hiérarchie de l'influence des variables de manière optimale, chacune des règles est considérée comme d'efficacité égale puisqu'elle permet de

localiser au moins une cellule urbanisée dans les données observées. Ce point est discuté plus en détail dans la sous-section suivante. L'influence individuelle d'une variable est donc moyennée d'après le nombre total de variables composant la règle. Par exemple, une règle composée de 32 variables, verra chacune d'entre-elles imputée d'un poids de 1/32 = 0,03125.

L'utilisation d'une moyenne est basée sur l'appréhension de la règle de transition suivant son aspect logique « Si …Alors », dont les conditions sont connectées à travers une fonction « Et ». Dans ce cadre, la position des variables (première, seconde, etc.) au sein de la règle n'a absolument aucune importance hiérarchique. L'utilisation d'une moyenne permet d'intégrer au score à la foi l'influence de la taille de la règle et celle de l'itération de sélection lors du processus d'apprentissage (amont du processus ou feuille). En effet, une variable appartenant à une règle plus courte est considérée comme porteuse d'un pouvoir explicatif plus important puisqu'elle permet de prédire au moins une transition urbaine, au sein d'un groupe de variables plus limité. Chacune des variables composant une règle courte peut donc être considérée comme plus influente individuellement que les variables composant une règle plus longue. Les variables sélectionnées en amont du processus d'apprentissage, quant à elles, sont présentes dans un nombre plus important de règles, et se verront donc attribuer un poids à chacune de leur mobilisation. En effet, la première variable sélectionnée pour effectuer la partition du jeu de données se verra attribuer un score dans chacune des règles. Cet aspect est particulièrement intéressant mettre en lumière le processus de sélection des variables qui est *in fine* hiérarchique. La Figure 7. 1 illustre par le biais d'un exemple fictif, le processus d'attribution du score. Trois règles y sont présentées.

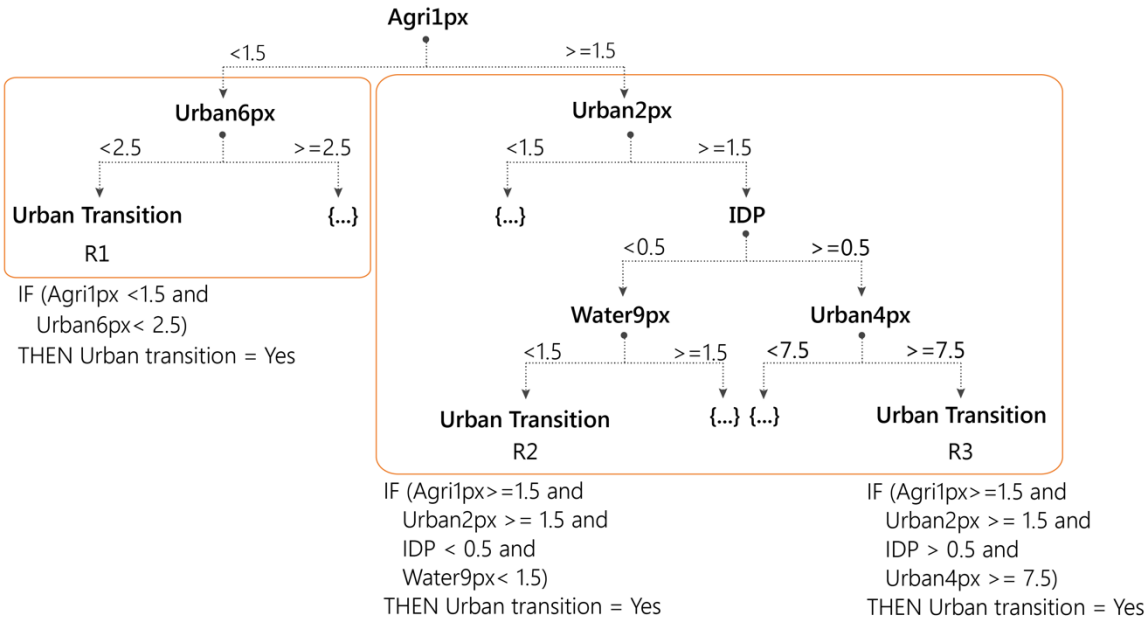


Figure 7. 1 : Illustration de l'attribution du score des différentes variables

Dans cet exemple, chacune des trois règles permet de localiser au moins une cellule ayant évoluée vers un état urbain. Cependant, R2 et R3 nécessitent l'utilisation de deux fois plus

de variables (4 au lieu de 2). Toutes choses égales par ailleurs, nous estimerons donc que l'influence des variables Urbain2px, IDP et Water9px est égale à celle de Urban2px, IDP et Urban4px, qui elles-mêmes sont tout aussi influente que la seule variable Urban6px. En conséquence, les variables mobilisées dans la règle 3, se verront chacune attribuer ¼ de l'influence de d'Urbain6px. L'importance d'une variable présente dans une règle plus courte est donc favorisée. La sélection d'une variable en amont du processus de création de l'arbre, c'est-à-dire en priorité au regard de celles sélectionnées par la suite, est interprétée comme une marque d'importance. Toujours suivant notre exemple, Agri1px est la variable qui permet la meilleure classification des individus ayant ou non effectué une transition vers l'urbain, suivie d'Urban6px et d'Urban2px au second niveau, etc. La moyenne proposée précédemment permet également de faire ressortir l'importance relative de la variable d'après son itération de sélection dans le processus de création de l'arbre de décision, tout en gardant l'influence de la taille des règles dont elle est une condition de validité. En effet, la variable Agri1px se verra attribué 3 poids, puisqu'elle est présente dans 3 règles. Ainsi, l'attribution du score des variables mobilisé dans l'exemple est :

R1 : Agri1px -> ½ = 0,5
Urbain6px -> ½ = 0,5

R2 : Agri1px -> ¼ = 0,25
Urbain2px -> ¼ = 0,25
IDP -> ¼ = 0,25
Eau9px -> ¼ = 0,25

R3 : Agri1px -> ¼ = 0,25
Urbain2px -> ¼ = 0,25
IDP -> ¼ = 0,25
Urbain4px -> ¼ = 0,25

Agri1px = 1
Urbain6px = 0,5
Urbain2px = 0,5
IDP = 0,5
Eau9px = 0,25
Urbain4px = 0,25

Le score reflète donc correctement l'importance des variables d'après les critères établis précédemment. Ici, la variable Agri1px est la première sélectionnée et participe à trois règles. Elle est donc la plus influente. Urbain6px, Urbain2px et IDP présentent une influence égale, pour autant les raisons sous-jacentes divergent. En effet, Urbain2px et IDP sont favorisées puisqu'elles sont mobilisées dans 2 règles. Même si les règles sont plus longues, cette double appartenance les place au même niveau qu'Urban6px qui elle a plus d'influence puisqu'elle est condition d'une règle plus courte (R1). Pour finir, Eau9px et Urbain4px présentent le plus petit poids puisqu'elles sont sélectionnées à la fin du processus de création de l'arbre de décision. Elles font partie des règles les plus longues et n'interviennent que dans une seule règle, respectivement R2 et R3.

En conclusion, l'attribution d'un poids moyen tel qu'il est calculé permet indirectement de retranscrire à la fois la hiérarchie de sélection de la variable ainsi que son importance en matière d'occurrence dans le jeu total.

2. L'INFLUENCE DU VOISINAGE

Plusieurs approches ont été développées dans l'objectif de synthétiser l'information contenue dans le jeu de règles. Pour rappel, le jeu de règles a été créé automatiquement par le processus de décision mobilisé lors de la construction de l'arbre. L'intérêt de cette démarche réside dans l'analyse des choix effectués par la machine, pour décrire le processus de développement urbain. Ainsi, cette sous-section permet de présenter la démarche élaborée pour étudier l'influence de l'occupation du sol présente dans le voisinage d'une cellule urbanisée, et celle de sa localisation géographique. Dans un second temps, c'est la polarité de cette influence qui est étudiée. La structure et la composition du jeu de règles est analysée à travers l'attribution d'un score à chacune de ses variables. Cette section présente une analyse du jeu de règles dans sa totalité. Comme mentionné précédemment, aucune hiérarchie n'est retenue pour évaluer la performance d'une règle d'après son potentiel prédictif. Chaque règle est considérée comme de performance équivalente puisque chacune d'entre-elles permet de prédire au moins une transition observée. La première sous-section se focalise sur l'analyse des différents types de transitions urbaines retrouvées dans le jeu issu de l'apprentissage automatique. En effet, différents types de transitions sont observés au regard de l'état initial d'occupation du sol de la cellule urbanisée. La deuxièmes sous-section présente, elle, l'influence des différents types d'occupation du sol présents dans le voisinage d'une cellule urbanisée à l'échelle de la zone d'étude. La troisième sous-section porte, pour finir, sur la polarité de l'influence des différents types d'occupation du sol présents dans le voisinage.

2.1. L'occupation du sol

Dans un premier temps, le jeu de règles est analysé dans sa totalité. La première information exploitable issue de cette analyse concerne le type de transition. Le type de transition est défini en fonction de l'état initial nécessaire à l'application de la règle. En effet, d'après l'application de la logique automate cellulaire, la transition se base et dépend, en premier lieu, de l'état initiale de la cellule interrogée. Il est intéressant d'observer les types de transitions régies par les règles élaborées à travers l'utilisation de l'arbre de décision (Table 7. 4).

Table 7. 4 : Règles et types de transitions

Transitions	Nombre de règles	Vrais Positifs	Faux Positifs
Agricole -> Urbain	262	1355	20
Tout type -> Urbain	96	57	13
Vignes -> Urbain	23	128	5
Industriel et commercial -> Urbain	7	27	0
Forêt -> Urbain	5	17	0
Zones récréatives -> Urbain	2	8	0
Total général	**395**	**2112**	**38**

La majorité des règles traduisent la transition d'un état agricole à un état urbain. Ce constat correspond aux tendances observées à l'échelle européenne (Chapitre 1). C'est particulièrement le deuxième type le plus importante qui attire notre attention. En effet, « tous types » précise que l'état initial ne rentre pas en compte dans l'application d'une règle de transition à la cellule. Ce point est en contradiction avec la définition classique de l'automate cellulaire. Il souligne que la configuration du voisinage prime sur l'état initial de la cellule, ce, dans 96 règles sur 395, soit dans 24% des cas (Table 7. 4). Ce point traduit un caractère plus « généraliste » de la règle produite, puisqu'elle s'affranchit de l'état initial. Toute cellule est, dès lors, susceptible de répondre positivement à ce type de règles. Ce point particulier sera discuté plus en détails dans le chapitre 9.

L'analyse du jeu dans sa totalité permet également d'appréhender l'influence des types d'occupation du sol à travers leur score. En effet, chaque règle illustre une configuration favorable à la transition urbaine. L'attribution du score à chaque variable composant le jeu de données permet de hiérarchiser l'influence des différents types d'occupation du sol (Table 7. 5).

Table 7. 5 : Illustration de l'influence des différents types d'occupation du sol

	Types d'occupation du sol	Occurences	Occurences (%)	Scores	Scores (%)
	Urbain	2884	23,95	97,35	24,65
	Agricole	1934	18,65	67,09	16,99
	Forêt	1583	15,27	58,84	14,90
	Vignes	1083	10,44	49,17	12,45
	Industrie et commerce	826	7,97	29,89	7,57
	Zone récréative urbaine	778	7,50	29,13	7,38
	Zone immergée	672	6,48	25,21	6,38
	Localisation géographique	627	6,05	23,86	6,04
	Transport	383	3,69	14,45	3,66
	Total	10370	100	395	100

Le score et l'occurrence des variables témoignent d'une tendance générale cohérente avec les connaissances accumulées au fils des décennies au regard du développement urbain (Table 7. 5). En effet, dans la hiérarchie, les occupations du sol urbaines, agricoles et forestières sont majoritairement retrouvées dans le voisinage des cellules urbanisées. Ces observations concordent donc avec la réalité des processus (Chapitre 1 et 3) et soulignent la cohérence du modèle développé. L'influence cumulée de ces trois types d'occupation du sol dans la construction de l'arbre de décision est supérieure à 50%. Nous ferons donc ici la distinction entre les types principaux (urbain, agricole, forêt) et secondaires (vignes, zones récréatives urbaines, zones en eau, zones industrielles et commerciales, transport).

L'occurrence et le score sont fortement corrélés, et pour cause, le score dépend directement de l'occurrence. Il est néanmoins également possible d'observer l'appartenance des variables à des règles plus ou moins longues, à travers la variation entre la part que représente le score en pourcentage et celle que représente l'occurrence pour le même type

d'occupation du sol. Par exemple, l'influence des vignes est supérieure de 2% quand le score est considéré (12,45 %), par rapport à la part qu'elles représentent au vu de leur occurrence (10,44 %). Par conséquent, les variables vignes sont présentes dans des règles plus courtes. Cette analyse globale de l'influence des diverses occupations du sol a permis de hiérarchiser leur importance dans le voisinage. La description précise, tous les 100m, des types d'occupation du sol présents dans le voisinage des cellules urbanisées, permet également d'apprécier la variation de l'influence de chaque type le long d'un transect de voisinage grâce au processus d'apprentissage automatique. L'analyse de l'influence des différents types de sols suivant leur distance dans le voisinage est également réalisée d'après leurs scores. La taille du voisinage étudié est de 30 pixels soit 3km (Chapitre 6).

L'attribution du score à chaque variable permet donc d'établir l'influence d'un type d'occupation du sol à une distance donnée (Figure 7. 2). Comme la table 5, la figure 2 permet d'établir le rôle prépondérant de l'urbain dans le voisinage. Elle permet de préciser que son influence majeure est localisée dans un périmètre de 200m. Son score est de loin supérieur à la majorité des scores observés pour les autres variables urbaines le long du transect. Il est également supérieur à celui des autres types d'occupation du sol. Un second pic majeur se distingue à 1,7km mais pour les vignes.

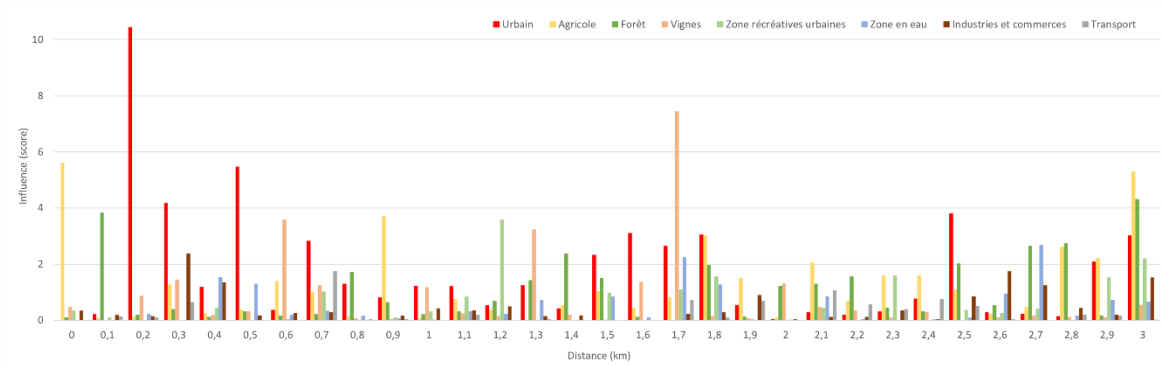


Figure 7. 2 : Influence des différents types d'occupation du sol entre 0 et 3km

Outre une description plus précise du voisinage, le graphique présenté ci-dessus permet également de souligner, une nouvelle fois, l'influence de l'état initial de la cellule suivant la règle de transition (influence à 0 km). L'agricole est principalement à la base de la transition urbaine puisqu'il est mobilisé à la distance 0 km soit en tant qu'état initial. Le graphique ci-dessus permet donc à la fois de synthétiser l'information contenue dans les tables 4 et 5, et présente l'avantage de fournir une observation de l'évolution de l'influence de chaque type d'occupation du sol tout au long du transect de voisinage.

2.2. La distance

La structure du jeu de données offre donc l'opportunité d'analyser l'influence des types d'occupation du sol en fonction de leurs distances dans le voisinage des cellules urbanisées. Le score de l'influence est appréhendé suivant deux référentiels. Dans un premier temps, brut, il permet de comparer l'influence des différents types de sols de manière générale sur l'ensemble de la zone étudiée (Figure 7. 2). Dans un second temps, l'influence est étudiée au sein de chaque classe d'occupation du sol. Poursuivant cet objectif, le score est converti en pourcentage pour plus de lisibilité. Ainsi, l'influence relative d'un type d'occupation du sol en fonction de la distance peut être appréhendée au sein de sa propre classe d'occupation du sol. Par exemple, le score de l'occupation du sol de type urbain à 200m (2px) est de 10,45 (Table 7. 6). Cette valeur représente une influence de 4,95 % de la totalité de l'influence des variables composant l'ensemble du jeu de règles. Ce même score représente part contre 19,22% de l'influence de la variable occupation du sol de type urbain, au regard de la contribution de chaque variable urbaine. C'est l'influence intra-classe. L'influence intra-classe de chaque variable d'occupation du sol est projetée suivant sa distance dans le voisinage. Cette projection permet d'observer plus précisément l'évolution de l'influence de chaque type d'occupation du sol en fonction de sa distance dans le voisinage. Elle permet également de faire abstraction des différences d'amplitudes qui existent entre types d'occupation du sol et limite la visibilité de l'évolution le long du transect d'un type de sol peu influent au départ. Par exemple, le transport n'apparait pas clairement sur la figure 2 en raison de l'influence très supérieure des types d'occupation principaux.

Table 7. 6 : Attribution du score et conversion en pourcentage de l'influence des variables

Influence de l'occupation du sol de type urbain dans le voisinage

Distances (km)	Score	Influence Total (%)	Pourcentage Intra-classe (%)
0	0	0	0
0,1	0,22	0,11	0,41
0,2	10,45	4,95	19,22
0,3	4,18	1,98	7,69
0,4	1,2	0,57	2,2
0,5	5,48	2,59	10,07
0,6	0,37	0,18	0,68
0,7	2,83	1,34	5,2
0,8	1,29	0,61	2,38
0,9	0,81	0,38	1,49
1	1,23	0,58	2,26
1,1	1,22	0,58	2,24
1,2	0,53	0,25	0,98
1,3	1,24	0,59	2,28
1,4	0,43	0,2	0,79
1,5	2,33	1,11	4,29
1,6	3,1	1,47	5,7
1,7	2,65	1,26	4,87
1,8	3,05	1,45	5,62
1,9	0,55	0,26	1,01
2	0,03	0,02	0,06
2,1	0,29	0,14	0,53
2,2	0,2	0,09	0,36
2,3	0,32	0,15	0,58
2,4	0,77	0,37	1,42
2,5	3,82	1,81	7,02
2,6	0,29	0,14	0,53
2,7	0,23	0,11	0,43
2,8	0,15	0,07	0,28
2,9	2,09	0,99	3,84
3	3,02	1,43	5,55
Total	***54,39***	***25,76***	***100***

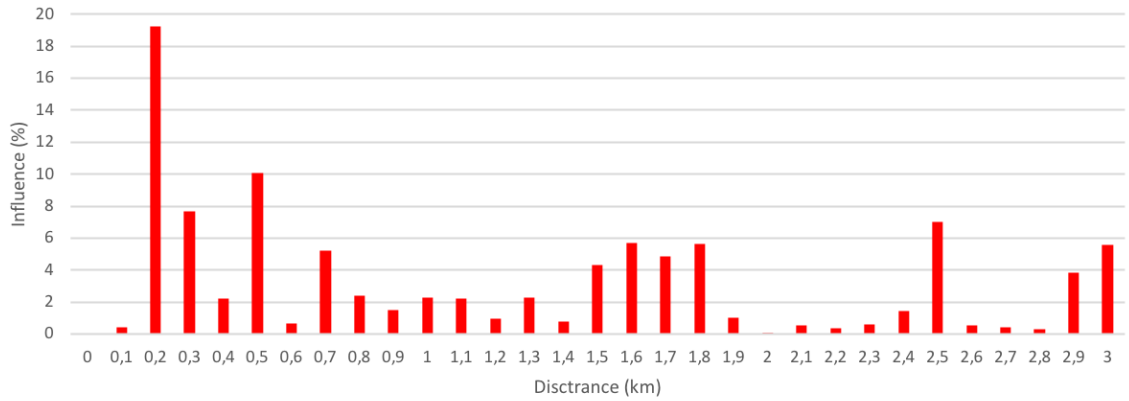


Figure 7. 3 : Influence de l'occupation du sol urbaine dans le voisinage en fonction de la distance (%)

Comme vu précédemment, le type urbain est parmi les plus influents dans le voisinage des cellules urbanisées. La Figure 7. 3 permet une observation relative de l'évolution de son influence en fonction de sa distance dans le voisinage. L'influence du type urbain présente une tendance globale décroissante. Elle est plus importante dans un voisinage proche (inférieur à 2 km). Elle est majeure dans les 500 premiers mètres. Un premier pic d'influence majeure est localisé dans un périmètre de 200 m, suivit d'un pic moins important, néanmoins toujours élevé à 500 m. Une seconde zone d'influence se dessine entre 1,5 et 1,8 km. Cette dernière est légèrement moins importante que le pic observé à 2,5 km, mais plus stable dans cet intervalle.

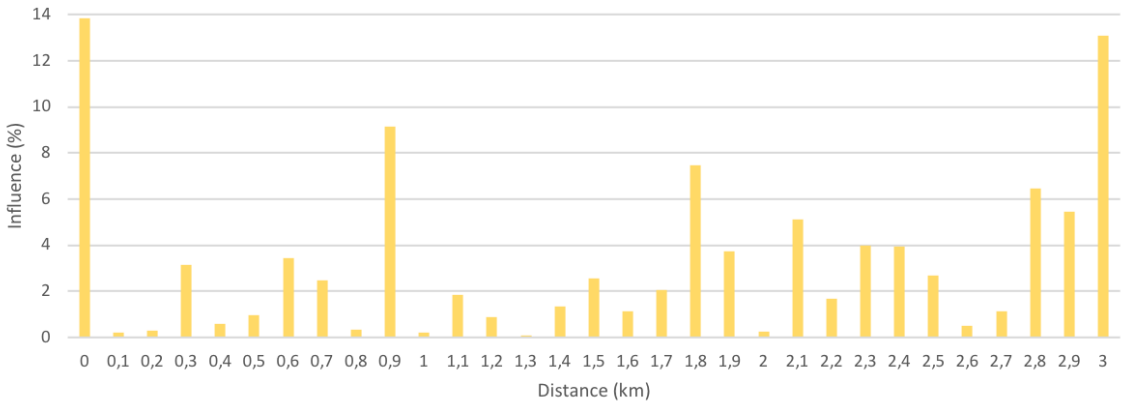


Figure 7. 4 : Influence de l'occupation du sol agricole dans le voisinage en fonction de la distance (%)

L'occupation du sol agricole se distingue par sa large mobilisation pour la distance nulle, c'est-à-dire en tant qu'état initiale favorisant la transition urbaine (Figure 7. 4). Pour le reste, l'importance de son influence relative est plus distribuée sur la totalité du transect. Contrairement à l'occupation du sol type urbain, elle est moins importante dans un périmètre très proche et augmente en s'éloignant. Si un pic est observé à 900 m, la majorité des pics suivant sont observés à partir d'une distance d'1,8 km. L'influence de l'occupation du sol

type agricole est donc globalement contraire à celle présentée par l'occupation du sol type urbain.

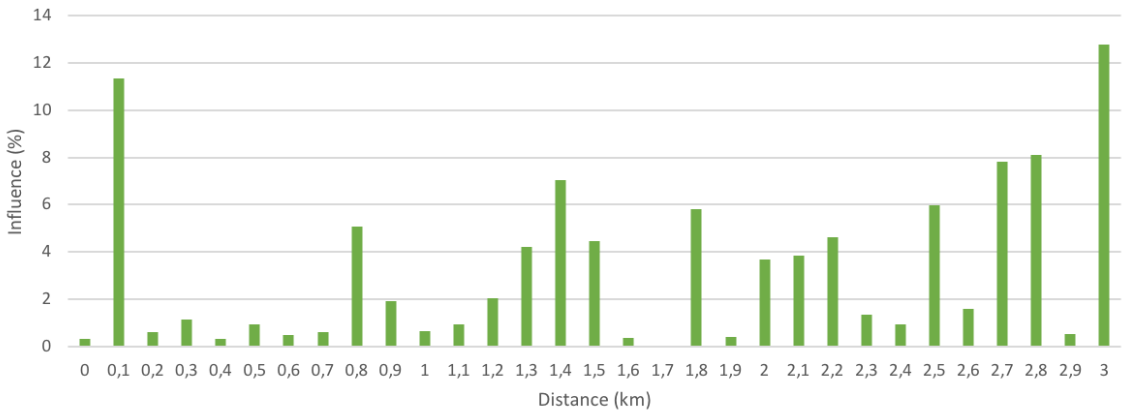


Figure 7. 5 : Influence de l'occupation du sol agricole dans le voisinage en fonction de la distance (%)

La forêt est le troisièmes type d'occupation du sol le plus important d'après l'attribution du score de l'influence dans le voisinage, ce quel que soit le type de règles. Deux pics majeurs se distinguent (Figure 7. 5). Le premier dans un périmètre très proche (100 m) ; le second à la limite du voisinage considéré (3 km). Une influence moyenne est observée aux alentours de 1,4 km puis de 2km, avant d'augmenter progressivement de nouveau à mesure de l'éloignement.

Les types d'occupation du sol principaux se distinguent des types d'occupation du sol secondaires par la mobilisation de leurs variables sur la totalité ou quasi-totalité du transect de voisinage. L'analyse de l'occupation du sol de types secondaires permet donc d'observer l'absence d'influence à certaines distances.

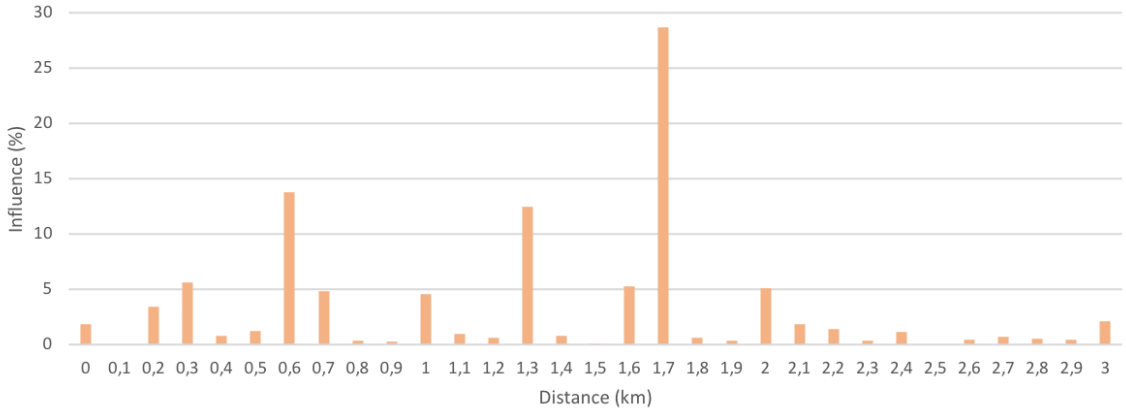


Figure 7. 6 : Influence de l'occupation du sol vignes dans le voisinage en fonction de la distance (%)

Les vignes sont moins mobilisées dans la construction de l'arbre de décision (Figure 7. 6). Elles présentent néanmoins un pic majeur à 1,7km. D'autre pics moins importants peuvent

être distingués notamment à 600m, 1,3 km ainsi que dans le voisinage proche (300 m). L'évolution de l'influence est plus disparate sur la longueur du transect au regard de celles observées précédemment pour les types d'occupation du sol. Elle diminue drastiquement après 2,3 km.

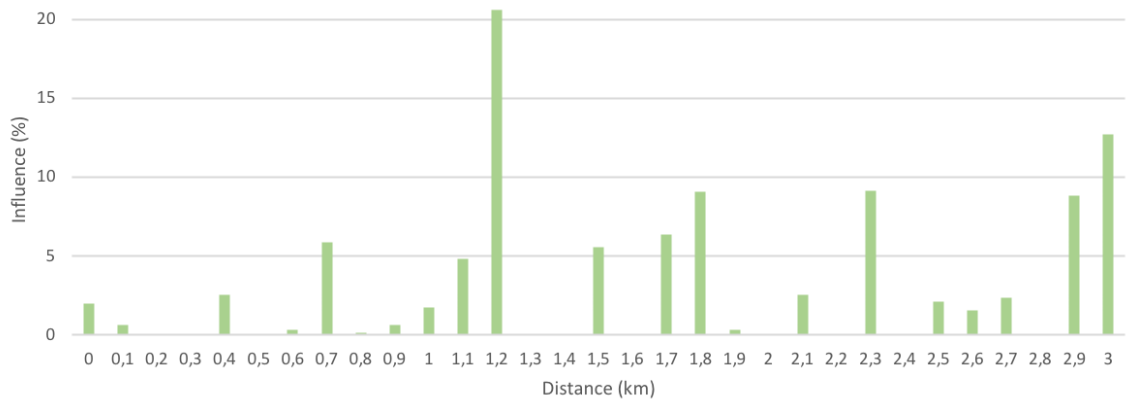


Figure 7. 7 : Influence de l'occupation du sol zones récréatives urbaines dans le voisinage en fonction de la distance (%)

Les zones récréatives urbaines présentent une distribution aussi éparse que celle observée pour les vignes. Le pic majeur est lui aussi présent après 1 km (1,2 km). Des pics de moindre importance apparaissent régulièrement le long du transect : 700 m, 1,8 km, 2,3 km, 2,9 km et 3 km (Figure 7. 7).

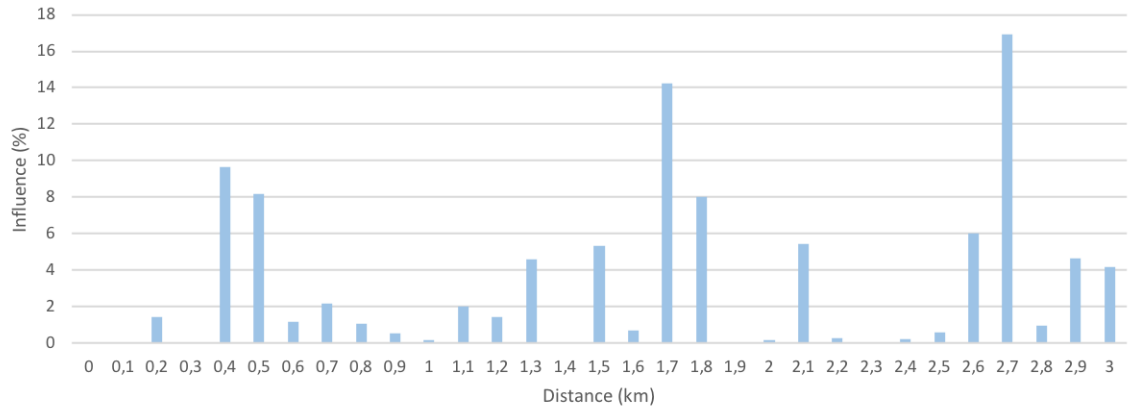


Figure 7. 8 : Influence de l'occupation du sol zones immergées dans le voisinage en fonction de la distance (%)

La question des zones en eau revêt quant à elle un intérêt particulier de par le double enjeu qu'elles représentent. La composante environnementale est indéniable, tout en étant complétée par l'aspect politique et culturel. En effet, le tracé de la frontière étatique suit celui du Rhin. Nous reviendrons ultérieurement sur ce point particulier (chapitre 8). Plusieurs pics majeurs se distinguent (Figure 7. 8) : 400 m, 500 m, 1,7 km, 1,8 km et 2,7 km.

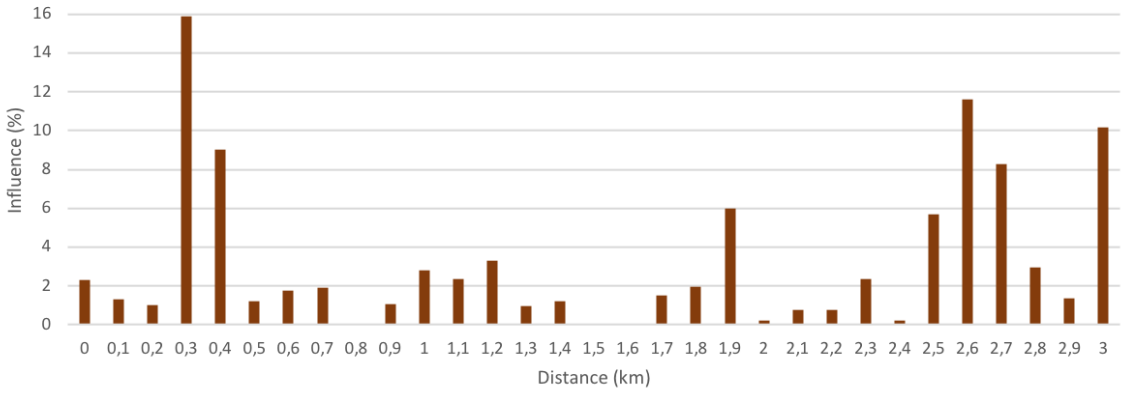


Figure 7. 9 : Influence de l'occupation du sol industries et commerces dans le voisinage en fonction de la distance (%)

Les zones industrielles et commerciales font également partie du type d'occupation du sol secondaire. Sa distribution s'apparente néanmoins à celles observées dans le cas des occupations principales, comme celle de la forêt ou des zones urbaines. Elle présente un pic d'influence majeur dans la partie proche du voisinage (300m). Un second pic est ensuite observé dans un voisinage plus éloigné à 2,6km (Figure 7. 9).

Finalement, le transport permet d'observer très nettement les distances mobilisées dans la construction de l'arbre pour ce type d'occupation du sol (Figure 7. 10). L'influence du transport n'est pas prépondérante dans un périmètre proche, et apparaît principalement après 1,7 km. Le pic majeur est cependant localisé à 700m.

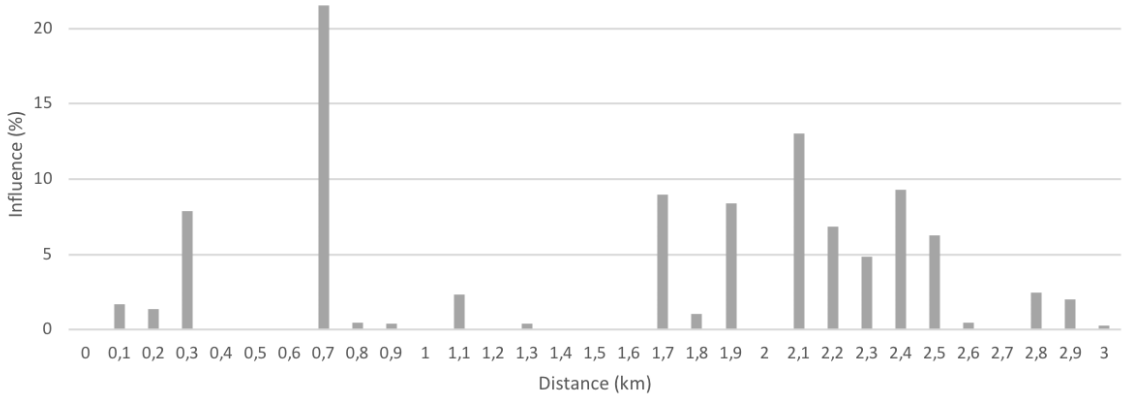


Figure 7. 10 : Influence de l'occupation du sol vignes dans le voisinage en fonction de la distance (%)

Cette sous-section a permis de mettre en lumière l'influence des différents types de sols en fonction de leur distance dans le voisinage des cellules urbanisées. L'influence prépondérante des types principaux d'occupations du sol est mise en évidence. Les graphiques permettent de localiser plus précisément les distances significatives dans le voisinage au regard du type d'occupation du sol analysé. Pour autant, la méthode développée nécessite une mise en relief de l'information obtenue indépendamment pour chaque

occupation du sol, au regard de l'influence générale des autres types d'occupation du sol. Si un pic d'influence du transport peut être observé à 700 m (Figure 7. 10) ce dernier reste relatif et doit être considéré parallèlement aux données représentant l'influence de tous les types d'occupation du sol en fonction du score brut (figure 2). L'influence de chaque type d'occupation du sol est donc définie pour une distance spécifique sur notre zone d'étude. Il convient néanmoins de s'interroger sur la polarité de cette influence : est-elle positive ? ou négative ? Ce type de sol à cette distance spécifique est-il recherché ou évité ? Ces questions font l'objet de la sous-section suivante.

2.3. Variables favorables, variables restrictives

L'influence générale puis relative des différents types d'occupation du sol, en fonction de leur localisation dans le voisinage des cellules urbanisées, est établie sur la base du score des variables sollicitées lors de la construction de l'arbre de décision. L'aspect logique sur lequel se base le processus de décision permet également d'établir la polarité de l'influence : la présence de ce type d'occupation du sol est-elle nécessaire à la transition ? Si oui, à quelle distance et dans quelle proportion ? Au contraire, ce type d'occupation du sol doit-il être limité ? doit-il être absent ?

Ces questions sont abordées à travers l'étude des pics majeurs d'influence observés sur les graphiques représentant l'influence intra-classe de chaque type d'occupation du sol. En effet, l'objectif ici n'est pas d'étudier chaque pic, mais bien d'étudier la polarité de l'influence des pics majeurs, dans le but d'établir des tendances généralisables sur la zone d'étude de l'occupation du sol présente dans le voisinage des cellules urbanisées.

La polarité est appréhendée, dans ce travail, d'après la portée restrictive (≤) ou favorable (>) de la présence d'un type d'occupation du sol à une distance spécifique dans le voisinage. La présence d'une restriction sera particulièrement intéressante quand elle révèle le caractère prohibitif d'une contrainte, donc qui impose l'absence d'un type d'occupation du sol (≤ 0). L'analyse de l'évolution des seuils, qui représentent la proportion (en hectare) d'un type d'occupation du sol dans le voisinage, permet également de dégager des tendances majeures dans l'évolution des surfaces favorisant la transition urbaine, le long du transect de voisinage.

C'est, dans un premier temps, l'influence totale des polarités restrictives ou favorable, pour chaque type d'occupation du sol qui est analysée.

Table 7. 7 : Influence de la polarité des différents types de variables

Types d'occupation du sol	Polarité - Scores			Polarité - Scores (%)	
	<=	>	Total	<=	>
Urbain	46,79	50,56	97,35	48,07	51,93
Agricole	47,96	10,21	58,17	82,45	17,55
Forêt	26,35	30,24	56,59	46,57	53,43
Vignes	40,88	6,03	46,91	87,14	12,86
Industrie et commerce	26,77	2,78	29,55	90,58	9,42
Zone récréative urbaine	22,23	5,99	28,22	78,76	21,24
Zone immergée	22,09	3,13	25,21	87,60	12,40
Localisation géographique	15,08	8,78	23,86	63,20	36,80
Transport	12,06	2,39	14,45	83,44	16,56
Total	260,21	120,11	380,32	68,42	31,58

Sur la totalité du jeu de règles, l'aspect restrictif prédomine puisque la mobilisation des variables d'occupation du sol en général est limitante dans plus de 68 % des cas (Table 7. 7). À l'échelle du type d'occupation du sol, ce caractère varie. En effet, le caractère favorable à la présence d'urbain ou d'agricole est, lui, plus important. Si le caractère restrictif prédomine, il ne doit pas être directement associé à une incompatibilité. L'incompatibilité existe seulement dans le cas d'un seuil inférieur ou égal à 0. La restriction stipule que la présence d'un certain type d'occupation du sol ne doit pas excéder une proportion définie par le seuil sélectionné lors de la construction de l'arbre. Par exemple, le seuil de tolérance de la forêt peut être inférieur à 80% de la surface du voisinage à 500m. Cette limite ne serait, par ailleurs, pas considérée comme extrêmement contraignante. Cependant, l'objectif, au-delà de l'extrapolation des tendances représentatives de l'influence des différents types de sol, est bien la localisation, en termes de distance, des seuils d'influence majeure des différents types d'occupation du sol qu'il est question d'étudier. Par exemple, en partant du postulat que la vigne joue un rôle dans le voisinage, il est question de définir plus précisément à quelle distance cette occupation du sol est recherchée ou rejetée (≤ 0 ou > 0).

Une occupation du sol qui influence dans une moindre mesure la construction de l'arbre présentera des tendances moins marquées. Par moindre mesure, nous entendons ici une occurrence faible dans le jeu de règles. Elle ne présente pas, pour autant, moins d'intérêt pour notre analyse. En effet, nous raisonnons ici en nombre de règles, indépendamment du nombre de cellules concernées par les règles en question. Ce choix trouve son fondement dans la volonté de ne pas restreindre l'importance de la règle à son appréhension dans un contexte rétrospectif. En effet, si une règle concerne un nombre restreint de cellules à un instant t-1, rien n'indique que cette tendance soit conservée à un instant t+1. Ainsi, une règle concernant un nombre restreint de cellules, pourrait être amenée à en concerner davantage à court du temps. L'intérêt est donc bien d'extraire l'information de distances et de proportion de sol, ce même dans le cas où un nombre restreint de règles est concerné. L'important est de localiser les seuils nécessitant une certaine proportion du type d'occupation du sol étudié ($0 <$) ainsi que les seuils exclusif (≤ 0) interdisant la présence.

Dans ce cadre, pour chaque type d'occupation du sol, l'analyse des seuils et proportions a été focalisé sur les pics majeurs d'influence favorable, ainsi que sur le voisinage proche (100m-300m). En effet, ce dernier sera le plus favorable aux seuils interdisant la présence d'un type de sol en particulier. La suite de cette sous-section permet d'analyser plus précisément l'évolution de la polarité des pics d'influence de chaque type d'occupation du sol ainsi que la présence ou non d'un caractère prohibitif.

Urbain

Comme mentionné précédemment, le type d'occupation du sol urbain est mobilisé à maintes reprises dans la construction de l'arbre de décision dont le modèle de simulation du développement urbain est issu.

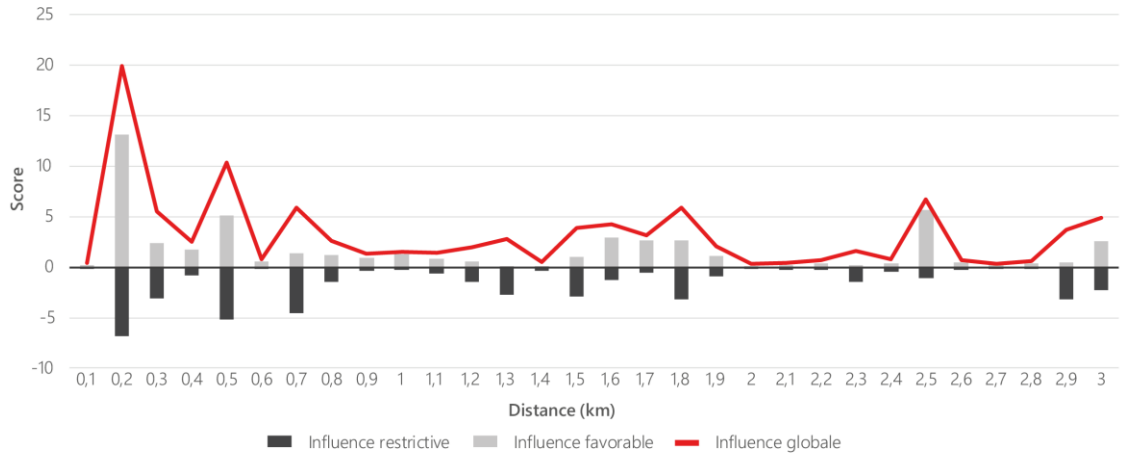


Figure 7. 11 : Polarité de l'influence du type urbain d'occupation du sol

Le graphique ci-dessus permet de mettre en relief la polarité de la contribution à l'influence totale observée et décrite par la courbe. Ainsi, pour le pic d'influence majeur observé à 200m, les deux polarités contribuent. L'influence favorable est légèrement supérieure (barre jaune). Le modèle impose donc, à travers certaines règles, une proportion d'urbain à 200m des cellules nouvellement urbanisées. Ce seuil retient par ailleurs particulièrement notre attention. En effet, des organismes comme l'INSEE, ou encore l'ONU, présentent cette distance comme favorable à l'urbanisation. Le modèle créé illustre également cette situation. Le recours aux variables urbaines se traduit par leur score. La Table 7. 8 permet également d'en analyser la polarité.

Table 7. 8 : Score et polarité des seuils de l'urbain

	Influence de l'urbain - score des variables			
	Unités	≤	>	Total
Total	Score	46,79	50,56	97,35
	Pourcentage	48	52	100
Pics	Score	19,87	32,05	51,92
	Pourcentage	38	62	100

Les scores des pics sélectionnés sont représentatifs de la tendance observée à l'échelle du jeu de règles total (Table 7. 8). La sélection des pics est donc cohérente avec la tendance globale. Ils accentuent néanmoins légèrement l'écart d'influence entre les variables restrictives et favorables (± 10%). Le jeu de règles est majoritairement composé de variables contraignant positivement la présence d'urbain au-dessus de leur seuil de référence dans le voisinage (>). Certaines permettent cependant de restreindre la surface urbanisée sous un certain seuil (≤) voire de l'exclure lorsque ce seuil est égal à 0. Les contraintes imposant un seuil minimum sont légèrement plus présentes (52%) que celles limitant la présence d'urbain (48%). La polarité et les différents seuils mobilisés sont analysés plus précisément, par le biais du score, pour les pics majeurs observés (Table 7. 9).

Table 7. 9 : Polarité des pics majeurs d'influence de l'urbain dans le voisinage

	Influence de l'urbain					
	Scores bruts			Scores %		
Distances (m)	<=	>	Total général	<=	>	Total général
200	6,80	13,12	19,91	34	66	100
500	5,20	5,12	10,32	50	50	100
1600	1,29	2,95	4,24	30	70	100
1800	3,19	2,68	5,87	54	46	100
2500	1,12	5,62	6,74	17	83	100
3000	2,27	2,56	4,84	47	53	100
Total	**19,87**	**32,05**	**51,92**	**38**	**62**	**100**

L'analyse des proportions d'occupation du sol de type urbain au niveau des seuils d'influence permet de rendre compte plus précisément de l'effet favorable ou au contraire, défavorable de la proximité urbaine des parcelles nouvellement urbanisées (Table 7. 10). L'absence de donnée signifie l'absence d'influence (tant positive que négative) dudit paramètre à la distance spécifiée. Le contenu de la table ci-dessous est restreint comme précédemment à la présentation des seuils au niveau des pics observés sur la Table 7. 10 (0,2km, 0,5km, 1,6km, etc.). Cette restriction permet une plus grande lisibilité des résultats. Le jeu total contenant 395 règles faisant intervenir l'urbain, seul un extrait concernant les seuils des 30 premières est présenté.

Table 7. 10 : Extrait des seuils urbains en fonction de la distance

Totalité des Règles - urbain dans la voisinage (%)

Règles	0,2 km		0,5 km		1,6 km		1,8 km		2,5 km		3 km	
	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>
1		0		24								5
2	0							3		4		
3		0		24								
4	0									4		5
5		0		24							19	
6	0					8	39				19	
7	0						39					
8		0		24								19
9	0						39					
10	0											
11		0		40					4			
12	0											
13		0	24									
14	0											
15		0	40	24					4			
16		0		24				3		4		10
17		25		24				3		5		4
18		0		24				3		4		
19		0		24								
20		0		24								
21		0	24			11						
22		0		24				3		4		
23		0		24								
24		0										
25	0						39				19	
26	0						39					19
27		0	24			7					9	
28		0		24						4		5
29		0		24						4		5
30		0		24				3		4		

Un des enjeux majeurs dans l'exploitation des résultats d'un point de vu thématique se trouve dans la synthèse et l'extraction de tendance dans la distribution des divers types d'occupation du sol présents dans le voisinage d'une cellule urbanisée. Dans l'analyse de la distribution de l'urbain dans le voisinage des cellules urbanisées, 6 tendances majeures peuvent être dégagées à partir de l'analyse de la totalité des règles de transitions (Table 7. 11). Ainsi, 56 règles interdisent la présence d'urbanisation à 200 m. Au contraire, une centaine de règles imposent une présence (tendances 2 à 6). La tendance 5 présente même un intervalle. L'urbain doit être présent à 200 m, mais ne doit pas dépasser 25%. Si l'urbain ne doit pas être présent à 200 m dans la tendance 1, cette dernière restreint aussi légèrement sa présence à 1,8km puisqu'elle doit être inférieure à 39%, mais également à 3km (≤ 19%). En résumé, la tendance 1 décrit un voisinage moins favorable à la présence d'urbanisation que les autres tendances majeures. La tendance 6, contrairement à la 5, impose un intervalle de tolérance mais à 500 m.

Table 7. 11 : Distribution de l'urbain dans le voisinage – tendances majeures à l'échelle de la zone

Totalité des Règles - urbain dans la voisinage (%)

Tendances	0,2 km	0,5 km	1,6 km	1,8 km	2,5 km	3 km	Nombre de règles
1	≤ 0			≤ 39		≤ 19	56
2	0 <	24 <		3 <	4 ou 5 <		47
3	0 <	≤ 24	7 <		4 <		43
4	0 <	≤ 24	≤ 7				10
5	0< x ≤ 25	24 <		3<	5 <	4 <	8
6	0 <	24< x ≤ 40		3<	≤ 4		8

L'analyse du voisinage urbain a permis de présenter en détails chaque étape effectuée pour expliciter l'élaboration des tendances générales extraites par l'analyse des règles de développement urbain crées automatiquement. La présentation des résultats pour les autres types d'occupation du sol est restreinte à celle du diagramme de l'influence polarisé ainsi que celle des tendances dégagées par l'application de la même démarche d'analyse. L'ordre de présentation des types de sol suit la hiérarchie établie précédemment d'après leurs scores d'influence générale à l'échelle de la zone (Table 7. 7).

L'analyse des seuils fournis par les pics majeurs permet donc de synthétiser l'information contenue dans le jeu de règles, issu du processus de décision. Cette même démarche est répliquée pour chaque type d'occupation du sol.

Forêt

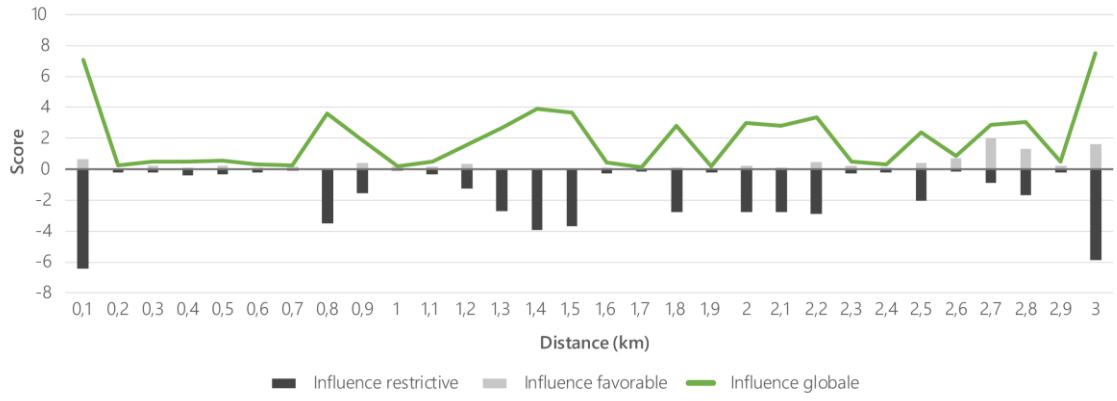


Figure 7. 12 : Polarité de l'influence du type forêt d'occupation du sol

La forêt intervient dans 378 règles sur les 395 présentent dans la totalité du jeu. Cette occupation du sol porte un caractère prohibitif à 100m interdisant la présence de forêt (56 règles). Par ailleurs, quand le seuil est restrictif, il reste ouvert à la présence de 2 ou 3 ha de forêt (≤ 37%) dans le voisinage adjacent des cellules urbanisées (91 règles), surface en somme toute importante puisque le voisinage à 100m est composé de 8 ha. Au contraire

seulement 7 règles imposent la présence de forêt à 100m dont la moitié impose par ailleurs plus de 2 ha.

La majorité des règles de l'échantillon limite à 50% la présence de forêt dans un périmètre de 800m. Il faut néanmoins souligner que plusieurs règles imposent également une présence minimum de forêt dans un périmètre de 1km (500m, 700m, 900m). Si la forêt n'est pas nécessaire dans le voisinage très proche, elle est cependant recherchée et nécessaire un peu plus loin. Quelques règles imposent également l'absence de forêt à 200m, 300m, 400m ou 500m, même si ce phénomène reste anecdotique.

Dans le cas des influences positives, c'est à partir de 2,2 km que l'occupation du sol prend plus d'importance. Ainsi, le développement urbain ne sera possible que si le seuil dépasse 60% dans certains cas. Pour d'autres, la limite sera plus basse (14%). C'est à partir de 2,7 km que l'influence de la présence de forêt augmente considérablement : 55 règles imposent la présence de forêt, dont 5 à plus de 50%. À 2,8 km, apparaît la présence d'un intervalle dans les proportions de forêt présente dans le voisinage, ainsi si dans une cinquantaine de règles elle doit dépasser 0 ou 30%, elle doit également ne pas dépasser 50%. Pour finir, plusieurs règles imposent un seuil supérieur à 70% de forêt à 3km. Quand ce dernier est plus faible, il impose seulement plus de 20% de forêt, par ailleurs encadré par une limite à 70%. Les grandes tendances dans la composition du voisinage considérant la présence de forêt sont résumées dans la table Table 7. 12.

Table 7. 12 : Distribution de la forêt dans le voisinage – tendances majeures à l'échelle de la zone

Tendances	0,1 km	0,2 km	0,3 km	2,2 km	2,6 km	2,7 km	2,8 km	3 km	Nombre de règles
1	≤ 37				0 <			≤ 74	91
2							0 < x ≤ 52		63
3	≤ 0			≤ 64					56
4						0 <			55
5								74 <	13
6		0 <				≤ 0			9
7				0 < ou 64 <					9
8	37 <								6
9						69 <			6
10			0 <						4
11	0 <			≤ 64					3

Agricole

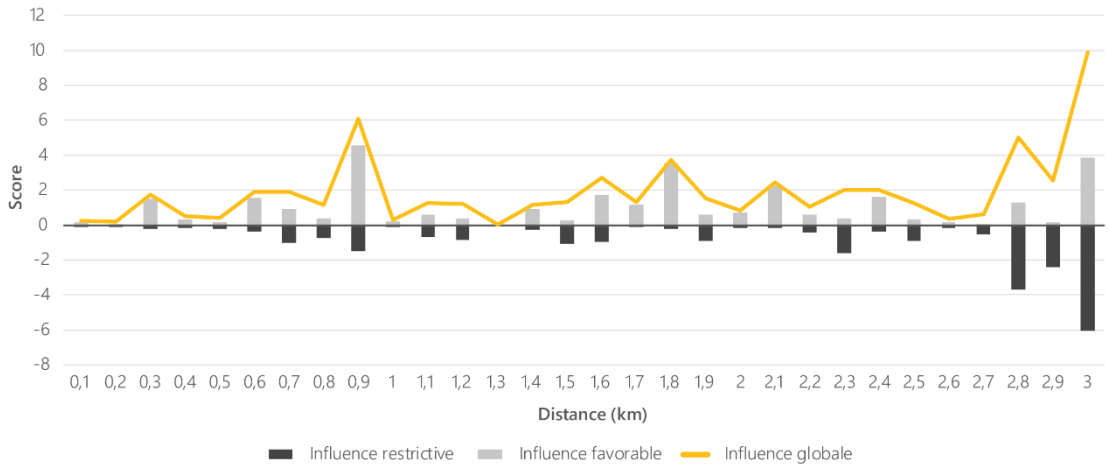


Figure 7. 13 : Polarité de l'influence du type agricole d'occupation du sol

L'influence favorable de la présence d'occupation du sol agricole dans le voisinage est également plus présente que l'influence restrictive (Table 7. 7). L'agricole intervient dans 277 règles sur les 395 présentent dans la totalité du jeu. Un avantage de la méthode illustré par l'analyse des seuils d'influence de l'agriculture observée dans le voisinage des cellules urbanisée, est celui de la présence d'encadrement des proportions d'occupation du sol agricole nécessaires à certaines distances. Un encadrement qui impose la présence d'occupation du sol agricole entre 21 et 84 % à 3km est retrouvé dans plusieurs règles par exemple (Table 7. 13). Les tendances majeures montrent des limites restrictives relativement élevées, comprises entre 84 et 89%. Elles ne sont donc pas excessivement limitantes. Au contraire, imposer une limite supérieure à 42% à 300m (tendances 3 et 4) revient à rechercher une présence de plus de 20 ha sur les 48 ha formant le voisinage de la cellule. L'agricole est alors recherché dans un voisinage proche. Dans l'ensemble, l'agricole est recherché dans le voisinage des cellules urbanisées, ceux quelques soit la distance étudiée au sein du transect. Dans la majorité des tendances, les limites supérieures sont absentes ou hautes, laissant beaucoup de flexibilité au regard de la présence de ce type de sol.

Table 7. 13 : Distribution de l'agricole dans le voisinage – tendances majeures à l'échelle de la zone

Tendances	0,1 km	0,3 km	0,9 km	1,6 km	1,8 km	2,1 km	2,4 km	2,8 km	3 km	Nombre de règles
1					31 < ou < 36					60
2		21 < et ≤ 84			32 <			≤ 85		54
3		42 <				21 <		≤ 85	21 < et ≤ 89	37
4		42 <				33 <		≤ 85	22 < et ≤ 89	28
5			21 <	46 < et ≤ 88				23 < et ≤ 88	23 < et ≤ 89	25
6			64 <	47 < et ≤ 88				23 < et ≤ 88	24 < et ≤ 89	12
7	62 <		20 <						14 < et ≤ 89	4

Vignes

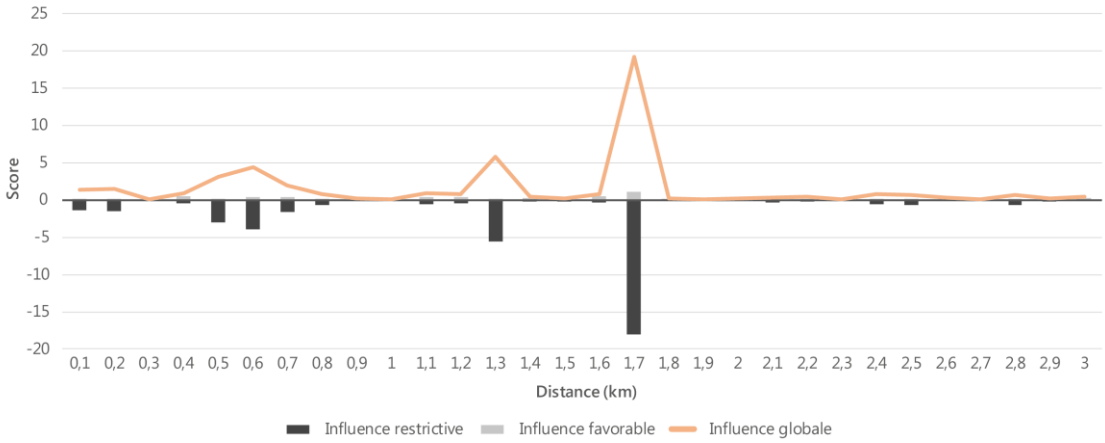


Figure 7. 14 : Polarité de l'influence du type vigne d'occupation du sol

La vigne est présente dans 383 règles sur les 395 règles composant le jeu, soit la presque totalité. Comme établi précédemment (Figure 7. 7), les occupations sur sol secondaire, qui comprennent la vigne, présentent une influence restrictive devient prépondérante (Table 7. 7). Ce point est également visible sur le graphique Figure 7. 14. Comme mentionné précédemment, l'analyse de l'influence des types de sols secondaire est principalement intéressante pour la restriction prohibitive (≤ 0) ainsi sur les seuils imposant la présence du type d'occupation du sol en question. Le graphique portant sur l'influence de la vigne dans le voisinage souligne également la divergence de comportement au regard de l'occupation agricole en générale.

Contrairement à l'occupation du sol agricole, un certain nombre de règles (55 règles) imposent l'absence de vignes dans le voisinage adjacent (Table 7. 14). Une règle impose même un pourcentage supérieur à 87% à 100m, en d'autres termes d'être presque entièrement entouré de vignes. La restriction imposant l'absence de vignes est également présente à 200m, néanmoins dans une moindre mesure (3 règles). La vigne est recherchée dans 40 règles à hauteur de 4% à 200 m. Ce seuil est généralement associé à un seuil limitant de 36% à 600 m. La tendance 3 présente quelques règles qui imposent la présence de vignes à 200m, 400m, 600m, et 700m. Le nombre de règles qui présentent ces seuils fait référence à un encadrement de 10 à 20 règles. En effet, ces seuils n'apparaissent pas nécessairement au sein des mêmes règles. Pour autant, chacun d'entre eux est présent dans un échantillon d'au moins 10 règles, portant le caractère incitatif de la vigne aux distances spécifiées. Par exemple, la présence de vigne nécessaire à 0,2 km est présente dans un certain nombre de règles pouvant aller de 10 à 20. Pour finir, 8 règles imposent l'absence de vignes à 1,1km.

Table 7. 14 : Distribution des vignes dans le voisinage – tendances majeures à l'échelle de la zone

Tendances	0,1 km	0,2 km	0,3 km	0,4 km	0,6 km	0,7 km	1,1 km	1,2 km	1,6 km	1,7 km	Nombre de règles
1	≤ 0					≤ 19				≤ 72	55
2		≤ 4			≤ 36						40
3		0 <		0 <	0 <	0 <			0 <		entre 10 et 20
4							> 27				8
5								> 8 ou > 20			
6										5 < x ≤ 72	6
7										72 <	5

Zones récréatives urbaines

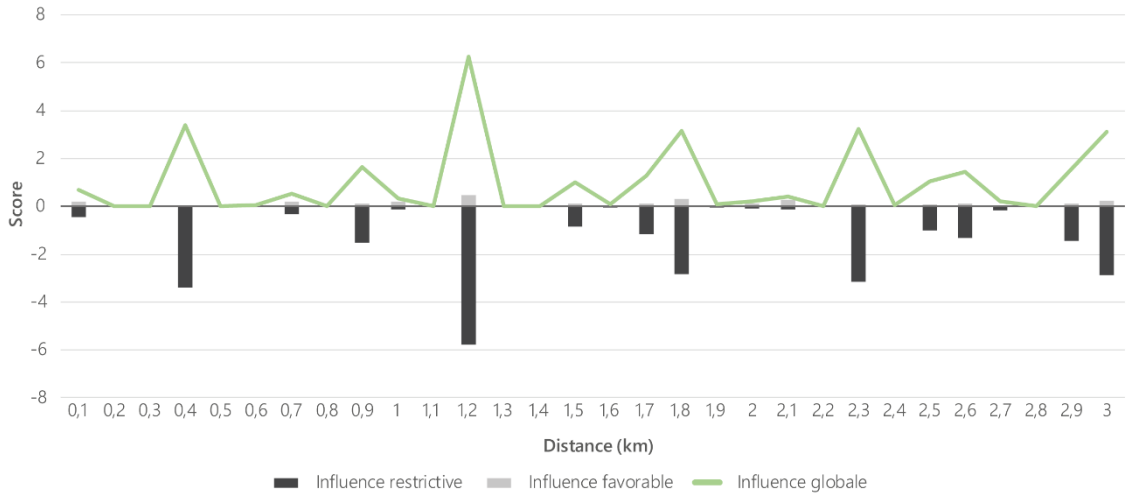


Figure 7. 15 : Polarité de l'influence du type zone récréative urbaine d'occupation du sol

La distribution de l'influence des zones récréatives urbaines est légèrement plus variée que celle de la vigne. Son influence positive s'observe dans un voisinage proche (100m) ou aux alentours du kilomètre. Le point intéressant est l'influence positive retenue jusqu'à 2 km. Dans les faits, la présence de zones récréatives urbaines a donc une certaine importance, même à plus de 2 km.

Plus précisément, les zones récréatives urbaines jouent un rôle dans 325 règles des 395 règles composant le jeu. Une seule règle impose la présence de ce type d'occupation du sol à 100m, contre 4 qui n'interdisent pas strictement leur présence avec un seuil inférieur ou égal à 1 (Table 7. 15). La colonne en question n'a donc pas été reportée dans la table ci-dessous. La plupart des règles faisant intervenir les zones récréatives imposent leur absence à 400m (tendance 1). Ce point peut paraître surprenant puisque 400m représente une distance accessible notamment à pied donc adéquate. Dans ce cas de figure, l'arbre de décision peut également servir à étudier les tendances de construction au regard de la structure préexistante de la ville. Ainsi, les espaces verts et stades incorporés ponctuellement

dans le tissu urbain ne sont pas accessibles dans un certain périmètre des nouvelles zones urbanisées. Ce constat peut soutenir une réflexion sur ce point, à savoir la définition de stratégies dans le cas où la situation en question ne remplirait pas les objectifs en termes d'aménagements. En effet, les tendances suivantes soulignent néanmoins la présence de ces zones dans un périmètre plus large. Ainsi, à 1,2km, la tendance 2 impose une présence supérieure à 14% soit plus de 3 ha dans un voisinage de 24 ha. Ce type d'occupation du sol est donc recherché. C'est également le cas pour les tendances suivantes, dans un périmètre plus large.

Table 7. 15 : Distribution des zones récréatives urbaines dans le voisinage – tendances majeures à l'échelle de la zone

Tendances	0,4 km	1,2 km	1,8 km	2,1 km	2,3 km	3 km	Nombre de règles
1	≤ 0					≤ 3	125
2		14 <					4
3			2 <				3
4				0 <			5
5					0 <		2
6						0 <	5

À travers l'étude des tendances majeures dans la distribution des zones récréatives dans le voisinage, apparaît un second point positif de l'application de notre méthodologie, soit l'analyse détaillée des règles créer par le biais de l'arbre de décision. Ces dernières permettent d'observer d'éventuelles lacunes à travers la description du processus tel qu'il s'opère d'après les données observées. Nous pouvons ici nous questionner sur la présence en nombre suffisant des zones récréatives urbaines limitant l'accessibilité dans un périmètre proche, pour les nouvelles zones urbanisées.

Zones immergées

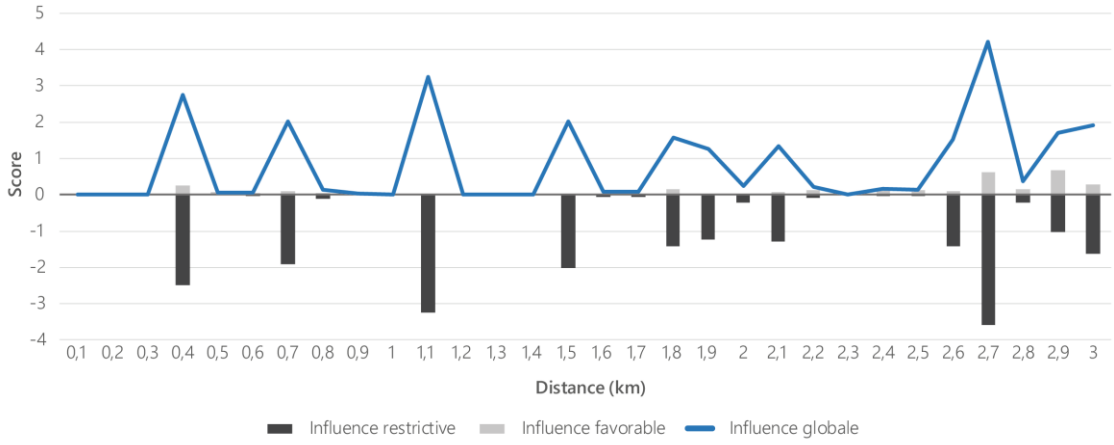


Figure 7. 16 : Polarité de l'influence du type zone immergée d'occupation du sol

L'intérêt des zones immergées avait été abordé précédemment suivant deux angles d'approche : environnement et frontière. En effet, la majeure partie des zones immergées sont représentées par le lit du Rhin, soit la délimitation spatiale de la frontière étatique. L'influence observée à 400m peut être mise en parallèle avec celle observée précédemment dans le cas des zones récréatives urbaines. Ainsi la caractéristique environnement associée à ce type d'occupation du sol peut être soulignée. Dans un second temps, le graphique précédent (Figure 7. 16) localise l'influence favorable dans un voisinage plus éloigné. Toujours accessible dans le cadre de l'utilisation d'un mode doux de déplacement (vélo, etc.) l'accès est néanmoins limité. L'influence des zones immergées dans ce cadre pourrait alors également être associée à la une proximité de la frontière présentant d'autres avantages que les avantages environnementaux.

Les zones immergées interviennent dans 173 règles des 395 règles composant le jeu. Parmi elles, 35 règles imposent la présence d'eau dans le voisinage d'une cellule urbanisée (Table 7. 16). Ce nombre est extrait des pics sélectionnés seuls donc ne peut qu'être sous-estimé par rapport à la considération de toutes les variables. La contrainte principale est celle d'une limitation maximale à 11% dans un périmètre de 400m (39 règles). Certaines règles qui imposent l'absence à 400m, imposent également la présence d'un pourcentage minimum de zones immergées à 3km. Les zones immergées revêtent de l'importance dans un voisinage plus éloigner (à partir de 2,7 km principalement). Pour autant, quelques règles imposent également la présence d'eau à différentes distances dans le voisinage (tendance 2).

Table 7. 16 : Distribution des zones immergées dans le voisinage – tendances majeures à l'échelle de la zone

Tendances	0,4 km	1,8 km	2,7 km	2,9 km	3 km	Nombre de règles
1	≤ 11		≤ 4			39
2	0 <	0 <			0 <	entre 2 et 6
3			4 <			10
4				1 <		16
5			4 <	≤ 13	10 <	2

Zones industrielles et commerciales

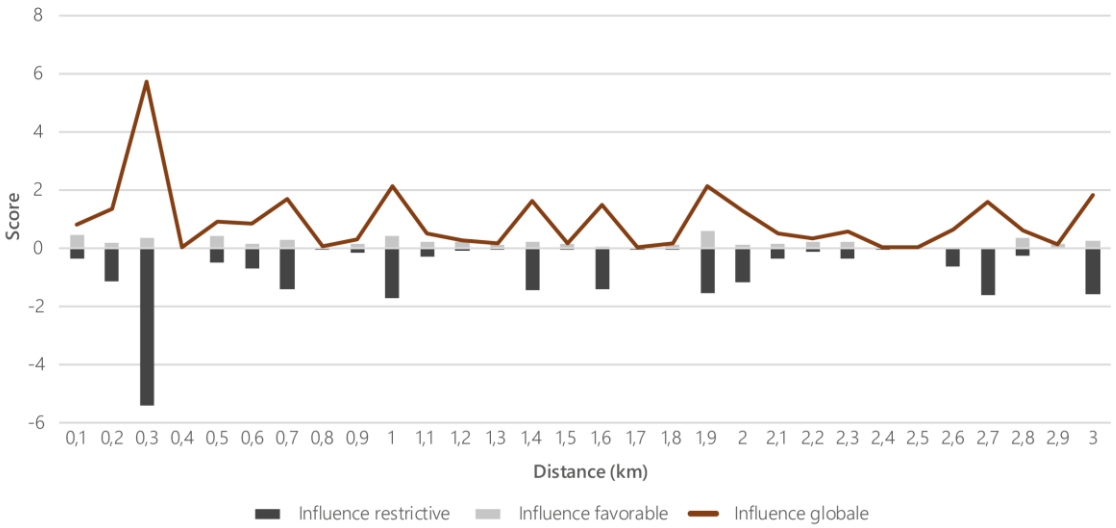


Figure 7. 17 : Polarité de l'influence du type industrie et commerce d'occupation du sol

L'influence positive des zones industrielles et commerciales est relativement homogène le long du transect. Par contre, l'influence restrictive apparaît majoritairement dans un voisinage proche (Table 7. 17). Cette occupation du sol rentre également dans la catégorie de celles influençant dans une moindre mesure la construction de l'arbre (Table 7. 5 : moins de 10%). Les tendances extraites de l'analyse du jeu de règles sont donc plus nombreuses tout en concernant moins de règles (Table 7. 17).

Les zones industrielles et commerciales interviennent dans 256 règles des 395 règles composant le jeu. Les seuils sont pour le moins intéressant à plusieurs niveaux. Ainsi, dans un voisinage proche, quelques règles relatent une proportion particulièrement élevée de zones industrielles et commerciales (plus de 50%). Ces règles pourraient être associées à celles de la reconversion des anciens bâtiments industriels. Si la majorité des règles s'avèrent limitante, certaines imposent néanmoins leur présence dans un certain périmètre. Notamment les tendances 5, 10 ou 12, respectivement à 300m, 500m, et 2,8km. Dans le cas de la tendance 12, une proportion minimum relativement importante est imposée, puisque que cette dernière doit soit être supérieure à 5% soit supérieure à 160ha. Les limites restrictives présentent également un intérêt puisque qu'elles traduisent une certaine tolérance, par exemple, moins de 1% à 500m, ou moins de 10% à 1km.

Table 7. 17 : Distribution des zones industrielles et commerciales dans le voisinage – tendances majeures à l'échelle de la zone

Tendances	0,1 km	0,2 km	0,3 km	0,5 km	1 km	1,9 km	2,8 km	Nombre de règles
1		≤ 0	≤ 16					46
2			≤ 4			≤ 9		46
3			≤ 2		≤ 10			41
4			≤ 8					28
5			8 <					8
6					≤ 2			8
7				≤ 1				6
8							≤ 1	6
9	≤ 7							5
10				0 <				5
11						≤ 0		5
12							5 <	5
13	50 <			≤ 38				4

Transport

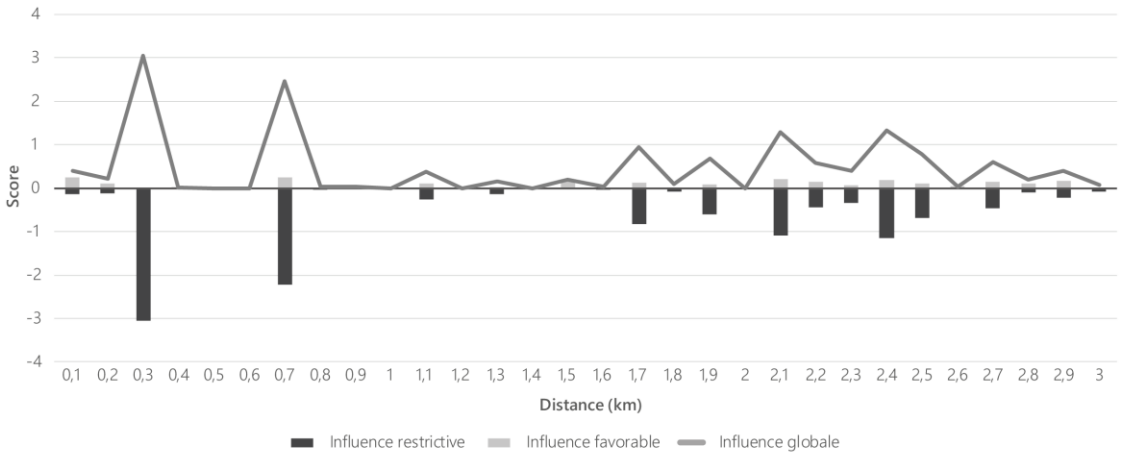


Figure 7. 18 : Polarité de l'influence du type transport d'occupation du sol

L'influence positive des zones de transports s'observe aux extrémités du voisinage : dans un périmètre proche, inférieur à 300m, puis dans un périmètre éloigné, supérieur à 1,5 km. Comme pour la majorité des occupations du sol secondaire, c'est principalement des pics d'influence restrictive qui sont observables sur la Figure 7. 18.

Les transports quant à eux interviennent dans 174 règles des 395 composant le jeu dans sa totalité. Leur profil suit la tendance de celui des zones industrielles et commerciales. L'influence positive est cependant légèrement plus concentrée dans le voisinage proche. Si les seuils restent majoritairement limitant, cette limite n'est que très rarement prohibitive (Table 7. 18). Seule la tendance 3 présente une restriction interdisant la présence de transport à 700m. La tendance 1 précise, comme précédemment, la localisation des contraintes

concernant le transport. Ces dernières apparaissent à différents niveaux sur le transect caractérisant le voisinage (100m, 200m, 700m, 1,5km, et 2,1km). Elles apparaissent cependant dans un nombre limité de règles chacune (entre 2 et 6 règles). L'interdiction ou la présence restreinte peut s'expliquer par la volonté de s'éloigner ou de limiter la présence de ce type d'occupation du sol pour les nuisances qui lui sont associées (notamment le bruit). Pour autant, sa présence dans une certaine proportion notamment associée à la taille de l'infrastructure peut présenter un avantage au niveau de l'accessibilité ou des déplacements. Ce point explique les conditions d'accessibilité des transports dans le voisinage.

Table 7. 18 : Distribution des transports dans le voisinage – tendances majeures à l'échelle de la zone

Tendances	0,1 km	0,2 km	0,3 km	0,7 km	1,5 km	2,1 km	2,4 km	Nombre de règles
1	0 <	0 <		0 <	0 <	0 <		entre 2 et 6
2			≤ 2					93
3				≤ 0		≤ 6		46
4							≤ 0 ou ≤ 6	24
5				≤ 5			≤ 6	20
6			≤ 2			≤ 6		5
7				5 <		≤ 4		4

La méthode appliquée permet donc d'évaluer les tendances majeures dans la composition du voisinage des cellules urbanisées, par type d'occupation du sol. Des seuils et proportions sont extraites et illustrent le cadre favorable au processus de développement urbain sur la zone d'étude.

En résumé, les occupations du sol principales présentent des limites incitatives quant à la présence de leur type d'occupation du sol dans le voisinage des cellules urbanisées. L'occupation du sol urbaine est particulièrement retrouvée dans un voisinage proche, et présente une tendance incitative décroissante le long du transect. Certaines règles favorisent néanmoins un environnement moins urbain, particulièrement dans un voisinage proche. Si la forêt n'est pas nécessaire dans le voisinage très proche, elle est généralement recherchée et nécessaire un peu plus loin. Dans l'ensemble, l'agricole est, lui, recherché dans le voisinage des cellules urbanisées, ceux quelques soit la distance étudiée au sein du transect. Dans la majorité des tendances, les limites supérieures sont absentes ou hautes, laissant beaucoup de flexibilité au regard de la présence de ce type de sol. Dans un voisinage proche, quelques règles relatent une proportion particulièrement élevée de zones industrielles et commerciales (plus de 50%). Leur influence positive est également observée à 2,8 km. Les zones immergées sont, elles, recherchées dans un voisinage proche. Elles sont également observées en limite de périmètre (3 km). Ce point soulève le questionnement thématique concernant la localisation au regard de la frontière d'État. En effet, si la localisation de proximité peut être associée à celle de la frontière et au cadre environnemental, une présence plus éloignée peut également définir l'intérêt de la proximité de la frontière donc un accès privilégié au pays voisin.

D'un point de vue analytique, cette démarche permet également d'étudier les tendances de construction au regard de la structure préexistante de la ville. Ainsi, les espaces verts et stades incorporés ponctuellement dans le tissu urbain ne sont pas accessibles dans un certain périmètre des nouvelles zones urbanisées. Ce constat peut soutenir une réflexion sur ce point, à savoir la définition de stratégies dans le cas où la situation en question ne remplirait pas les objectifs en termes d'aménagements. L'observation d'éventuelles lacunes à travers la description du processus tel qu'il s'opère d'après les données observées peut permettre de questionner la présence en nombre suffisant des zones récréatives urbaines favorisant l'accessibilité dans un périmètre proche des nouvelles zones urbanisées.

L'analyse des règles issues du processus de décision de l'arbre permet donc de remplir un des objectifs majeurs, à savoir interpréter thématiquement les résultats produits par apprentissage automatique, d'après des données observées disponibles à plusieurs dates. L'analyse des données d'occupation du sol est suivie de celles portant sur l'accessibilité sur la zone d'étude.

3. ANALYSE DU JEU DE RÈGLES : ACCESSIBILITÉ

La section précédente a introduit le score des variables d'occupation du sol qui permettent de décrire la composition du voisinage favorable à la transition urbaine. Le score permet également de hiérarchiser l'influence incitative ou restrictive de chaque variable. Le chapitre précédent a permis de mettre en lumière l'apport des variables de localisation géographique dans la performance des modèles issus de l'apprentissage. L'analyse de leur influence dans le jeu de règles est donc présentée dans la section qui suit d'après la même démarche que celle utilisée dans la section précédente.

3.1. Les variables de localisation géographique

Comme dans le cas des différents types d'occupation du sol, la structure et la composition du jeu de règles sont analysées à travers l'attribution d'un score aux différentes variables de localisation géographiques mobilisées lors du processus de décision (Table 7.19).

Table 7. 19 : Scores des variables de localisation géographique

Nombre de règles	Variables	Scores	Sans doublons
176	Accessibilité pondérée niveau 1	7,41	6,80
166	IDP	5,05	x
117	Accessibilité niveau 2	6,10	5,12
45	Accessibilité niveau 1	1,88	1,70
44	Accessibilité pondérée niveau 2	2,46	x
27	Nord - Sud	0,97	x

Avant de discuter et d'interpréter le score des variables de localisation, il est important d'évoquer le phénomène de doublons pouvant affecter l'influence des variables d'accessibilité au regard de celles qui précise la localisation au sein de la zone d'étude. En effet, les variables d'accessibilité peuvent être sélectionnées à plusieurs reprise lors de la création de l'arbre puisqu'elles ne sont pas binaires. Au contraire, les variables Indice de Pays et Nord-Sud ne peuvent plus s'avérer discriminantes après une sélection puisque la division du jeu formera des classes pures au regard de l'une ou l'autre des valeurs prises par la variable binaire. Par exemple, la sélection de l'Indice de Pays restreint l'application des règles à l'un ou l'autre des pays. Les règles en question ne peuvent donc pas redevenir transnationale ou changer d'appartenance dans la suite du processus. Ainsi une règle française ne pourra pas devenir allemande par la suite. Ce cas de figure n'apparaît pas pour les types de variables numériques continues qui représentent les différentes accessibilités. La même variable d'accessibilité pourra être sélectionnée plusieurs fois au cours de la construction de l'arbre. Une première sélection restreindra le périmètre d'accessibilité. Une deuxième l'affinera davantage. Par exemple, l'accessibilité au pôle de niveau 1 qu'est Strasbourg pourra dans un premier temps être définie comme inférieure à 30 min pour favoriser le développement urbain, puis, dans un second temps, comme supérieur à 20 min. Ainsi, les règles de développement caractériseront un environnement type périurbain, ou la banlieue proche du pôle principal de la zone d'étude, et seront caractéristiques

d'urbanisation d'une zone particulière. L'interprétation des scores attribués aux variables de localisation géographique peut être mise en relief d'après ces exemples concrets. Ce phénomène se retrouve à 5 reprises pour l'accessibilité au pôle de niveau 1, à 25 reprises pour l'accessibilité au pôle de niveau 2, à 21 reprises pour l'accessibilité pondérée au pôle de niveau 1 et ne se retrouve pas pour l'accessibilité pondérée au pôle de niveau 2. Un score sans doublons est calculé pour relativiser l'influence de l'accessibilité en prenant en compte les potentielles multi-sélections (Table 7.19).

Les scores présents dans la table ci-dessus reflètent dans tous les cas une influence majeure de l'accessibilité aux différents pôles. L'attribution du score le plus haut à l'accessibilité pondérée du pôle de niveau 1 est particulièrement intéressante. En effet, cette variable traduit indéniablement l'importance du pôle Strasbourgeois tout en mettant en lumière celle des pôles de niveau 2 dont la plupart, rappelons-le, sont situé en Allemagne. Ainsi, cette variable peut en partie représenter spatialement la préférence pour une localisation qui permet l'exploitation des deux côtés de la frontière. Cette analyse souligne donc une certaine composante transfrontalière des règles s'appuyant sur cette variable. La frontière étatique s'efface peut-être au profit d'une exploitation des ressources transfrontalières. Pour autant, une seconde tendance est observée à travers le score attribué à l'Indice de Pays. Comme expliqué précédemment, sa sélection localise la règle dans l'un ou l'autre des pays constituant la zone d'étude. La règle reflète alors le caractère national de la dynamique étudiée et inscrit pour sa part la frontière comme une limite séparant deux dynamiques dont les définitions varient en fonction du pays.

La variable Nord-Sud est quant à elle mobilisée épisodiquement. Son faible score traduit sa très faible influence dans la caractérisation du développement urbain sur la zone. Ce point souligne d'autant plus l'importance de l'Indice de Pays puisque les deux variables ont été créées sur le même motif. Les sections Nord et Sud au même titre que les sections Est et Ouest séparées par la frontière séparent des espaces de surfaces équivalentes qui présentent également un nombre similaire de cellules urbanisées. Les variables d'accessibilité niveau 1 et pondérée niveau 2 sont, quant à elles, mobilisées plus épisodiquement que les deux autres variables d'accessibilité.

L'influence des variables de localisation géographique s'analyse également à travers le nombre de règles auxquelles elles participent. Notons que l'Indice de Pays est mobilisé dans 176 d'entre-elles, soit dans 45% des cas. Ce nombre révèle la part importante du processus d'urbanisation localisé dans l'un ou l'autre des pays frontaliers. Cette part n'est pas transfrontalière. Réciproquement, ce résultat atteste également de la non présence de l'Indice de Pays dans la majorité des règles du jeu, à savoir 55% ou 219 règles.
En conclusion, l'analyse des scores et occurrences des variables de localisation géographique établit plus clairement l'influence de la frontière dans le processus d'urbanisation sur la zone frontalière de Strasbourg-Kehl. Son influence diffère au sein du jeu de règles. Dans un premier cas, l'accessibilité pondérée au pôle de niveau 1 temporise la perception du rôle de la frontière comme barrière limitante, puisqu'elle tend à localiser le

développement dans une zone intermédiaire entre Strasbourg-Kehl et les pôles de niveau 2, dont la majorité sont situés en Allemagne. Elle souligne l'influence du caractère transfrontalier inhérent à la pondération effectuée pour créer cette variable d'accessibilité (Chapitre 3).

L'accessibilité pondérée pôle de niveau 1 est mise à contribution dans la majorité des règles (176). Pour autant, dans un second cas de figure, l'indice de pays inscrit précisément le processus de développement urbain dans l'un ou l'autre des pays, donnant à la frontière un rôle de limite. Cette approche est donc particulièrement intéressante pour « mesurer » l'intensité des parts transfrontalière et nationale de l'urbanisation sur la zone d'étude. Dans une vision plus large, elle pourrait servir à caractériser la porosité de la frontière sur la zone d'étude, ce à travers le prisme des données spatiales illustrant le développement urbain sur la période 1990-2006.

3.2. L'accessibilités

Si chaque règle qui mobilise des variables de localisation géographique peut littéralement décrire les caractéristiques favorables à l'urbanisation, l'objectif est bien de synthétiser les tendances générales qu'elles expriment. Nous nous intéressons dans un premier temps aux caractéristiques d'accessibilité sur la zone d'étude, indépendamment de la localisation dans l'un ou l'autre des États. Nous analyserons le lien existant avec la localisation dans un pays ou dans un autre dans un second temps (chapitre 8).

L'accessibilité pondérée au niveau 1 est la variable de localisation géographique la plus influente dans la structure de l'arbre, suivit de celle caractérisant l'accessibilité aux niveaux 2 (Table 7. 20). Les variables d'accessibilité présentent globalement un caractère limitant. Le temps de trajet ne doit pas dépasser un certain seuil dans 68% des cas, contre 32% présentant un temps minimum de trajet pour rejoindre l'un ou l'autre type de pôle de niveau.

Table 7. 20 : Polarité de l'influence des différentes variables d'accessibilité

	Influence totale (%)	≤	>
Accessibilité Niveau 1 pondérée	41,53	31,06	10,47
Accessibilité Niveau 2	34,19	23,34	10,85
Accessibilité Niveau 2 pondérée	13,77	6,37	7,40
Accessibilité Niveau 1	10,52	6,93	3,58
Total	100,00	67,70	32,30

Comme dans le cas de l'occupation du sol, l'intérêt de l'analyse réside de l'extraction des seuils qui servent de limites, de bornes, à la définition d'une zone favorable pour l'urbanisation, cette fois, d'après l'accessibilité aux différents types de pôles urbains.

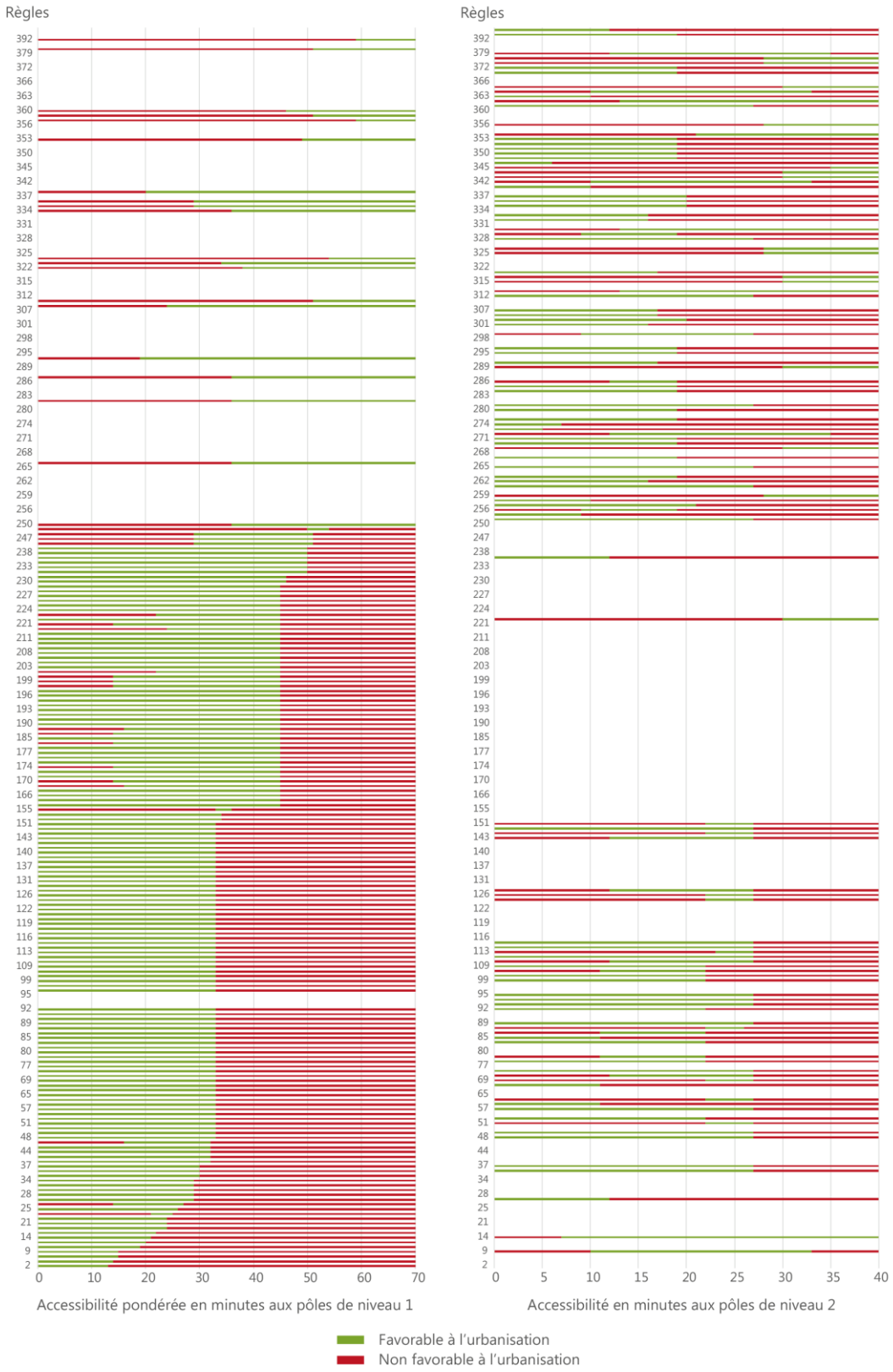


Figure 7. 19 : Contraintes liées à l'accessibilité pondérée pôle de niveau 1 et celles au pôle de niveau2

Figure 7. 20 : Contraintes liées à l'accessibilité pondérée pôle de niveau 2 et celles au pôle de niveau 1

L'extraction des seuils limites mobilisés dans la construction des règles qui s'appuient sur les variables d'accessibilité sont extraits pour chaque règle (Figure 7. 19 et Figure 7. 20). L'ordre des graphiques dépend de l'importance de la variable d'accessibilité en fonction de son score. Ainsi, le premier graphique présente les seuils retenus lors de la mobilisation de la variable d'accessibilité pondérée niveau 1, le second, ceux retenus pour l'accessibilité aux pôles de niveau 2, etc. Les seuils mobilisés pour l'accessibilité pondérée au pôle de niveau 1 ont été utilisés pour classer les règles. L'absence de mobilisation de certaines variables dans les règles est caractérisée par l'absence de données observables sur le graphique. Par exemple, la règle 2 s'appuie uniquement sur l'accessibilité pondérée niveau 1 alors que la règle 9 mobilise à la fois l'accessibilité pondérée niveau 1, mais aussi l'accessibilité niveau 2. Cette même règle n'est cependant pas dépendante des deux autres types d'accessibilité. Ces graphiques permettent donc une vision globale de l'utilisation de chaque variable d'accessibilité, au regard des règles basées sur l'accessibilité pour définir la transition urbaine. Ils permettent également une vision transversale de la mobilisation des différentes variables dans une même règle. En effet, d'après la Figure 7. 19, une majorité des règles influencées par l'accessibilité aux pôles sont contraintes par l'accessibilité pondérée au pôle de niveau 1.

L'accessibilité pondérée niveau 1 est la plus influente des variables d'accessibilité. La moitié d'entre-elles présentent une seule limite séparant une zone favorable d'une zone non-favorable à l'urbanisation. La proximité du pôle favorise donc l'urbanisation. Un certain temps de trajet rend la zone peu propice au développement urbain au regard de l'analyse de la localisation des cellules urbanisées entre 1990 et 2006. Au contraire, 20 règles décrivent un caractère opposé, puisque c'est dans un premier temps une zone défavorable à l'urbanisation, « trop près » des pôles, puis une zone favorable qui se dessine à partir d'une certaine distance. La proximité de l'un ou l'autre des types de pôle joue alors un rôle répulsif. Finalement, 21 règles présentent 2 seuils. Ce comportement est particulièrement intéressant puisqu'il encadre la zone d'urbanisation souhaitable, qui doit s'éloigner des villes majeures, tout en restant à l'intérieur d'un certain périmètre d'accessibilité. Cette zone est donc délimitée spatialement par un temps de trajet minimum, qui ne doit néanmoins pas dépasser un temps maximum considéré comme trop éloigné. Finalement, 176 règles ne s'appuient pas sur la variable d'accessibilité pondérée niveau 1. La Figure 7. 19 permet d'établir que la majeure partie du jeu de règles qui ne s'appuie pas sur l'accessibilité pondérée au pôle de niveau 1, implique l'accessibilité aux pôles de niveau 2. L'accessibilité aux pôles de niveau 2 est également parfois encadrée (25 règles). 72 règles présentent une zone favorable proche des pôles puis défavorable. 20 règles localisent au contraire une zone proche défavorable puis une zone plus éloignée favorable.

Les graphiques mis en parallèle Figure 7. 20 peuvent être décrits suivant le même protocole. L'intérêt de cette analyse, au-delà de la définition spatiale de zones favorables à l'urbanisation en fonction de l'accessibilité des cellules urbanisées entre 1990 et 2006, est la mise en parallèle des comportements traduits par l'exploitation de différentes variables d'accessibilité dans une même règle. Ainsi la Figure 7. 19 et la Figure 7. 20 permettent de

constater qu'une partie des règles qui incluent la variable d'accessibilité pondérée au pôle de niveau 1 semblent voir leur l'intérêt pour les pôles de niveau 2 renforcé par d'autres variables d'accessibilité qui favorisent, elles, leur proximité (par exemple règle 113). Au contraire certaines règles appuient l'influence du pôle de niveau 1 par la mobilisation de l'accessibilité au pôle de niveau 1 en parallèle de l'accessibilité pondérée au pôle de niveau 1. Une analyse peut être menée aussi bien à l'échelle des règles, qu'à l'échelle d'une variable spécifique ou d'un enchaînement de variables, qui précise une caractéristique spécifique de localisation (distance optimale de localisation vis-à-vis de l'un ou l'autre de pôles de niveau).

Les graphiques projetant les contraintes d'urbanisation en fonction de l'accessibilité permettent d'établir une première analyse des liens existants entre ces différentes variables. Néanmoins, comme dans le cas des occupations du sol, l'intérêt et de distinguer plus précisément les seuils majeurs des différentes contraintes au sein d'un même groupe de règles, et de décrire des tendances générales plus spécifiques. L'information de la mobilisation des différentes variables d'accessibilités est résumée pour décrire des tendances majeures présentes dans un certain nombre de règles (minimum une vingtaine). Plusieurs tendances majeures sont donc distinguées (Table 7. 21).

Table 7. 21 : Tendances d'accessibilités observées au sein du jeu de règles dans sa totalité

Tendances	Accessibilité pondérée Niveau 1	Accessibilité Niveau 2	Accessibilité pondérée Niveau 2	Accessibilité Niveau 1
1	≤ 29			
2	≤ 33	22 < x ≤ 27		≤ 41
3	≤ 33	12 < x ≤ 27		41 <
4	≤ 45			
5	14< x ≤ 45		≤ 27	
6	≤ 50			
7	29 < x ≤ 51			
8		≤ 21		
9		≤ 33		
10		12 < x ≤ 33		

La synthèse des seuils d'accessibilité privilégiés pour définir les zones favorables à l'urbanisation est présentée dans la Table 7. 21. Les seuils observés présentent majoritairement une accessibilité maximum de 30 min. Comme dans le cas de l'occupation du sol, un point intéressant est la définition d'intervalles. Par exemple, dans le cas de la tendance 5, l'accessibilité pondérée doit à la foi être supérieure à 14 min tout en restant inférieure à 45 min. Dans ce cas, une certaine distance d'éloignement des pôles est recherchée à travers un temps de trajet minimum, tout en ne dépassant pas la limite de 50 min d'accessibilité à un pôle majeur (indépendamment de son niveau).

En toute logique, la plupart des comportements sont contraints d'après l'accessibilité pondérée au pôle de niveau 1, puisque cette variable influence majoritairement la construction de l'arbre. Les tendances 2 et 3 montrent néanmoins une certaine importance

de la localisation au regard des pôles de niveau 2 en dépit du fait, qu'intrinsèquement, la variable d'accessibilité pondérée pôle de niveau 1 donne initialement plus de poids à Strasbourg-Kehl. Ces tendances montrent un intervalle favorable de 12 à 27 minutes d'accessibilité au pôle de niveau 2. La tendance 2 précise néanmoins que le trajet pour rejoindre le pôle strasbourgeois ne doit pas excéder 41 min. Un tiers des tendances sont caractérisées par l'accessibilité au pôle de niveau 2 uniquement. La tendance 10 présente, elle, un intervalle éloignant légèrement la localisation optimale à 12 min du centre d'un pôle de niveau 2.

La méthodologie mise en place sur la base de l'utilisation d'arbres de décision permet d'extraire les seuils d'accessibilités optimaux en termes de localisation géographique, au regard des différents niveaux hiérarchiques des pôles urbains de la zone d'étude. Certaines tendances font appel à une accessibilité influencée dans sa définition par des localisations de part et d'autre de la frontière (accessibilité pondérée). D'autres tendances ne sont dépendantes que de la localisation à un pôle de niveau spécifique. Elles ne tiennent donc pas directement compte d'une préférence pour une localisation permettant d'atteindre des pôles urbains situés de part et d'autre de la frontière. Ces tendances pourraient, par ailleurs, être associées à des comportements en lien avec le type d'habitat recherché, lui-même directement lié au cycle de vie ou encore au niveau de vie des individus (tendance 5). La sous-section suivante présente la projection spatialisée des zones favorables à l'urbanisation en fonction de leur localisation géographique dans l'espace de la zone d'étude.

3.3. Une cartographie

La synthèse des informations présentées dans la sous-section précédente permet de mettre en lumière les principales zones privilégiant l'urbanisation au regard de leur accessibilité aux pôles urbains. Ces zones sont donc plus favorables à l'urbanisation d'après leur temps de trajet pour rejoindre les différents pôles. Leur spatialisation permet d'élaborer une carte de zonages hiérarchisés décrivant le caractère favorable à l'urbanisation en fonction de la localisation au regard des pôles de différents niveaux.

La carte a été créée en attribuant un poids positif aux zones favorables à l'urbanisation, zones délimitées par l'extraction des seuils des différentes tendances présentées (Table 7. 21). Dans un premier temps, chaque zone favorable s'est vue attribuer un point positif. Pour exemple, dans le cadre de la tendance 1, les cellules qui présentent une accessibilité pondérée niveau 1 inférieure ou égale à 29 min se sont vues attribuer 1 point, celles qui ne répondent pas positivement à ce critère, 0. L'opération a été répétée pour chaque variable d'accessibilité de chaque tendance majeure (Table 7. 21). Dans un second temps, si la tendance présente un caractère favorable d'après différentes variables, les points sont additionnés. Seules les cellules qui obtiennent le maximum de points sont considérées comme favorables à l'urbanisation. Ainsi, la pondération est recalculée pour obtenir une description binaire des zones favorables ou défavorables à l'urbanisation, ce pour chaque tendance. En effet, l'enchaînement de variables dans le cadre des règles implique un

comportement logique qui ne permet pas à une cellule de changer d'état si elle ne répond pas à toutes les contraintes. Par exemple, la tendance 5 présente un premier zonage favorable entre 14 et 45 min d'après l'accessibilité pondérée au pôle de niveau. Elle présente également un zonage favorable pour les cellules localisées à moins de 27 min d'un pôle de niveau 2. Ainsi, une simple addition établirait d'ores et déjà une hiérarchie entre les cellules remplissant les deux conditions. Or, l'application d'une règle logique implique que seules les cellules répondant effectivement aux deux conditions sont considérées comme favorables. Un nouveau score a donc été réattribué, 1 pour les cellules remplissant les deux conditions et 0 pour les autres. La même démarche a été appliquée dans le cadre des tendances 2 et 3, pour les trois variables concernées.

Pour finir, le calcul de la somme des points obtenus pour chaque tendance permet de hiérarchiser le caractère favorable de la cellule en fonction de sa réponse positive aux seuils imposés par différentes tendances. Ainsi, une cellule répondant positivement à trois tendances sera considérée comme plus propice au développement urbain qu'une cellule ne répondant qu'à une tendance. La carte obtenue permet donc de hiérarchiser les cellules, suivant leur caractère favorable à l'urbanisation au regard de l'accessibilité des différents pôles (Figure 7. 21). Cette carte fait donc complètement abstraction de l'occupation du sol.

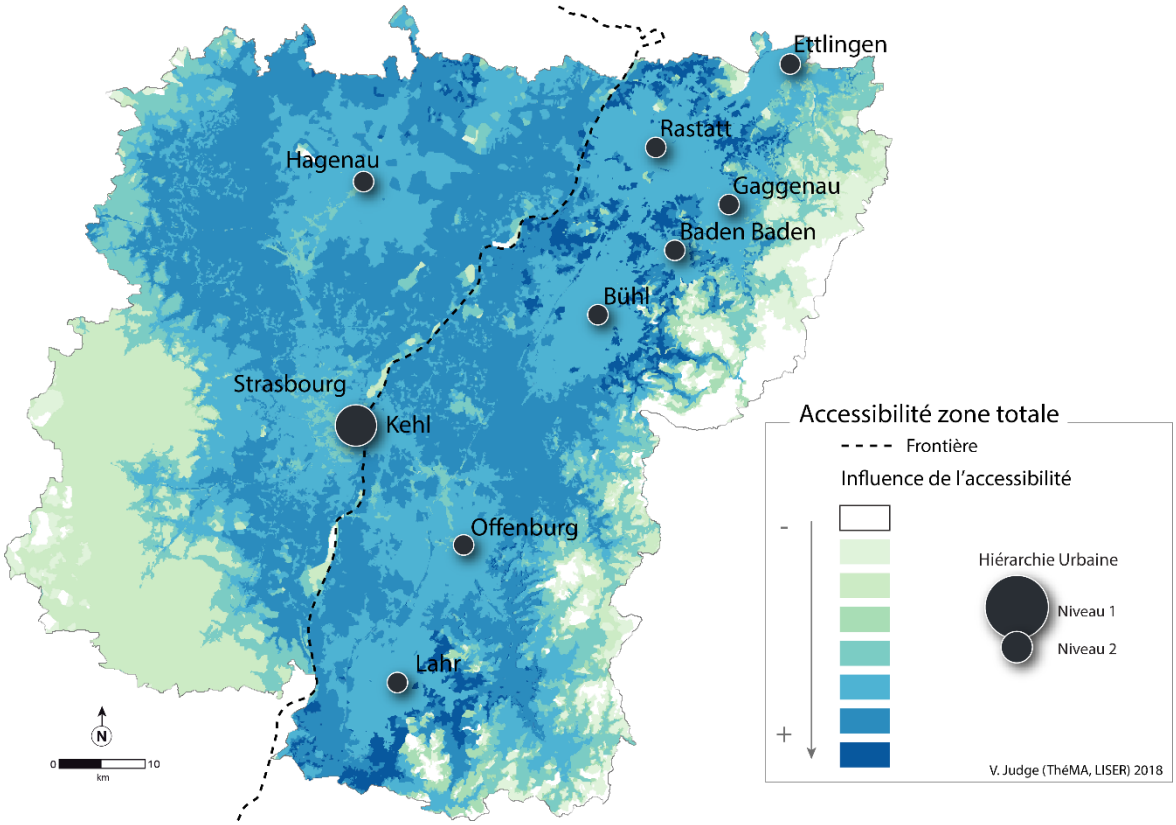


Figure 7. 21 : Zones préférentielles d'urbanisation en regard de l'accessibilité aux différents pôles urbains

La synthèse cartographiée des tendances d'accessibilités plus ou moins favorables au développement urbain doit être appréhendée comme une information à mettre en parallèle avec les tendances en matière d'occupation du sol dans le voisinage. L'accessibilité la moins favorable (blanc) est associée à un comportement prohibitif lors de l'application de règles faisant intervenir l'accessibilité sur la zone (soit 261 sur 395). Ainsi, la zone blanche contient les cellules ne répondant pas favorablement aux caractéristiques dégagées dans les tendances d'accessibilité observées. Pour autant, ces zones présentent peut-être les combinaisons favorables d'occupation du sol et seront donc urbanisées, mais en considérant l'environnement proche, non l'accessibilité aux pôles.

De l'influence des variables d'accessibilité, il ressort un caractère très favorable dans les zones intermédiaires entre les pôles de différents niveaux. Les zones périurbaines sont donc favorisées. Les pôles de niveau 2 présentent les zones les plus favorables dans leurs alentours. L'influence positive de l'accessibilité sur l'urbanisation en Allemagne est répartie de façon plus homogène que sur la partie française de la zone d'étude.

En résumé, cette carte définie hiérarchiquement les zones plus ou moins favorables à l'urbanisation, au regard des principaux seuils d'accessibilité extraits de l'analyse des règles de développement urbain, créés automatiquement par l'arbre de décision. Cette carte permet également de soutenir le choix de ne pas intégrer le réseau routier dans les données d'entrée puisque ce dernier est extrêmement corrélé à la donnée d'accessibilité, en plus de ne pas être adapté à l'échelle de la donnée utilisée (Corine Land Cover). La carte ci-dessus permet néanmoins de rendre compte de l'influence du réseau lors de l'observation des limites des différents zonages.

Conclusion du chapitre 7

Ce chapitre a permis de présenter une analyse du jeu de règles dans sa globalité. Les différents types de transitions urbaines, illustrées par différentes règles, ont été mis en évidences. Si la majeure partie des règles caractérise une urbanisation des cellules agricoles, le point à souligner ici est la définition des règles qui s'affranchissent de l'état initial de la cellule. Ce point est particulièrement intéressant et permet d'enrichir notre réflexion sur le caractère général des règles produites par l'arbre de décision (Chapitre 9).

Nous avons également hiérarchisé l'influence des variables mises à contribution dans la construction de l'arbre par l'apprentissage automatique. Dans un premier temps, nous avons procédé à l'analyse de l'influence des variables d'occupation qui permettent la description du voisinage favorable à l'urbanisation. La méthode s'avère concluante pour décrire précisément l'influence positive ou restrictive d'un type d'occupation du sol à une distance spécifique. Les types d'occupations du sol principales participent majoritairement à la composition d'un voisinage favorable : l'urbain dans un voisinage proche, l'agricole de manière plus homogène le long du transect de voisinage, la forêt dans un voisinage plus éloigné. Les types d'occupations secondaires présentent eux des pics d'influence plus sporadiques que les types principaux. L'influence des zones immergées est, par exemple, particulièrement intéressante en raison de sa forte corrélation spatiale avec la localisation de la frontière entre les deux pays de notre zone d'étude. Pour autant, ce sont les variables de localisation géographique qui permettent de mettre en lumière une composante localisée du développement urbain à travers l'influence de la variable Index de Pays, qui localise les règles de développement dans l'un ou l'autre des États. Cette variable présente la seconde influence la plus importante, après l'accessibilité pondérée au pôle de niveau, donc à Strasbourg – Kehl. Le rôle de la frontière et son influence dans la description du développement urbain localisé sur la zone d'étude est alors davantage précisé.

La sélection de l'Indice de Pays dans une partie des règles seulement induit également une seconde interprétation : certaines règles ne sont pas dépendantes de cette variable, donc favorisent le caractère transfrontalier d'une partie du jeu. Il est alors possible de caractériser la porosité de la frontière, à travers le prisme de la modélisation du développement, par la portion du jeu de règles s'affranchissant de la spécificité nationale induite par l'utilisation de l'Indice de Pays dans la construction des règles.

Puisque l'étude de l'influence des variables de localisation géographique, et plus particulièrement celle de l'Indice de Pays, met en évidence l'émergence de trois composantes distinctes du jeu de règles, le prochain chapitre s'attache, dès lors, à étudier plus en détail les caractéristiques des différents processus en fonction de leur localisation (France, Allemagne, Zone Totale). Le chapitre suivant présente donc une analyse des règles en fonction de leur nature : française, allemande ou transfrontalière.

Chapitre 8

Le développement urbain : Analyse à travers la frontière

L'analyse de l'influence des variables de localisation géographique dans le modèle automatiquement généré par apprentissage sur des données observées, a permis d'établir l'importance de la localisation géographique dans le processus de développement urbain (chapitre 7). Elle a mis en lumière l'influence de la variable Indice de Pays, qui localise spécifiquement le processus décrit par les règles dans l'un ou l'autre des pays de notre zone. Ce point souligne l'influence de la localisation en France ou en Allemagne, donc une influence de la frontière elle-même, non pas en tant qu'objet mais en tant que limite séparant des pratiques ou des attentes différentes. Cette analyse a également mis en évidence une composante transfrontalière du développement urbain, puisqu'une part des règles ne s'appuie pas sur la variable Indice de Pays. Ces règles temporisent cette fois le rôle de la frontière, qui s'estompe au profit de celui de l'accessibilité aux pôles d'activité.

Le présent chapitre s'attache donc à extraire et analyser les similitudes ou divergences qui existent dans les règles, ce en fonction de leur nature. Dans ce travail, la « nature » d'une règle est définie par son appartenance au groupe des règles françaises, allemandes ou transfrontalières, caractérisée par la mobilisation de l'Indice de Pays, ou par sa valeur. La première section permet d'établir plus précisément l'influence de la frontière à l'échelle du jeu de règles. La deuxième section présente la composition du voisinage d'une cellule urbanisée, étudiée en fonction de la nature des règles qui régissent le développement urbain, suivant la démarche établie dans le précédent chapitre. La troisième section permet, quant à elle, de présenter cette même analyse au regard, cette fois, de l'influence de l'accessibilité en fonction de la nature des règles analysées.

1. LA PART FRONTALIÈRE : EXPLOITATION DU JEU DE RÈGLES

L'analyse du processus d'urbanisation sur la zone frontalière franco-allemande a pour objectif de caractériser l'influence de la frontière à travers des données d'occupation du sol et de localisation géographique. L'influence relative de « l'Indice de Pays » (IDP), au regard des autres variables de localisation géographique, a été discuté précédemment (Chapitre 7 – Section 3.1). Cette analyse est cette fois accomplie en fonction de la nature des différentes règles. Un premier regard sur la composition du jeu rend possible une évaluation préalable de l'importance des processus nationaux et transfrontaliers. La mobilisation de la variable « Indice de Pays » (IDP) permet d'établir le nombre de règles nationales (françaises ou allemandes) ainsi que celui des règles transfrontalières (Table 8. 1).

Table 8. 1 Nature des règles

Nature des règles	Nombres de règles	Vrais positifs	Faux positifs
Françaises	82	311	1
Allemandes	85	515	11
Transfrontalières	228	1286	26
Total	395	2112	38

La part de règles transfrontalières est légèrement plus importante que celle occupée par les règles nationales puisqu'elle représente 228 règles contre 167 (82 règles françaises, 85 règles allemandes). Les proportions de règles françaises et allemandes sont néanmoins similaires. Ce premier élément souligne l'importance des composantes nationales, mais également la part du processus pouvant être exprimée en s'affranchissant de la localisation étatique. Autrement dit, une partie du jeu de règles est positivement influencée par de la spécification nationale alors qu'au contraire une autre partie s'en affranchit. Si la porosité de la frontière pouvait être mesurée d'après le jeu de règles, elle serait alors égale à l'influence des règles transfrontalières dans le jeu de données : 58%. L'influence de la frontière est au contraire représentée par la proportion de règles nationales, soit 42%. L'analyse de la nature des règles permet donc une approche quantitative de l'homogénéité de la zone d'étude. Elle montre également que le processus d'apprentissage est plus performant pour la partie française du jeu qui ne présente qu'un faux positif, mais reste néanmoins extrêmement performant pour l'ensemble des différentes parties du jeu, comme discuté Chapitre 6.

Il est important de rappeler que ce résultat est à mettre en perspective avec le choix de privilégier le nombre de règles pour caractériser l'importance des différentes natures de règles. Comme précédemment, l'analyse est effectuée en partant du postulat que chaque règle qui prédit correctement au moins une transition présente une efficacité égale. En effet, la qualité de la prédiction estimée à travers la règle est dépendante du pas de temps. Il est possible d'envisager qu'une règle ne concernant qu'un nombre restreint de cellules à un instant t, en concerne d'avantage ou au contraire beaucoup moins à un instant t+1. Nous estimons néanmoins que la logique issue de la règle est, elle, plus stable dans le temps. C'est donc cette dernière que nous analysons, indépendamment du nombre de vrais ou faux positifs concernés.

L'importance de l'influence des variables de localisation géographique en fonction de la nature des règles (Table 8. 2) varie suivant le caractère national ou transfrontalier des règles. Les règles nationales montrent des similitudes indépendamment dans l'influence de certaines variables d'accessibilité. Ainsi l'accessibilité pondérée niveau 1, ou encore l'accessibilité niveau 2, ont une influence plus ou moins similaire et représentent à elles deux plus de 70% de l'influence des variables de localisation géographique. Pour autant, le jeu de règles allemand privilégie l'intervention de l'accessibilité niveau 1 alors que celui des règles françaises privilégie lui, l'accessibilité pondérée niveau 2. Nous reviendrons sur ce point Section 2. L'influence de la variables Nord-Sud est, d'un point de vu relatif, plus importante dans les règles nationales, que dans celles qui s'affranchissent de la frontière. Elle reste néanmoins à la marge d'un point de vue global, avec un score très faible. Ce point soulignerait néanmoins une légère différenciation du processus Nord de celui prenant place au Sud, qui ne sera pas approfondi au vu de sa faible influence globale. Les règles transfrontalières présentent une variation moins importante de l'influence de chacune des variables d'accessibilité puisque les deux premières ont la même influence (35%). L'accessibilité pondérée niveau 2 et l'accessibilité niveau 1 révèlent, elles, une influence similaire autour de 12%. L'importance de la variable Nord-Sud est d'autant plus faible dans cette partie du jeu de règles.

Table 8. 2 Influence des variables de localisation géographique en fonction de la nature des règles

Influence des variables de localisation géographique suivant la nature des règles (Score)

	Règles françaises	Règles allemandes	Règles tranfrontalières	Somme
Accessibilité pondérée Niveau 1	1,23	1,13	5,05	7,41
Accessibilité Niveau 2	0,55	0,54	5,00	6,10
Accessibilité pondérée Niveau 2	0,30	0,20	1,96	2,46
Accessibilité Niveau 1	0,03	0,17	1,68	1,88
Nord - Sud	0,20	0,17	0,60	0,97
Somme	2,32	2,20	14,29	

Influence des variables de localisation géographique suivant la nature des règles (%)

	Règles françaises	Règles allemandes	Règles tranfrontalières
Accessibilité pondérée Niveau 1	52,94	51,12	35,37
Accessibilité Niveau 2	23,88	24,69	35,00
Accessibilité pondérée Niveau 2	13,05	8,99	13,68
Accessibilité Niveau 1	1,35	7,58	11,74
Nord - Sud	8,78	7,62	4,20

Comme précédemment, l'analyse de l'influence générale d'une variable, bien qu'apportant une première information sur son rôle au sein du jeu de règles, doit être mise en relation avec sa polarité et la valeur du seuil mobilisé pour prendre tout son sens. Ainsi, en dépit du fait que l'accessibilité niveau 1 apparaisse comme plus influente dans le jeu de règles allemand que dans celui composée des règles françaises, ce stade de l'analyse ne permet pas de définir si l'influence de la proximité de Strasbourg est « positive », privilégiant l'urbanisation dans un périmètre restreint, ou au contraire, si elle impose une certaine distance à travers le temps de trajet. La polarité de l'influence doit donc être analysée pour répondre à ces questions.

C'est l'aspect restrictif des variables de localisation géographique qui prédomine dans les règles au regard de la présence accrue des contraintes « inférieure ou égale » (Table 8. 3). Néanmoins, l'extraction brute des polarités permet de constater une fois de plus des tendances nationales similaires dans le ratio restrictif/incitatif, de l'ordre du 70 – 30 %. Ainsi l'influence des variables porte à 70% une polarité restrictive, donc imposent un temps de trajet maximum pour rejoindre le pôle qu'elles représentent. Cette polarité peut soutenir une certaine proximité. Les 30% restant présente une polarité incitative qui impose, elle, un temps de trajet minimum, donc un certain éloignement du pôle. Les polarités des règles transfrontalières tendent plus vers l'équilibre des deux polarités avec un ratio de 60 - 40% (Table 8. 3).

Table 8. 3 Polarité de l'influence des variables de localisation géographique en fonction de la nature des règles

	Règles françaises		Règles allemandes		Règles transfrontalières	
	<=	>	<=	>	<=	>
Nord - Sud	0,11	0,10	0,02	0,15	0,27	0,33
Accessibilité pondérée Niveau 1	1,02	0,21	0,77	0,36	0,77	0,36
Accessibilité pondérée Niveau 2	0,17	0,13	0,17	0,03	0,79	1,17
Accessibilité Niveau 1	0,00	0,03	0,07	0,09	1,16	0,51
Accessibilité Niveau 2	0,31	0,24	0,52	0,02	3,33	1,67
Somme Scores	1,61	0,71	1,55	0,65	6,33	4,04
Pourcentages	69,46	30,54	70,43	29,57	61,05	38,95

Quel que soit la nature des règles, une prédominance de la restriction est observée. Cependant, c'est l'extraction du seuil d'accessibilité qui permet de rendre réellement compte de la significativité de cette restriction. En effet, une restriction imposée, sur un seuil très haut, comporte de fait un caractère restrictif relativement restreint (Chapitre 7). L'influence de l'accessibilité à travers l'extraction des seuils est discutée spécifiquement dans la section 3, notamment à travers la cartographie des tendances majeures. Auparavant, la composition du voisinage est analysée en fonction de la nature des règles (section 2). L'intérêt de ce partitionnement du jeu est principalement de mettre en relief les similitudes et les divergences dans la description du voisinage ou de l'accessibilité favorable au processus d'urbanisation français, allemand, ou encore transfrontalier. L'intensité de l'influence d'un type de sol ou d'une variable d'accessibilité ainsi que le seuil de distance ou temps retenu peut, dès lors, être comparée entre les différents groupes de règles.

2. INFLUENCE CULTURELLE

Après avoir étudié l'influence des variables de localisation géographique, qui justifie la partition du jeu de règles d'après la mobilisation de l'Indice de Pays (IDP), le jeu de règles est donc subdivisé en trois parties : transnationale, française et allemande. La description des processus est spécifique en fonction de la nature des règles. Cette section s'intéresse plus particulièrement à la compostions du voisinage favorable à l'urbanisation.

Types de transition et composition du voisinage

La composition de chacune des parties du jeu est abordée indépendamment pour permettre la comparaison des voisinages et caractéristiques mobilisées dans les différentes situations.

Table 8. 4 Tendance d'occupation du sol en fonction des types de règles

Règles Françaises		
Types d'occupation du sol	Scores	Scores (%)
Urbain	19,42	23,68
Agricole	15,56	18,98
Forêt	12,78	15,58
Agricole protégé	9,18	11,19
Zone verte urbaine	7,17	8,74
Industrie et commerce	5,52	6,73
Eau	4,93	6,01
Transport	2,68	3,27

Règles Allemandes		
Types d'occupation du sol	Scores	Scores (%)
Urbain	18,42	22,19
Agricole	17,12	20,62
Forêt	13,80	16,63
Industrie et commerce	7,62	9,18
Eau	5,62	6,77
Agricole protégé	5,59	6,74
Zone verte urbaine	5,25	6,32
Transport	4,92	5,93

Règles transfrontalières		
Types d'occupation du sol	Scores	Scores (%)
Urbain	35,97	24,72
Agricole	23,47	16,13
Forêt	19,99	13,74
Agricole protégé	18,32	12,59
Zone verte urbaine	11,74	8,07
Eau	10,30	7,08
Industrie et commerce	9,77	6,71
Transport	3,27	2,25

Comme précédemment (Chapitre 7) nous nous appuyons sur la distinction entre les types d'occupations du sol principaux (urbain, agricole, forêt) et secondaires (vigne, zones récréatives urbaines, zones en eau, zones industrielles et commerciales, transport) pour décrire les compositions de voisinage. La hiérarchie établie grâce au score imputé à chaque occupation principale est similaire quel que soit la nature des règles. Ainsi aucune différence notable n'apparaît à ce stade. C'est au niveau des occupations du sol « secondaires » qu'elle est modifiée. Les règles transfrontalières et françaises privilégient la présence de vignes et de zones vertes urbaines. Au contraire les règles allemandes privilégient, elles, les zones industrielles et commerciales ou la présence du fleuve. Ce point est particulièrement intéressant puisqu'il rapproche le phénomène d'urbanisation en Allemagne, du Rhin, autrement dit du tracé de la frontière. La hiérarchie des types d'occupations du sol secondaires met en lumière l'importance des zones industrielles et commerciales en Allemagne, reléguant l'agriculture protégée et les zones vertes urbaines à la fin du classement. Au contraire, ces deux types d'occupation du sol sont privilégiées dans les règles françaises et transfrontalières. Cette première analyse de la mobilisation des variables représentant les différents types d'occupation du sol montre les similitudes et divergences existantes au regard de la nature des règles (françaises, allemandes, transfrontalières).

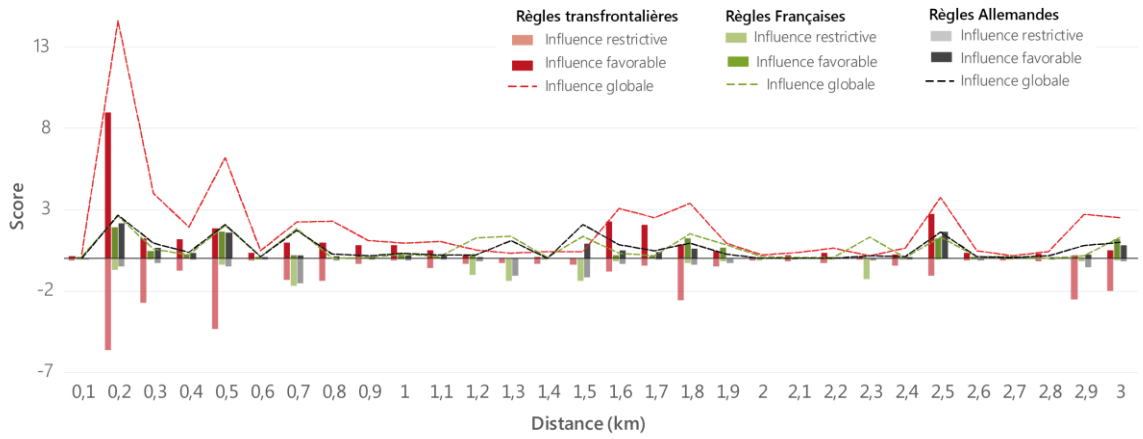
Influence des occupations principales du sol dans les règles de différentes natures

L'influence des différents types d'occupation est précisée à travers l'exploration de leurs proportions dans le voisinage, relevées en fonction de la distance. À travers cette démarche, la nature des relations entre l'urbanisation et les zones en eau, elles-mêmes associées au tracé de la frontière, peut, par exemple, être établie. Dans la suite de cette sous-section, la présence des types principaux d'occupation du sol est analysée en fonction de la polarité de son influence dans le voisinage. L'analyse s'appuie, dans un premier temps, sur la production du graphique de polarité des influences, qui permet de distinguer les contributions spécifiques des règles de différentes natures pour chaque type d'occupation du sol. Dans un second temps, les tendances majeures d'occupation du sol dans le voisinage sont construites sur la base des pics principaux d'influences. Une sélection des six pics est

réalisée : les pics incitatifs les plus influents d'après leur score, ou ceux dont le seuil mobilisé interdit la présence d'un type d'occupation du sol à une distance spécifique. Réduire leur nombre facilite l'interprétation du jeu de règles, auparavant compliquée par la multiplication des variables mobilisées dans la construction des règles.

Type d'occupation du sol urbaine

Figure 8. 1 : Influence et polarité de l'occupation du sol urbaine dans le voisinage en fonction de la distance



L'influence de l'occupation du sol urbaine, dans le voisinage des cellules urbanisées présente une tendance globalement décroissante (Figure 8. 1) : l'urbain proche est plus influent. Au-delà des tendances générales similaires observées précédemment entre les règles de différentes natures (Table 8. 4), qui soulignait une influence majeure de l'urbain dans le voisinage des cellules urbanisées, la Figure 8. 1 montre également une variation similaire dans le premier kilomètre, ce quel que soit la nature des règles. Bien que le score de l'urbain soit plus important pour les règles transfrontalières, les variations de l'influence (croissantes/décroissantes) sont identiques pour l'ensemble des règles. La présence d'urbain proche est particulièrement remarquable dans les règles transfrontalières. Ce point peut souligner que l'environnement urbain proche est recherché en priorité dans le cas où la localisation nationale en France ou en Allemagne, elle, ne l'est pas. Les règles allemandes présentent un léger pic aux alentours de 1,5 km. L'intensité de ce même pic dans le cas des règles françaises est moindre. Ces dernières présentent néanmoins un pic à 2,3 km, où l'influence de l'urbain dans les règles allemandes et transfrontalières est presque inexistante. La Figure 8. 1 offre également l'opportunité d'analyser l'absence de recours à certaines variables. Si la quasi-totalité des variables représentant le type d'occupation du sol urbain sont mobilisées, la variable urbaine à 2,8 km ne l'est pas dans le processus de décision pour la construction des règles françaises. Les variables urbaines à 1,4 km, 2 km, 2,1 km, 2,2 km, elles, ne le sont pas pour le groupe de règles allemandes. La composition du voisinage urbain des cellules urbanisées présente donc des différences suivant la nature des règles.

Au regard du nombre important de variables urbaines d'occupation du sol et suivant la même démarche que celle mobilisée précédemment (Chapitre 7 – Section 2.3), les pics d'intérêts

sont extraits pour chaque groupe de règles. Leur analyse spécifique permet d'extraire les tendances majeures en matière de voisinage urbain en fonction de la nature des règles analysées. L'extraction des seuils est effectuée manuellement, et dépend du nombre de règles s'y référant.

Table 8. 5 Tendances majeures françaises, allemandes et transfrontalières d'occupation urbaine dans le voisinage

Règles françaises - Tendances urbaines majeures (%)

Tendances	0,2 km	0,3 km	0,5 km	1,8 km	2,5 km	3 km
1	≤ 25		24 <	3 <	5 <	4 <
2	≤ 0	2 <		≤ 39		
3	0 < x ≤ 25		24 <	3 <	5 <	4 <
4	≤ 0		24 <			
5	0 <		24 <	3 <	4 < x ≤ 5	≤ 4

Règles allemandes - Tendances urbaines majeures (%)

Tendances	0,2 km	0,3 km	0,5 km	1,5 km	2,5 km	3 km
1	≤ 0	2 <				6 <
2	0 <	≤ 48	24 <	10 <	4 <	5 < x ≤ 7
3	0 <		24 <		4 <	
4	0 <		≤ 24		4 <	
5	0 <		24 <	10 < x ≤ 46	4 <	5 <

Règles transfrontalières - Tendances urbaines majeures (%)

Tendances	0,2 km	0,3 km	0,5 km	1,6 km	1,7 km	2,5 km
1	0 <	(≤ 25)	≤ 24	7 <	7 <	4 <
2				≤ 6		3 <
3	≤ 0	≤ 2				
4	≤ 0	≤ 2	≤ 18			
5	0 <	≤ 25	≤ 24	7 <	7 <	6 <
6	0 <		24 < x ≤ 40			≤ 4

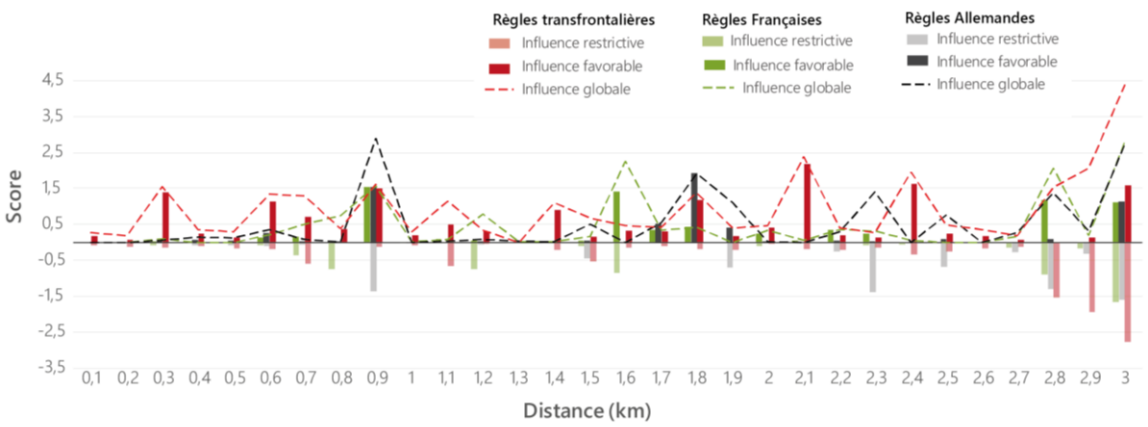
La Table 8. 5 met en lumière les similitudes mais aussi les divergences observées dans les tendances générales. Si les pics retenus sont majoritairement identiques quelques soit la nature des règles, celui localisé à 1,5 km n'apparaît qu'en Allemagne, tout comme celui localisé à 1,8 km en France, ou encore de ceux localisés à 1,6 km, 1,7 km et 2,5 km pour les règles transfrontalières. Il est légitime de s'interroger sur la portée de cette différence. En effet, d'un point de vue spatial, la variation est seulement de 200 m ou 300 m. Toutefois, les autres pics présentent une similitude parfaite. Ce point souligne donc une différence, même de légère ampleur lors de l'analyse de l'influence de l'occupation du sol urbaine dans le voisinage.

La divergence peut s'exprimer également au niveau des seuils. En effet, l'occupation du sol en France semble être plus « tolérante » face à la présence d'urbain dans un périmètre de 200m puisque le seuil de restriction, quand restriction il y a, n'est pas uniquement un seuil strict interdisant la présence d'urbain (≤ 0). Une tolérance jusqu' à 25% est observée. Ce n'est pas le cas dans les règles allemandes ou transfrontalières. Une similitude est également observée entre la France et l'Allemagne où le seuil incitatif supérieur à 24% apparaît, contrairement au transfrontalier ou les seuils restent majoritairement limitant, à l'exception de ceux relevés dans la tendance 6, puisque la présence d'urbain dans le voisinage doit être comprise entre 24% et 40%. La limite minimale imposée de la présence d'urbain à 2,5km est lui du même ordre de grandeur quel que soit la partie du jeu de règles étudiée (autour de 4%).

Type d'occupation du sol agricole

La Figure 8. 2 reprend la même démarche mise en place pour l'analyse de l'influence de l'occupation agricole du voisinage des cellules urbanisées.

Figure 8. 2 : Influence de l'occupation du sol agricole dans le voisinage en fonction de la distance



L'influence de l'agriculture est légèrement moins homogène à travers les règles de différentes natures. Les règles allemandes présentent un pic important à 900m. Un pic presque équivalent est retrouvé dans les règles françaises 700m plus loin (1,6 km). Les règles transfrontalières ont un voisinage dont les proportions d'occupation du sol agricole sont plus linéaires à mesure de l'augmentation de la distance dans le voisinage. Toutefois, l'agriculture se retrouve dans un voisinage proche (entre 600m et 900m) et éloigné (supérieur ou égale à 2,8 km) pour la totalité des règles.

Table 8. 6 Tendances majeures françaises, allemandes et transfrontalières d'occupation agricole dans le voisinage

Règles françaises - Tendances agricoles majeures (%)

Tendances	0,3 km	0,9 km	1,6 km	1,8 km	2,8 km	3 km
1	≤ 71	21 <	47 <		23 <	21 < x ≤ 89
2	71 <			31 <		
3		21 <	47 < x ≤ 88		23 < x ≤ 88	21 < x ≤ 89
4		64 <	47 < x ≤ 88		23 < x ≤ 88	21 < x ≤ 89
5				36 <		

Règles allemandes - Tendances agricoles majeures (%)

Tendances	0,3 km	0,9 km	1,7 km	1,8 km	1,9 km	3 km
1	35 <	21 < x ≤ 84		32 <	≤ 75	21 < x ≤ 89
2		21 < x ≤ 84		32 <	75 <	≤ 89
3		21 < x ≤ 84	47 <		≤ 75	21 < x ≤ 89
4						
5						

Règles transfrontalières - Tendances agricoles majeures (%)

Tendances	0,3 km	0,3 km	0,9 km	2,1 km	2,4 km	3 km
1	≤ 50					
2	63 <		21 <			15 < x ≤ 89
3				45 <		
4		42 <		33 <	33 <	≤ 85
5			23 <			15 < x ≤ 89
6			21 <	26 <		21 < x ≤ 89
7		42 <		36 <	33 < x ≤ 84	≤ 85
8						54 <

L'évaluation des tendances agricoles présente un intérêt particulier avec l'apparition d'encadrements dans les proportions de l'occupation attendues dans un certain périmètre. Ce point est illustré quel que soit la nature des règles envisagées. Ainsi les règles françaises encadrent entre 47% et 88 % la proportion d'occupation du sol agricole dans un périmètre de 1,6 km, alors qu'un encadrement plus large apparaît à 900m pour les règles allemandes (entre 21% et 84%). Dans le cas des règles transfrontalières, l'encadrement apparaît en fin de transect, principalement vers 3 km.

Les seuils extraits sont, de manière générale, plus élevés que ceux obtenus dans l'analyse de la présence urbaine. La restriction absolue, qui interdit la présence de terres agricoles, ne ressort pas dans les tendances majeures extraites. Ces deux points sont en cohérence avec la répartition de ce type d'occupation du sol sur la zone d'étude, à la fois très étendue, et dans une proportion majeure au regard des autres occupations du sol.

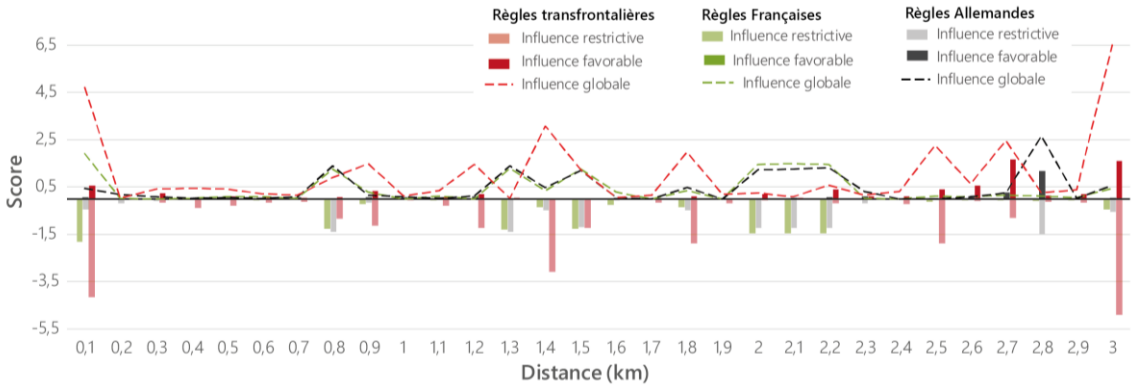
En résumé, la méthode permet d'extraire des intervalles de proportion d'occupation du sol agricole dans le voisinage qui influencent favorablement la transition urbaine. Les tendances majeures sont similaires entre les règles de différentes natures et présentent une influence

plus importante de l'occupation du sol agricole dans un voisinage éloigné. Pour autant, ce constat est davantage vérifié dans les règles transfrontalières.

Type d'occupation du sol forêt

La forêt est le troisièmes type d'occupation du sol le plus important d'après l'attribution du score d'influence dans le voisinage, ce quel que soit la nature des règles (Figure 8. 3).

Figure 8. 3 : Influence de l'occupation du sol forêt dans le voisinage en fonction de la distance



Deux pics majeurs se distinguent. Le premier dans un voisinage très proche pour les règles françaises et transfrontalières. La forêt est donc présente dans un périmètre proche et peut être visible (100m). Ce n'est pas le cas des règles allemandes dont cette composante est présente dans une moindre mesure. Le second s'observe vers la fin du transect de voisinage pour les règles allemandes et transfrontalières (2,8 km – 3 km). Des pics similaires d'influence moindre s'observent également entre 1,3 km et 1,5 km indépendamment de la nature des règles. L'influence de la forêt présente des caractéristiques qui permettent de faire la différence entre les voisinages proches allemands qui ne lui accordent pas d'importance remarquable et ceux décrit dans les règles françaises ou transfrontalières qui au contraire lui donne de l'importance. A l'inverse, les règles allemandes donnent de l'importance à la forêt dans un périmètre plus éloigné, tout comme celles de nature transfrontalière. Le point intéressant ici est donc la capacité des règles transfrontalières à retranscrire des caractéristiques générales différentes dans les règles nationales.

Table 8. 7 Tendances majeures françaises, allemandes et transfrontalières d'occupation forêt dans le voisinage

Règles françaises - Tendances forestières majeures (%)

Tendances	0,1 km	0,5 km	0,9 km	1,1 km	1,6 km	2,7 km
1	≤ 0					
2	≤ 38					
3	0 <					
4	≤ 0		≤ 10 ou 10 <		≤ 26	
5						1 <

Règles allemandes - Tendances forestières majeures (%)

Tendances	1,2 km	2,2 km	2,3 km	2,6 km	2,7 km	2,8 km
1	7 <	≤ 64				≤ 52
2		≤ 64				1 < x ≤ 52
3			≤ 19			15 < x ≤ 89
4		≤ 64	9 <			≤ 52
5				54 <		
6					1 <	

Règles transfrontalières - Tendances forestières majeures (%)

Tendances	0,1 km	2,2 km	2,5 km	2,6 km	2,7 km	3 km
1	≤ 38		≤ 52 ou 52 <			≤ 70
2	38 <					
3	≤ 38	14 <				≤ 70
4		64 <				
5	≤ 38		≤ 74		70 <	70 <
6	≤ 38		≤ 52	0 <		≤ 70
7					≤ 1 ou 1 <	
8						≤ 74

La divergence générale de composition du voisinage, notamment entre les règles françaises et allemandes, s'observe nettement à travers la sélection des pics majeurs. Seulement 5 pics incitatifs sont extraits des règles françaises. La majorité d'entre eux ne dépassent pas le kilomètre. Au contraire, les règles allemandes présentent 6 pics, tous localisés plus loin que la limite du kilomètre. Cette observation est également vraie lors de l'analyse des règles transfrontalières, où tous les pics sont localisés après 2 km, à l'exception d'un seul. Il apparaît donc que l'influence de la forêt est plus importante dans le voisinage Allemand.

Une différence plus marquée au regard de la nature des règles analysées apparaît clairement dans l'analyse de la présence de forêts dans le voisinage des cellules urbanisées. Si des divergences existent entre les règles de différentes natures, certaines s'observent également à l'échelle des tendances extraites pour les règles de même nature. Ainsi pour les règles transfrontalières deux compositions peuvent être observées à 100m : celles restreignant la présence de forêt en dessous du seuil de 38% (entre 2 et 3 ha) ; celles contraignant la présence à plus de 38%. La première continue d'imposer en partie une certaine restriction à 2,5 km plus à 5 km. En partie, car certaines règles changent la polarité en imposant cette fois la présence de plus de 50% de forêt à 2,5 km. Dans un sens, un type de voisinage impose une présence modérée au long du transect alors qu'un second impose dans un premier temps

une restriction (voisinage proche) puis une présence accrue dans un second temps (2,5 km). Pour finir, la satisfaction d'une présence accrue de forêt à 100m (plus de 38%) permet de s'affranchir de toute autre contrainte dans un périmètre plus éloigné.
En résumé, une différence plus marquée dans la composition du voisinage forestier apparaît dans les règles au regard de leur nature. L'influence de la forêt présente des caractéristiques qui permettent de faire la différence entre les voisinages français, allemands et transfrontaliers. Elle permet également d'établir des variations de comportement dans les tendances des règles de même nature.

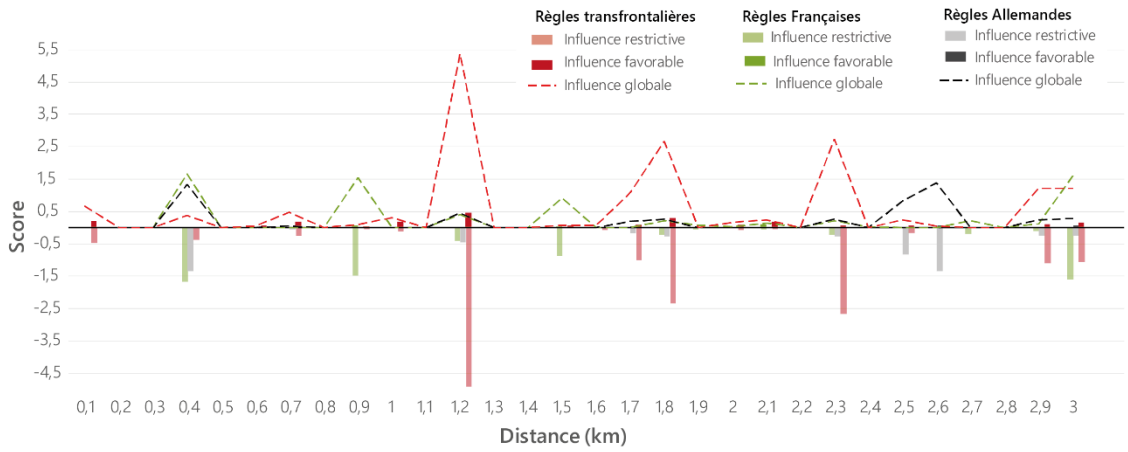
Influence des occupations du sol secondaires dans les règles de différentes natures

L'observation des tendances d'occupation du sol des types principaux est complétée par celle des types secondaires.

Type d'occupation du sol zone récréative urbaine

Il est moins évident de dégager des tendances générales similaires concernant les zones récréatives urbaines en observant leur influence dans les différents types de règles (Figure 8. 4). En effet, les pics d'influences apparaissent de manière plus sporadique le long du transect, et sont principalement restrictifs.

Figure 8. 4 : Influence de l'occupation du sol zone récréative urbaine dans le voisinage en fonction de la distance



Les voisinages français et allemand sont influencés par les zones récréatives à 400m, donc plus proche que dans le cadre des règles transfrontalières dont l'influence majeure se situe après le kilomètre (1,2 km). Un second pic est observé à 2,6 km pour les règles allemandes et 2,9 km pour les règles françaises. En générale, les règles transfrontalières font état d'une différence dans l'appréhension des zones récréatives urbaines.

L'extraction des pics majeurs souligne ce constat (Table 8. 8). Il n'est pas possible d'extraire plus de 3 pics incitatifs pour les règles allemandes, tous localisés en fin de transect (supérieur à 1,7 km). Cette tendance est également observée pour les règles françaises dont les pics incitatifs retenus sont principalement localisés après 1,5 km, à l'exception du premier pic (0,9 km). Les seuils observés sont beaucoup plus bas que ceux extraits pour les occupations du sol principales, notamment pour les règles nationales. Les règles transfrontalières se démarquent par une influence plus proche (en dessous du kilomètre) et plus importante (supérieure à 13% à 100m, supérieure à 17% à 1 km).

Table 8. 8 Tendances majeures françaises, allemandes et transfrontalières d'occupation zones récréatives urbaines dans le voisinage

Règles françaises - Tendances majeures zones récréatives urbaines (%)

Tendances	0,9 km	1,5 km	1,9 km	2 km	2,1 km	2,4 km
1	≤ 4	≤ 3				
2	4 <			3 <		
3		3 <				
4			≤ 0			
5			0 <			
6					2 <	
7						0 <

Règles allemandes - Tendances majeures zones récréatives urbaines (%)

Tendances	1,7 km	2,6 km	3 km
1	≤ 3		
2	3 <		
3		0 <	
4		≤ 1	1 <
5			≤ 2

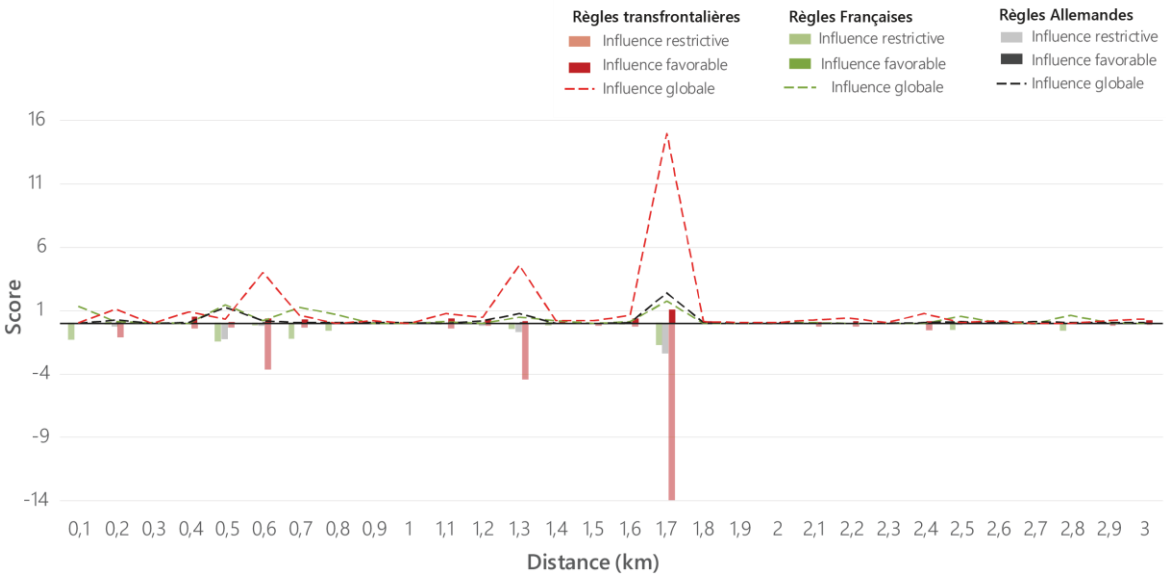
Règles transfrontalières - Tendances majeures zones récréatives urbaines (%)

Tendances	0,1 km	0,7 km	1 km	1,2 km	1,8 km	2,1 km
1	≤ 13					
2	13 <					
3		≤ 0				
4		0 <				
5			≤ 12	≤ 14		
6			17 <			
7				≤ 14		0 <
8					2 <	

Type d'occupation du sol vigne

La place des vignes dans le voisinage est différente de celle de l'occupation agricole en générale (Figure 8. 5). S'il n'y avait pas de similarité majeure entre les règles transfrontalières et les autres pour l'occupation du sol agricole, l'influence des vignes est presque équivalente quel que soit la nature des règles. L'intensité de l'influence est néanmoins supérieure pour les règles transfrontalières, dont trois pics majeurs sont très distinctement observables.

Figure 8. 5 : Influence de l'occupation du sol vigne dans le voisinage en fonction de la distance



La présence incitative est plus marquée dans le cadre des règles nationales (Table 8. 9). Cette observation est particulièrement vraie dans la première tendance allemande qui préconise plus de 63% de vigne à 200m et plus de 32% à 3 km. Les autres tendances sont moins extrêmes, requérant la présence de vignes mais n'imposant pas de seuil particulier. La vigne est également appréciée dans un périmètre plus proche pour le voisinage français (inférieur à 1,6 km). Les règles transfrontalières ne permettent que rarement d'apporter une précision sur la présence de vigne dans le voisinage. En effet, l'extraction des pics, même focalisée sur les incitatifs, se trouve submergée par des seuils restrictifs relativement hauts, ne permettant donc pas clairement d'apprécier la présence des vignes. La prédominance de la restriction inférieure à 72% à 3 km illustre ce phénomène.

Table 8. 9 Tendances majeures françaises, allemandes et transfrontalières d'occupation vigne dans le voisinage

Règles françaises - Tendances majeures vignes (%)

Tendances	0,2 km	0,8 km	1,1 km	1,4 km	1,6 km	2,9 km
1	≤ 4					
2	0 <			≤ 5		
3		≤ 0				
4		0 < x ≤ 18				
5		≤ 18		5 <		
6		18 <				33 <
7			≤ 0			
8					≤ 1 ou 38 <	

Règles allemandes - Tendances majeures vignes (%)

Tendances	0,7 km	0,8 km	1,6 km	2,6 km	2,8 km	3 km
1	63 <					32 <
2		16 <				
3			0 <			
4				0 <		
5					0 <	

Règles transfrontalières - Tendances majeures vignes (%)

Tendances	0,4 km	0,6 km	0,7 km	1,1 km	1,6 km	1,7 km
1	≤ 4					≤ 72
2	≤ 1	≤ 76			≤ 5	5 < x ≤ 72
3		≤ 36				≤ 72
4		24 <				
5			≤ 13 ou 13 <			
6				≤ 0		≤ 72
7				27 <		
8					0 <	≤ 72

Type d'occupation du sol zone industrielle et commerciale

Comme pour les types principaux d'occupation du sol, l'industriel et commercial est influent dans le voisinage proche (300m). Ce pic est présent quel que soit la nature des règles (Figure 8. 6). Le caractère incitatif est plus présent tout au long du transect dans les règles transfrontalières. Au contraire, plusieurs pics restrictifs d'intensité moyenne se retrouvent dans les règles allemandes, à mesure de l'augmentation de la distance avec la cellule urbanisée (700m, 1km, 1,6 km, 2 km). Les règles françaises, elles, ne présentent qu'un seul autre pic de forte intensité à 1,4 km. Ce dernier est restrictif. L'influence de l'industriel et commercial permet donc également de différencier les types de règles tout en soulignant une composante commune à 300 m.

Figure 8. 6 : Influence de l'occupation du sol industriel et commercial dans le voisinage en fonction de la distance

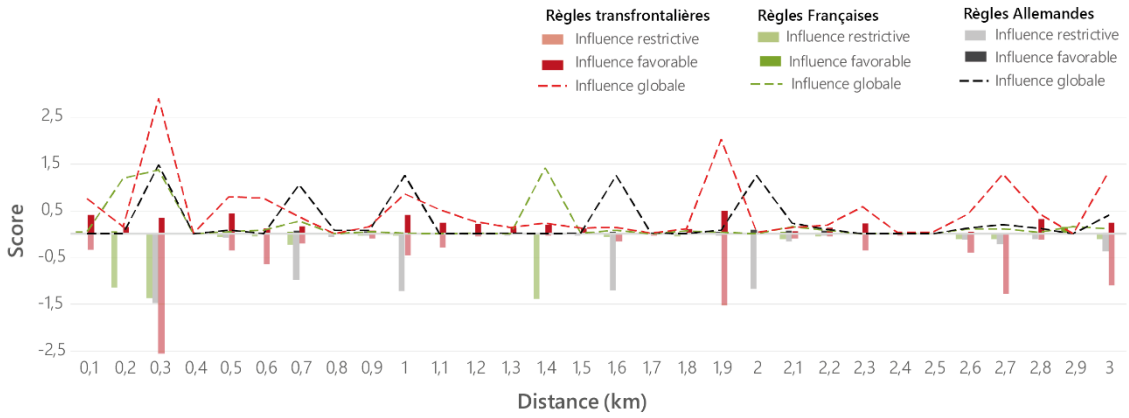


Table 8. 10 Tendances majeures françaises, allemandes et transfrontalières d'occupation zone industrielle et commerciale dans le voisinage

Règles françaises - Tendances majeures zones industrielles et commerciales (%)

Tendances	0,1 km	0,2 km	0,7 km	1,9 km	2,1 km	2,9 km
1	38 <					
2		≤ 0				
3		≤ 0	≤ 2			
4		≤ 0	2 <			
5				7 <		
6					≤ 2	
7		≤ 0			0 <	
8						2 <

Règles allemandes - Tendances majeures zones industrielles et commerciales (%)

Tendances	0,7 km	0,9 km	1,6 km	2 km	2,1 km	2,2 km
1	≤ 6		≤ 7	≤ 8		
2	≤ 0		≤ 7	≤ 8		
3	6 <		≤ 7	≤ 8		
4		≤ 0				
5		4 <			2 <	
6					0 < x ≤ 2	
7						5 <

Règles transfrontalières - Tendances majeures zones industrielles et commerciales (%)

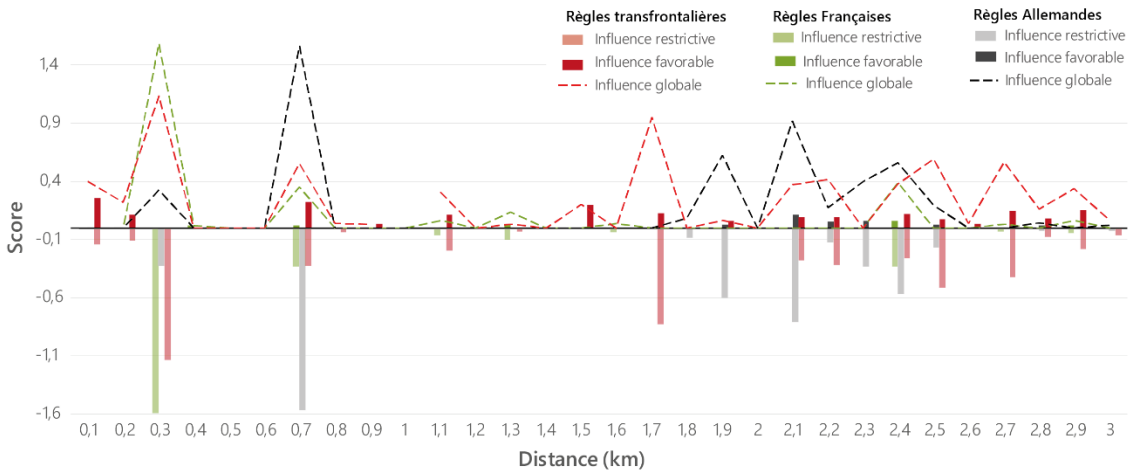
Tendances	0,1 km	0,3 km	0,5 km	1 km	1,9 km	2,8 km
1	50 <					
2	≤ 38					
3		≤ 4			≤ 9	
4		≤ 8				
5		≤ 17				
6		8 <				
7			≤ 0			
8			38 <			
9				≤ 6 ou ≤ 0		
10				0 <		
11					≤ 0	
12					0 <	
13						7 <

La présence de zones industrielles et commerciales à 200 m est « interdite » dans plusieurs tendances françaises. Elle est également interdite à 700 m et 900 m dans certaines règles allemandes, mais aussi à 500 m et 1,9 km dans les règles transfrontalières. La présence est épisodiquement recherchée quel que soit la nature des règles : à 700 m puis aux alentours de 2 km pour les règles françaises ; à 700 m, 900 m et après 2 km pour les règles allemandes ; à partir de 300 m pour les règles transfrontalières.

Type d'occupation du sol transport

L'influence des zones de transport est plus anecdotique au regard de celles des autres occupations du sol. Ce point est illustré par une amplitude plus faible des scores caractérisant les variables de transport mobilisées dans la construction de l'arbre de décision (Figure 8. 7).

Figure 8. 7 : Influence de l'occupation du sol transport dans le voisinage en fonction de la distance



Ces dernières ne semblent que revêtir une importance dans la cadre des règles françaises et transfrontalières aux alentours de 300 m. Ce pic restrictif est également présent mais à 700 m pour les règles allemandes. Le graphique ci-dessus montre une dépendance moindre des règles françaises vis-à-vis du transport, que les règles d'autres natures. Par ailleurs, la part française du jeu de règles ne présente pas plus de 5 pics incitatifs pour composer les tendances générales (Table 8. 11). Le transport sera éventuellement recherché épisodiquement à 400 m ou encore à partir de 2,9 km. Une influence incitative plus marquée est remarquable dans le cadre des règles allemandes. Elle n'apparaît pas avant 2 km, mais présente des seuils légèrement plus importants (supérieurs à 6 % ou 7 %). La situation est plus mitigée dans l'analyse de la partie transfrontalière du jeu de règles et permet difficilement de s'avancer sur l'élaboration d'un comportement générique d'après les tendances déjà généralisées lors de l'observation des seuils mobilisés dans les pics incitatifs majeurs.

Table 8. 11 Tendances majeures françaises, allemandes et transfrontalières d'occupation transport dans le voisinage

Règles françaises - Tendances majeures transport (%)

Tendances	0,4 km	0,7 km	1,3 km	2,4 km	2,9 km
1	1 <				
2		≤ 5		≤ 6	
3			≤ 1		
4					≤ 0
5					0 <

Règles allemandes - Tendances majeures transport (%)

Tendances	1,9 km	2,1 km	2,2 km	2,3 km	2,5 km	2,8 km
1	≤ 6	≤ 7		0 <		1 <
2			≤ 1			
3	6 <	≤ 7				
4		7 <				
5				≤ 0		
6					≤ 0	

Règles transfrontalières - Tendances majeures transport (%)

Tendances	0,1 km	0,7 km	1,5 km	1,7 km	2,7 km	2,9 km
1	≤ 0					
2	0 <					
3		≤ 5 ou 5 <				
4			0 <			
5				≤ 4 ou 4 <		
6					≤ 6	
7				≤ 4		≤ 2
8						≤ 0 ou 0 <

Type d'occupation du sol zone en eau

Comme mentionné dans le chapitre précédent, la question des zones en eau revêt un intérêt particulier de par le double enjeu qu'elles représentent. La composante environnementale est indéniable, cependant cette dernière est complétée par un aspect politique/culturel puisque que le tracé de la frontière étatique suit celui du Rhin.

L'influence de ces zones n'est pas homogène suivant l'appréhension des règles de différentes natures (Figure 8. 8). Ainsi les règles transfrontalières présentent un pic majeur limitant à 2,7 km accompagné d'un pic secondaire plus proche (400 m). Les règles françaises présentent elles deux pics majeurs limitants d'intensité similaire à 1,1 km et 1,9 km. Les règles allemandes ne présentent pas de pic aussi intense mais bien plusieurs pics d'intensité moyenne ou faible.

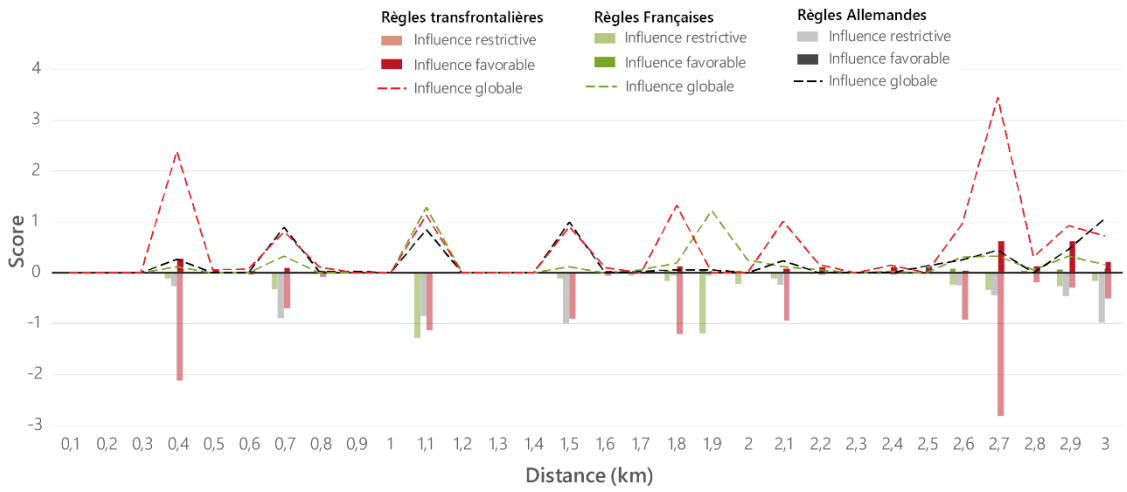


Figure 8. 8 : Influence de l'occupation du sol zone immergée dans le voisinage en fonction de la distance

Table 8. 12 Tendances majeures françaises, allemandes et transfrontalières d'occupation zone immergée dans le voisinage

Règles françaises - Tendances majeures eau (%)

Tendances	1,8 km	1,9 km	2 km	2,6 km	2,8 km	2,9 km
1	1 <	≤ 8				
2		≤ 8	1 <			
3			≤ 1			
4				≤ 0		
5				0 <		
6						≤ 4

Règles allemandes - Tendances majeures eau (%)

Tendances	0,9 km	1,7 km	2,5 km	3 km
1	2 <		≤ 5	1 <
2		0 <		≤ 1
3			5 <	
4				1 <
5				4 <

Règles transfrontalières - Tendances majeures eau (%)

Tendances	0,4 km	1,8 km	2,7 km	2,8 km	2,9 km	3 km
1	≤ 11	≤ 2	≤ 4			
2			≤ 0		≤ 2	
3				≤ 1		
4		7 <	4 <	≤ 14		10 <
5					2 < ou 4 <	
6						≤ 0

Les pics incitatifs présents dans une moindre mesure permettent néanmoins d'observer des schémas différents suivant la nature des règles analysées. Ainsi, le jeu de règles allemand ne présente que 4 pics incitatifs mais leur influence apparaît à 900 m. Au contraire, la France présente plusieurs pics incitatifs, cependant l'influence commence à 1,8 km. L'influence des

zones immergées est donc plus proche dans le cadre des règles allemandes que dans celui des règles françaises. C'est dans le cadre des règles transfrontalières que l'influence est la plus significative. Les seuils sont plus élevés (supérieur à 7% à 1,8 km, supérieur à 10 % à 3 km).

Conclusion

L'intérêt d'extraire les seuils associés aux différentes tendances d'occupation du sol dans le voisinage, en fonction de la nature des règles, réside moins dans le fait de s'appuyer sur des proportions précises, que dans celui de souligner l'apparition de divergences dans la réponse obtenue, entre les règles spécifiques à la France, celles spécifiques à l'Allemagne ainsi que les règles s'affranchissant d'une localisation géographique spécifique au regard de la frontière. L'analyse des tendances d'occupation du sol secondaire est rendue moins explicite en raison de la superficie moindre qu'elle représente. En effet, elles complètent la continuité du voisinage en comblant les espaces non-occupés par les trois types d'occupation principales, à elles seules représentant plus de 50 % de la surface de la zone d'étude. Ainsi, les seuils les plus représentatifs des tendances observées sont des seuils restrictifs assez haut, qui ne permettent pas une description précise. Ce point permet de comprendre la démultiplication de tendances majeurs observées pour les occupations du sol secondaires en dépit d'une présence moins importante sur la zone d'étude.

Cette analyse du voisinage des cellules urbanisées entre 1990 et 2006, cette fois en fonction de la nature des règles, montre :

- La capacité des règles de type transfrontalier à résumer certaines tendances exprimées dans les deux autres types de règles. La présence de forêt en est un exemple. Cette dernière souligne également une différence France – Allemagne : proximité française, éloignement allemand ;
- La dissociation des règles transfrontalières et nationales concernant la présence agricole dans le voisinage. Elles présentent un comportement beaucoup plus linéaire alors que des pics sont observés pour les règles françaises ou allemandes ;
- Une importance accrue dans la proximité de zones industrielles et commerciales dans les règles allemandes ;
- Une influence prépondérante des zones immergées est observée dans les règles transfrontalières. Ce point est intéressant au regard de la présence de la frontière qui suit le Rhin est peut donc avoir influencé l'urbanisation ;
- La présence de vignes est privilégiée dans un périmètre proche pour les règles françaises. Au contraire plus éloigné pour les règles transfrontalières et dans une moindre mesure pour les règles allemandes.

La mise en relation des observations précédentes permet également d'étendre la réflexion à une analyse entre occupations du sol. Par exemple, quand un voisinage de proximité fortement influencé par l'urbain est requis dans les règles transfrontalières, l'influence de la forêt est également importante dans un périmètre plus éloigné.

3. INFLUENCE DES PÔLES URBAINS

Cette sous-section présente l'analyse de l'accessibilité des cellules urbanisées suivant la nature des règles. Elle permet ainsi de statuer sur les différences ou similitudes relevées entre la France et l'Allemagne, au regard de l'influence des pôles de différents niveaux, présents sur la zone d'étude. En effet, les variables d'accessibilité jouent un rôle majeur dans la construction de l'arbre. Quel que soit la nature des règles envisagées, ces variables sont mobilisées dans plus 50 % des cas, pour finalement, être retrouvé dans 263 règles sur les 395 composant le jeu total (Table 8. 13).

Table 8. 13 Règles mobilisant les variables d'accessibilité en fonction de leur nature

	Nombre de règles	Règles basées sur l'accessibilité
Règles françaises	82	70,7 %
Règles allemandes	85	56,4 %
Règles transfrontalières	228	68,9 %
Total	395	66,6 %

L'objectif de cette sous-section est d'étudier plus en détail l'influence de chaque variable d'accessibilité, en fonction de la nature des règles concernées, et d'en extraire les caractéristiques principales. Dans un premier temps, la section revient sur la polarité des seuils d'accessibilité sélectionnés par l'arbre de décision, dans le but de définir une limite à ne pas dépasser ou au contraire une limite minimum à observer au regard des cellules urbanisées. En d'autres termes, il s'agit de définir à quelles distances des pôles il est possible d'habiter. Dans un second temps, des tendances d'accessibilité sont extraites et cartographiées pour être appréciées spatialement en fonction de la nature des règles.

Caractéristiques générales de l'accessibilité suivant la nature des règles

C'est une fois encore une polarité restreignant l'accessibilité en dessous d'un certain temps de trajet, qui influence principalement le jeu de règle au regard de l'amplitude des scores (Table 8. 14). Ce cas de figure est observé indépendamment de la nature des règles. L'influence restrictive impose donc de ne pas dépasser un périmètre établi en matière de temps de trajet. La hiérarchie de l'influence des variables d'accessibilité ne varie pas non plus en fonction du groupe de règles. Ainsi, pour les règles françaises, allemandes et transfrontalières, c'est l'accessibilité pondérée niveau 1 qui présente l'influence la plus importante, suivit de l'accessibilité niveau 2. Les règles françaises semblent néanmoins accorder plus d'importance que les autres groupes à un certain éloignement des pôles de niveaux 2 puisque leur score est de 1,35, soit presque équivalent à celui de l'influence restrictive, alors que ceux des règles allemandes et transfrontalières sont respectivement 3 fois et 2 fois plus faibles.

Table 8. 14 Influence de l'accessibilité en fonction de la nature des règles

Règles Françaises (%)

	≤	>	Total
Accessibilité Niveau 1 pondérée	5,71	1,17	6,89
Accessibilité Niveau 2 pondérée	0,98	0,72	1,70
Accessibilité Niveau 1	0,00	0,18	0,18
Accessibilité Niveau 2	1,75	1,35	3,11
Total	8,44	3,42	11,86

Règles Allemandes (%)

	≤	>	Total
Accessibilité Niveau 1 pondérée	3,83	2,02	5,84
Accessibilité Niveau 2 pondérée	0,97	0,14	1,11
Accessibilité Niveau 1	0,40	0,53	0,94
Accessibilité Niveau 2	2,92	0,13	3,05
Total	8,12	2,82	10,94

Règles Transfrontalières (%)

	≤	>	Total
Accessibilité Niveau 1 pondérée	21,52	7,28	28,80
Accessibilité Niveau 2 pondérée	4,42	6,54	10,96
Accessibilité Niveau 1	6,53	2,88	9,41
Accessibilité Niveau 2	18,67	9,37	28,03
Total	51,13	26,06	77,20

L'analyse des seuils extrêmes qui représentent l'amplitude maximum des seuils d'accessibilité, en fonction de la nature des règles, est intéressante à plusieurs niveaux. Dans le cas de l'accessibilité niveau 1, en effet, un seul seuil est mobilisé pour les règles françaises. Ce dernier est obligatoirement supérieur à 39 min quand il intervient dans une règle. Au contraire, dans le cas Allemand, l'intervalle est compris entre 26 et 35 min. Les règles transfrontalières présentent, elles, une plus grande tolérance puisque leur intervalle est, lui, compris entre 6 et 61 min. Les cas de l'accessibilité niveau 2 et de l'accessibilité pondérée niveau 1 sont particuliers pour les règles françaises, puisque les seuils des limites minimum ou maximum sont uniquement restrictifs. Par exemple, une partie des règles imposent un développement urbain dans un périmètre restreint de 10 min (niveau 2) ou de 13 min (niveau 1 pondéré). Certaines règles étendent cependant cette limite à respectivement 35 et 50 min. Il n'y a donc pas de règles qui imposent l'absence d'urbanisation de proximité. Au contraire, certaines règles allemandes imposent un temps de trajet supérieur à 19 min (accessibilité niveau 2 en Allemagne) voir à 51 min (accessibilité pondérée niveau 1 en Allemagne). C'est également le cas des règles transfrontalières.

Table 8. 15 Seuils extrêmes mobilisés pour chaque variable d'accessibilité

		Règles françaises	Règles allemandes	Règles transfrontalières
Accessibilité Niveau 1	Min	> 39	> 26	> 6
	Max	> 39	< 35	< 61
Accessibilité Niveau 2	Min	< 10	< 5	> 5
	Max	< 35	< ou > 19	< ou > 33
Accessibilité Niveau 1 pondérée	Min	< 13	> 14	> 11
	Max	< 50	< ou > à 51	< 54
Accessibilité Niveau 2 pondérée	Min	< 14	< 16	< ou > 10
	Max	< ou > 28	< ou > 27	> 36

Les règles transfrontalières semblent offrir une plus grande flexibilité que les règles nationales en matière d'accessibilité. Les règles allemandes peuvent présenter une restriction plus stricte sur les pôles de niveau 2, qui impose une certaine proximité, ce qui reste cohérent au regard de leur répartition sur la zone d'étude. En effet, quatre des cinq pôles de niveau 2 se situent en Allemagne.

Dans le même objectif que celui entrainant l'extraction des seuils dégagés pour les tendances générales d'occupation du sol dans le voisinage, les seuils d'accessibilités sont utilisés dans l'objectif de comparer les tendances majeures en fonction de la nature des règles. La valeur absolue du seuil en lui-même doit être considérée au regard des contraintes et limites méthodologiques discutées par la suite.

Analyse de tendances et spatialisation de l'accessibilité suivant la nature des règles

Règles françaises

L'accessibilité pondérée niveau 1 est l'accessibilité majoritairement mobilisée dans la construction des règles françaises, comme dans le cas des règles allemandes et transfrontalières. Les règles françaises présentent majoritairement une limite restrictive, localisant les cellules urbanisées dans une zone préférentielle dont l'accessibilité pondérée niveau 1 est inférieure à 33 min (Figure 8. 9). Certaines étendent cette limite à 45 min. Ce seuil devient moins restrictif dans l'absolu puisqu'il englobe la majorité de la zone d'étude. Plus épisodiquement, les règles présentent parfois une contrainte imposant un éloignement des pôles, au regard de l'accessibilité pondérée niveau 1. Les cas d'intervalle sont quant à eux sporadiques (5 règles), mais restent cependant intéressants puisque leur double contrainte localise l'urbanisation dans une bande spatiale pouvant être assimilée à l'environnement périurbain. Ce point sera abordé à travers la cartographie des tendances d'accessibilité (Table 8. 16).

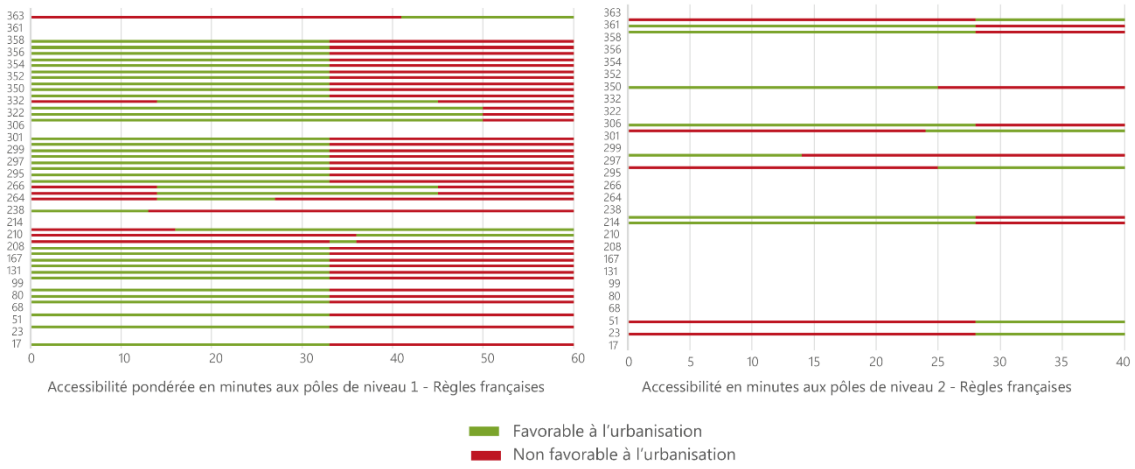


Figure 8. 9 : Contraintes liées à l'accessibilité pondérée pôle de niveau 1 et celles aux pôles de niveau 2 dans les règles françaises

L'accessibilité niveau 2 présente quant à elle différents seuils permettant de dégager un intervalle préférentiel puisque la majorité des règles permettent l'urbanisation dans une bande spatiale dont l'accessibilité est comprise entre 12 et 35 min (Figure 8. 9). L'accessibilité niveau 2 pondérée est mobilisée dans une moindre mesure dans les règles françaises. Aucun intervalle n'y est observé. Pour autant, deux comportements peuvent être dégagé. Le premier limite l'urbanisation à un périmètre inférieur à 28 min. Au contraire, le second impose un éloignement d'au minimum 28 min (Figure 8. 10).

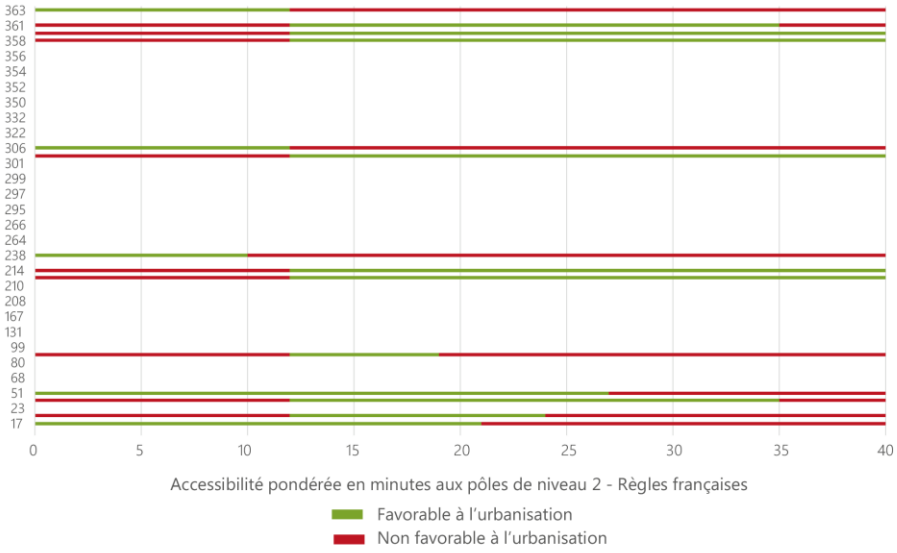


Figure 8. 10 : Contraintes liées à l'accessibilité pondérée pôle de niveau 2 dans les règles françaises

L'exploitation et la mise en relation des différentes contraintes d'accessibilité permettent l'extrapolation de quatre tendances majeures concernant l'accessibilité préférentielle française sur la zone d'étude (Table 8. 16).

Table 8. 16 Tendances d'accessibilités observées au sein du jeu de règles français

Tendances	Accessibilité pondérée Niveau 1	Accessibilité Niveau 2	Accessibilité pondérée Niveau 2	Accessibilité Niveau 1
1	0 < x ≤ 33			
2	14 < x ≤ 45			
3		12 < x ≤ 35	≤ 28	
4		12 < x ≤ 35	28 <	

Les différents zonages présents sur la carte ci-dessous hiérarchisent le potentiel d'urbanisation des différentes zones au regard des seuils d'accessibilité dégagés précédemment. Pour une meilleure lisibilité et une analyse plus confortable du comportement française vis-à-vis des pôles de différents niveaux, cette carte présente les préférences françaises à l'échelle de la zone totale (Figure 8. 11). Dans les faits, ces préférences sont restreintes à la zone française. La visualisation des préférences françaises mises en parallèle avec celles apparaissant dans les règles allemandes, sont présentées à la fin de la sous-section pour une comparaison avec les préférences transfrontalières (Figure 8. 18).

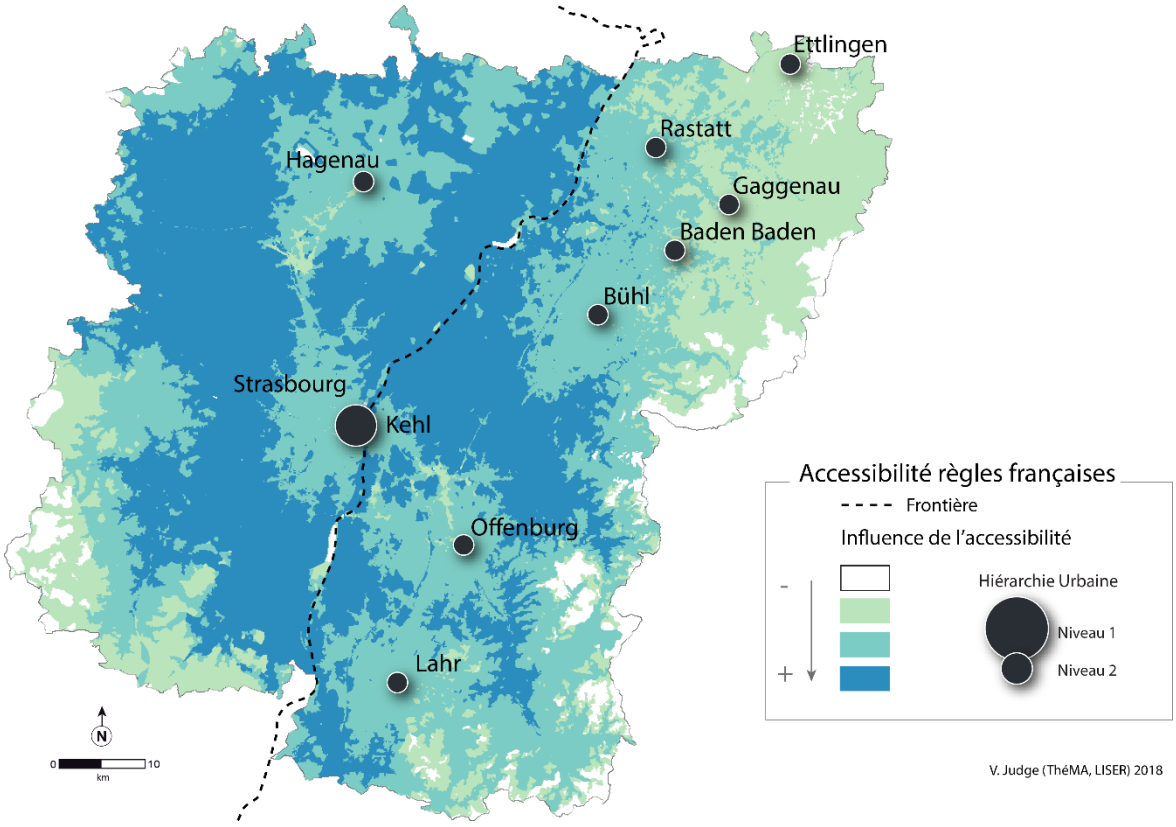


Figure 8. 11 : Zones préférentielles d'urbanisation françaises au regard de l'accessibilité aux différents pôles urbains

Spatialement, les tendances extraites traduisent une attractivité préférentielle du « périurbain » français. Ce point donne de l'importance au périmètre frontalier, localisé entre les pôles urbains de différents niveaux, ainsi qu'en périphérie ouest de Strasbourg ou d'Haguenau. Les centres urbains ne sont pas considérés comme les plus attractifs du point de vue accessibilité.

Règles allemandes

Sur les 85 règles allemandes, 48 mobilisent l'accessibilité pour décrire l'environnement favorable à l'urbanisation. Toutes les variables décrivant l'accessibilité sont mobilisées dans cette partie du jeu. L'accessibilité au pôle de niveau 1 n'apparaît cependant que dans un nombre de règles restreint (6 règles), tout comme l'accessibilité aux pôles de niveau 2 moins présente que dans le jeu de règles français. L'influence positive de la proximité des pôles s'observe sur la plupart des graphiques d'accessibilité, excepté pour celui représentant le pôle de niveau 1, ou l'attractivité est plus mitigée : 3 règles limitent l'éloignement, 3 règles imposent un éloignement minimum. L'accessibilité présentant l'occurrence maximale est, comme dans le cas des règles françaises, l'accessibilité pondérée niveau 1. Une bande préférentielle entre 14 min et 45 min se retrouve également dans les règles allemandes. Le seuil unique est lui légèrement plus proche que précédemment en limitant la zone préférentielle à un périmètre dont l'accessibilité est inférieure à 30 min. Par contre, une autre bande préférentielle apparaît, imposant cette fois un temps de trajet minimum de 30 min tout en le limitant à quelques exceptions près, en dessous de la limite horaire (51 min). L'accessibilité niveau 2 est par contre beaucoup plus mobilisée dans le cas allemand que dans celui français. La limite du temps de trajet en dessous de 20 min se retrouve dans la plupart des règles. Cette observation peut s'expliquer par une présence accrue des pôles de niveau 2 dans l'espace allemand de la zone d'étude. Il est néanmoins intéressant de souligner que malgré l'importance du nombre de pôles de niveau 2 sur la partie allemande de la zone d'étude, l'accessibilité niveau 1 pondérée est celle ayant le plus d'influence sur le jeu de règle.

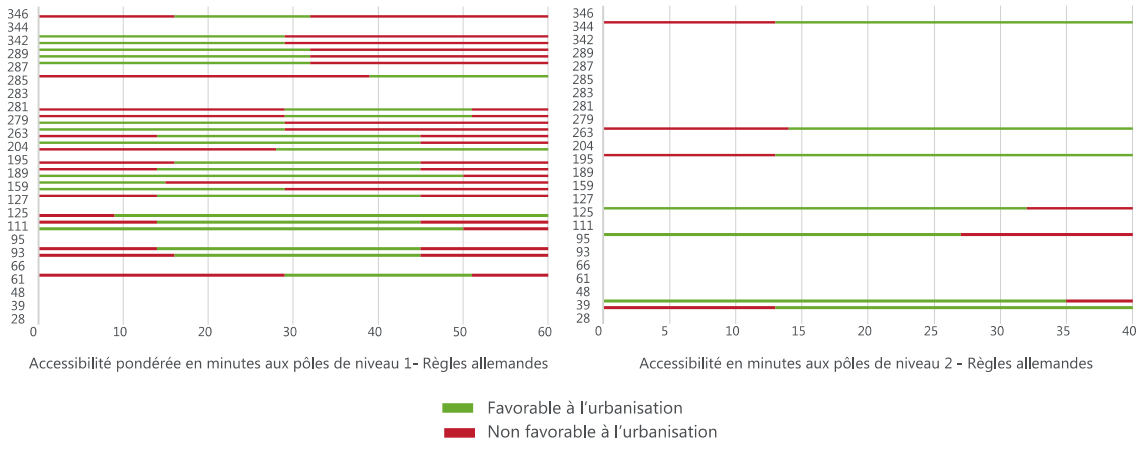


Figure 8. 12 : Contraintes liées à l'accessibilité pondérée pôle de niveau 1 et celle au pôle de niveau 2 dans les règles allemandes.

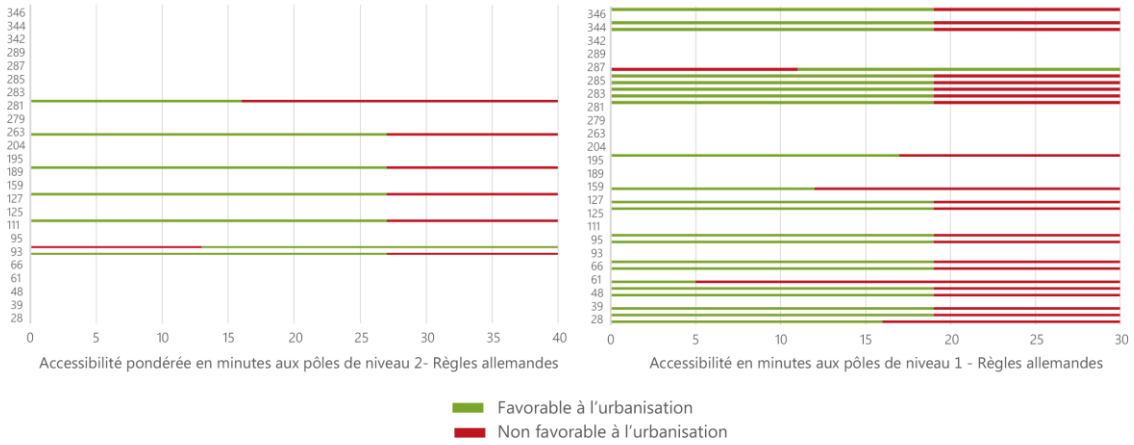


Figure 8. 13 : Contraintes liées à l'accessibilité pondérée pôle de niveau 2 et celles au pôle de niveau 1 dans les règles allemandes.

L'accessibilité niveau 2 est souvent retrouvée en complément de l'accessibilité pondérée niveau 1. Si l'importance des pôles de niveau 2 est réduite au regard de celle du pôle de niveau 1 dans le premier cas, cette dernière est donc appuyée dans un second temps. La hiérarchie de cette influence est également intéressant dans la définition des zones préférentielles. Pour finir, l'accessibilité niveau 1 est mobilisée épisodiquement. Ce point illustre la pluralité des comportements pouvant être observés au sein d'un même espace, et dans une mesure plus ou moins importante. Au vu de sa faible influence dans le jeu de règles allemand, et de l'absence de couplage à une autre variable, l'accessibilité niveau 1 n'as pas été prise en compte dans la définition de tendances générales d'accessibilité en Allemagne. Les tendances sont regroupées dans la table Table 8. 17.

Table 8. 17 Tendances d'accessibilités observées au sein du jeu de règles allemand

Tendances	Accessibilité pondérée Niveau 1	Accessibilité Niveau 2	Accessibilité pondérée Niveau 2	Accessibilité Niveau 1
1	≤ 32			
2	14 < x ≤ 45		≤ 27	
3	29 < x ≤ 51			
4		≤ 19		

La représentation cartographique des tendances fait ressortir l'influence positive des pôles de niveau 2. Comme précédemment, pour une meilleure lisibilité, les tendances extraites du jeu de règles allemand ont été appliquées à l'ensemble de la zone.

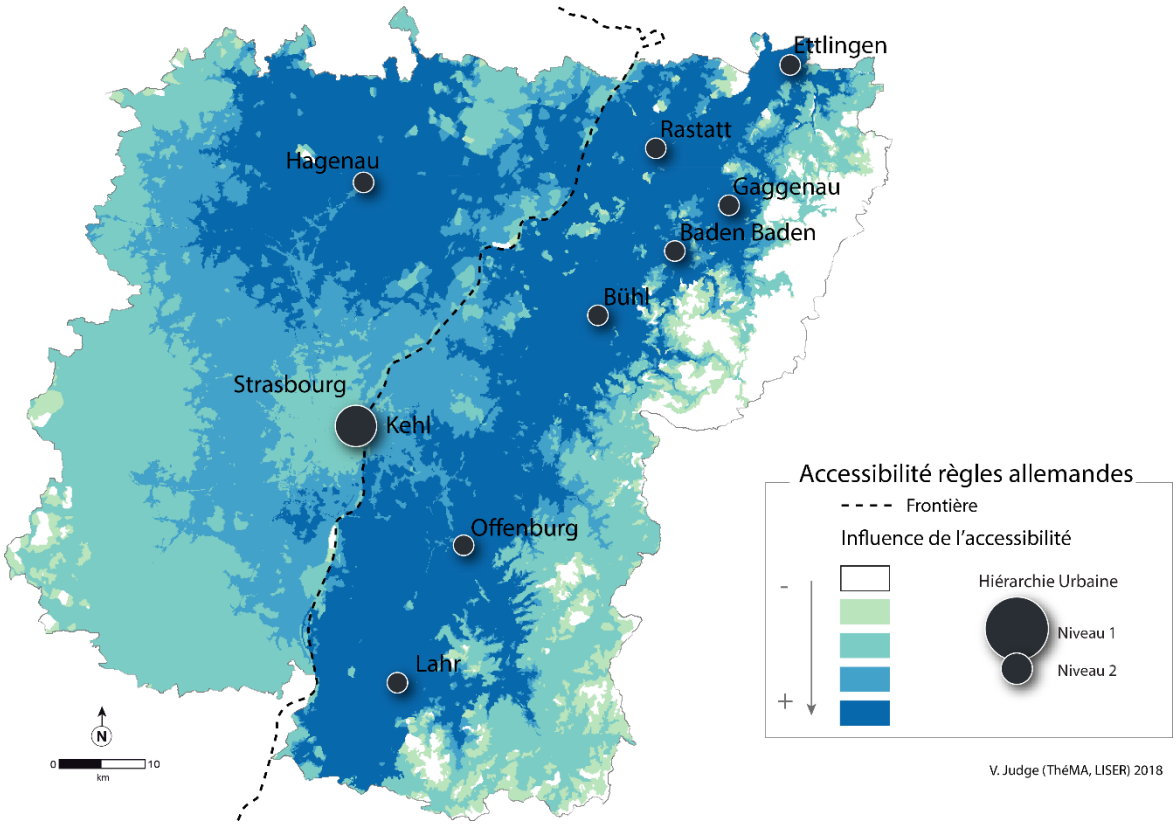


Figure 8. 14 : Zones préférentielles d'urbanisation allemandes au regard de l'accessibilité aux différents pôles urbains.

L'attractivité périurbaine observée, notamment autour de Strasbourg, apparaît également dans un second temps (Figure 8. 14). L'accès aux pôles de moindre importance est donc privilégié. Ce point ne limite pas totalement l'accès au pôle principal mais traduit une acceptation d'un temps de trajet plus conséquent pour l'atteindre.

Règles transfrontalières

Sur les 228 règles transfrontalières, 157 mobilisent des variables d'accessibilité, soit plus de 60 %. L'accessibilité niveau 1 pondérée influence également en majorité les règles transfrontalières, suivit comme dans le cas allemand de l'accessibilité niveau 2, niveau 2 pondérée puis niveau 1. Les seuils d'accessibilité pondérée niveau 1 sont principalement des limites minimum ou maximum (Figure 8. 17) et sont rarement mis à contribution en temps qu'intervalle contrairement au cas Allemand. Le seuil des 33 min apparaît, comme dans le cas des règles françaises. Cette contrainte est cependant complétée par une limite au regard du temps de trajets vers le pôle de niveau 1 ou encore vers un pôle de niveau 2 (Figure 8. 16). Ainsi, la contrainte sur le pôle de niveau 1 peut imposer un temps minimum de trajet, interdisant la trop forte proximité, tout en restant à moins de 40 min de Strasbourg. La contrainte sur les pôles de niveaux 2 est également renforcée à moins de 22 min.

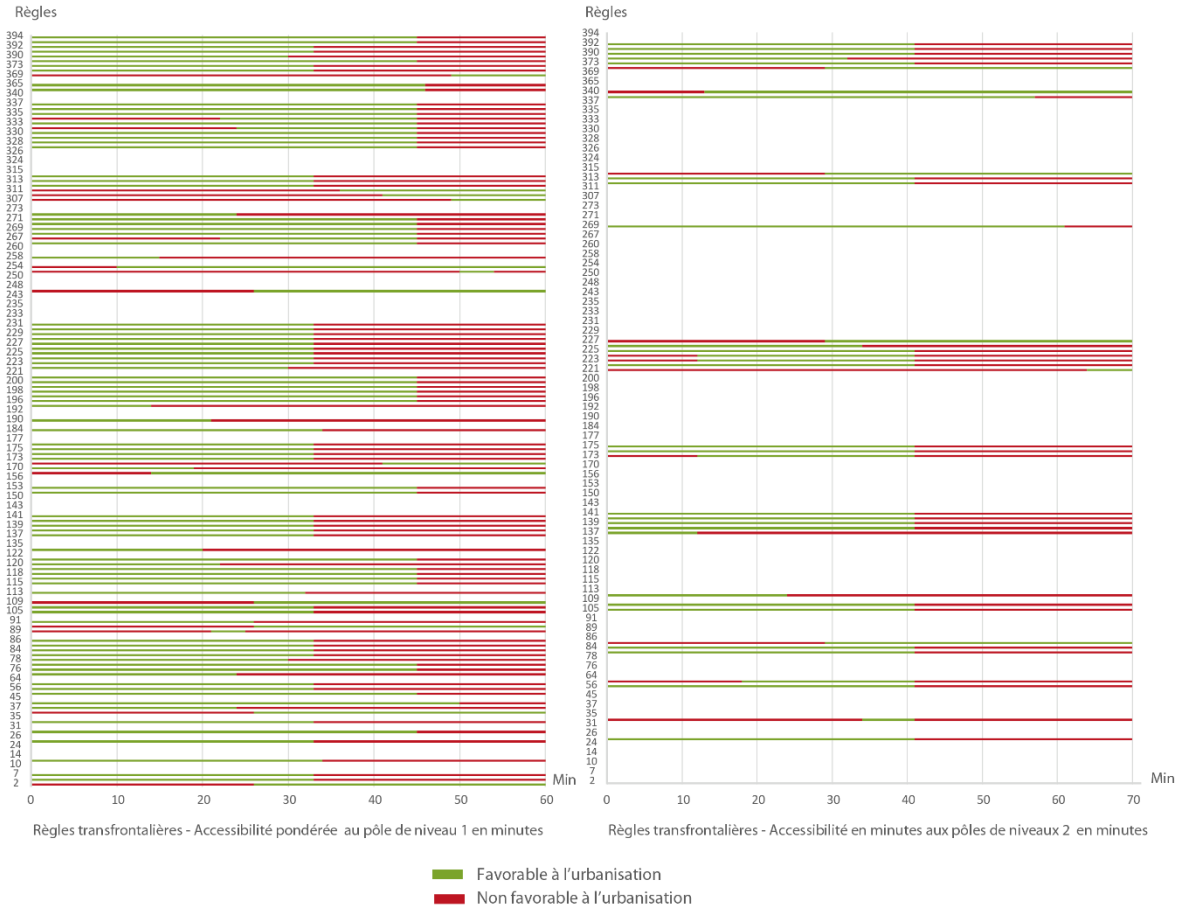


Figure 8. 15 : Contraintes liées à l'accessibilité pondérée pôle de niveau 1 et au pôle de niveau 2 dans les règles transfrontalières.

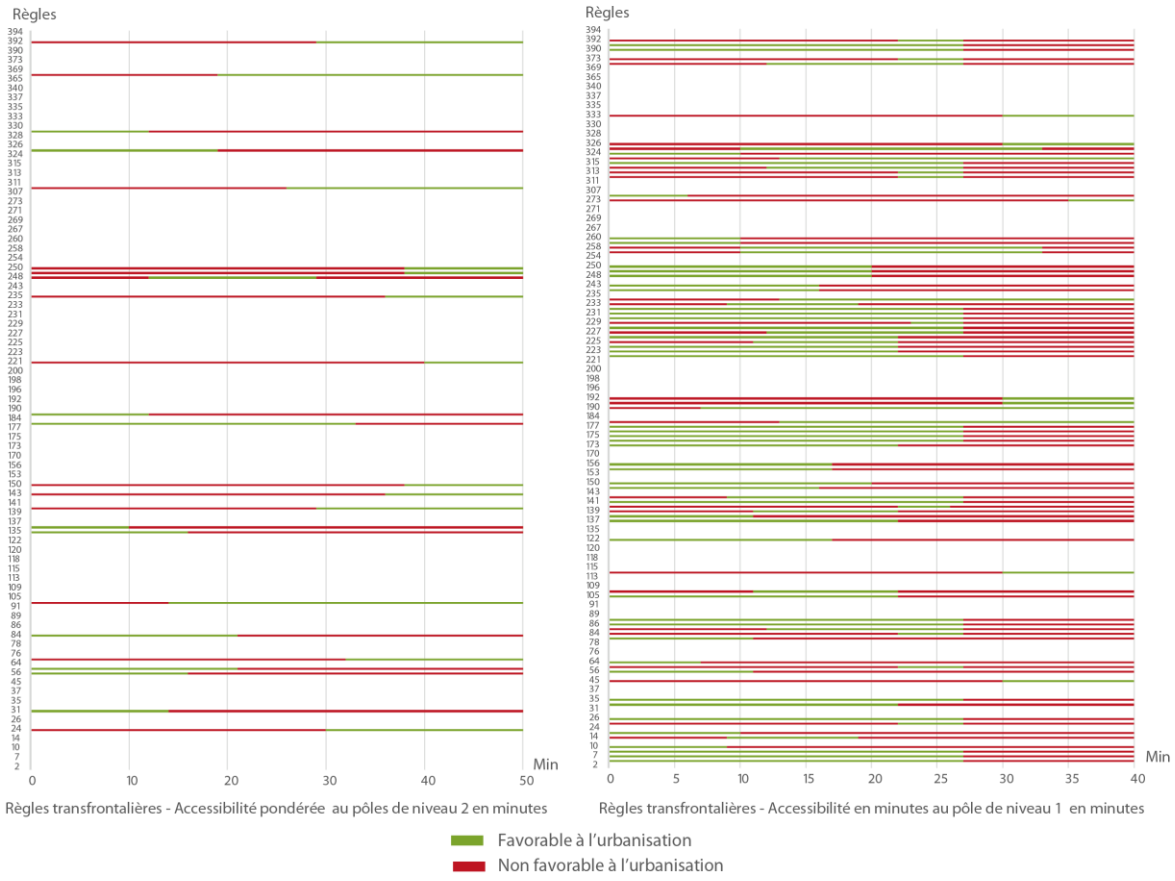


Figure 8. 16 : Contraintes liées à l'accessibilité pondérée pôle de niveau 2 et au pôle de niveau 1 dans les règles transfrontalières.

Il est plus délicat de résumer les seuils extraits pour l'accessibilité pondérée niveau 2. Une partie des règles présente des seuils limitant, imposant une certaine proximité, variant entre 10 min et 30 min. L'autre partie impose un trajet minimum, s'appuyant là aussi sur une limite unique variant de 10 min à 36 min. Chaque seuil étant mobilisé une seule et unique fois. Une des différences principales relève donc de la tolérance dans la longueur du trajet maximum. La tendance 2 illustre une localisation stratégique quel que soit le niveau du pôle à atteindre. Cet élément traduit bien un intérêt prédominant pour l'accessibilité au pôle, relayant à un second plan la localisation nationale. Les tendances décrites précédemment d'après les figures 15 et 16 sont résumées dans la table ci-dessous.

Table 8. 18 Tendances d'accessibilités observées au sein du jeu de règles transfrontalier

Tendances	Accessibilité pondérée Niveau 1	Accessibilité Niveau 2	Accessibilité pondérée Niveau 2	Accessibilité Niveau 1
1	≤ 32 ?			
2	≤ 33	22 <		12 < x ≤ 41
3	≤ 33			41 <
4	≤ 45			
5	> 20			
6		≤ 33		
7		10 <		
8			≤ 30	

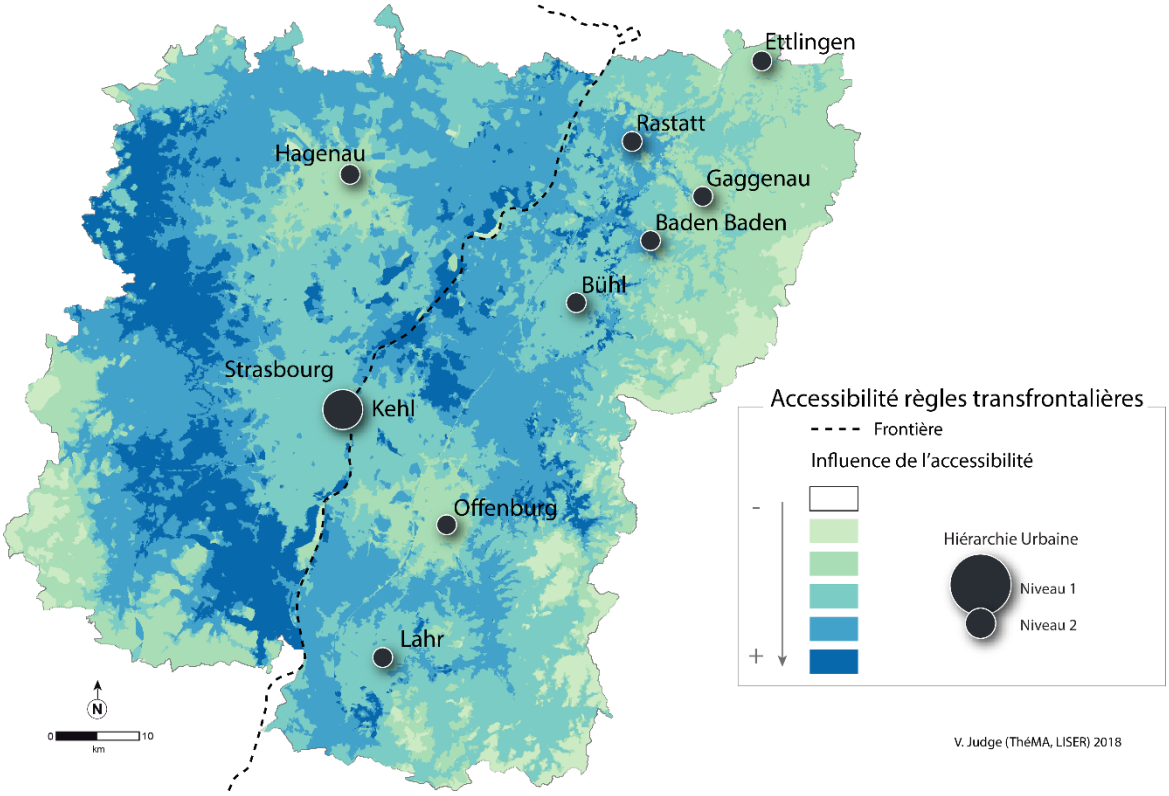


Figure 8. 17 : Zones préférentielles d'urbanisation transfrontalière au regard de l'accessibilité aux différents pôles urbains.

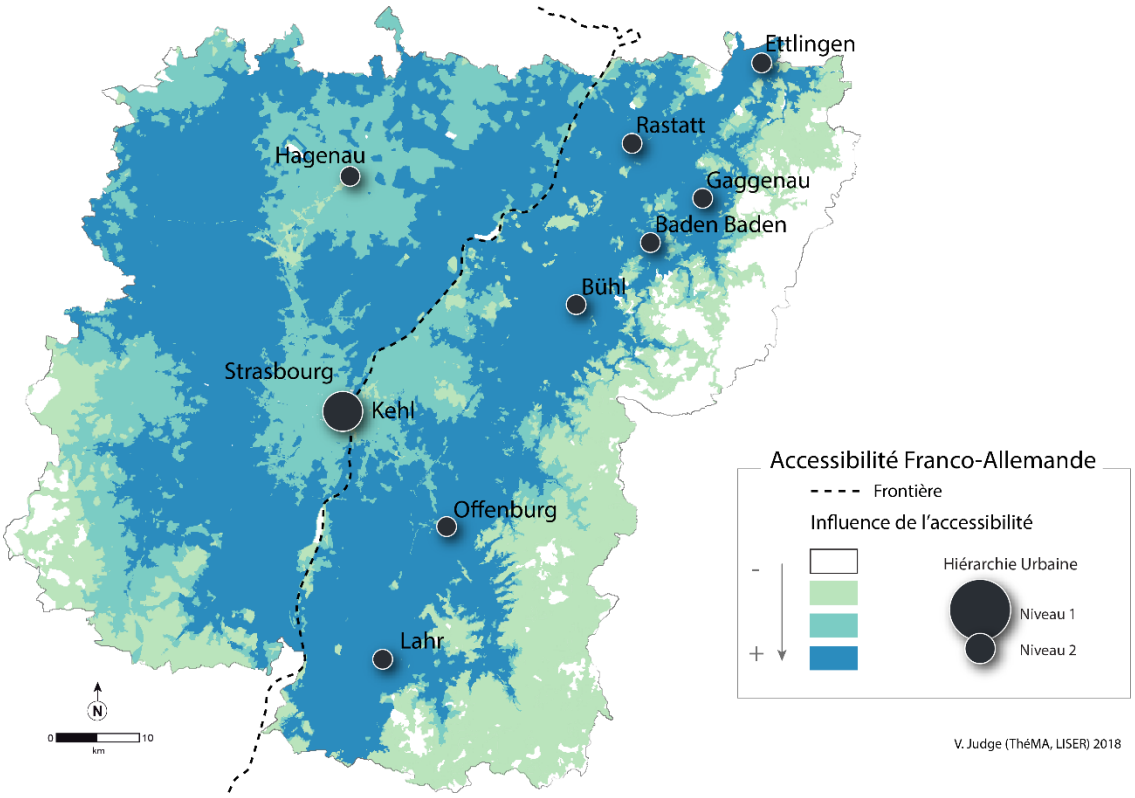


Figure 8. 18 : Zones préférentielles d'urbanisation franco-allemande au regard de l'accessibilité aux différents pôles urbains.

La représentation spatiale des zones préférentielles d'urbanisations extraites de la mobilisation des variables d'accessibilité dans les règles transfrontalières bénéficie d'une discrétisation plus riche (6 classes contre 4 pour les règles nationales). Le nombre de règles composant les jeux nationaux peut être directement lié à ce résultat. Cette représentation permet de hiérarchiser l'intérêt des pôles de niveau 2, certains comme Offenburg présentant un intérêt moindre. La périphérie urbaine est toujours plus attractive, une hiérarchie apparaît néanmoins. La zone frontalière localisée entre Strasbourg – Kehl, Haguenau et Bühl présente un intérêt particulier, égal aux périphéries Nord-Ouest et Sud de Strasbourg.

Les préférences en matière d'urbanisme au regard de l'accessibilité sur la zone d'étude doivent être considérée dans le cadre établi par les pôles sélectionnés, et dans le choix d'étudier une zone qui reste potentiellement sous l'influence de la frontière étatique. Ainsi, les zones appartenant à la première catégorie, présentant un potentiel d'urbanisation nul au regard de l'accessibilité aux pôles sélectionnés, pourrait être envisagé pour urbanisation en fonction des critères d'occupation du sol dans le voisinage (certaines règles ne dépendent pas de l'accessibilité), mais également en raison de la localisation d'un pôle de niveau 2 à proximité, mais hors périmètre d'étude.

La mise en relation des tendances transfrontalières (Figure 8. 17) et nationales (Figure 8. 18) permet d'observer les zones préférentielles portant un enjeu important par la mise en concurrence des deux comportements. Notamment pour la France, les zones péri-urbaines strasbourgeoises sud et ouest, ainsi que l'espace adjacent à la frontière, localisé au sud-est d'Haguenau. Pour l'Allemagne, les zones d'intérêt majeur sont, dans un cas comme dans l'autre, moins nombreuses. Pour autant, elles existent dans le quart nord de la partie allemande de la zone d'étude.

Conclusion

L'analyse des zones préférentielles d'urbanisation en fonction de l'accessibilité permet d'établir un zonage hiérarchique du potentiel en matière de développement urbain, basé sur l'apprentissage automatique. Ce dernier montre un zonage qui avantage le péri-urbain ou les pôles de seconde zone, et souligne des différences dans la hiérarchie des zones favorables à l'urbanisation suivant la nature des règles.

Conclusion chapitre 8

Le présent chapitre établi explicitement les types de voisinages favorables à l'urbanisation en fonction de la nature des règles (françaises, allemandes, transfrontalières). Il permet de mettre en lumière l'influence des divers types d'occupation du sol en fonction de leur distance dans le voisinage d'une cellule urbanisée et d'établir le constat suivant : les distances et seuils retenus lors de l'apprentissage automatique varient suivant la présence ou non de contrainte de localisation au sein d'un État en particulier, et suivant l'État en question. Ainsi, la composition du voisinage favorable à l'urbanisation en France varie de celle retenu en Allemagne dans le cadre des règles qui accordent le l'importance à la localisation étatique. Les règles transfrontalières présentent, elles aussi, des caractéristiques propres. Néanmoins, l'analyse permet également d'extraire des caractéristiques communes.

La cartographie issue de l'analyse des variables d'accessibilités fournit un zonage hiérarchique représentant le potentiel d'urbanisation, toujours en fonction de la nature des règles analyser. Ces résultats soulignent ici aussi les différences émergeant de cette analyse, ou les zones de forts potentiels sont plus étendus (zonages nationaux) ou au contraire restreints (zonage transfrontalier). Ainsi, les règles ne privilégiant pas la préférence nationale privilégieraient une accessibilité optimisée aux pôles d'activités.

Chapitre 9

Les enjeux de la modélisation : Analyse à travers la prospective

Un approche modélisatrice a été privilégiée pour étudier le processus de développement urbain sur la zone d'étude dans le but d'étudier l'homogénéité des pratiques de part et d'autre de la frontière franco-allemande. L'aspect novateur de ce travail réside dans le couplage d'un automate cellulaire et d'un arbre de décision et dans l'analyse approfondie du jeu de règles créé et calibré automatiquement. En effet, les modélisations réalisées sur la base de méthodes d'apprentissages automatiques et statistiques ne permettent pas toujours une caractérisation aussi transparente des processus dynamiques de changements d'occupation du sol. Les règles issues de l'apprentissage automatique par arbre de décision présentent donc l'avantage de pouvoir être analysées en détails et de manière explicite. Cette démarche a permis de répondre aux objectifs de notre travail de recherche. L'étude de la structure du jeu de règles (nombre de règles, types de transitions, types de variables, etc.) et de sa composition (variables et seuils) permet d'appuyer l'influence de la localisation des cellules urbanisées dans l'un ou l'autre des États frontaliers. Elle permet également d'extraire et caractériser un développement urbain transfrontalier, s'affranchissant d'une localisation étatique spécifique en France ou en Allemagne. Il devient, de ce fait, également possible de caractériser l'homogénéité de la zone d'étude. Ce point est abordé dans la première sous-section. Mais si la méthode permet de répondre à nos objectifs, comme toute démarche modélisatrice elle présente également des limites. Ces limites et perspectives sont présentées d'après trois piliers permettant d'élaborer et d'exécuter la méthode : les données, l'échelle, l'algorithme. Elles sont discutées et mises en perspectives avec les pistes d'amélioration inhérentes à leur présence, pour compléter la première sous-section. Comme plusieurs pistes de recherches sont envisageables sur la base de ce travail, la section 2 permet de présenter une réflexion pour approfondir l'approche thématique. La section 3 quant à elle se penche sur les autres cadres qui pourraient bénéficier de l'utilisation de notre outil et de la méthode de modélisation.

1. RÉSULTATS ET LIMITES

1.1. Mesurer la valeur des résultats

L'apprentissage par arbre de décision permet de décrire thématiquement des processus de développement urbain à travers l'analyse du jeu de règles qui en est issu. Dans notre travail, l'évolution des cellules est définie d'après l'occupation du sol dans leur voisinage et leur accessibilité aux agglomérations majeures. L'utilisation d'un AC pour la simulation du processus dynamique de développement urbain, impose, d'après la définition classique, de spécifier l'état initial de la cellule. Il contraint donc la transition par ce biais. L'ajout des variables de localisation géographique à notre jeu de données, donc disponibles lors du processus d'apprentissage, permet également de contraindre l'application de la règle à une zone géographique particulière, notamment dans l'un ou l'autre des pays frontaliers. Dès lors, nous pouvons considérer que l'évolution des cellules sur la zone d'étude peut présenter deux contraintes majeures :

- Contrainte de l'état initial qui favorise la transition
- Contrainte spatiale étatique qui permet la transition

Quatre cas de figures peuvent donc être retrouvés dans notre jeu de règles :

- aucune contrainte,
- contrainte sur l'état initial uniquement,
- contrainte sur la localisation étatique uniquement,
- contrainte sur l'état initial et la localisation étatique.

L'analyse des contraintes peut ainsi nous permettre d'évaluer le caractère général des règles et de le définir par un certain degré (Figure 9. 1). Ainsi certaines règles ne présentent aucune des deux contraintes (1er degré). Elles peuvent, par conséquent, être considérées comme les plus générales. Leur caractère universel est donc maximal à l'échelle de la zone d'étude.

Le chapitre 7 section 2.1, mettait en lumière, grâce à l'étude des différents types de transitions urbaines, un échantillon de règles qui s'affranchit de l'état initial de la cellule (96 règles). En d'autres termes, une transition urbaine devient possible quel que soit l'occupation du sol initiale. S'il traduit une description générale du processus, ce résultat est d'autant plus intéressant qu'il permet également de s'éloigner de la définition classique de l'automate cellulaire imposant par définition un état initial spécifique. Sur les 96 règles s'affranchissant d'un état initial, 80 ne dépendent pas non plus d'une localisation spécifique au regard de la frontière. Elles représentent donc la généralisation optimale (1er degré) du processus au regard des contraintes étudiées dans ce travail de recherche, notamment la localisation vis-à-vis de la frontière. Une partie des règles présente un degré de liberté moins élevé, sans contrainte sur l'état initial mais localisé dans l'un ou l'autre des pays frontaliers (16 règles). Au contraire, le cas de figure inverse est également retrouvé (148 règles) : l'état initial est contraint mais pas la localisation dans l'un ou l'autre des pays de la zone d'étude (2^{e} degré). Pour finir, une part du jeu de règles (151 règles) est contrainte suivant les deux

critères (3e degré). La gradation du caractère universel des règles issues de l'apprentissage automatique par arbre de décision est résumée à travers les différents degrés de contraintes exprimés précédemment (Figure 9. 1). Ainsi, 80 règles présentent le degré le plus important d'universalité sur notre zone d'étude, contre 151 règles qui sont, elles, contraintes suivant la localisation dans l'un ou l'autre des États, et par l'état initial de l'occupation du sol nécessaire à la transition urbaine.

	Contrainte spatiale étatique	*Contrainte état initial*	
1er degré	Non	Non	*80 règles*
2ème degré	Non	Oui	*148 règles*
	Oui	Non	*16 règles*
3ème degré	Oui	Oui	*151 règles*

Universalité

Spécificité

Figure 9. 1 : Caractère spécifique des règles de transitions

L'analyse des données, par le biais de l'apprentissage, permet donc de hiérarchiser le degré de généralisation des règles obtenues. En d'autres termes, elle permet d'évaluer l'importance du caractère « universel » du processus de développement urbain sur la zone d'étude.

Cette dernière analyse de la composition du jeu de règles permet une fois encore de souligner l'apport d'une analyse par arbre de décision. Cet outil, mettant en lumière les différences et similitudes dans la description thématique du processus de développement urbain, permet également d'entrevoir la part généralisable de ce même processus à travers l'élaboration de règles répondant au minimum de contraintes dans le cadre d'étude établi. Si l'apport de l'apprentissage automatique par arbre de décision s'avère enrichissant thématiquement et efficace techniquement au regard de l'aspect simple de sa mise en œuvre, son utilisation concrète fait également apparaître certaines limites de notre approche.

1.2. Prendre conscience des limites

L'apprentissage automatique est un outil d'analyse puissant, permettant l'extraction de tendances. L'analyse de ces tendances permet d'acquérir des connaissances au regard du cadre méthodologique développé. Les arbres de décisions, comme les automates cellulaires, sont particulièrement appréciés car ils sont simples à manipuler, au regard d'autres outils ou méthodes, et permettent d'acquérir des résultats extrêmement explicites sur la base des données d'entrée implémentées. Néanmoins, la principale limite de ce travail réside dans la combinaison de deux biais probablement importants :

- La qualité des données d'occupation du sol utilisées pour l'apprentissage. En effet, comme discuté précédemment (Chapitre 3) le choix a été fait de sélectionner une donnée libre et disponible sur l'espace transfrontalier. Si Corine Land Cover présente

une source extrêmement riche pour l'analyse de données d'occupation du sol à l'échelle européenne, l'importance du biais est proportionnelle à la diminution de l'étendue de la zone d'étude envisagée. Il se peut alors que certaines transitions soient sur- ou sous- estimées en raison de la précision de l'acquisition de la donnée de base. Il n'est pas à exclure donc que certaines règles traduisent en partie des transitions fictives.

- Le choix du meilleur modèle. En effet, il dépend directement des critères d'évaluations utilisés dans le cadre de cette étude (Chapitre 6). Il est donc lié au choix du modélisateur lui-même, comme toute modélisation.

La combinaison des deux points évoqués ci-dessus rend indispensable une prise de recul sur les résultats discutés précédemment. Il convient notamment d'envisager les valeurs absolues des seuils qui présentent les proportions adéquates des différents types d'occupation du sol présents dans les voisinages des règles de différentes natures, extraits en pourcentages, avec précaution. En effet, les choix méthodologiques peuvent l'avoir impacté (hiérarchie des critères de performances du modèle, etc.). Pour autant, si la valeur absolue du seuil peut effectivement être questionnée, la généralisation des tendances extraites à travers leur analyse, elle, ne doit pas faire l'objet d'une remise en cause. En effet, en partant du postulat que le modèle est faux, il reste faux de la même manière de part et d'autre de la frontière étatique. Toute chose égale par ailleurs, ce constat souligne alors que la différence observée au regard de l'environnement favorable au développement urbain entre ces deux espaces est fondamentalement effective. Toutefois, les différences observées entre la France et l'Allemagne, sont atténuées par la présence de tendances applicables à la totalité de la zone d'étude quelle que soit la localisation étatique envisagée.

Avoir conscience des limites que présente ce travail permet d'envisager des pistes pour corriger, améliorer, aller plus loin dans le développement du modèle (qualité du jeu de règles) ou dans l'exploitation des résultats (analyse).

1.3. Améliorer la qualité du jeu de règles

Plusieurs pistes d'améliorations peuvent donc d'ores et déjà être envisagées pour approfondir ou élargir le champ des possibles, notamment lors de la phase de création du jeu de règles, intrinsèquement liées à la donnée source et à l'utilisation de l'algorithme d'apprentissage. Les perspectives sont présentées selon 8 orientations :

- L'application de la méthodologie à partir de données plus fines.

Une telle application permettrait de définir plus précisément si une part des contre-performances peuvent être imputées à la qualité de la donnée. Nous entendons ici par contre-performances, le nombre et la taille des règles nécessaires à l'élaboration d'un modèle permettant de retrouver la quasi-totalité des transitions urbaines tout en minimisant la surestimation des transitions. Ce test permettrait d'évaluer l'influence de la qualité des données d'occupation du sol en entrée et la différence entre l'utilisation de données Corine Land Cover et celle de données plus fines sur une même zone.

Par ailleurs, l'acquisition d'une donnée plus fine pourrait permettre d'intégrer les changements d'utilisation du sol pour étudier l'influence des conversions. Un changement d'utilisation du sol local, par exemple une modification de denrée cultivée pourra avoir un impact significatif à l'échelle supérieur (dans la région), qu'une transition d'une occupation du sol à une autre comme celle d'une occupation du sol agricole à une occupation du sol urbaine. L'arrosage nécessaire à certaines cultures peut avoir une influence significative sur les modes de gestion de la ressource en eau. L'esthétique du paysage peut également être modifié, rendant la zone plus ou moins attractive, etc.

- L'introduction de nouvelles données.

Le travail de thèse réalisé ici a volontairement été focalisé sur l'utilisation des données d'occupation du sol et d'accessibilité en raison des points évoqués et défendus dans le chapitre 3 (contexte transfrontalier, élaboration d'un modèle simple et d'une méthode transposable). L'introduction d'autres données dans le processus d'apprentissage est également une piste d'amélioration du modèle, certainement d'un point de vue thématique, mais potentiellement du point de vue de la performance du modèle également. L'ajout de variables explicatives, comme dans le cas des modèles présentés chapitre 4, est une piste à explorer. Par ailleurs, l'analyse d'autre facteurs dont on suspecte une influence sur le développement urbain, est facilité par la flexibilité de notre outil. Ainsi, des données concernant la géographie physique (topographie, etc.), permettant une description socio-économique (zonage caractérisant les types d'habitats, etc.) ou encore fournissant une description réglementaire des possibilités en termes d'urbanisation, peuvent être testées.

- L'application de la méthode sur une autre zone d'étude.

L'utilisation de données libres disponibles à l'échelle européenne rend la méthodologie transposable. Les résultats produits sur une autre zone transfrontalière peuvent par conséquent être comparés, notamment au regard de la composition du jeu de règles, suivant la même démarche que celle développée dans ce travail. Cette application permettrait également d'établir plus précisément la robustesse des résultats thématiques. Par exemple, une application sur les frontières franco-belge, ou franco-luxembourgeoise, permettrait de comparer les règles françaises élaborées sur une zone frontalière franco-allemande, et celles

élaborées sur une autre zone frontalière française. Dès lors, il sera possible d'établir si ces règles présentent des similitudes qui permettent de caractériser un comportement français en général.

- L'utilisation d'un autre algorithme d'apprentissage par arbre de décision.

Plusieurs algorithmes peuvent être utilisés pour effectuer l'apprentissage par arbre de décision. La littérature n'établit pas clairement de hiérarchie sur la performance de l'apprentissage des différents algorithmes (Chapitre 5). Pour autant, il n'est pas exclu que le choix de l'algorithme ait une incidence sur le résultat de l'apprentissage, notamment sur le choix des variables mises à contribution. Le test d'un autre algorithme pourrait, dès lors, appuyer ou non les résultats thématiques présentés dans ce travail.

- La comparaison des résultats obtenus avec un autre procédé d'apprentissage automatique.

L'utilisation d'une autre méthode d'apprentissage pose la question de la comparaison des résultats produits. Il serait intéressant d'être en mesure de comparer les résultats issus d'un modèle basé sur l'apprentissage par réseaux de neurones par exemple. Pour autant la comparaison des résultats restera limitée. Il sera par exemple possible de comparer la performance globale de prédiction du changement et du non-changement, ou encore d'un point de vue technique, le temps nécessaire à l'apprentissage ou à la simulation. Pour autant il n'est, par exemple, pas possible d'analyser les règles issues d'un apprentissage par réseaux de neurones, sur le même modèle que l'analyse réalisée pour le jeu de règles issue de l'apprentissage par arbre de décision. Toutefois, une évaluation de la performance des modèles suivant plusieurs méthodes permet une meilleure appréhension de la robustesse des résultats produits. Cette démarche est particulièrement pertinente dans le cas d'une utilisation prospective puisque c'est alors un pari sur l'avenir qui est lancé.

- La comparaison des résultats et des données d'enquêtes

L'élaboration d'un modèle simple permet de « discuter » du processus de développement urbain à travers une approche thématique. Cette démarche a été initiée pour avoir l'opportunité d'extraire des informations sur les comportements et besoins individuels et collectifs reflétés par le développement urbain observé sur la zone d'étude. L'application sur une zone d'étude transfrontalière permet, par ailleurs, de souligner les différences ou similitudes dans les caractéristiques favorables au développement urbain en fonction de la localisation dans l'un ou l'autre des pays frontaliers, donc dans un contexte géographique particulier. Il serait alors intéressant de comparer les seuils d'accessibilités, ou le voisinage favorable au développement urbain, extrait par l'algorithme, et une donnée d'enquête qui interroge directement les individus sur leurs préférences vis-à-vis de la localisation ou de l'environnement. Le parallèle entre l'information obtenue de façon « objective » par

l'analyse des résultats fournis par la machine et celle issue directement des individus ou acteurs pourrait ainsi être fait.

- L'intégration des changements d'occupation du sol entre les autres catégories d'occupation du sol.

Dans le cadre d'une approche prospective, une limite de ce travail est la modélisation des transitions urbaines uniquement. L'objectif de ce travail était d'étudier spécifiquement le développement urbain, en tant que processus à même de refléter les différences culturelles. Pour autant, une approche prospective, pour être cohérente, devra s'attacher à modéliser les changements présents entre les autres classes d'occupation du sol. Ainsi, l'évolution dynamique complète des voisinages serait effective. La méthode intégrée dans LucSim permet, par ailleurs, cette approche. Pour parfaire l'approche dynamique globale, tout en restant focalisé uniquement sur la donnée d'occupation du sol et celle d'accessibilité, il serait également intéressant d'intégrer des réflexions sur l'évolution du réseau routier. Ainsi, l'accessibilité ne serait plus, non plus, statique au court de la simulation. L'évolution du réseau pourrait être inspirée des projets de planification ou d'aménagement du territoire, ou des plans réglementaires.

- La sélection des règles en fonction de leur « efficacité »

Une classification des règles d'après leur efficacité peut faciliter l'analyse et l'exploitation du jeu de règles. Il est alors possible de concentrer son attention sur un échantillon, à travers la sélection de celles présentant un intérêt majeur. Une conséquence directe sera par exemple la possibilité de réduire le nombre de règles contenues dans le modèle, par la sélection des règles les plus importantes. La performance globale du modèle peut, par exemple, être améliorée en éliminant les règles au faible pouvoir prédictif (ratio « vrais positifs/faux positifs » défavorable). L'évaluation quantitative du nombre de cellules urbanisées en fonction des règles utilisées peut également servir de base pour établir un objectif de développement en fonction de la surface urbanisée. Il est alors possible de calibrer le développement urbain sur la base d'une projection de Markov ou d'objectif en matière d'aménagement, établi par les plans directeurs. Il peut s'agir de limiter la surface urbanisée, ou de favoriser un type d'urbanisation en particulier tel l'urbanisation des dents creuses par exemple. D'un point de vue technique, la réduction du nombre de règles contenues dans un jeu permet encore l'amélioration du temps de simulation. Ce point est néanmoins négligeable au regard d'un temps de simulation déjà très court.

Cette sous-section montre la multitude des pistes de réflexion envisageables, sur la base du travail réalisé dans le cadre de cette thèse. Celle de la sélection d'un échantillon de règles en fonction du type de développement urbain modélisé, est par ailleurs présenté dans la sous-section suivante. Cette démarche a pour objectif d'illustrer une des perspectives offertes par l'utilisation de notre méthode, dans le cadre, par exemple, d'une application prospective.

Cette application est présentée en parallèle de celle d'un scénario basé sur une projection de tendance de l'urbanisation.

2. PERSPECTIVE ET MODÉLISATION

Comme mentionné dans la section précédente, toute analyse menant à une interprétation plus approfondie du jeu de règles doit être mise en perspective avec les choix de modélisation, notamment ceux impliquant le paramétrage de l'algorithme d'apprentissage. Certaines pistes pour améliorer l'interprétation peuvent néanmoins être envisagées en l'état. Cette sous-section présente une réflexion complémentaire sur une exploitation plus approfondie du jeu de règles, notamment à travers la classification des règles en fonction de leur efficacité.

2.1. Qu'est-ce qu'une bonne règle ?

Dans ce travail de recherche, le parti pris a été de considérer chaque règle permettant de prédire au moins une transition urbaine des données observées, comme d'efficacité égale (Chapitre 6). Ce postulat s'appuie sur l'approche de la modélisation dans une démarche de simulation prospective. En effet, une règle ne représentant qu'un nombre limité de transitions observées à un instant t, peut être amenée à concerner un nombre plus important de cellules à un instant t+1 d'après la totalité des changements effectifs sur la zone d'étude à l'étape de simulation précédente.

La classification des règles en fonction de leur efficacité dépend néanmoins d'une réflexion sur la définition « d'efficacité ». Comme dans le cas de la caractérisation de l'influence d'une variable au sein du jeu de règles (chapitre 7), la notion d'efficacité doit être définie avant de pouvoir sélectionner les « meilleures » règles visant à contraindre la simulation. La notion d'efficacité est bien souvent relative, et la plupart du temps subjective. Relative car une performance considérée comme efficace à un certain niveau ou dans un certain cadre ne l'est pas forcément quand ce dernier change. Subjective car elle sera de fait défini par un individu, un groupe d'individus voire par une communauté scientifique, en accord avec un certain cadre et des objectifs spécifiques. Deux approches peuvent soutenir la classification de règles en fonction de leur efficacité, et retiendront particulièrement notre attention :

- Une sélection pour réduire l'échantillon de règles analysées et favoriser les règles courtes. L'objectif serait de faciliter l'analyse thématique par la diminution du nombre de règles et par celle de leur complexité souvent corrélée à leur longueur. Une classification peut permettre de restreindre l'analyse aux règles les plus efficaces. Dans ce cadre, une règle efficace serait définie en considérant une limite maximale du nombre de conditions la composant, puisqu'une règle courte est plus simple à analyser thématiquement, qu'une règle longue composée d'une plus grande quantité de variables.

- Une sélection pour atteindre un nombre déterminé de transitions, sans contraindre la simulation au préalable (Chapitre 4). Si l'utilisation du modèle s'inscrit dans une démarche prospective, le modélisateur ne peut se satisfaire d'une surestimation trop importante du nombre cellules urbanisées. Le nombre de cellules à urbaniser peut, par exemple, être défini d'après une projection de Markov basée sur l'observation de l'urbanisation sur une période passée. Dans ce cas, la définition de l'efficacité est alors basée sur la capacité à prévoir les transitions observées, en limitant la surestimation. Une règle efficace n'est donc plus une règle courte en priorité, mais une règle limitant les erreurs de prédiction.

La question soulevée ici porte donc sur le moyen le plus adapté de réaliser la sélection des règles les plus pertinentes au regard des objectifs de modélisation. En effet, notre démarche permet, par exemple, de caractériser une règle d'après sa capacité à localiser précisément une transition observée, en parallèle de celle à localiser une transition n'ayant pas été observée dans les données de références (surestimée/faux positifs). Certaines règles présentent une efficacité maximale pour localiser des transitions observées, d'autres pour minimiser le rapport entre les transitions correctement simulées et les erreurs de prédiction. Dans ce cadre, en considérant une règle au nombre de variables égal, deux questions se posent : est-il préférable d'avoir recours à une règle parfaitement juste, sans erreurs, mais dont le potentiel de prédiction des transitions observées est minime ? Ou est-il préférable d'avoir le meilleur ratio vrais positifs, faux positifs, qui permettra de maximiser la prédiction des transitions observées au risque d'accroître l'erreur ?

Par ailleurs, ces indicateurs de performance (vrai positif ou ratio vrais positifs/faux positifs) peuvent à leur tour être relativisés par rapport à la taille de la règle nécessaire. L'un ou l'autre des positionnements peut être pertinent suivant les objectifs liés à la modélisation. L'intention ici est bien de souligner la pluralité des approches pouvant être mobilisées suivant les besoins.

Au-delà de l'analyse technique concernant la structure de la règle, le type de développement qu'elle illustre peut également faire l'objet d'une attention particulière, notamment dans la construction de scénarios orientés. Ce point est illustré dans la sous-section suivante.

2.2. Quelles règles pour quels développements ?

La sélection des règles sur la base du développement spatial qu'elles génèrent est également une option envisageable pour étudier le développement urbain simulé par rapport à un type d'urbanisation particulier. Cette démarche peut, par exemple, être initiée pour étudier l'impact d'une politique d'aménagement spécifique au regard d'un objectif dans la consommation de l'espace par l'urbanisation. Pour ce faire, LucSim (Chapitre 4) a été doté d'un module spécifique permettant de lier la règle à la zone ou aux zones urbanisées qui en dépendent. Dans ce cadre, une classification des règles en fonction de leur type de

développement urbain peut soutenir l'élaboration d'un modèle à visée prospective. Il est alors possible d'illustrer l'impact d'un scénario orienté pour privilégier un type de développement (avantages et inconvénients). Dès lors, il est possible de sélectionner des règles qui permettent de combler les dents creuses, et limitent par la même occasion l'étalement urbain au rayonnement des zones déjà urbanisées, par exemple, pour élaborer un scénario en particulier. Cette démarche est illustrée à travers l'analyse du développement urbain simulé d'après la règle 1 de notre jeu de règles (Figure 9. 2).

Les zones urbaines concernées par le développement urbain issu de la règle 1 sont localisées au sein de la zone d'étude. Les zones concernées (en rouge) sont toutes localisées en Allemagne. Pour cause, son application est imposée à l'Allemagne uniquement, à travers la présence de l'Indice de Pays (Chapitre 6). Pour indication, la structure de la règle spécifie qu'elle n'est pas influencée par l'accessibilité. Son voisinage proche est principalement urbain. Cette occupation du sol influence positivement la transition à 200 m, 500 m, 1,5 km, 1,7 km et 3 km. Les terrains agricoles sont moins recherchés et dans un périmètre plus éloigné (à partir du kilomètre). La forêt et l'eau sont localisées autour de 2,5 km. Cette dernière contrainte rapproche donc les zones urbanisables de la frontière étatique, en imposant une certaine proximité aux zones immergées, principalement représentées sur notre zone d'étude par le Rhin dont la frontière suit le tracé. La règle 1 présente un intérêt dans le cadre d'un aménagement visant à réduire l'étalement urbain. En effet, elle caractérise le comblement de dents creuses particulièrement visible sur le zoom 4, mais également sur les zooms 1 et 3 (Figure 9. 2).

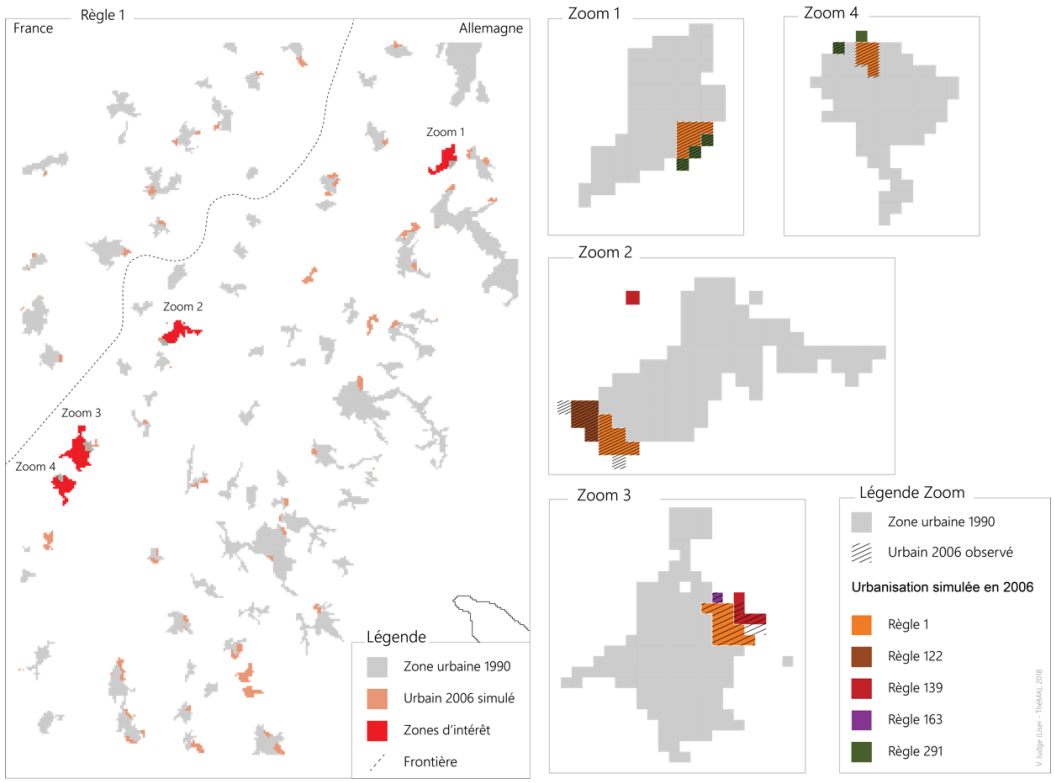


Figure 9. 2 : Focalisation sur le développement urbain illustré par la règle 1

D'après la Figure 9. 2, la règle 1 permet donc principalement d'illustrer le remplissage des dents creuses, ou un étalement urbain (Zoom 2). Cette règle peut donc être privilégiée dans le cadre de l'élaboration d'un scénario prospectif visant à orienter le développement futur dans cette direction.

L'analyse des autres règles présente également un intérêt. Une partie du développement urbain simulé, et visible dans les différents zooms, dépend d'autres règles de notre jeu. La règles 139 illustre plutôt le phénomène de mitage d'une urbanisation discontinue (zoom 2). Pour autant, elle génère également une urbanisation continue dans le zoom 3. Ce point est particulièrement intéressant à étudier dans le cadre d'une approche prospective puisqu'il permet de questionner la pertinence de l'urbanisation liée à la règle 139 dans le zoom 2. Si cette cellule peut être appréhendée comme une erreur (faux positif), elle présente néanmoins des caractéristiques favorables au développement urbain, en matière de voisinage et d'accessibilité, puisqu'elle répond positivement aux conditions d'une règle issue de l'apprentissage automatique. Dans ce cas, le modèle permet d'attirer l'attention sur des zones dont le potentiel est souligné et qui ne sont pas encore urbanisées. Cette approche de la modélisation est retrouvée dans les travaux de J.P. Antoni (2018). Dans le cadre de l'amélioration du modèle, une analyse plus approfondie des divergences et similitudes dans le voisinage ou l'accessibilité de ces deux zones peut permettre de sélectionner et d'introduire une donnée discriminante précisant le modèle et la description du processus d'urbanisation. La règle 291 observable sur les zooms 1 et 4 illustre quant à elle un étalement urbain concentré.

L'analyse spatiale du développement urbain issu des différentes règles semble donc exploitable dans la création de scénarios prospectifs d'urbanisation, orientés sur un type de développement en particulier. La simulation de la totalité du jeu de règles sur plusieurs étapes supplémentaires présente également une piste plausible dans le cadre de l'exploration d'un scénario basé sur une projection de tendances. Cette piste est présentée dans la suite de ce chapitre.

3. ADAPTATION À LA PROSPECTIVE ET À L'AIDE À LA DÉCISION

Les pistes de développements théoriques et techniques sont nombreuses. Une utilisation de l'outil et de la méthodologie dans le cadre d'une exploration prospective est, par exemple, une voie qui peut être envisagée dans la continuité de ce travail. Cette application pourra être développée dans un environnement qui privilégie l'approche académique, ou en partenariat avec des acteurs dont les besoins sont concrets et ancrés dans la réalité de l'aménagement d'un territoire. Ces deux approches sont bien souvent conjointes (Hewitt *et al.*, 2014).

Comme discuté dans la sous-section 1.2, la résolution et la qualité des données utilisées, l'étendue de la zone d'étude, ou encore l'échelle de l'analyse peuvent s'avérer limitantes dans le cadre d'une application opérationnelle. En effet, l'enjeu d'aménagement des

territoires n'était pas l'objectif principal de ce travail. Pour autant, le développement méthodologique a été envisagé en tenant compte de cette évolution potentielle. L'utilisation d'un automate cellulaire dont les règles sont issues de l'apprentissage automatique par arbres de décision n'est pas limitant dans le dialogue avec un public expert ou non. Au contraire, l'outil possède une interface visuelle, et les règles qui permettent de simuler le processus de développement sont simples et explicites. Si les caractéristiques de notre outil facilitent le travail sur une zone d'étude transfrontalière, et influencent le caractère transposable de la méthode, la volonté de transparence en vue d'une application opérationnelle a aussi été prise en considération en amont du développement de la méthode et des choix nécessaires à son élaboration.

L'utilisation opérationnelle des outils de modélisation en aménagement du territoire repose souvent sur une démarche prospective. Les bases et objectifs qui mènent à l'élaboration et à la simulation de scénarios prospectifs sont multiples. La sous-section suivante permet donc de revenir sur la démarche prospective et son application dans l'étude des changements d'occupation du sol, plus précisément dans celle des zones urbaines.

3.1. Modélisation et démarche prospective

J. de Courson (1999) dans son ouvrage sur « La prospective des territoires » cite l'acteur W. Allen : « la prévision est un art difficile, surtout quand il s'agit de l'avenir. » Cette citation, sans être dénuée d'humour, souligne pourtant un élément moteur de nombreuses études visant la compréhension de phénomènes ou processus spatiaux, parmi eux celui du développement urbain : la compréhension du présent pour envisager le futur. La prospective tente, alors, d'apporter des éléments de réponse lors de la démarche qui consiste à envisager le futur. Cette exploration, effectuée par le biais d'une modélisation prospective, peut être envisagée suivant trois orientations (Antoni, 2010) : i – Le prolongement dans le futur des tendances passées observées ; ii – La projection de tendances ajustées à des objectifs de planification ou à des objectifs institutionnels prévus ou suggérés réglementairement, et qui tente d'estimer une ou plusieurs issues possibles ou envisageables d'après les connaissances acquises au moment de la modélisation ; iii – La prospective qui a pour objectif principal d'illustrer ou de fournir des éléments de réflexion sur le cheminement nécessaire pour atteindre un ou plusieurs futurs bien définis.

En effet, de par son étymologie latine *pro-jicere*, littéralement « jeter en avant », la prospective « apparaît comme une représentation (et donc une interprétation) symbolique et/ou visuelle du futur » (Antoni, 2016). Or G. Berger, introduit dès 1957, l'idée que l'avenir est aussi ce « que nous aurons voulu qu'il soit » (Berger, 1957). La question consiste donc à définir ce que nous voulons qu'il soit, et pour cela à étudier plusieurs alternatives. Les futurs possibles deviennent, dès lors, un objet d'étude, et le regard que l'on y porte, un levier d'action. Néanmoins, B Cadet précise que « l'attitude prospective privilégie la compréhension des mécanismes qui gouvernent la dynamique des phénomènes étudiés ». Il insiste sur l'idée que « la prospective ne cherche pas à prédire mais à donner un sens, à la

fois en indiquant les grandes directions du changement et en essayant de répondre au pourquoi de l'action ». Et pour cause, si la démarche modélisatrice laisse entrevoir un potentiel extrêmement riche, notamment au titre de l'exploration des systèmes qui composent le monde dans lequel il nous est donné d'évoluer, il faut cependant garder à l'esprit ses contraintes et limites. Les résultats qu'elle permet d'obtenir sont indissociables des performances des outils et sont également impactés par la subjectivité du modélisateur, « à la fois acteur et spectateur de [ses] propres démarches » (Antoni, 2010).

En résumé, dans une approche prospective de l'étude des processus spatiaux en géographie, et notamment lors de l'élaboration de scénarios, l'observé se mêle au supposé, imaginé sur la base d'une réalité subjective. Les résultats produits par leur modélisation, s'ils ne donnent pas de réponses faisant office de « vérité », fournissent cependant de nombreuses pistes de réflexions. Ces dernières peuvent à la fois présenter une contribution en recherche fondamentale, sans objectif opérationnel mais plutôt avec celui « d'aller plus loin » dans la compréhension des objets d'études, ou bien justement présenter une contribution en recherche appliquée dans un but opérationnel, comme cela peut être le cas à des échelles spatio-temporelles plus fines, en aménagement du territoire ou dans la cadre de la mise en place de politiques spécifiques à objectifs concrets. Quel que soit l'échelle spatio-temporelle ou les objectifs sous-jacents, les modèles peuvent donc servir de support pour l'exploration des processus, qu'ils appartiennent au passé, soient contemporains ou encore qu'ils soient le fruit de l'imagination. Bien que ce travail se soit appuyé sur une analyse rétrospective du développement urbain, LucSim a été développé pour répondre à ces diverses perspectives.

LucSim a, en effet, été développé pour être flexible. Les simulations produites peuvent être contraintes à différentes étapes, ou n'en présenter d'autres que celles de voisinages imposés par les règles elles-mêmes. La structure du jeu de règles permet par exemple de faire référence aux données spatiales contenues dans d'autres couches que celle contenant l'occupation du sol. Il est donc possible d'intégrer des aménagements au cours de la simulation, contraindre le développement au regard de l'intégration de nouvelles structures ou d'objectifs politiques d'urbanisation. Si la possibilité existe, elle n'a cependant pas été explorée dans ce travail. En effet, l'intégration d'un plan réglementaire à l'échelle de la zone transfrontalière est par exemple rendu délicate en raison de l'étendue de la zone et ne présente pas d'adéquation performante avec la résolution des données utilisées. La démarche prospective est parfois liée dans les discours et dans les faits, à la démarche de projection, qui « apparaît comme un exercice purement mécanique qui consiste à prolonger dans le futur des évolutions passées toutes choses égales par ailleurs, sans soucis de réalisme » (Antoni, 2016). La sous-section suivante s'attache donc à illustrer les perspectives de l'outil en présentant la projection à long terme[45] des tendances observées entre 1990 et 2006.

[45] Comme dans Antoni *et al.* Le long terme est fixé par l'application d'un processus de Markov en 2038 (Antoni *et al.*, 2018).

3.2. Scénario projection de tendance

La réflexion quant à l'évolution future, que ce soit en matière d'aménagements urbains ou de changements d'occupation du sol est généralement basée sur l'élaboration de différents scénarios. Ils prendront des formes variées et seront orientés différemment au regard des objectifs de l'exploration effectuée à travers l'outil de modélisation (Section 2.2). L'objectif, ici, n'est pas de comparer, ni d'interpréter le fond des résultats produits lors d'une simulation des tendances futures, mais bien de montrer le potentiel de l'outil à travers leur projection à plus long terme. En effet, il est important de souligner une fois de plus que l'objectif premier de ce travail était principalement de mettre en exergue les différences et similitudes pouvant exister dans la caractérisation du processus de développement urbain sur une zone d'étude frontalière. Néanmoins, l'élaboration d'un modèle de simulation peut être associé à celle de scénarios basés sur une connaissance préalable, ou acquise spécifiquement, dans le but d'explorer le futur et les divers chemins de développement envisageables. Sans en faire l'apport principal de ce travail de recherche, l'outil et la méthode permettent, notamment à travers un prolongement de tendance, la localisation des zones potentiellement urbanisées dans un futur proche.

La série de cartes suivantes présentent donc l'urbanisation en 2022 et 2038 sur la base de la projection des tendances observées entre 1990 et 2006. Le choix d'une projection jusqu'en 2038 se base sur les seuils d'une projection de Markov (Antoni *et al.*, 2018). Les résultats sont présentés sous forme de cartes. L'intérêt du prolongement de tendance est double :

- mettre en relation le développement urbain projeté, avec les erreurs de la modélisation rétrospective,
- localiser les zones à plus fort enjeux au regard de l'analyse rétrospective du développement urbain,

La mise en relation du développement urbain projeté avec les imprécisions du modèle de simulation est possible à travers l'observation à un niveau très fin, celui de la cellule. Pour ce faire, il est nécessaire de restreindre la zone analysée à l'échelle d'une agglomération par exemple (Figure 9. 3). La précision du modèle est soulignée puisque la quasi-totalité des cellules urbanisées d'après les données observées, le sont également dans les données simulées. La projection de l'urbanisation jusqu'en 2038 permet de constater que certaines imprécisions sont lissées dans le temps de simulation. Une omission (faux négatif) en 2006 est parfois corrigée au pas de simulation suivant (Figure 9. 3). Le développement observé est par ailleurs cohérent dans la globalité avec un phénomène d'étalement urbain classique, à l'exception des trois hectares isolés au sud-est du village, dans la projection 2038.

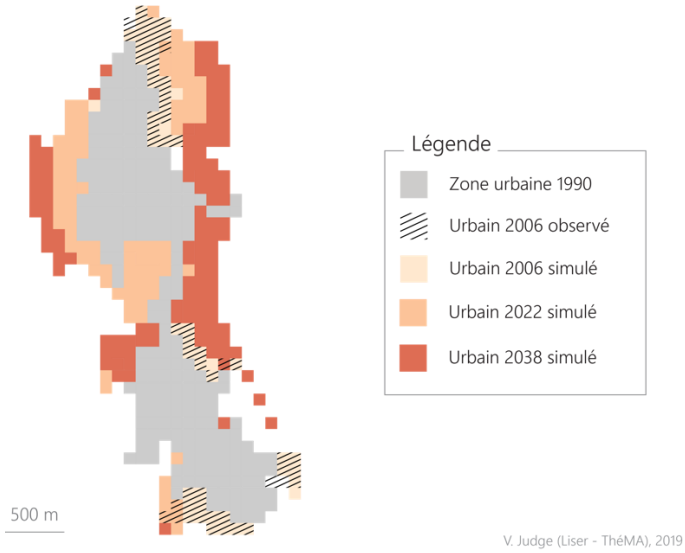


Figure 9. 3 : Simulation dynamique du développement urbain observé autour d'un village de la zone d'étude

La localisation des zones qui présentent des enjeux en matière d'urbanisation est aussi rendue possible à travers un scénario basé sur la projection des tendances observées. Une réflexion peut ainsi être mise en place en amont de l'élaboration des plans d'aménagement, avec pour objet l'anticipation, ou la mise en lumière de zones pouvant faire l'objet d'une attention particulière, et ainsi servir de support d'aide à la décision (Antoni *et al.*, 2018).

Ainsi, les résultats de la projection de tendances montrent le caractère favorable de la périphérie de certaines agglomérations allemandes, ou le développement est particulièrement visible jusqu'en 2038 (Figure 9. 4). L'émergence de nouvelles zones urbanisées s'observe également le long de la frontière, sur l'espace allemand de la zone d'étude. Le développement urbain est moins perceptible sur la zone française.

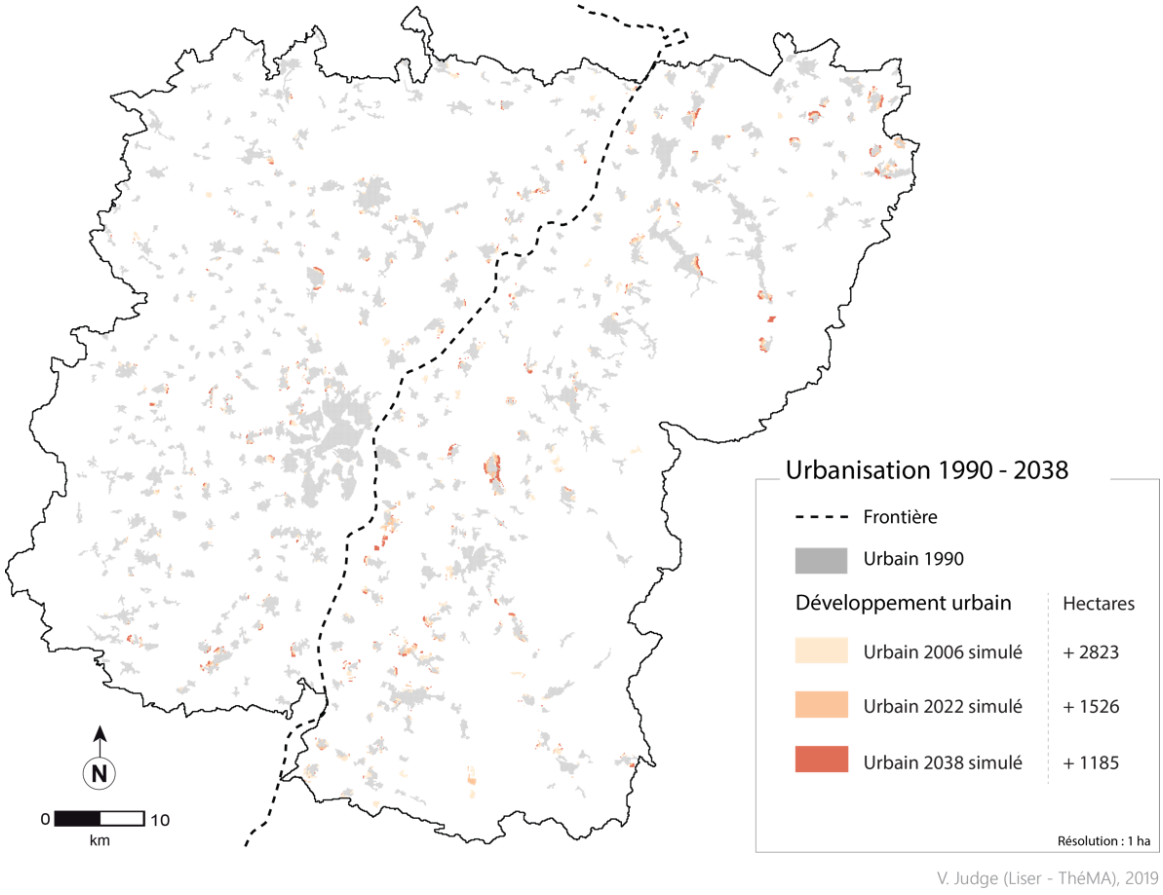
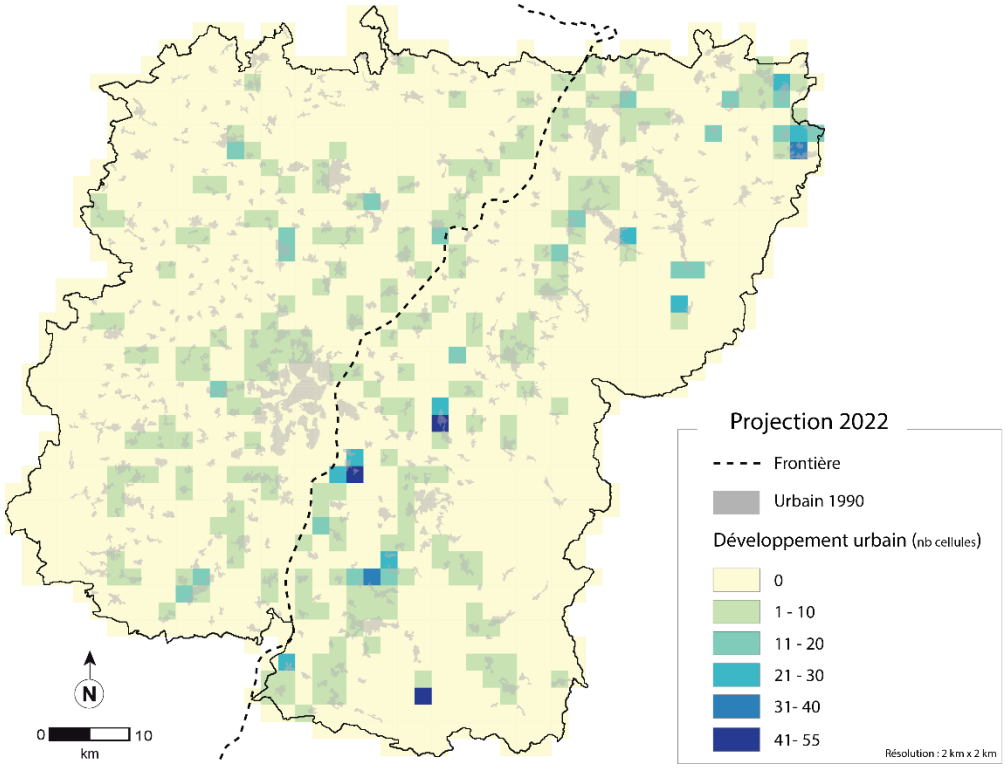


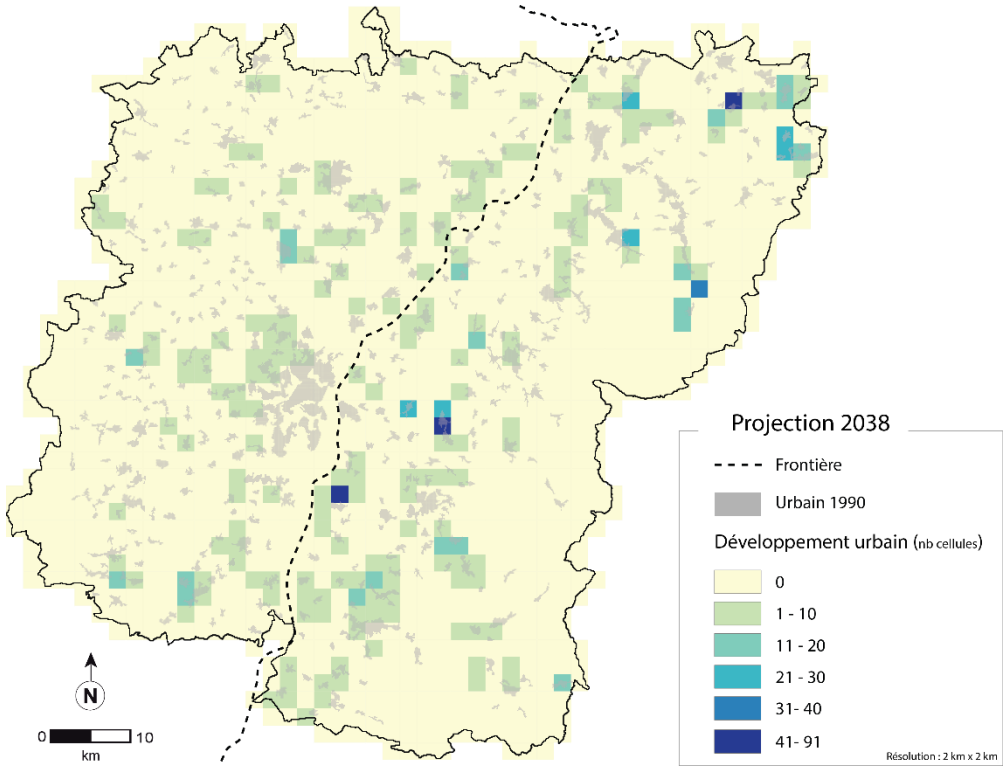
Figure 9. 4 : Développement urbain simulé de 1990 à 2038

La visualisation de la projection du développement urbain jusqu'en 2038 est toutefois limitée par le rapport entre la résolution des cellules et l'étendue de la zone d'étude. Les résultats obtenus pour la période 1990 – 2038 sont donc agrégés, sur une grille de résolution 2 km x 2 km, suivant la même démarche que celle adoptée précédemment (Chapitre 6 Section 3.3). Cette solution optimise la visualisation du développement urbain simulé par le modèle, et permet de se focaliser plus facilement sur les zones touchées en priorité. Les projections 2022 et 2038 localisent différentes zones d'intérêt (Figure 9. 5, Figure 9. 6). Elles sont parfois proches de la frontière, mais aussi proches des limites externes de la zone d'étude. Le modèle souligne un caractère favorable légèrement supérieur en Allemagne.



V. Judge (Liser - ThéMA), 2018

Figure 9. 5 : Développement urbain simulé à l'horizon 2022



V. Judge (Liser - ThéMA), 2018

Figure 9. 6 : Développement urbain simulé à l'horizon 2038

La présentation du développement simulé entre 1990 et 2038 définit plus clairement les zones d'intérêt (Figure 9. 7). L'agrégation peut à la fois permettre une discussion sur les localisations préférentielles d'après le modèle, mais aussi être appréhendée en tant que carte localisant les zones à plus fort potentiel orientant un développement optimal au regard du voisinage et de l'accessibilité aux pôles de différents niveaux (équivalent des *suitability map* ou *zoning* en anglais).

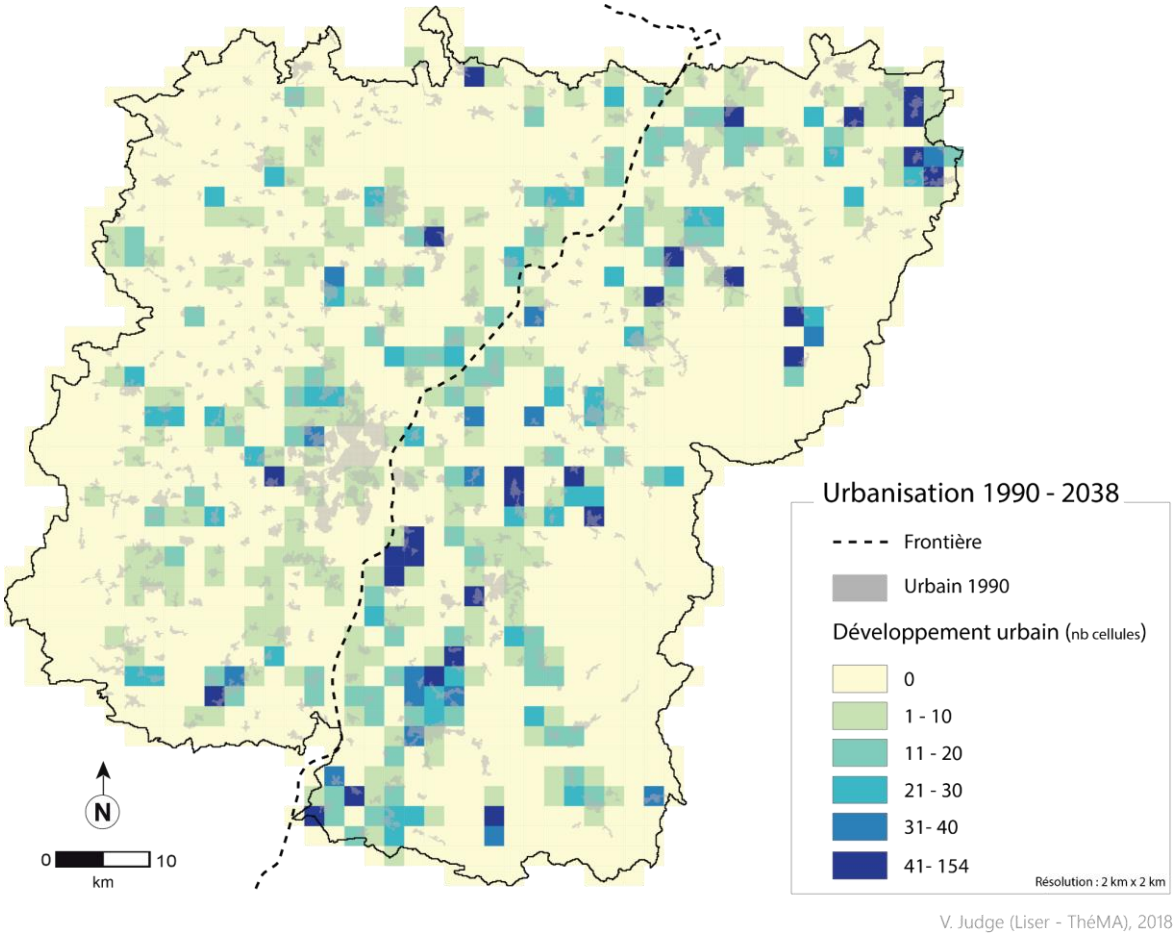


Figure 9. 7 : Intensité du développement urbain simulé entre 1990 et 2006

Cette sous-section a permis de présenter le potentiel de l'outil dans la projection des tendances observées. Un des avantages de la modélisation spatiale en géographie est la visualisation spatiale, voire dynamique, de l'évolution du processus étudié. LucSim permet de représenter l'évolution du processus de développement urbain étape par étape. Il peut également permettre une visualisation dynamique dans le cadre de réunions entre différents acteurs. Un premier pas vers l'opérationnel est donc envisageable.

3.3. Aide à la décision et opérationnalisation

« Les résultats de simulation finissent pas offrir des images tangibles de l'avenir des territoires » (Antoni, 2016). J.P. Antoni souligne qu'il n'existe toujours pas de consensus pour préciser les modalités qui intégreraient, ou non, le développement d'un outil dans cette

démarche (Antoni, 2016). Cette dernière doit néanmoins permettre de répondre à trois questions générales : Comment ? Pourquoi ? Quoi ? Ainsi, le travail réalisé dans cette recherche permet bien de définir le « comment » en justifiant d'une méthode adaptée à la modélisation et l'analyse des processus de développement urbain d'un point de vu spatial, mais également d'un point de vu thématique. Cette caractéristique favorise la discussion autour des résultats. L'objet d'étude, le « quoi », est par ailleurs clairement identifiable et peut également être étendu aux changements d'occupation du sol dans leur globalité si nécessaire. Le passage dans l'univers de l'aide à la décision nécessite alors de répondre à la question « pourquoi ? » en fonction des besoins institutionnels ou politiques en matière d'aménagement. Nous nous éloignons alors de l'exploration théorique de l'homogénéité de la zone d'étude à travers la modélisation du développement urbain, qui permet de souligner ses différences ou similitudes de part et d'autre de la frontière, pour aller vers une démarche qui vise à étudier l'impact d'un projet ou d'une politique d'aménagement, en fonction d'un besoin opérationnel spécifique. Par exemple, la faisabilité d'un objectif politique de consommation d'espace sur une période donnée et la réponse adéquate en matière d'aménagement (Frémond, 2015). Ce besoin opérationnel peut concerner plusieurs types d'acteurs et notamment des institutions et organismes nationaux.

Des organismes comme l'Agence De L'Environnement et de la Maitrise de l'Énergie (ADEME) porte, par exemple, un intérêt particulier sur la modélisation opérationnelle. Cet établissement public cherche, en effet, à « connaître et comprendre l'offre d'outils de modélisation territoriale intégrée existante » (ADEME, 2016). Pour se faire, l'ADEME a procédé à l'élaboration d'une grille d'analyse des outils de modélisation territoriale. Leur objectif est d'orienter la « conception des outils » tant dans l'intérêt de l'utilisateur que dans celui du modélisateur. Les outils développés par les laboratoires scientifiques font partie des instruments cités comme utiles dans une approche d'aide à la décision. Notre modèle répond la plupart du temps aux critères de sélection envisagés pas l'ADEME (Figure 9. 8). Les outils doivent présenter des caractéristiques spécifiques au regard de trois axes : type d'outil, périmètre thématique, échelle territoriale. LucSim s'intègre donc parfaitement dans la partie « outil de modélisation de processus », à l'échelle territoriale « régionale » et rejoint la thématique « bâtiments » à travers la modélisation des changements d'occupation du sol.

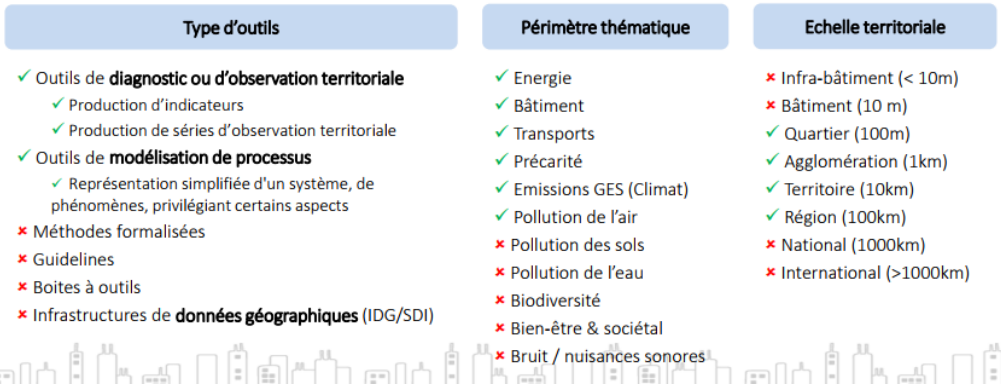


Figure 9. 8 : Critères de sélection (1er filtre) d'un outil d'aide à la décision à vocation territoriale d'après l'ADEME (2016)

La démarche développée dans le cadre de ce travail de recherche permet de passer le premier filtre établi par l'ADEME dans la sélection des outils d'intérêt. Le second filtre précise la variété des caractéristiques des outils et le degré d'approche intégré et multicritères.

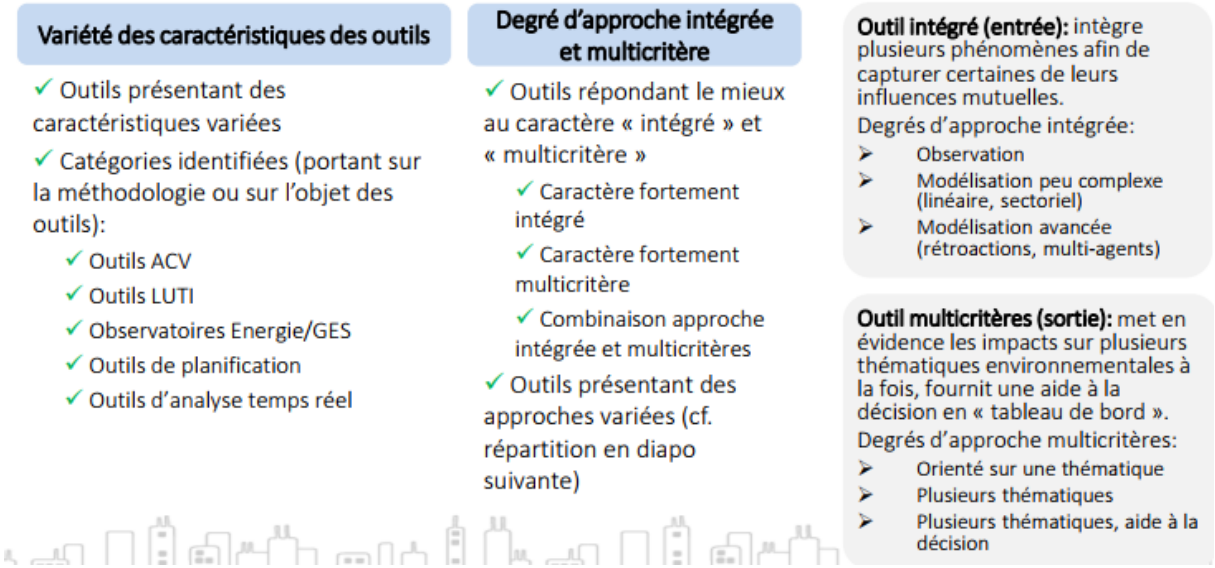


Figure 9. 9 : Critères de sélection (2ème filtre) d'un outil d'aide à la décision à vocation territoriale d'après l'ADEME (2016)

LucSim présente également des caractéristiques variées. Il peut, par ailleurs, être inclus parmi les outils de planification suivant les données d'entrées mobilisées et la précision de la question qui est abordée par la modélisation. Sa flexibilité permet l'ajout de critères variés tout en conservant l'opérabilité. Dans la limite d'un volume de données raisonnable, le modèle et la simulation peuvent être adaptés à un besoin immédiat, sous couvert de la disponibilité de la donnée. Des tests peuvent être réalisés dans un temps restreint variant de quelques minutes à quelques dizaines de minutes. La thématique abordée dépendra de l'utilisateur et de la disponibilité des données. Il est également possible d'inclure manuellement des rétroactions pendant la simulation. Il peut néanmoins fournir une base dans l'utilisation d'une approche LUTI (*Land Use and Transport Integrated model)* à travers la modélisation de l'occupation du sol. Les caractéristiques intrinsèques de LucSim, qui comprennent l'aspect flexible de l'implémentation des données en fait un outil pertinent pour étudier des thématiques variées. Notre outil passe donc le 2ème filtre de l'ADEME.

La figure ci-dessus présente la liste des modèles sélectionnés comme répondant aux critères présentés dans les Figure 9. 8 et 9.9.

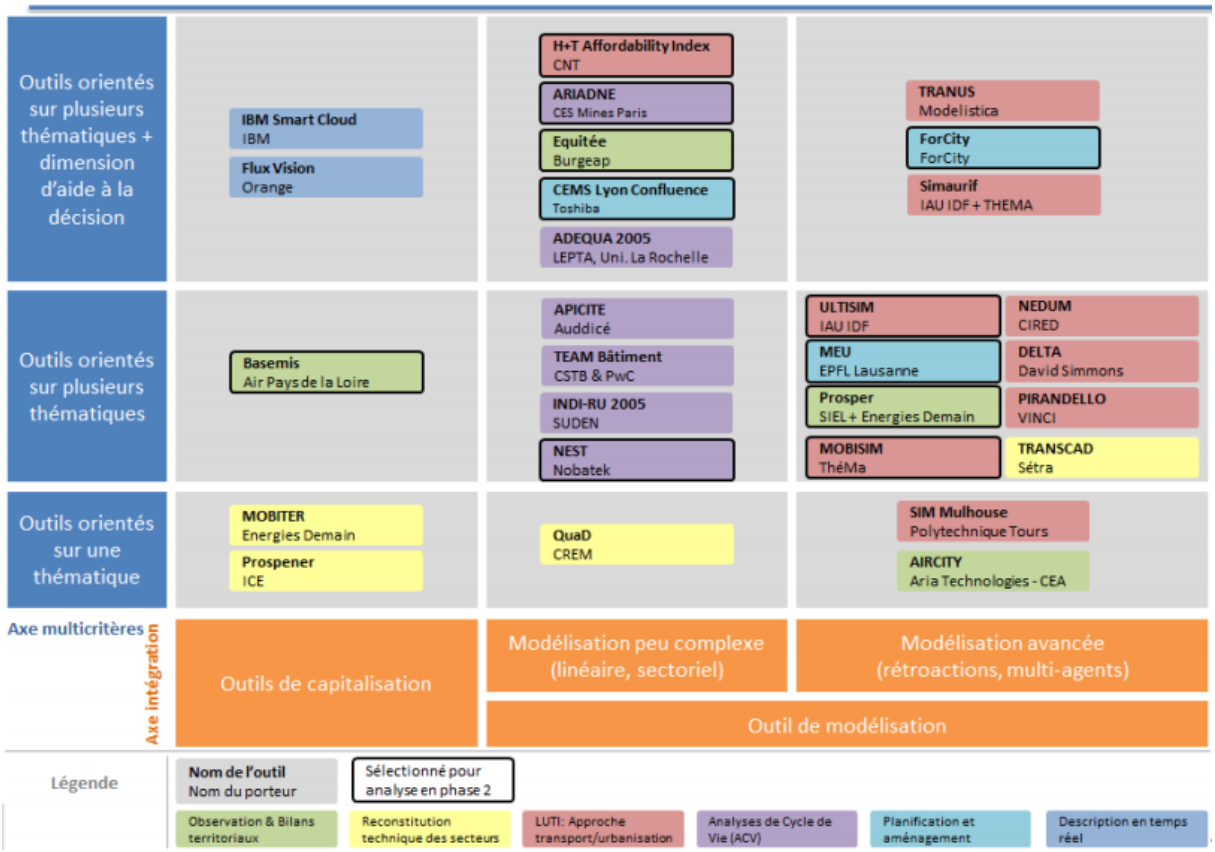


Figure 9. 10 : Sélection des outils à analyser de manière approfondie : répartition selon les axes Intégré et Multicritère (2016)

Le document de l'ADEME met en lumière la diversité des outils potentiellement mobilisables dans le cadre d'une approche de l'aménagement par la modélisation opérationnelle. Dans le cadre d'une exploitation future orientée vers l'opérationnel, il est intéressant de constater que la méthodologie élaborée et implémentée dans LucSim permet d'apporter une autre pièce dans le tableau. LucSim pourrait en effet être intégré dans la case centrale, dans une catégorie associant *Observation et bilan territoriaux* et *Approche de l'urbanisation* (déconnectée du transport comme il est envisagé dans les modèles LUTI).

L'ADEME sélectionne plus spécifiquement onze outils susceptibles d'être associés aux supports d'aide à la décision. Cette liste non-exhaustive permet de souligner que certains outils de la liste ont vu leur développement avorter pour diverses raisons, notamment le coût de développement. Réalisme et efficacité des temps de calculs sont la plupart du temps antithétiques (White *et al.*, 2015). Les avancées récentes en matière de technologie offrent, certes, la possibilité d'élaborer des modèles extrêmement fournis et complexes. Cette richesse est néanmoins couteuse à la fois en argent mais également en temps. Par ailleurs, « trop souvent, le réalisme est privilégié au dépend de la simplicité » (White *et al.*, 2015). Cet élément permet également de souligner les points forts de la méthode élaborée et intégrée dans LucSim dans le cadre de ce travail de recherche. Cette méthode, rapide et peu coûteuse, permet une première approche du territoire et favorise le dialogue autour des

processus dynamiques structurant les changements, notamment ceux de l'occupation du sol et de l'évolution des zones urbanisées.

En conclusion, LucSim intègre toutes les caractéristiques techniques souhaitables pour un outil d'aide à la décision. La modélisation est simple, rapide et transparente, tout en offrant la possibilité de visualiser les résultats, favorisant donc la discussion autour de scénarios d'aménagement. L'ajout d'une fonctionnalité d'apprentissage, qui constituait l'un des enjeux majeurs de ce travail doctoral, a permis d'enrichir indéniablement son potentiel pour la prospective territoriale et pour l'aide à la décision en aménagement.

Conclusion du Chapitre 9

L'analyse finale de la composition du jeu de règles a permis de présenter une mesure de la porosité de la frontière, au regard de l'étude des processus de développement urbain sur notre zone. Ce résultat, et ceux produits dans les précédents chapitres de cette partie, sont ensuite mis en relief en soulignant certaines limites du présent travail. Ainsi, ce chapitre a été l'occasion de s'interroger sur les limites de notre travail et de proposer plusieurs pistes pour améliorer la qualité du jeu de règles ou celle de la méthode, en incluant les interprétations qui en sont issues. Les pistes d'améliorations concernent notamment la qualité, la résolution ou la diversité des données utilisées, l'utilisation de la méthode sur d'autres terrains d'études (frontaliers ou non), ou encore le test d'autres algorithmes d'apprentissages. Nous avons finalement proposé des exemples d'application qui illustrent le potentiel de l'outil et de la méthode dans l'optique d'une utilisation dans un cadre prospectif de l'aménagement du territoire, ou encore en tant que support d'aide à la décision. Il était en effet question de proposer une méthode et un outil pouvant à terme être adaptés dans ces perspectives d'évolution, en dépit du fait que ces objectifs n'étaient pas, initialement, au cœur du travail de cette thèse.

Conclusion de la troisième Partie

La troisième partie de notre travail a permis de présenter en détail l'analyse et l'interprétation des résultats extraits par l'étude du jeu de règles. Elle constitue donc l'apport principal de ce travail de recherche et permet d'apporter des éléments de réponse au regard de l'évaluation de l'homogénéité de la zone frontalière d'après du processus de développement urbain.

L'importance de la variable Indice de Pays, qui localise le processus de développement urbain dans l'un ou l'autre des pays frontaliers, a tout d'abord été mis en évidence au regard de celle des autres variables de localisation géographique. Une première analyse de la totalité des règles a permis de faire ressortir l'influence générale des différents types d'occupation du sol dans le voisinage des cellules urbanisées sur notre zone d'étude. Elle a aussi mis en évidence le caractère incitatif ou restrictif de cette influence et permis de déterminer des seuils favorables ou défavorables à l'urbanisation. Cette analyse a également permis de tracer un zonage hiérarchique des zones favorables à l'urbanisation au regard de leur accessibilité aux différents pôles urbains.

L'existence de groupes de règles de différentes natures a donc été mis en lumière. Plus de la moitié des règles sont transfrontalières, elles s'affranchissent donc de la contrainte de localisation dans l'un ou l'autre des pays uniquement et sont donc principalement influencées par l'accessibilité aux pôles urbains. Le reste du jeu de règles est composé dans des proportions similaires par des règles qui s'appliquent uniquement en France ou uniquement en Allemagne. Elles décrivent alors le voisinage et l'accessibilité favorable à l'urbanisation dans l'un ou dans l'autre des pays frontaliers. Ce résultat nous a permis d'affiner l'interprétation des tendances dans la composition du voisinage ou de l'accessibilité, au regard de la nature des règles. Ainsi, les différences et similitudes qui existent entre les règles françaises, allemandes, ou transfrontalières ont pu être discutées. A une échelle encore plus fine, les différences entre les tendances majeures, issues des règles d'une même nature, ont également fait l'objet d'une attention particulière. D'un point de vue plus général, la porosité de la frontière au regard de l'étude du processus d'urbanisation, a été évaluée d'après la proportion de règles transfrontalières dans la composition du jeu.

Suite à ces analyses, plusieurs perspectives ont été proposées concernant l'interprétation ou l'utilisation du jeu de règles produit, ou orientées vers une approche qui appondit encore les tests d'un point de vue technique. L'une des pistes, et celle qui présente un intérêt majeur en l'état, est de privilégier l'application de notre méthode et du protocole d'analyse des règles, sur d'autres territoires. Il est par ailleurs également possible d'envisager un protocole de validation des résultats, par des données d'enquêtes. Ces deux perspectives permettraient de renforcer la robustesse de nos résultats.

Enfin, les exemples de perspectives quant à l'utilisation de notre outil et de la méthode développée dans cette thèse montrent le potentiel de l'outil en tant que support d'aide à la décision ou encore lors de l'étude de scénarios prospectif avec une portée opérationnelle.

Conclusion générale

La démarche initiée dans ce travail a permis de caractériser l'homogénéité d'une zone frontalière, à travers la modélisation du développement urbain. L'hypothèse initiale repose en effet sur l'idée selon laquelle le développement urbain, quel que soit son lieu d'expression, présente une caractéristique commune liée à sa fonction : en tous lieux, il s'agit de construire des bâtiments visant à accueillir habitations et activités humaines. Les formes d'urbanisation observées à travers le monde présentent évidemment des différences et des variations, qui sont en partie attribuées à l'environnement naturel dans lequel elles s'insèrent, ainsi qu'au contexte culturel. Ces formes peuvent également différer alors même que l'urbanisation se développe dans des espaces aux contraintes environnementales similaires. Au contraire, elles peuvent être similaires, dans un cadre environnemental comparable mais dans des contextes culturels éloignés. Ce phénomène s'observe d'ailleurs à proximité de certaines frontières d'États. Il s'agit alors de s'interroger sur l'influence de la frontière en tant que limite et séparateur réelle des pratiques, notamment s'agissant du développement urbain.

Pour répondre à cette question, nous avons choisi d'analyser la zone frontalière franco-allemande, centrée sur Strasbourg-Kehl. Cette zone présente effectivement des caractéristiques pertinentes pour tester nos hypothèses à travers l'application de la méthode développée dans ce travail. Plusieurs éléments permettent d'envisager une frontière poreuse. C'est un lieu d'histoire riche, où la mobilité spatiale du tracé de la frontière d'État à travers le temps permet d'interroger son rôle en tant que limite. C'est aussi un lieu symbolique de l'Union Européenne, puisque Strasbourg accueille certaines institutions, de l'amitié Franco-Allemande, symbole qui tend ainsi à temporiser l'influence de la frontière en tant que limite. C'est enfin une zone de voisinage où s'utilise une monnaie unique européenne, où la ratification de traités européens a favorisé l'ouverture, l'échange et la circulation. Certains éléments permettent au contraire d'ancrer une différence marquée dans l'espace par la frontière : la langue, le fonctionnement des institutions, ou encore la culture. Il s'agissait donc, à travers l'analyse de la zone frontalière franco-allemande, de répondre à la question posée dans l'introduction générale du présent travail :

Est-il possible de dissocier une composante locale du processus d'urbanisation, liée à une variation culturelle intrinsèque au sein d'un espace séparé par une frontière étatique, d'une composante plus générale, qui caractérise le processus indépendamment de sa localisation ?

L'analyse de la zone frontalière franco-allemande

L'application de la méthode développée dans cette thèse a mis en évidence des processus d'urbanisation différenciés suivant leur localisation dans la zone d'étude. La composition du voisinage en termes d'occupation du sol, ou encore les caractéristiques d'accessibilité aux pôles urbains ont été étudiées en fonction de la nature des règles qui permettent de

modéliser le développement urbain (règles françaises, allemandes, ou transfrontalières). Ces règles permettent ainsi de faire émerger des processus spécifiques à la France et à l'Allemagne. La méthode a également mis en évidence un processus de développement urbain transfrontalier, davantage contraint par l'accessibilité aux pôles urbains, que par la localisation au sein de l'un ou l'autre des États. Dès lors, l'analyse des règles qui définissent le développement urbain sur la zone d'étude conduit à les isoler selon leur degré d'universalité : 20% du jeu de règles s'applique quel que soit le pays ou l'état initial d'occupation du sol, 40% présente au moins une des deux contraintes, alors que 40% de ces règles sont à la fois localisées spécifiquement en France ou en Allemagne et présentent une contrainte concernant l'état initial d'occupation du sol. Au total, plus de 50% des règles obtenues sont transfrontalières, illustrant la part du processus commune à toute la zone d'étude et permettant d'en caractériser l'homogénéité. Si la porosité de la frontière pouvait être estimée de manière quantitative d'après le jeu de règles, elle serait alors égale à l'influence des règles transfrontalières dans le jeu de données : soit 58%.

Une démarche modélisatrice à la fois classique et innovante

L'évolution des performances informatiques en matière de développement technologique ou d'acquisition de données pousse la recherche à s'intéresser davantage à l'élaboration de modèles dont la complexité croît, et nécessite la mobilisation de ressources techniques de plus en plus importantes. Cette perspective oriente la démarche de modélisation vers l'utilisation de méthodes adaptées à la prise en charge de volumes conséquents et d'une diversité croissante. La performance de modélisation devient l'objectif principal de ces méthodes complexes, qui doivent se rapprocher au maximum de la réalité, optimiser le détail de la modélisation et arriver au mieux à reproduire le phénomène étudié. Ce faisant, elles perdent néanmoins le caractère explicite et transparent, qui permet une description plus thématique du processus et en simplifie la compréhension.
Dès lors, notre approche apparaît fondamentalement novatrice, tout en étant basée sur une démarche classique en Land Use Sciences (Camacho Olmedo *et al.*, 2018). Bien que de nombreux modèles aient été développés pour simuler les changements d'occupation du sol et les dynamiques de développement urbain, tout l'enjeu de notre démarche réside dans l'apport thématique issu de l'analyse des règles d'urbanisation créées par apprentissage automatique. En d'autres termes, il s'agit d'évaluer l'apport d'une forme d'intelligence artificielle dans la caractérisation de la dynamique de l'occupation du sol, mais également, en définitive, des comportements humains.

La base méthodologique de ce travail repose sur une approche utilisée de manière très classique dans le domaine de la géographie théorique et quantitative ainsi que dans les *Land Use Sciences*, à savoir une démarche modélisatrice qui implique l'utilisation d'un automate cellulaire (AC) évolué. Elle s'appuie sur l'utilisation d'un minimum de données spatiales, et sur le concept de base de l'AC pour simuler le processus dynamique de développement urbain : une cellule qui présente un certain état d'occupation du sol initial, effectuera une transition vers un état d'occupation du sol final, en fonction de l'état d'occupation du sol

des cellules présentes dans son voisinage. La définition de caractéristiques supplémentaires spécifiques à chaque cellule, notamment leur accessibilité aux agglomérations majeures de la zone d'étude, nous éloigne quelque peu du concept original de l'AC, tel que défini par S. Ulam ou S. Wolfram, et se rapproche de la réappropriation géographique de l'outil telle qu'elle a été proposée par Tobler, avant d'être généralisée par R. White et G. Engelen. L'innovation de notre travail a consisté à apporter une étape supplémentaire à cette évolution avec l'utilisation d'un algorithme d'apprentissage par arbre de décision (AD) pour définir et calibrer automatiquement les règles de transitions nécessaires à la simulation du processus de développement urbain dans l'AC. Le couplage de ces deux instruments a été inspiré par un paradigme commun. En effet, l'AC et l'AD sont généralement privilégiés et appréciés en raison de leur caractère simple et explicite. D'un certain point de vue, ce couplage a permis de répondre, par l'utilisation de l'outil accessible qu'est l'arbre de décision, à la question éminemment complexe de la définition et du calibrage des règles de transition pour la simulation du développement urbain, telle qu'elle a été posée par P. Torrens (2011). Ces règles définies par une forme d'intelligence artificielle permettent, à travers le caractère simple et explicite de l'AC, d'extraire des informations sur les processus de développement urbain de notre zone d'étude, donc de la caractériser dans son homogénéité, et de remplir l'objectif principal de notre travail.

Bien que cette démarche ait essentiellement eu un caractère fondamental, la méthode a été développée dans l'optique d'être transposable à tout terrain d'étude dans n'importe quel contexte. Dans ce cadre, il semble aujourd'hui pertinent d'une part de multiplier les applications sur différents terrains d'études, et d'autre part d'orienter l'utilisation de l'outil vers une démarche plus opérationnelle. Ce premier axe permet d'approfondir l'approche théorique, proposée dans ce travail de thèse, sur la part généralisable ou localisée du processus de développement urbain. Cela est d'autant plus envisageable qu'une application de la méthode qui viserait à reproduire la démarche initiée dans ce travail de thèse nécessiterait uniquement de réitérer les phases de paramétrage de l'algorithme d'apprentissage. Le second axe a vocation d'élargir le champ d'application de l'outil, qui peut être mobilisé en tant que support de discussion dans un processus d'aide à la décision, en raison de sa simplicité et de sa rapidité d'exécution.

En d'autres mots, le couplage proposé ici ouvre des perspectives importantes pour intégrer les technologies nouvelles, fondées sur l'intelligence artificielle et l'émergence de données libres et massives, aux acquis de la modélisation de l'occupation du sol en géographie. Ces perspectives ont été ici testées dans un cadre académique et théorique. La question qui se pose en ouverture est désormais de savoir dans quelles mesures elles peuvent être appropriées par les acteurs et les décideurs locaux pour une prise de décision raisonnée visant à construire un aménagement durable tourné vers une coopération transfrontalière avec un bagage méthodologique mis en commun. La simplicité relative de la démarche, sa transposabilité et sa rapidité d'exécution ouvrent une voie optimiste, en même temps qu'elles nécessitent de nouvelles investigations dans le champ de l'aménagement des territoires.

Table des figures

Chapitre 1 :

Figure 1. 1 : Couple Humain-Environnement – Impact – interactions – rétroactions 17
Figure 1. 2 : Images satellites de la forêt amazonienne acquises respectivement en 1986, 1998 et 2005. ©CNES – SPOT 19
Figure 1. 3 : Développement de l'urbanisation autour de Paris entre 1900 et 2013(Source : Géoportail – Carte Topographique 1900 – 2013)..... 20
Figure 1. 4 : Le développement spatial dans les régions urbaines (Hersperger et al., 2018)..... 23
Figure 1. 5 : Physionomie du développement urbain à travers le monde (Source : Pinterest)..... 25
Figure 1. 6 : Densité de population en 2010 par NUTS de niveau 3 (habitant par km^2)..... 27
Figure 1. 7 : Variation de population en 2010 par NUTS de niveau 3 (pour 1000 habitants). 28
Figure 1. 8 : Urbanisation à proximité de la frontière franco-allemande (source : GoogleEarth – 2009) 30
Figure 1. 9 : Émergence de la notion de la frontière comme la construction résultant de différentes visions..... 42

Chapitre 2 :

Figure 2. 1 : Du modèle conceptuel au modèle de simulation (Source : C. Tannier dans (Antoni et al., 2011)..... 50
Figure 2. 2 : Global Map Project source ESA/ ESA CCI Land Cover dirigé par UC Louvain 58
Figure 2. 3 : Corine Land Cover adapté de Mücher 2015 59
Figure 2. 4 : Illustration portant sur les choix de classifications. A) Carte présentant la zone d'étude centrée sur Besançon dans le cadre de la thèse de M. Bourgeois (2015) intitulée « Zones constructibles de l'air urbaine de Besançon » b) Carte présentant la même zone avec un périmètre différent dans de le cadre de la thèse de Y. Sahraoui (2016) intitulée « Occupation du sol simplifiée du terrain de Besançon en 1984 et 2010 »..... 59
Figure 2. 5 : Nomenclature Corine Land Cover - 44 classes (Source : Copernicus, CLC 2012) 64
Figure 2. 6 : Classification détaillées des différents niveaux de Corine Land Cover (Source : Copernicus)..... 65

Chapitre 3 :

Figure 3. 1 : Limites entre le Royaume de France et le Saint-Empire romain germanique précédent la signature des Traités de Westphalie (Zajec & Antoni, 2017)..... 77
Figure 3. 2 : Frontière franco-allemande en 1809 (Source : Encyclopédie Larousse)..... 78
Figure 3. 3 : L'unité Allemande en 1871 inclus l'Alsace et la Lorraine (Encyclopédie Larousse)..... 78
Figure 3. 4 : La frontière franco-allemande en 1871 (Brun, 2015)..... 79
Figure 3. 5 : La France occupée (Zajec & Antoni, 2017)..... 80
Figure 3. 6 : Retour au tracé de la frontière précédent la Seconde Guerre Mondiale (Source : Encyclopédie Larousse)..... 81
Figure 3. 7 : : Hiérarchie des documents liés à l'aménagement du territoire en France (Sénat, 2018)(d'après Sénat, 2018)..... 86
Figure 3. 8 : Système de planification en Allemagne (Scholl et al., 2007)..... 88
Figure 3. 9 : Projets de territoires et organes institutionnels de coopération (MOT, 2015) 90
Figure 3. 10 : Limites de l'Eurodistrict (Eurodistrict, 2013)..... 95
Figure 3. 11 : Limites de la zone d'étude..... 96
Figure 3. 12 : Occupation du sol en 1990 après fusion des classes..... 99
Figure 3. 13 : Répartition des différents types d'occupation du sol par pays..... 100
Figure 3. 14 : Occupation du sol en 2006 (CLC fusionné)..... 102
Figure 3. 15 : Urbanisation entre 1990 et 2006 (CLC fusionné)..... 103
Figure 3. 16 : Répartition démographique (INSEE 2012 ; Destatis 2012)..... 105
Figure 3. 17 : Rang des unités urbaines en fonction de leur population (Judge, 2017).... 107
Figure 3. 18 : Loi Rang-Taille - distribution Log-Log sur la zone d'étude (Judge, 2017) 107
Figure 3. 19 : Localisation et hiérarchisation des pôles de niveau 1 et 2 sur la zone de Strasbourg-Kehl..... 108
Figure 3. 20 : Accessibilité réseau routier au pôle de niveau 1..... 110
Figure 3. 21 : Accessibilité réseau routier au pôle de niveau 1 pondérée niveau 2..... 111
Figure 3. 22 : Accessibilité réseau routier au pôle de niveau 2..... 111
Figure 3. 23 : Accessibilité réseau routier au pôle de niveau 1 pondérée niveau 2..... 112

Chapitre 4 :

Figure 4. 1 : Le jeu de la Vie de Conway (Bredèche, 2017)..... 123
Figure 4. 2 : Automate cellulaire à 1 dimension selon S. Wolfram (Bredèche, 2017)..... 124
Figure 4. 3 : Les 4 classes d'automate d'après Wolfram : a) Classe I règle 32 - Structure uniforme ; b) Classe II règle 10 – Structure périodique simple ; c) Classe III règle 126 – Structure apériodique, imprédictible, chaotique ; d) Classe IV règle 110 – Structure complexe et imprédictible (Martinez et al., 2013)..... 124
Figure 4. 4 : Simulation de Fredkin (Calicode, 2018)..... 127
Figure 4. 5 : Boucle de Langton (Calicode, 2018)..... 128

Figure 4. 6 : Généralisation commune des automates cellulaires (Couclelis, 1997)..... 131
Figure 4. 7 : Échantillon de carroyage utilisé dans le cadre de simulations d'automates cellulaires. a) grille polygonale anisotrope basée sur les limites des différentes occupations du sol (Pinto et al., 2017) ; b) Carroyage hexagonal isotrope utilisé pour la simulation de la diffusion des feux de forêt (Hernández Encinas et al., 2007) ; c) Carroyage triangulaire anisotrope pour simuler la diffusion d'un phénomène géographique (Ortigoza, 2015).... 132
Figure 4. 8 : Automate cellulaire à 8 états (Judge, 2017)..... 133
Figure 4. 9 : Les voisinages classiques et leurs extensions (Judge, 2017)..... 133
Figure 4. 10 : Structure du jeu de données..... 146
Figure 4. 11 La calibration - un processus itératif (Judge)..... 148
Figure 4. 12 : : Développements méthodologiques appliqués aux AC contraints par les chaînes de Markov, dans la modélisation de l'environnement géospatial (Ghosh et al., 2017) 150

Chapitre 5 :

Figure 5. 1 : Sous- et sur-apprentissage d'un modèle de classement (d'après Bousquet dans Tufféry 2017)..... 162
Figure 5. 2 Taux d'erreur en fonction de la complexité du modèle (d'après (Tufféry, 2017)..... 163
Figure 5. 3 Division récursive de l'échantillon apprentissage..... 169
Figure 5. 4 : Démarche effectuée lors des tests systématiques..... 178

Chapitre 6 :

Figure 6. 1 Six situation ou les cartes de gauche et droite sont comparées, en considérant l'aspect flou de la localisation (Hagen, 2002):..... 189
Figure 6. 2 :Voisinage propre..... 192
Figure 6. 3 :Voisinage cumulé..... 192
Figure 6. 4 : Évolution du nombre de cellules dans le voisinage en fonction du rayon en pixel 192
Figure 6. 5 : Influence de l'intervalle de confiance sur le CCM..... 193
Figure 6. 6 : Influence du nombre minimum d'individus par feuille sur le Coefficient de corrélation de Matthews (apprentissage occupation du sol seulement)..... 194
Figure 6. 7 : Évolution du coefficient de corrélation de Matthews en fonction rayon du voisinage..... 195
Figure 6. 8 : Évolution du coefficient de corrélation de Matthews en fonction du poids attribué aux transitions exceptionnelles..... 195
Figure 6. 9 : Évolution du coefficient de corrélation de Matthews en fonction du nombre de règles présentes dans le modèle..... 196

Figure 6. 10 : Évolution du coefficient de corrélation de Matthews en fonction du nombre de variables 196
Figure 6. 11 : Évolution du nombre de règles en fonction du rayon du voisinage utilisé 198
Figure 6. 12 : Influence du nombre minimum d'individus par feuille sur le Coefficient de Corrélation de Matthews (apprentissage sur totalité des variables)..... 199
Figure 6. 13 : Comparaison des performances des modèles obtenus après apprentissage sur le jeu de données partiel ou total..... 199
Figure 6. 14 : Relation entre les MCC entraînement et test lors de l'apprentissage 75% . 202
Figure 6. 15 : Données observées et simulées du développement urbain entre 1990 et 2006..... 207
Figure 6. 16 : Une observation plus localisée 208
Figure 6. 17 : Transitions agrégées sur une grille de résolution 2 km2 209
Figure 6. 18 : Transitions agrégées sur une grille de résolution 4km2 210
Figure 6. 19 : Évaluation spatiale des performances du modèles. Introduction d'un seuil de tolérance plus flexible 211

Chapitre 7

Figure 7. 1 : Illustration de l'attribution du score des différentes variables 224
Figure 7. 2 : Influence des différents types d'occupation du sol entre 0 et 3km 228
Figure 7. 3 : Influence de l'occupation du sol urbaine dans le voisinage en fonction de la distance (%)..... 230
Figure 7. 4 : Influence de l'occupation du sol agricole dans le voisinage en fonction de la distance (%)..... 230
Figure 7. 5 : Influence de l'occupation du sol agricole dans le voisinage en fonction de la distance (%)..... 231
Figure 7. 6 : Influence de l'occupation du sol vignes dans le voisinage en fonction de la distance (%)..... 231
Figure 7. 7 : Influence de l'occupation du sol zones récréatives urbaines dans le voisinage en fonction de la distance (%)..... 232
Figure 7. 8 : Influence de l'occupation du sol zones immergées dans le voisinage en fonction de la distance (%)..... 232
Figure 7. 9 : Influence de l'occupation du sol industries et commerces dans le voisinage en fonction de la distance (%)..... 233
Figure 7. 10 : Influence de l'occupation du sol vignes dans le voisinage en fonction de la distance (%)..... 233
Figure 7. 11 : Polarité de l'influence du type urbain d'occupation du sol 236
Figure 7. 12 : Polarité de l'influence du type forêt d'occupation du sol..... 239
Figure 7. 13 : Polarité de l'influence du type agricole d'occupation du sol 241
Figure 7. 14 : Polarité de l'influence du type vigne d'occupation du sol 242
Figure 7. 15 : Polarité de l'influence du type zone récréative urbaine d'occupation du sol 243

Figure 7. 16 : Polarité de l'influence du type zone immergée d'occupation du sol 244
Figure 7. 17 : Polarité de l'influence du type industrie et commerce d'occupation du sol 246
Figure 7. 18 : Polarité de l'influence du type transport d'occupation du sol 247
Figure 7. 19 : Contraintes liées à l'accessibilité pondérée pôle de niveau 1 et celles au pôle de niveau2 253
Figure 7. 20 : Contraintes liées à l'accessibilité pondérée pôle de niveau 2 et celles au pôle de niveau 1 254
Figure 7. 21 : Zones préférentielles d'urbanisation en regard de l'accessibilité aux différents pôles urbains 258

Chapitre 8 :

Figure 8. 1 : Influence et polarité de l'occupation du sol urbaine dans le voisinage en fonction de la distance 266
Figure 8. 2 : Influence de l'occupation du sol agricole dans le voisinage en fonction de la distance 268
Figure 8. 3 : Influence de l'occupation du sol forêt dans le voisinage en fonction de la distance 270
Figure 8. 4 : Influence de l'occupation du sol zone récréative urbaine dans le voisinage en fonction de la distance 272
Figure 8. 5 : Influence de l'occupation du sol vigne dans le voisinage en fonction de la distance 274
Figure 8. 6 : Influence de l'occupation du sol industriel et commercial dans le voisinage en fonction de la distance 276
Figure 8. 7 : Influence de l'occupation du sol transport dans le voisinage en fonction de la distance 277
Figure 8. 8 : Influence de l'occupation du sol zone immergée dans le voisinage en fonction de la distance 279
Figure 8. 9 : Contraintes liées à l'accessibilité pondérée pôle de niveau 1 et celles aux pôles de niveau 2 dans les règles françaises 284
Figure 8. 10 : Contraintes liées à l'accessibilité pondérée pôle de niveau 2 dans les règles françaises 284
Figure 8. 11 : Zones préférentielles d'urbanisation françaises au regard de l'accessibilité aux différents pôles urbains 285
Figure 8. 12 : Contraintes liées à l'accessibilité pondérée pôle de niveau 1 et celle au pôle de niveau 2 dans les règles allemandes. 286
Figure 8. 13 : Contraintes liées à l'accessibilité pondérée pôle de niveau 2 et celles au pôle de niveau 1 dans les règles allemandes. 287
Figure 8. 14 : Zones préférentielles d'urbanisation allemandes au regard de l'accessibilité aux différents pôles urbains. 288

Figure 8. 15 : Contraintes liées à l'accessibilité pondérée pôle de niveau 1 et au pôle de niveau 2 dans les règles transfrontalières. 289
Figure 8. 16 : Contraintes liées à l'accessibilité pondérée pôle de niveau 2 et au pôle de niveau 1 dans les règles transfrontalières. 290
Figure 8. 17 : Zones préférentielles d'urbanisation transfrontalière au regard de l'accessibilité aux différents pôles urbains. 291
Figure 8. 18 : Zones préférentielles d'urbanisation franco-allemande au regard de l'accessibilité aux différents pôles urbains. 291

Chapitre 9 :

Figure 9. 1 : Caractère spécifique des règles de transitions 297
Figure 9. 2 : Focalisation sur le développement urbain illustré par la règle 1 304
Figure 9. 3 : Simulation dynamique du développement urbain observé autour d'un village de la zone d'étude 309
Figure 9. 4 : Développement urbain simulé de 1990 à 2038 310
Figure 9. 5 : Développement urbain simulé à l'horizon 2022 311
Figure 9. 6 : Développement urbain simulé à l'horizon 2038 311
Figure 9. 7 : Intensité du développement urbain simulé entre 1990 et 2006 312
Figure 9. 8 : Critères de sélection (1er filtre) d'un outil d'aide à la décision à vocation territoriale d'après l'ADEME (2016) 314
Figure 9. 9 : Critères de sélection (2ème filtre) d'un outil d'aide à la décision à vocation territoriale d'après l'ADEME (2016) 314
Figure 9. 10 : Sélection des outils à analyser de manière approfondie : répartition selon les axes Intégré et Multicritère (2016) 315

Liste des tables

Chapitre 3 :

Table 3. 1 : Classification 8 classes élaborée sur la base des 44 classes Corine Land Cover (Copernicus) 98
Table 3. 2 : Évolution de l'occupation du sol entre 1990, 2006 et 2012 en hectare (CORINE Land Cover fusionné) 101
Table 3. 3 : Transitions d'occupation du sol 2006-2012 (Source : CORINE Land Cover fusionné) 101
Table 3. 4 : Changement d'occupation du sol entre 1990 et 2006 (CLC fusionné) 103

Chapitre 4 :

Table 4. 1 Calibrations et modèles de changement d'occupation du sol (van Vliet et al., 2016) 149

Chapitre 5 :

Table 5. 1 : Deux grandes familles de méthodes statistiques en sciences sociales (Tufféry, 2017) 164
Table 5. 2 : Méthodes prédictives mobilisables en fonction des types des variables explicatives ou à expliquer (Tufféry, 2017) 165
Table 5. 3 Synthèse des forces et faiblesses des méthodes statistiques mobilisables 167
Table 5. 4 Matrice de confusion 175
Table 5. 5 : Liste des paramètres et intervalle test 179

Chapitre 6 :

Table 6. 1 : Rappel de la liste des paramètres et intervalle test 191
Table 6. 2 : Paramétrages et performances de l'apprentissage sur les données d'occupation du sol uniquement 197
Table 6. 3 : Paramétrages et performances de l'apprentissage sur la totalité des variables 200
Table 6. 4 : Variation du CCM test pour un CCM apprentissage de 0,98 202
Table 6. 5 : Évaluation de la sensibilité du modèle suivant l'échantillonnage entraînement - test 203
Table 6. 6 : Les 10 répétitions du modèle de voisinage 30 et de pondération 160 204

Table 6. 7 : Matrice de confusion du paramétrage 2 599 205

Chapitre 7 :

Table 7. 1 : Paramétrages les plus performants d'après la somme des erreurs de prédictions 221
Table 7. 2 : Occurrence des variables de localisation géographique dans les 15 meilleurs paramétrages. 221
Table 7. 3 : Répétition du modèle contenant toutes les variables de localisation géographique 222
Table 7. 4 : Règles et types de transitions 226
Table 7. 5 : Illustration de l'influence des différents types d'occupation du sol 227
Table 7. 6 : Attribution du score et conversion en pourcentage de l'influence des variables 229
Table 7. 7 : Influence de la polarité des différents types de variables 235
Table 7. 8 : Score et polarité des seuils de l'urbain 236
Table 7. 9 : Polarité des pics majeurs d'influence de l'urbain dans le voisinage 237
Table 7. 10 : Extrait des seuils urbains en fonction de la distance 238
Table 7. 11 : Distribution de l'urbain dans le voisinage – tendances majeures à l'échelle de la zone 239
Table 7. 12 : Distribution de la forêt dans le voisinage – tendances majeures à l'échelle de la zone 240
Table 7. 13 : Distribution de l'agricole dans le voisinage – tendances majeures à l'échelle de la zone 241
Table 7. 14 : Distribution des vignes dans le voisinage – tendances majeures à l'échelle de la zone 243
Table 7. 15 : Distribution des zones récréatives urbaines dans le voisinage – tendances majeures à l'échelle de la zone 244
Table 7. 16 : Distribution des zones immergées dans le voisinage – tendances majeures à l'échelle de la zone 245
Table 7. 17 : Distribution des zones industrielles et commerciales dans le voisinage – tendances majeures à l'échelle de la zone 247
Table 7. 18 : Distribution des transports dans le voisinage – tendances majeures à l'échelle de la zone 248
Table 7. 19 : Scores des variables de localisation géographique 250
Table 7. 20 : Polarité de l'influence des différentes variables d'accessibilité 252
Table 7. 21 : Tendances d'accessibilités observées au sein du jeu de règles dans sa totalité 256

Chapitre 8 :

Table 8. 1 Nature des règles 262
Table 8. 2 Influence des variables de localisation géographique en fonction de la nature des règles..... 263
Table 8. 3 Polarité de l'influence des variables de localisation géographique en fonction de la nature des règles 264
Table 8. 4 Tendance d'occupation du sol en fonction des types de règles 265
Table 8. 5 Tendances majeures françaises, allemandes et transfrontalières d'occupation urbaine dans le voisinage..... 267
Table 8. 6 Tendances majeures françaises, allemandes et transfrontalières d'occupation agricole dans le voisinage..... 269
Table 8. 7 Tendances majeures françaises, allemandes et transfrontalières d'occupation forêt dans le voisinage..... 271
Table 8. 8 Tendances majeures françaises, allemandes et transfrontalières d'occupation zones récréatives urbaines dans le voisinage..... 273
Table 8. 9 Tendances majeures françaises, allemandes et transfrontalières d'occupation vigne dans le voisinage..... 275
Table 8. 10 Tendances majeures françaises, allemandes et transfrontalières d'occupation zone industrielle et commerciale dans le voisinage 276
Table 8. 11 Tendances majeures françaises, allemandes et transfrontalières d'occupation transport dans le voisinage 278
Table 8. 12 Tendances majeures françaises, allemandes et transfrontalières d'occupation zone immergée dans le voisinage 279
Table 8. 13 Règles mobilisant les variables d'accessibilité en fonction de leur nature 281
Table 8. 14 Influence de l'accessibilité en fonction de la nature des règles 282
Table 8. 15 Seuils extrêmes mobilisés pour chaque variable d'accessibilité 283
Table 8. 16 Tendances d'accessibilités observées au sein du jeu de règles français..... 285
Table 8. 17 Tendances d'accessibilités observées au sein du jeu de règles allemand..... 287
Table 8. 18 Tendances d'accessibilités observées au sein du jeu de règles transfrontalier 290

Références

A

ADEME, (2016), Approche intégrée et multicritères dans les outils d'aide à la décision à vocation territoriale. Eléments de cadrage pour la gestion des données territoriales.

AGHIMIEN, D., AIGBAVBOA, C. & OKE, A., (2019), A Review of the Application of Data Mining For Sustainable Construction in Nigeria, Energy Procedia, vol. 158, p. 3240–3245.

AGUEJDAD, R., DOUKARI, O., HOUET, T., AVNER, P. & VIGUIÉ, V., (2016), Etalement urbain et géoprospective : apports et limites des modèles de spatialisation., Cybergeo : European Journal of Geography, n°782, p. 15.

ALFASI, N., ALMAGOR, J. & BENENSON, I., (2012), The actual impact of comprehensive land-use plans: Insights from high resolution observations, Land Use Policy, vol. 29, n°4, p. 862‑877.

ALONSO, W., (1964), Location and land use: toward a general theory of land rent, Harvard University Press, Cambridge (Mass.), Etats-Unis d'Amérique, xi+204 p.

AMILHAT SZARY, A.-L. & FOURNY, M.-C., (2006), Après les frontières, avec la frontière. Nouvelles dynamiques transfrontalières en Europe, Editions de l'Aube, 169 p.

ANDERSON, J., (1996), Cartography at the 1996 Association of American Geographers Meetfng, Cartographic Perspectives, n°24, p. 24–25.

ANDERSON, J. R., (1976), A Land use and land cover classification system for use with remote sensor data, U.S. Govt. Print. Off., Washington, Etats-Unis d'Amérique, iii+28 p.

ANTONI, J.-P., (2003), Modélisation de la dynamique de l'étalement urbain : Aspects conceptuels et gestionnaires.Application à Belfort (Thèse de doctorat).

ANTONI, J.-P., (2009), Lexique de la ville, Ellipses, Paris, France, 184 p.

ANTONI, J.-P. (ÉD.), (2010), Modéliser la ville: formes urbaines et politiques de transport, Economica, DL 2010, Paris, France, x+438; 12 p.

ANTONI, J.-P., (2016), Concepts, Méthodes et Modèles pour l'aménagement et les mobilités, Paris, 250 p.

ANTONI, J.-P., JUDGE, V., VUIDEL, G. & KLEIN, O., (2018), Constraint Cellular Automata for Urban Development Simulation: An Application to the Strasbourg-Kehl Cross-Border Area, In : M. T. CAMACHO OLMEDO, M. PAEGELOW, J.-F. MAS et F. ESCOBAR (éd.), Geomatic Approaches for Modeling Land Change Scenarios, Springer, p. 293‑306.

ARBARET-SCHULZ, C., (2013), La question du continu et du discontinu à l'épreuve de la dimension technique des sociétés, In : A. GÉNIN et A. FRÉDERIC (éd.), Continu et

discontinu dans l'espace géographique, Presses universitaires François-Rabelais, Tours, p. 409‑416.

AUMANN, C. A., (2007), A methodology for developing simulation models of complex systems, Ecological Modelling, vol. 202, n°3‑4, p. 385‑396.

B

BANOS, A., (2016), Modéliser, c'est apprendre: itinéraire d'un géographe, Editions Matériologiques.

BATTY, M., (1997), Cellular Automata and Urban Form: A Primer, Journal of the American Planning Association, vol. 63, n°2, p. 266‑274.

BATTY, M., (1998), Urban Evolution on the Desktop: Simulation with the Use of Extended Cellular Automata, Environment and Planning A: Economy and Space, vol. 30, n°11, p. 1943‑1967.

BATTY, M., XIE, Y. & SUN, Z., (1999), Modeling urban dynamics through GIS-based cellular automata, Computers, Environment and Urban Systems, vol. 23, n°3, p. 205‑233.

BAUD, M. & VAN SCHENDEL, W., (1997), Toward a Comparative History of Borderlands, Journal of World History, vol. 8, n°2, p. 211‑242.

BAVOUX, J. J., BEAUCIRE, F., CHAPELON, I. & ZEMBRI, P., (2005), Géographie des transports, Paris.

BEN-AKIVA, M. & DE PALMA, A., (1986), Analysis of a dynamic residential location choice model with transaction costs, Journal of Regional Science, vol. 26, n°2, p. 321‑341.

BENENSON, I. & TORRENS, P. M., (2004), Geosimulation: Automata-based Modeling of Urban Phenomena, Wiley.

BIERMANN, F., BETSILL, M. M., GUPTA, J., KANIE, N., LEBEL, L., LIVERMAN, D., … ZONDERVAN, R., (2010), Earth system governance: A research framework, International Environmental Agreements: Politics, Law and Economics, vol. 10, n°4, p. 277‑298.

BLAIKIE, P. & BROOKFIELD, H., (1986), Land degradation and society, Land degradation and society.

BOUGHORBEL, S., JARRAY, F. & EL-ANBARI, M., (2017), Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric, PLOS ONE, vol. 12, n°6, p. e0177678.

BOX, G. E. P. & DRAPER, N., (1987), Empirical Model-Building and Response Surfaces., 424 p.

BREDÈCHE, N., (2017), Introduction aux Automates Cellulaires.

BREIMAN, L., FRIEDMAN, J., STONE, C. J. & OLSHEN, R. A., (1984), Classification and Regression Trees, Taylor & Francis.

BRIASSOULIS, H., (2000), Analysis of land use change: theoretical and modeling approaches, In : Introduction to Regional Science, West Virginia University, MorgantownThe Web Book of Regional Science.

BRIASSOULIS, H., (2008), Land-Use Policy and Planning, Theorizing, and Modeling: Lost in Translation, Found in Complexity?, Environment and Planning B: Planning and Design, vol. 35, n°1, p. 16‑33.

BROUWER, F. M. & CHADWICK, M. J., (1991), Future land use patterns in Europe, Land use changes in Europe, p. 49‑78.

BROWN, S. D. & MYLES, A. J., (2009), 3.17 - Decision Tree Modeling in Classification, In : S. D. BROWN, R. TAULER et B. WALCZAK (éd.), Comprehensive Chemometrics, Elsevier, Oxford, p. 541–569.

BRUN, G., (2015), 1900 : L'Alsace à la veille de la première Guerre Mondiale, Canopé - Académie de Strasbourg : Base numérique du patrimoine d'Alsace. http://www.crdp-strasbourg.fr/data/histoire/1GM_combats/alsace_avant_conflit.php (page consultée le 16/04/19)

BRUNET, R., THÉRY, H. & FERRAS, R., (1993), Les mots de la géographie : dictionnaire critique., Reclus, 470 p.

BUCHANAN, A. & MOORE, M., (2003), States, nations, and borders: The ethics of making boundaries, Cambridge University Press, 1 p.

BURLEY, T. M., (1961), Land Use or Land Utilization?, The Professional Geographer, vol. 13, n°6, p. 18‑20.

C

CALICODE, (2018), Automate cellulaire en Python · Snarkturne. https://callicode.fr/blog/post/dmiold-automate-cellulaire.html (page consultée le 15/05/19)

CAMACHO OLMEDO, M. T., PAEGELOW, M., MAS, J. F. & ESCOBAR, F., (2018), Geomatic Approaches for Modeling Land Change Scenarios. An Introduction, In : M. T. CAMACHO OLMEDO, M. PAEGELOW, J.-F. MAS et F. ESCOBAR (éd.), Geomatic Approaches for Modeling Land Change Scenarios, Springer International Publishing, Cham, p. 1–8.

CARPENTIER, S., EPSTEIN, D. & GERBER, P., (2013), Implications de la mobilité quotidienne dans les stratégies résidentielles transfrontalières, Espace, populations, sociétés, vol. 2012, p. 95‑115.

CHARTE DU G8 POUR L'OUVERTURE DES DONNÉES PUBLIQUES, (2013), Mission Etalab.

CHAUVEAU, E., CHADENAS, C., COMENTALE, B., POTTIER, P., BLANLŒIL, A., FEUILLET, T., … TROUILLET, B., (2011), Xynthia : leçons d'une catastrophe, Cybergeo : European Journal of Geography.

CHIGNARD, S., (2013), A brief history of Open Data - Paris Innovation Review, Paris Innovation review. http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data (page consultée le 13/05/19)

CHRISTALLER, W., (1933), Die zentralen Orte in Süddeutschland. Eine ökonomisch-geographische Untersuchung über die Gesetzmä igkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen (PhD Thesis), Jena.

CIESLAK, D. A. & CHAWLA, N. V., (2008), Learning Decision Trees for Unbalanced Data, In : W. DAELEMANS, B. GOETHALS et K. MORIK (éd.), Machine Learning and Knowledge Discovery in Databases, Springer Berlin Heidelberg, Berlin, Heidelberg, p. 241–256.

CLARKE, K. C., (2018), Cellular Automata and Agent-Based Models, In : M. M. FISCHER et P. NIJKAMP (éd.), Handbook of Regional Science, Springer Berlin Heidelberg, Berlin, Heidelberg, p. 1‑16.

CLARKE, K. C. & GAYDOS, L. J., (1998), Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore, International Journal of Geographical Information Science, vol. 12, n°7, p. 699‑714.

CLARKE, K. C., HOPPEN, S. & GAYDOS, L. J., (1997), A Self-Modifying Cellular Automaton Model of Historical Urbanization in the San Francisco Bay Area, Environment and Planning B: Planning and Design, vol. 24, n°2, p. 247‑261.

CLARKE-LAUER, M. D. & CLARKE, K. C., (2011), Evolving Simulation Modeling: Calibrating SLEUTH Using a Genetic AlgorithmPrésenté à 11th International Conference on GeoComputation, , London, p. 5.

CLAVAL, P., (1966), Chronique de géographie économique. La théorie des lieux centraux., Revue Géographique de l'Est, vol. 6, n°1, p. 131–152.

CNRS, (2O18), CNRS Images - Les Sciences de la Terre - Lexique. https://www.cnrs.fr/cnrs-images/sciencesdelaterreaulycee/contenu/lexique.html (page consultée le 10/05/19)

COMMISSARIAT GÉNÉRAL AU DÉVELOPPEMENT DURABLE, (2009), CORINE Land Cover France Guide d'utilisation, Ministère de l'écologie, de l'énergie, du développement durable et de l'aménagement du territoire.

CONSEIL DE L'EUROPE, Convention-cadre européenne sur la coopération transfrontalière des collectivités ou autorités territoriales (1980).

COUCLELIS, H., (1985), Cellular Worlds: A Framework for Modeling Micro—Macro Dynamics, Environment and Planning A: Economy and Space, vol. 17, n°5, p. 585‑596.

COUCLELIS, H., (1988), Of Mice and Men: What Rodent Populations Can Teach Us about Complex Spatial Dynamics, Environment and Planning A: Economy and Space, vol. 20, n°1, p. 99‑109.

COUCLELIS, H., (1997), From Cellular Automata to Urban Models: New Principles for Model Development and Implementation, Environment and Planning B: Planning and Design, vol. 24, n°2, p. 165‑174.

COUCLELIS, H., (2005), "Where has the Future Gone?" Rethinking the Role of Integrated Land-Use Models in Spatial Planning, Environment and Planning A: Economy and Space, vol. 37, n°8, p. 1353‑1371.

D

DAUSSIN-BENICHOU, J.-M. & SALA, M., (2013), Pourquoi le chômage a-t-il continué de baisser en Allemagne après 2007 ? − Éclaircie mondiale, l'Europe encore dans l'ombre | Insee, INSEE.

DE ANGELIS, D. & GROSS, L. J., (1992), Individual based models and approaches in ecology, De Angelis D., Gross L.J., New York, 522 p.

DECOVILLE, A. & DURAND, F., (2019), Exploring cross-border integration in Europe: How do populations cross borders and perceive their neighbours?, European Urban and Regional Studies, vol. 26, n°2, p. 134‑157.

DELDEN, H. VAN & HURKENS, J., (2011), A generic Integrated Spatial Decision Support System for urban and regional planning, Perth, Australia.

DEUTSCH, A. & DORMANN, S., (2004), Cellular Automaton Modeling of Biological Pattern Formation. Characterization, Applications and Analysis, 331 p.

DI NAPOLI-MARCHETTI, M., (2009), Représentation des frontières : la spécificité de la figuration cartographique des frontières aux États-Unis au xixe siècle, L'Information géographique, vol. 73, n°4, p. 24‑34.

DIETZEL, C. & CLARKE, K. C., (2007), Toward Optimal Calibration of the SLEUTH Land Use Change Model, Transactions in GIS, vol. 11, n°1, p. 29‑45.

DUBOS-PAILLARD, E. & LANGLOIS, P., (2018), Simulation et validation de la dynamique urbaine par automate cellulaire, HAL slhs, n°01694570, p. 17.

DUPUY, G., (1999), La dépendance automobile : Symptômes, analyses, diagnostic, traitements, Anthropos, 160 p.

E

EASTMAN, J. R., (2009), DRISI Taiga. Guide to GIS and Remote Processing, Clark University: Worcester, MA, USA, 234–256 p.

ENGELEN, G., WHITE, R. & TIMMERMANS, H. J. P., (1997), Integrating Constrained Cellular Automata Models, GIS and Decision Support Tools for Urban Planning and Policy Making.

ENGELEN, G., WHITE, R., ULJEE, I. & DRAZAN, P., (1995), Using cellular automata for integrated modelling of socio-environmental systems, Environmental Monitoring and Assessment, vol. 34, n°2, p. 203–214.

EPSTEIN, J. M., (2008), Why model?, Journal of Artificial Societies and Social Simulation, vol. 11, n°4, p. 12.

EURODISTRICT, (2013), Rapport d'activité.

EUROPEAN ENVIRONMENT AGENCY, (1985), Corine land Cover, European Environment Agency.

EUROPEAN UNION & EUROSTAT, (2000), Manuel des concepts relatifs aux sytèmes d'information sur l'occupation et l'utilisation des sols, EUR-OP, Luxembourg.

F

FAWCETT, T., (2006), An introduction to ROC analysis, Pattern Recognition Letters, vol. 27, n°8, p. 861‑874.

FRANKHAUSER, P., TANNIER, C., HOUOT, H. & VUIDEL, G., (2010), Développement urbain fractal sous contraintes d'accessibilités - Modèles et outils d'aide à la décision pour l'aménagement urbain, 85 p.

FREDKIN, E. & TOFFOLI, T., (1982), Conservative logic, International Journal of Theoretical Physics, vol. 21, n°3, p. 219–253.

FRÉMOND, M., (2015), Une approche normative de l'aménagement au Luxembourg. Evaluation par la simulation (Thèse de doctorat), Besançon, Franche-Comté.

FUJITA, M., (1986), Urban Land Use Theory, In : Location Theory, J. Gabszewicz, J.-J. Thisse, M. Fujita, U. Scheiwzer, SwitzerlandHarwood Academic.

G

GARDNER, M., (1970), Mathematical games, Scientific American, vol. 222, n°6, p. 132–140.

GHOSH, P., MUKHOPADHYAY, A., CHANDA, A., MONDAL, P., AKHAND, A., MUKHERJEE, S., … HAZRA, S., (2017), Application of Cellular automata and Markov-chain model in geospatial environmental modeling- A review, Remote Sensing Applications: Society and Environment, vol. 5, p. 64‑77.

GIBBS, H. K., RUESCH, A. S., ACHARD, F., CLAYTON, M. K., HOLMGREN, P., RAMANKUTTY, N. & FOLEY, J. A., (2010), Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s, Proceedings of the National Academy of Sciences, vol. 107, n°38, p. 16732‑16737.

GIEC, (2000), L'utilisation des terres, le changement d'affectation des terres et la foresterie: résumé à l'intention des décideurs : rapport spécial du Groupe d'experts intergouvernemental sur l'évolution du climat, OMM (Organisation météorologique mondiale) : PNUE (Programme des Nations Unies pour l'environnement), Genève.

GILBERT, A. & BROSSEAU, M., (2011), La frontière asymétrique : Franco-Ontariens et Anglo-Québécois dans la région de la capitale nationale, The Canadian Geographer, vol. 55, .

GRIFFOND-BOITIER, A., MARIANI-ROUSSET, S., FRANKHAUSER, P., VALENTIN, J., ALEXANDRE, V. & NICOT, B., (2012), Habiter : où et comment ?, In : La décision d'habiter ici ou ailleurs, Economica, p. 43‑74.

H

HAGEN, A., (2003), Fuzzy set approach to assessing similarity of categorical maps, International Journal of Geographical Information Science, vol. 17, n°3, p. 235‑249.

HAGEN-ZANKER, A. & LAJOIE, G., (2008), Neutral models of landscape change as benchmarks in the assessment of model performance, Landscape and Urban Planning, vol. 86, n°3‑4, p. 284‑296.

HAGERSTRÄND, T., (1967), Innovation and diffusion as a spatial process traduction Allan Pred, University of Chicago Press., p. 334.

HÄGERSTRAND, T., (1970), What about people in Regional Science?, Papers in Regional Science, vol. 24, n°6, .

HÄGERSTRAND, T., (1985), Time-geography: focus on the corporeality of man, society, and environment, The science and praxis of complexity, p. 193–216.

HAGGETT, P., CLIFF, A. D. & FREY, A., (1977), Locational analysis in human geography., Tijdschrift Voor Economische En Sociale Geografie, vol. 68, n°6, .

HAIBO, H. & GARCIA, E. A., (2009), Learning from Imbalanced Data, IEEE Transactions on Knowledge and Data Engineering, vol. 21, n°9, p. 1263‑1284.

HAIXIANG, G., YIJING, L., SHANG, J., MINGYUN, G., YUANYUE, H. & BING, G., (2017), Learning from class-imbalanced data: Review of methods and applications, Expert Systems with Applications, vol. 73, p. 220‑239.

HAMEZ, G., AMILHAT-SZARY, A.-L., PARIS, D., REITEL, B. & WALTHER, O., (2013), Guest editorial : Modèles de frontières, Belgeo. Revue belge de géographie, n°1, .

HANSEN, W. G., (1959), How Accessibility Shapes Land Use, Journal of the American Institute of Planners, vol. 25, n°2, p. 73–76.

HEMA LABORATORY, (2006), Land Transformation Model.

HERRMANN, L., (2018), Le lotissement en France : histoire réglementaire de la construction d'un outil de production de la ville, In : Géoconfluences.

HERSPERGER, A. M., OLIVEIRA, E., PAGLIARIN, S., PALKA, G., VERBURG, P., BOLLIGER, J. & GRĂDINARU, S., (2018), Urban land-use change: The role of strategic spatial planning, Global Environmental Change, vol. 51, p. 32‑42.

HEWITT, R., VAN DELDEN, H. & ESCOBAR, F., (2014), Participatory land use modelling, pathways to an integrated approach, Environmental Modelling & Software, vol. 52, p. 149‑165.

HILLIER, J., (2007), Stretching beyond the horizon: a multiplanar theory of spatial planning and governance, Routledge, London.

HINFRAY, N., (2010), Les espaces transfrontaliers, nouveaux territoires de projet , nouveaux pôles de développement au sein d'une Europe polycentrique ? (Thèse de doctorat, Tours).

HOLMGREN, P., (2006), Global Land Use Area Change Matrix - Input to the Fourth Global Environmental Outlook (GEO-4), Forestry Department, FAO.

HOUET, T., AGUEJDAD, R., DOUKARI, O., BATTAIA, G. & CLARKE, K. C., (2016), Description and validation of a "non path-dependent" model for projecting contrasting urban growth futures, Cybergeo : European Journal of Geography, vol. 759, .

HOUSSEMAND, J., GRIFFOND-BOITIER, A., YOUSSOUFI, S. & ANTONI, J.-P., (2018), Caractériser l'offre d'habitat par la demande résidentielle. Éléments pour un renouvellement conceptuel du couple logement/environnement, Cybergeo : European Journal of Geography.

HUTZLER, G., (2015), Automates cellulaires, Evry.

I

INSEE, (2015), Dossier complet - Résultats pour toutes les communes, départements, régions, intercommunalités, ..., INSEE.

INSEE, (2016), Définition - Unité urbaine, Définition, méthodes et qualité. https://www.insee.fr/fr/metadonnees/definition/c1501 (page consultée le 23/03/19)

J

JENIK, C., (2017), Infographie: Les plus longs trajets domicile-travail du monde.

JONGMAN, R. H. G., (1995), Nature conservation planning in Europe: developing ecological networks, Landscape and Urban Planning, vol. 32, n°3, p. 169‑183.

JUIGNET, P., (2015), Le concept d'émergence.

K

KLEIN, O., (2017), Mesurer la mobilité tranfrontalièrePrésenté à Quelles gouvernances pour l'information géographique ?, , Strasboourg.

KLEINBAUM, D., (1994), Logistic Regression; A Self-Learning Text, New York, USA, 312 p.

KOCABAS, V. & DRAGICEVIC, S., (2006), Assessing cellular automata model behaviour using a sensitivity analysis approach, Computers, Environment and Urban Systems, vol. 30, n°6, p. 921‑953.

KOCEVA, M. M., BRANDMÜLLER, T., LUPU, I., ÖNNERFORS, Å., CORSELLI-NORDBLAD, L., COYETTE, C., … EUROPÄISCHE KOMMISSION (ÉD.), (2016), Urban Europe: statistics on cities, towns and suburbs, Publications Office of the European Union, Luxembourg, 282 p.

KOLOSOV, V., (2017), Préface - Etudier les frontières, un enjeu contemporain, In : FrontièresPresses Universitaires de Bordeaux.

KOTSIANTIS, S. B., (2011), Decision trees: a recent overview, Artificial Intelligence Review, vol. 39, n°4, p. 261‑283.

KRZANOWSKI, W. J., HAND, D. J. & HAND, D. J., (2009), ROC Curves for Continuous Data, Chapman and Hall/CRC.

L

LAMBIN, E. F. & GEIST, H. (ÉD.), (2006), Land-Use and Land-Cover Change, Springer Berlin Heidelberg, Berlin, Heidelberg.

LAMBIN, E. F., TURNER, B. L., GEIST, H. J., AGBOLA, S. B., ANGELSEN, A., BRUCE, J. W., … XU, J., (2001), The causes of land-use and land-cover change: Moving beyond the myths, Global Environmental Change, vol. 11, n°4, p. 261‑269.

LANGLEY, P. & SIMON, H. A., (1995), Applications of Machine Learning and Rule Induction, Commun. ACM, vol. 38, n°11, p. 54–64.

LANGTON, C. G., (1986), Studying artificial life with cellular automata, Physica D: Nonlinear Phenomena, vol. 22, n°1‑3, p. 120‑149.

LELOUP, F. & GAGNOL, L., (2017), Présentation. De la barrière à la coopération transfrontalière : frontière, développement et gouvernance de l'environnement, Mondes en developpement, vol. n° 177, n°1, p. 7‑12.

LELOUP, F. & MOYART, L., (2006), La Région frontalière : vers quels nouveaux modes de développement et de gouvernance?, In : Après les frontières, avec la frontière : nouvelles dynamiques transfrontalières en Europe, Amilhat-Szary A.-L, Fourny M-C, La Tour d'Aigues, FranceEditions de l'Aube, , p. 38‑54.

LETOURNEAU, A., VERBURG, P. H. & STEHFEST, E., (2012), A land-use systems approach to represent land-use dynamics at continental and global scales, Environmental Modelling & Software, vol. 33, p. 61‑79.

LÉVY, J. & LUSSAULT, M., (2013), Dictionnaire de géographie et de l'espace des sociétés, La Documentation Française, 1137 p. p.

LEWES, G. H., (1875), Problems of Life and Mind, A M S Press, Incorporated.

LI, X. & YEH, A. G.-O., (2000), Modelling sustainable urban development by the integration of constrained cellular automata and GIS, International Journal of Geographical Information Science, vol. 14, n°2, p. 131‑152.

LIM, T.-S., LOH, W.-Y. & SHIH, Y.-S., (2000), A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-three Old and New Classification Algorithms, Machine Learning, n°40, p. 203‑229.

LOH, C. G., (2011), Assessing and Interpreting Non-conformance in Land-use Planning Implementation, Planning Practice & Research, vol. 26, n°3, p. 271‑287.

LÓPEZ, V., FERNÁNDEZ, A., GARCÍA, S., PALADE, V. & HERRERA, F., (2013), An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics, Information Sciences, vol. 250, p. 113‑141.

LÖSCH, A., (1940), Die räumliche Ordnung der Wirtschaft, G. Fischer, Jena.

M

MAGLIOCCA, N. R., VAN VLIET, J., BROWN, C., EVANS, T. P., HOUET, T., MESSERLI, P., … YU, Q., (2015), From meta-studies to modeling: Using synthesis knowledge to build broadly applicable process-based land change models, Environmental Modelling & Software, vol. 72, p. 10‑20.

MARTINEZ, G. J., SECK-TUOH-MORA, J. C. & ZENIL, H., (2013), Wolfram's Classification and Computation in Cellular Automata Classes III and IV, In : Irreducibility and Computational Equivalence 10 Years After Wolfram's A New Kind of Science, Hector ZenilSpringer, , p. 237–259.

MARTINEZ, O. J., (1994), Border people: life and society in the US-Mexico borderlands, Border people: life and society in the US-Mexico borderlands.

MAS, J. F., PAEGELOW, M. & CAMACHO OLMEDO, M. T., (2018), LUCC Modeling Approaches to Calibration, In : M. T. CAMACHO OLMEDO, M. PAEGELOW, J.-F. MAS et F. ESCOBAR (éd.), Geomatic Approaches for Modeling Land Change Scenarios, Springer International Publishing, Cham, p. 11‑25.

MATTHEWS, B. W., (1975), Comparison of the predicted and observed secondary structure of T4 phage lysozyme, Biochimica et Biophysica Acta (BBA) - Protein Structure, vol. 405, n°2, p. 442‑451.

MCCULLOCH, W. S. & PITTS, W., (1943), A logical calculus of the ideas immanent in nervous activity, The Bulletin of Mathematical Biophysics, vol. 5, n°4, p. 115–133.

MEDINA-NICOLAS, L., (2004), L'étude des frontières : un état des lieux à travers la production doctorale française, Annales de Géographie, vol. 113, n°635, p. 74–86.

MEHTA, M., AGRAWAL, R. & RISSANEN, J., (1996), SLIQ: A fast scalable classifier for data mining, In : P. APERS, M. BOUZEGHOUB et G. GARDARIN (éd.), Advances in Database Technology — EDBT '96, Springer Berlin Heidelberg, p. 18‑32.

MEINEL, G. & WINKLER, M., (2002, janvier 1), Spatial analysis of settlement and open land trends in urban areas on basis of RS data - studies of five European cities over a 50-year period.

MERLIN, P., (2007), L'aménagement du territoire en France, La Documentation française, DL 2007, Paris, France, 174 p.

METROPOLIS, N. & ULAM, S., (1949), The Monte Carlo Method, Journal of the American Statistical Association, vol. 44, n°247, p. 335‑341.

MEYER, W. B. & TURNER, B. L., (1992), Human Population Growth and Global Land-Use/Cover Change, Annual Review of Ecology and Systematics, vol. 23, n°1, p. 39‑61.

MIMEUR, C., (2016), Les traces de la vitesse entre réseau et territoire : approche géohistorique de la croissance du réseau ferroviaire français (Thèses de doctorat, Université de Bourgogne).

MINSKY, M., (1965), Matter, Mind and Models.

MISHRA, V., CHERKAUER, K. A., NIYOGI, D., LEI, M., PIJANOWSKI, B. C., RAY, D. K., … YANG, G., (2010), A regional scale assessment of land use/land cover and climatic changes on water and energy cycle in the upper Midwest United States, International Journal of Climatology, vol. 30, n°13, p. 2025‑2044.

MITCHELL, T. M., (1997), Does Machine Learning Really Work?, AI Magazine, vol. 18, n°3, p. 11‑11.

MOLLICONE, D., ACHARD, F., EVA, H., BELWARD, A., FEDERICI, S., LUMICISI, A., … VALENTINI, R., (2003), Land Use Change Monitoring in the framework of the UNFCCCC and its Kyoto Protocol: Report on Current Capabilities of Satellite Remote Sensing Technology, vol. EUR 20867 EN, p. 48.

MON PAYS C'EST LA FRONTIÈRE, (2017, juillet 23), LCI : Grands reportages.

MONNOYER-SMITH, L., (2018), La lettre du Commissariat général au développement durable.

MOSER, S. C., (1996), A partial instructional module on global and regional land use/cover change: assessing the data and searching for general relationships, GeoJournal, vol. 39, n°3, p. 241–283.

MOT, (2015), espaces-transfrontaliers.org: espaces-transfrontaliers.org. http://www.espaces-transfrontaliers.org/ (page consultée le 15/05/19)

MOULLÉ, F., (2013), La frontière et son double. Un modèle à partir de l'expérience européenne, Belgeo. Revue belge de géographie, n°1, .

MOULLÉ, F., (2017), Frontières, Presses Universitaires de Bordeaux.

MURTHY, S. K., (1998), Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey, Data Mining and Knowledge Discovery, vol. 2, n°4, p. 345‑389.

MUTH, R. F., (1969), Cities and housing: the spatial pattern of urban residential land use, Chicago, Etats-Unis d'Amérique, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, xxii+355 p.

N

NAGEL, K. & SCHRECKENBERG, M., (1992), A cellular automaton model for freeway traffic, Journal de Physique France, vol. 2, n°12, p. 2221‑2229.

NEWMAN, G., LEE, J. & BERKE, P., (2016), Using the land transformation model to forecast vacant land, Journal of Land Use Science, vol. 11, n°4, p. 450‑475.

NORDMAN, D., (1998), Frontières de France. De l'espace au territoire, XVIe-XIXe siècles, París, Gallimard.

O

O'CONNOR, T. & WONG, H. Y., (2015), Emergent Properties, In : E. N. ZALTA (éd.), The Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, Stanford UniversitySummer 2015.

OJIMA, D., MORAN, E., MCCONNELL, W., STAFFORD SMITH, M., LAUMANN, G., MORAIS, J. & YOUNG, B., (2005), Science plan and implementation strategy, IGBP Secretariat.

OTTERMAN, J., (1974), Baring High-Albedo Soils by Overgrazing: A Hypothesized Desertification Mechanism, Science, vol. 186, n°4163, p. 531‑533.

P

PADUA, L., SCHULZE, H., MATKOVIĆ, K. & DELRIEUX, C., (2014), Interactive exploration of parameter space in data mining: Comprehending the predictive quality of large decision tree collections, Computers & Graphics, vol. 41, p. 99‑113.

PAEGELOW, M., CAMACHO OLMEDO, M. T. & MAS, J. F., (2018), Techniques for the Validation of LUCC Modeling Outputs, In : Geomatic Approaches for Modeling Land Change ScenariosSpringer, , p. 53‑80.

PAL, M. & MATHER, P. M., (2003), An assessment of the effectiveness of decision tree methods for land cover classification, Remote Sensing of Environment, vol. 86, n°4, p. 554‑565.

PARETO, V., (1897), Cours d'économie politique, F. Rouge, Lausanne, Suisse, 904 p.

PEPE, M. S., (2003), The Statistical Evaluation of Medical Tests for Classification and Prediction, Oxford University Press.

PHIPPS, M., (1989), Dynamical Behavior of Cellular Automata under the Constraint of Neighborhood Coherence, Geographical Analysis, vol. 21, n°3, p. 197‑215.

PINTO, N. N. & ANTUNES, A. J. P., (2010), A Cellular Automata Model Based on Irregular Cells: Application to Small Urban Areas, Environment and Planning B: Planning and Design, vol. 37, n°6, p. 1095‑1114.

PLIENINGER, T., DRAUX, H., FAGERHOLM, N., BIELING, C., BÜRGI, M., KIZOS, T., … VERBURG, P. H., (2016), The driving forces of landscape change in Europe: A systematic review of the evidence, Land Use Policy, vol. 57, p. 204‑214.

PONTIUS, R. G., CORNELL, J. D. & HALL, C. A. S., (2001), Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica, Agriculture, Ecosystems & Environment, vol. 85, n°1, p. 191‑203.

PUMAIN, D., (2004), Lieux centraux, In : Hypérgéo.

PUMAIN, D., (2008), The Socio-Spatial Dynamics of Systems of Cities and Innovation Processes: a Multi-Level Model, In : S. ALBEVERIO, D. ANDREY, P. GIORDANO et A. VANCHERI (éd.), The Dynamics of Complex Urban Systems: An Interdisciplinary Approach, Physica-Verlag HD, Heidelberg, p. 373‑389.

Q

QUINLAN, J. R., (1986), Induction of decision trees, Machine Learning, vol. 1, n°1, p. 81‑106.

QUINLAN, R., (1993), C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA.

R

RATTI, R. & SCHULER, M., (2013), Typologie des espaces-frontières à l'heure de la globalisation, Belgeo (En ligne), n°1, .

RAY, B., BROSSEAU, M., VERONIS, L. & GILBERT, A., (2014), La frontière au quotidien, Université d'Ottawa, Ottawa, Canada, 227 p.

REIMER, M., GETIMIS, P. & BLOTEVOGEL, H., (2014), Spatial Planning Systems and Practices in Europe: A Comparative Perspective on Continuity and Changes, Routledge, 332 p.

REITEL, B., (2017), La frontière, une approche polysémique et multi-scalaire, In : Frontières, Moullé F., BordeauxPresses Universitaire de Bordeaux, , p. 53‑71.

RENARD, J.-P., (2000), Introduction aux Automates Cellulaires.

RIKS, (2009), Metronamica - Model descriptions., RIKS.

RINDFUSS, R. R., ENTWISLE, B., WALSH, S. J., AN, L., BADENOCH, N., BROWN, D. G., … VERBURG, P. H., (2008), Land use change: complexity and comparisons, Journal of Land Use Science, vol. 3, n°1, p. 1‑10.

ROKACH, L. & MAIMON, O., (2010), Data mining with decision trees: theory and applications, World Scientific, New Jersey, London, Singapore, ..., Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, xviii+244 p.

ROUNSEVELL, M. D. A., PEDROLI, B., ERB, K.-H., GRAMBERGER, M., BUSCK, A. G., HABERL, H., … WOLFSLEHNER, B., (2012), Challenges for land system science, Land Use Policy, vol. 29, n°4, p. 899‑910.

S

SAFAVIAN, S. R. & LANDGREBE, D., (1991), A survey of decision tree classifier methodology, IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, n°3, p. 660‑674.

SANGERMANO, F., EASTMAN, J. R. & ZHU, H., (2010), Similarity Weighted Instance-based Learning for the Generation of Transition Potentials in Land Use Change Modeling, Transactions in GIS, vol. 14, n°5, p. 569‑580.

SANTAMARIA, F., (2016), Origines et fondements de l'aménagement du territoire, In : L'aménagement du territoire en France, La documentation française.

SCHELLING, T. C., (1971), Dynamic models of segregation, Journal of Mathematical Sociology, vol. 1, p. 143‑186.

SCHOLL, B., ELGENDY, H. & NOLLERT, M., (2007), Raumplanung in Deutschland - Formeller Aufbau und zukünftige Aufgaben, Schriftenreihe des Instituts für Städtebau und Landesplanung, vol. 35, .

SÉNAT, (2018), Les documents d'urbanisme et les règles générales d'urbanisme | Collectivités locales. https://www.collectivites-locales.gouv.fr/documents-durbanisme-et-regles-generales-durbanisme (page consultée le 15/05/19)

SETO, K. C., FRAGKIAS, M., GÜNERALP, B. & REILLY, M. K., (2011), A Meta-Analysis of Global Urban Land Expansion, PLoS ONE, vol. 6, n°8, p. e23777.

SETO, K. C., KAUFMANN, R. K. & WOODCOCK, C. E., (2000), Landsat reveals China's farmland reserves, but they're vanishing fast, Nature, vol. 406, n°6792, p. 121‑121.

SHAFER, J., AGRAWAL, R., MEHTA, M., ROAD, H. & JOSE, S., (1996), SPRINT: A Scalable Parallel Classi er for Data MiningPrésenté à Proceedings of the 22 nd VLDB Conference, , Inde, p. 12.

SHMUELI, G., (2010), To Explain or to Predict?, Statistical Science, vol. 25, n°3, p. 289–310.

SOHL, T. & SAYLER, K., (2008), Using the FORE-SCE model to project land-cover change in the southeastern United States, Ecological Modelling, vol. 219, n°1, p. 49‑65.

SONQUIST, J. & MORGAN, J., (1964), The detection of interaction effects. A report on a computer program for the selection of optimal combinations of explanatory variables, Ann Arbor, MI: Survey Research Centre, The Institute for Social Research, University of Michigan, n°35, .

STATISTISCHES BUNDESAMT, ([s.d]), Destatis, Federal Statistical Office. http://www.destatis.de/EN/Home/_node.html (page consultée le 12/04/19)

STEHMAN, S. V., (1997), Selecting and interpreting measures of thematic classification accuracy, Remote Sensing of Environment, vol. 62, n°1, p. 77‑89.

SUN, Z., LORSCHEID, I., MILLINGTON, J. D., LAUF, S., MAGLIOCCA, N. R., GROENEVELD, J., … BUCHMANN, C. M., (2016), Simple or complicated agent-based models? A complicated issue, Environmental Modelling & Software, vol. 86, p. 56–67.

T

TANNIER, C., VUIDEL, G., FRANKHAUSER, P. & HOUOT, H., (2010), Simulation fractale d'urbanisation. MUP-city, un modèle multi-échelle pour localiser de nouvelles implantations résidentielles, Revue Internationale de Géomatique, vol. 20, n°3, p. 303‑329.

TERRIER, C., (2011), La valeur des données géographiques, L'Espace géographique, vol. 40, n°2, p. 103‑108.

TESEMA, D. & DINSA SORI, N., (2017), Evaluating land-use plan using conformance-based approach in Adama city, Ethiopia, Spatial Information Research, vol. 25, .

THERNEAU, T. M., ATKINSON, E. J. & FOUNDATION, M., (1997), An Introduction to Recursive Partitioning Using the RPART Routines.

TOBLER, W. R., (1970), A Computer Movie Simulating Urban Growth in the Detroit Region, Economic Geography, vol. 46, n°sup1, p. 234‑240.

TOMLINSON, R. F., (1962), An introduciton to the use of electronic computers in the storage, compilation and assessment of natural and economic data for the evaluation of marginal lands, Canada Department of Agriculture.

TORRENS, P. M., (2001), Can geocomputation save urban simulation ? Throw some agents into the mixture, simmer, and wait..., CASA.

TORRENS, P. M., (2009), Cellular automata, In : International Encyclopedia of Human Geography, Elsevier, London, p. 1‑4.

TORRENS, P. M., (2011), Calibrating and Validating Cellular Automata Models of Urbanization, In : X. YANG (éd.), Urban Remote Sensing, John Wiley & Sons, Ltd, Chichester, UK, p. 335‑345.

TRIANTAKONSTANTIS, D. & MOUNTRAKIS, G., (2012), Urban Growth Prediction: A Review of Computational Models and Human Perceptions, Journal of Geographic Information System, vol. Vol.04No.06, p. 33.

TUFFÉRY, S., (2017), Data mining et statistique décisionnelle : l'intelligence des données, Éditions Technip.

TURING, A., (1936), On Computable Numbers, with an Application to the Entscheidungsproblem, Proceedings of the London Mathematical Society, vol. 42, n°2, p. 230‑265.

TURNER, B. L. I., MOSS, R. H. & SKOLE, D. L. (EDS), (1993a), Relating land use and global land-cover change: a proposal for an IGBP-HDP core project. A report from the IGBP/HDP Working Group on Land-Use/Land-Cover Change, Global Change Report (Sweden).

TURNER, B. L., II (ÉD.), (1990), The earth as transformed by human action: global and regional changes in the biosphere over the past 300 years, Cambridge University Press, Cambridge, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, xvi+713 p.

TURNER, B. L., II, (1994), Local faces, global flows: The role of land use and land cover in global environmental change, Land Degradation & Development, vol. 5, n°2, p. 71‑78.

TURNER, B. L., II, LAMBIN, E. F. & REENBERG, A., (2007), The emergence of land change science for global environmental change and sustainability, Proceedings of the National Academy of Sciences of the United States of America, vol. 104, n°52, p. 20666‑20671.

TURNER, B. L., II, MEYER, W. B. & KATES, R. W., (1994), The Earth as Transformed by Human Action in Retrospect, Annals of the Association of American Geographers, vol. 84, n°4, p. 711‑715.

TURNER, B. L., MOSS, R. H. & SKOLE, D., (1993b), Relating land use and global land-cover change, In : IGBP Report 24 ; HDP Report 5. IGBP, International Geosphere-Biosphere Programme, Stockholm.

U

UNEP (ÉD.), (2014), Assessing global land use: balancing consumption with sustainable supply, United Nations Environment Programme, Nairobi, Kenya, 131 p.

V

VAN DELDEN, H., STUCZYNSKI, T., CIAIAN, P., PARACCHINI, M. L., HURKENS, J., LOPATKA, A., … VANHOUT, R., (2010), Integrated assessment of agricultural policies with dynamic land use change modelling, Ecological Modelling, vol. 221, n°18, p. 2153‑2166.

VAN KARNENBEEK, L. & JANSSEN-JANSEN, L., (2018), Playing by the rules? Analysing incremental urban developments, Land Use Policy, vol. 72, p. 402‑409.

VAN VLIET, J., BREGT, A. K., BROWN, D. G., VAN DELDEN, H., HECKBERT, S. & VERBURG, P. H., (2016), A review of current calibration and validation practices in land-change modeling, Environmental Modelling & Software, vol. 82, p. 174‑182.

VAN VLIET, J., HAGEN-ZANKER, A., HURKENS, J. & VAN DELDEN, H., (2013a), A fuzzy set approach to assess the predictive accuracy of land use simulations, Ecological Modelling, vol. 261‑262, p. 32‑42.

VAN VLIET, J., NAUS, N., VAN LAMMEREN, R. J. A., BREGT, A. K., HURKENS, J. & VAN DELDEN, H., (2013b), Measuring the neighbourhood effect to calibrate land use models, Computers, Environment and Urban Systems, vol. 41, p. 55‑64.

VARENNE, F. & PUMAIN, D., (2017), Théories et modèles en sciences humaines : le cas de la géographie, Éditions Matériologiques, Paris, France, 643 p.

VELASCO-GRACIET, H., (2008), Des frontières et des géographes, Géoconfluences.

VELDKAMP, A. & VERBURG, P. H., (2004), Modelling land use change and environmental impact, Journal of Environmental Management, vol. 72, n°1, p. 1‑3.

VERBURG, P. H., DEARING, J. A., DYKE, J. G., LEEUW, S. VAN DER, SEITZINGER, S., STEFFEN, W. & SYVITSKI, J., (2016), Methods and approaches to modelling the Anthropocene, Global Environmental Change, vol. 39, p. 328‑340.

VERBURG, P. H., ERB, K.-H., MERTZ, O. & ESPINDOLA, G., (2013), Land System Science: between global challenges and local realities, Current Opinion in Environmental Sustainability, vol. 5, n°5, p. 433‑437.

VISSER, H. & DE NIJS, T., (2006), The Map Comparison Kit, Environ. Model. Softw., vol. 21, n°3, p. 346–358.

VON NEUMANN, J. & BURKS, A. W., (1966), Theory of Self-reproducing Automata, University of Illinois Press.

VON THÜNEN, J. H., (1826), Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie, Hamburg.

W

WANG, L.-G., HAN, H. & LAI, S.-K., (2014), Do plans contain urban sprawl? A comparison of Beijing and Taipei, Habitat International, vol. 42, p. 121‑130.

WANG, X., SONG, L., SUN, L. & GAO, H., (2019), Nonparametric estimation of the ROC curve based on the Bernstein polynomial, Journal of Statistical Planning and Inference.

WEBER, A. & FRIEDRICH, C. J., (1929), Theory of the location of industries., University of Chicago Press, Chicago, Ill.

WHITE, R. & ENGELEN, G., (1993), Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land-Use Patterns, Environment and Planning A: Economy and Space, vol. 25, n°8, p. 1175‑1199.

WHITE, R., ENGELEN, G. & ULJEE, I., (1997), The Use of Constrained Cellular Automata for High-Resolution Modelling of Urban Land-Use Dynamics, Environment and Planning B: Planning and Design, vol. 24, n°3, p. 323‑343.

WHITE, R., ENGELEN, G. & ULJEE, I., (2015), Modeling cities and regions as complex systems: From theory to planning applications, 354 p.

WHITE, R., ULJEE, I. & ENGELEN, G., (2012), Integrated modelling of population, employment and land-use change with a multiple activity-based variable grid cellular automaton, International Journal of Geographical Information Science, vol. 26, n°7, p. 1251–1280.

WOLFRAM, S., (1982), Cellular Automata as Simple Self-Organizing Systems, p. 12.

WOLFRAM, S., (1984), Cellular automata as models of complexity, Nature, vol. 311, n°5985, p. 419‑424.

WU, F. & WEBSTER, C. J., (1998), Simulation of Land Development through the Integration of Cellular Automata and Multicriteria Evaluation, Environment and Planning B: Planning and Design, vol. 25, n°1, p. 103‑126.

Y

YEH, A. G.-O. & LI, X., (2003), Error Propagation and Model Uncertainties of Cellular Automata in Urban Simulation with GIS, In : GeoDynamics, CRC Press, UK, p. 329‑345.

Z

ZAJEC, O. & ANTONI, J.-P., (2017), Frontières. Des confins d'autrefois aux murs d'aujourd'hui, 184 p.

ZHOU, X.-H., MCCLISH, D. K. & OBUCHOWSKI, N. A., (2009), Statistical Methods in Diagnostic Medicine, John Wiley & Sons.

ZILLONIZ, S., (2015), Les temps de déplacement entre domicile et travail, Dares - ministère du travail de l'emploi et de la solidarité sociale.

ZIPF, G. K., (1929), Relative Frequency as a Determinant of Phonetic Change, Harvard Studies in Classical Philology, vol. 40, p. 1–95.

ZIPF, G. K., (1949), Zipf, George K. Human behavior and the principle of least effort. Cambridge, (Mass.): Addison-Wesley, 1949, pp. 573, Journal of Clinical Psychology, vol. 6, n°3, p. 306–306.

Table des matières

REMERCIEMENTS ..... **III**

AVANT – PROPOS ..... **V**

INTRODUCTION GÉNÉRALE **1**

ÉTUDIER LE DÉVELOPPEMENT URBAIN EN ZONE FRONTALIÈRE **7**

CHAPITRE 1 DÉVELOPPEMENT URBAIN : UN CHANGEMENT D'OCCUPATION DU SOL 11
1. Changements d'occupation du sol *12*
2. Développement urbain *22*
3. Frontière *31*
CHAPITRE 2 ZONE FRONTALIÈRE : DES DONNÉES À MODÉLISER 45
1. Modéliser un espace frontalier *46*
2. Collecter des données frontalière *57*
CHAPITRE 3 DONNÉES LIBRES : LA ZONE D'ÉTUDE 75
1. Évolution historique *76*
2. Réalités géographiques *94*
3. Déterminants fonctionnels *104*

MODÉLISER LE DÉVELOPPEMENT URBAIN EN ZONE FRONTALIÈRE **117**

CHAPITRE 4 UNE BASE : LES AUTOMATES CELLULAIRES 121
1. Un outil spécifique *122*
2. Un objectif général : la modélisation *136*
3. Une implémentation adaptée : LucSim *144*
CHAPITRE 5 UN COUPLAGE : LES ARBRES DE DÉCISION 155
1. Principe du Machine Learning *156*
2. Principe des arbres des décision *167*
3. Principe du couplage AC - AD *174*
CHAPITRE 6 UNE DIFFICULTÉ : LE CALIBRAGE 183
1. Validation et calibrage *185*
2. Test des paramètres *191*
3. Sensibilité à la partition du jeu de données *206*

INTERPRÉTER LE DÉVELOPPEMENT URBAIN EN ZONE FRONTALIÈRE **215**

CHAPITRE 7 L'OCCUPATION DU SOL : ANALYSE À TRAVERS LE JEU DE RÈGLES 219
1. L'influence des variables *220*
2. L'influence du voisinage *226*
3. Analyse du jeu de règles : accessibilité *250*
CHAPITRE 8 LE DÉVELOPPEMENT URBAIN : ANALYSE À TRAVERS LA FRONTIÈRE 261
1. La part frontalière : Exploitation du jeu de règles *262*
2. Influence culturelle *264*
3. Influence des pôles urbains *281*
CHAPITRE 9 LES ENJEUX DE LA MODÉLISATION : ANALYSE À TRAVERS LA PROSPECTIVE 295
1. Résultats et limites *296*
2. Perspective et modélisation *302*

3. Adaptation à la prospective et à l'aide à la décision 305

CONCLUSION GÉNÉRALE 321

TABLE DES FIGURES 325

LISTE DES TABLES 331

RÉFÉRENCES 335

TABLE DES MATIÈRES 355

ECOLE DOCTORALE
SEPT

Titre : Apport de l'apprentissage automatique pour la modélisation et l'analyse des changements d'occupation du sol. Application au développement urbain de la zone frontalière franco-allemande

Mots clés : Occupation du sol - automate cellulaire - arbres de décision - modélisation - zone frontalière - Strasbourg Kehl

Le développement urbain apparait distinct selon sa localisation géographique, le contexte socio-économique, politique ou culturel dans lequel il s'ancre. Néanmoins, l'action d'urbaniser s'appuie sur un seul et même principe : construire des bâtiments pour l'habitation et les activités humaines. Par conséquent, l'objectif principal de ce travail de recherche est de déterminer si une part universelle du processus de développement urbain peut se distinguer d'une part spécifique, liée à une localisation géographique. Cette recherche s'intéresse plus précisément aux changements d'occupation du sol sur la zone transfrontalière Franco-Allemande. En effet, le contexte transfrontalier renforce les différences au sein de la zone d'étude. Cependant, l'importance des frontières internes européennes en tant que séparation, a diminué parallèlement au renforcement des accords et de la coopération entre les pays membres de l'Union Européenne. Un questionnement sur le caractère éventuellement homogène de cette zone à travers l'étude des processus de développement urbain à l'œuvre, est donc d'autant plus pertinent. Pour une meilleure compréhension du processus de développement urbain, une méthode spécifique basée sur l'utilisation de l'apprentissage automatique par arbres de décision (AD) a été élaborée. Elle vise à générer et calibrer les règles de transition nécessaires à la simulation par automate cellulaire (AC). L'apprentissage s'appuie sur les caractéristiques d'occupation du sol dans le voisinage des cellules urbanisées entre 1990 et 2006, ainsi que sur les caractéristiques intrinsèques de ces cellules (accessibilité aux pôles urbains majeurs, localisation dans l'un ou l'autre des pays frontaliers). Les résultats obtenus par l'analyse du jeu de règles mettent en évidence une influence de la spécificité nationale et de l'occupation du sol initiale dans certaines règles. Cependant, il apparait qu'une part du processus s'affranchit de l'une comme de l'autre des contraintes, tendant ainsi à une plus grande universalité. Une analyse plus détaillée de ces règles permet ensuite de dégager des critères de voisinage et d'accessibilité influençant positivement de l'urbanisation de chaque cellule, et de comparer leur importance suivant la nature des règles analysées (françaises, allemandes, transfrontalières). En résumé, l'apport de cette thèse se mesure à trois niveaux : (1) d'un point de vue théorique en proposant d'identifier un degré d'universalité dans les processus de développement urbain ; (2) d'un point de vue thématique en caractérisant l'influence de la localisation du développement urbain dans une zone frontalière ; et (3) d'un point de vue méthodologique en développant un modèle sur la base d'un couplage AD – AC qui a permis de générer et calibré automatiquement le modèles utilisé pour simuler le développement urbain dans ce travail de recherche.

Title: Machine Learning for Land Use Change Analysis and Modelling. An exploration of urban development on the French-German border area.

Keywords: Land use - cellular automata - decision tree - modeling - border area - Strasbourg Kehl

Urban development can take different forms or features, depending on its geographical location and its socioeconomic, political and cultural context. Nevertheless, the overall action relies on one fundamental principle: building construction in order to give people housing. Therefore, the main objective of this research is to determine whether an underlying universal aspect of the urban development process can be distinguished from a specific one, being the reflect of local specificities. Specifically, this research analyzes the land use change on the French-German cross-border area. Indeed, the border context enhances the difference within this territory. Nonetheless, the internal European border importance as a separation is getting lower while European Union agreement and cooperation are getting stronger. Consequently, we tend to question a potential homogeneity of such area according the urban development analysis. To gain a better understanding of the urban development in the French-German cross-border area, a specific method using a decision tree (DT) process to generate and calibrate urban transition rules for cellular automata (CA) has been developed. In order to define CA rules, the learning algorithm is provided with data of land use, from 1990 and 2006, accessibility to the main urban area and country location. The findings demonstrate that the rule set identified rules, which are constrained at different levels: from the initial land use state needed for urbanization to the geographical location in a specific country. The analysis therefore shows rules specific to France or Germany, as well as rules free from any state location constraints, which characterize a cross-border urbanization process. The latter is more influenced by the location toward main urban areas. Proper neighborhood and internal characteristic of urbanized cells depending of the nature of the rule (French, German or Cross-border) are explicitly defined. In conclusion, this research contribution can be summarized according to the following three orientations: (1) from a theoretical point of view, which propose to identify urban processes degree of universality, (2) from a thematic outlook, describing the influence of the urban development process location on the border area, (3) from a methodological point of view, pairing DT and AC to automatically design and calibrate the urban development model used in this research work.

Université Bourgogne Franche-Comté
32, avenue de l'Observatoire
25000 Besançon

UBFC
UNIVERSITÉ
BOURGOGNE FRANCHE-COMTE