



Détection automatique des parcelles sur les plans napoléoniens : comparaison de deux méthodes

Jean-Michel Follin, Elisabeth Simonetto, Anthony Chalais

► To cite this version:

Jean-Michel Follin, Elisabeth Simonetto, Anthony Chalais. Détection automatique des parcelles sur les plans napoléoniens : comparaison de deux méthodes. *Humanités numériques*, 2021, 3, 10.4000/revuehn.1779 . hal-03555293

HAL Id: hal-03555293

<https://hal.science/hal-03555293>

Submitted on 18 Feb 2022

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.



Distributed under a Creative Commons Attribution 4.0 International License

Détection automatique des parcelles sur les plans napoléoniens : comparaison de deux méthodes

Automatic Detection of Parcels on Napoleonic Maps: Comparison of Two Methods

Jean-Michel Follin, Élisabeth Simonetto et Anthony Chalais



Édition électronique

URL : <https://journals.openedition.org/revuehn/1779>

DOI : 10.4000/revuehn.1779

ISSN : 2736-2337

Éditeur

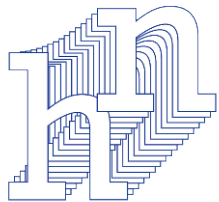
Humanistica

Référence électronique

Jean-Michel Follin, Élisabeth Simonetto et Anthony Chalais, « Détection automatique des parcelles sur les plans napoléoniens : comparaison de deux méthodes », *Humanités numériques* [En ligne], 3 | 2021, mis en ligne le 01 mai 2021, consulté le 16 juillet 2021. URL : <http://journals.openedition.org/revuehn/1779> ; DOI : <https://doi.org/10.4000/revuehn.1779>



Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la Licence Creative Commons Attribution 4.0 International.



Détection automatique des parcelles sur les plans napoléoniens : comparaison de deux méthodes

Automatic Detection of Parcels on Napoleonic Maps: Comparison of Two Methods

Jean-Michel Follin, Élisabeth Simonetto et Anthony Chalais

Résumés

Le cadastre napoléonien fournit la description la plus détaillée du territoire français dans sa globalité au XIX^e siècle. Il contient une mine d'informations (structure des parcelles cadastrales, numéro des parcelles, tracé des rivières et des routes, toponymie, bâti...) auxquelles les chercheurs en sciences humaines (archéologues, historiens, urbanistes...) s'intéressent de plus en plus, notamment pour mener des analyses historiques du territoire. Les travaux présentés ici s'inscrivent dans cette perspective et proposent une chaîne de traitement semi-automatique permettant de vectoriser, géoréférencer et assembler des planches scannées du cadastre ancien afin de construire une base de données multi-époques. Dans cet article, nous abordons plus particulièrement les solutions adoptées pour l'étape de vectorisation automatique des parcelles en comparant deux méthodes, l'une basée sur la transformée de Hough probabiliste (THP) et l'autre sur l'algorithme Line Segment Detector (LSD). Après avoir précisé comment ces méthodes ont été implémentées pour notre étude, nous présentons les résultats obtenus, qui montrent la supériorité de la méthode LSD sur la méthode THP pour les planches cadastrales les plus anciennes.

The most detailed geographic description of France's whole 19th-century territory is the Napoleonic cadastre. It contains a wealth of information (e.g. cadastre lot structure, lot numbering, rivers and road shapes, toponymy, buildings...) which underpin most historical studies

of land usage and transformation. We therefore decided to develop a semi-automatic toolchain able to vectorise, georeference and combine scanned old cadastre sheets in order to build a multi-epoch database describing land property. In this article, we focus on the automatic vectorisation of the parcels' shapes and assess the respective merits of two methods we tested: the first based on the Probabilistic Hough Transform (PHT) and the second on the Line Segment Detector (LSD) algorithm. After explaining how we implemented these two methods, we present the results, which show that the LSD approach performs better on the oldest cadastre sheets.

Entrées d'index

MOTS-CLÉS : humanités numériques spatialisées, imagerie, numérisation

KEYWORDS: spatial digital humanities, digitisation, image processing

Nous tenons à remercier les anciens étudiants de l'École supérieure des géomètres et topographes, stagiaires du laboratoire Géomatique et foncier, Maïté Fahrasmane, Charlotte Odie et Jean-Marc Beveraggi, sans qui ces travaux n'auraient pu être réalisés. Nous remercions également nos collègues Marie Fournier et Mathieu Bonnefond ainsi que Gilles Berteau de la direction générale des finances publiques de la Sarthe.

Introduction

¹ À l'heure actuelle, de plus en plus de documents anciens, en particulier géographiques (cartes, plans, clichés aériens), sont numérisés et mis à disposition sur Internet. Ils constituent une source d'information précieuse sur le territoire, notamment pour les chercheurs en sciences humaines. Ainsi, dans de nombreux départements français, les services des archives diffusent en ligne leurs fonds documentaires numérisés. On y retrouve entre autres les plans cadastraux anciens qui peuvent parfois être géoréférencés comme sur les plateformes cartographiques des régions Pays de la Loire (Géopal¹) et Provence-Alpes-Côte-d'Azur (CRIGE-PACA²). Selon la période à laquelle ils ont été dressés, ces plans contiennent des informations très variées, qui sont plus ou moins détaillées ou précises (Clergeot 2009).

² Notre projet vise à créer une base de données spatiotemporelles modélisant l'évolution des parcelles au cours du temps (depuis les années 1800) en proposant une chaîne méthodologique complète, reproductible et semi-automatique consistant à traiter des feuilles cadastrales scannées par vectorisation, géoréférencement et mosaïquage. Cette base sera utilisée pour mener des analyses telles que la quantification et la qualification des changements, la mise en relation avec d'autres facteurs, ou l'identification de motifs récurrents.

³ Dans cet article, nous nous intéressons à la vectorisation automatique des parcelles à partir de plans du cadastre napoléonien. De tels plans correspondent à la description la plus détaillée du territoire français dans sa globalité sur près d'un siècle avec des échelles variant de 1/2 500 à 1/500 et leur étude permet de reconstruire avec précision l'occupation du sol durant cette période (Dupouey *et al.* 2007). Le traitement de données plus récentes est aussi considéré pour mettre en exergue les difficultés et les limites de la vectorisation automatique des planches plus anciennes.

⁴ Notre application porte sur une commune rurale du sud de la Sarthe, Aubigné-Racan, pour laquelle nous disposons des planches cadastrales scannées de deux époques napoléoniennes (1813 et 1850) et d'une époque récente (1972-1974).

⁵ Dans une première partie, nous allons brosser un état de l'art succinct des projets de recherche proches de notre thème, avant de faire un bref historique du cadastre (en particulier au XIX^e siècle), puis de présenter notre démarche. Nous décrirons ensuite une étape clé de notre chaîne de traitement, celle de la vectorisation des parcelles en évaluant deux méthodes : l'une basée sur la transformée de Hough probabiliste (THP) et l'autre sur l'algorithme Line Segment Detector (LSD). Nous concluons par des pistes d'amélioration.

Les données cartographiques anciennes, un contexte porteur

De nombreux projets

⁶ Les études menées sur des données cartographiques anciennes se sont considérablement développées ces dernières années en France et dans le monde. Elles montrent l'intérêt récent pour les bases de données géohistoriques et leur utilisation à des fins analytiques ou prospectives.

⁷ Le projet de recherche ANR *GéoPeuple*³, porté par l'École des hautes études en sciences sociales (EHESS) et l'Institut national de l'information géographique et forestière (IGN), s'est notamment proposé de construire une base de données géohistorique spatiotemporelle en procédant au géoréférencement et à la vectorisation automatique de cartes topographiques anciennes (Cassini et État-Major, qui présentent des échelles de l'ordre de 1/80 000), afin d'étudier la densification des territoires (Costes, Grosso et Plumejeaud 2012). Autre fruit de la collaboration entre l'IGN et l'EHESS, le projet *GeoHistoricalData*⁴ a abouti à la création d'un outil de saisie collaborative de données géohistoriques. Ce dernier a permis de vectoriser l'intégralité du réseau routier et des villes françaises à partir de la carte de Cassini (Perret, Gribaudi et Barthelemy 2015).

⁸ On peut également citer le projet ANR *ALPAGE*⁵ (Analyse diachronique de l'espace urbain parisien : approche géomatique), porté par les laboratoires LAMOP (université Paris 1), ArScAn (université Paris-Nanterre), OTELO et L3i (université de La Rochelle), qui a réuni des historiens, géomaticiens et informaticiens pour construire un système d'information géographique (SIG) historique sur l'espace parisien préin-

dustriel. Il a notamment intégré à la base de données le plus ancien plan parcellaire de Paris en géoréférençant et en vectorisant le cadastre par îlots de Vasserot (1810-1836). L'échelle du cadastre par îlots, 1/200, a permis aux historiens du projet d'effectuer des analyses fines de la démographie urbaine de cette ville (Noizet et Grosso 2012).

9 Le projet ANR *MODE RESPYR*⁶ (Modélisation rétrospective et prospective des changements d'occupation des sols dans les Pyrénées), porté par quatre laboratoires toulousains (GEODE, CESBIO, DYNAFOR et LPT), avait quant à lui pour but, entre autres, de reconstituer l'histoire de l'occupation du sol dans la région des Pyrénées. La « Carte de la France » du XIX^e siècle a été pour cela utilisée. Son géoréférencement a été tout particulièrement étudié dans ce projet en déterminant la meilleure transformation géométrique par un protocole de validation croisée : il s'agit de la régression par noyau gaussien, définie comme méthode globale localement sensible (Herrault *et al.* 2013). Nous avons partiellement repris ces travaux dans l'étape de géoréférencement de notre chaîne de traitement.

10 Enfin, le projet ANR *ModelSpace* et son successeur *Architerre*⁷, portés par les laboratoires CITERES-Lat (université de Tours), FRAMESPA-Terrae et l'IMT (université Toulouse-Jean-Jaurès), se sont donné pour mission de retracer l'historique du parcellaire de certains départements au moyen de registres terriers et de compoix. Les terriers et compoix se présentent sous la forme de registres accompagnés parfois de plans permettant d'illustrer le parcellaire. Ils consignent le nom et la profession du propriétaire, parfois le nom de l'exploitant, mais surtout et très souvent la superficie, une localisation par un microtoponyme et les noms des propriétaires des parcelles adjacentes (Le Couédic *et al.* 2012). En recoupant toutes les données fournies par les registres et en s'aidant des plans terriers lorsqu'ils sont disponibles ou à défaut du cadastre napoléonien, ces projets proposent une solution pour modéliser et cartographier le parcellaire ancien sous la forme de graphes d'adjacence.

11 L'existence de revues spécialisées sur les documents cartographiques anciens comme le journal en ligne *e-Perimetron*, lié à la Commission sur le patrimoine cartographique dans le numérique de l'Association cartographique internationale (International Cartographic Association [ICA]), témoigne du dynamisme dans de nombreux pays de ce type de travaux notamment sur les problématiques de géoréférencement (Boutoura et Livieratos 2006) et de vectorisation automatique de cartes topographiques historiques (Baily 2007 ; Iosifescu, Tsorlini et Hurni 2016).

12 Comme nos travaux, ces différents projets s'inscrivent dans le domaine en plein essor du patrimoine culturel cartographique avec l'objectif de sauvegarder et si possible valoriser l'héritage patrimonial géohistorique. Cependant, ils ne s'intéressent pas nécessairement aux mêmes objets, échelles ou espaces d'étude et ne partagent pas forcément notre objectif de semi-automatisation du processus complet. La maîtrise de la chaîne complète nous permettra ainsi de conserver le lien entre la donnée d'origine, dans l'espace image, et la donnée objet en sortie, dans l'espace géoréférencé : il sera possible de caractériser (par des informations de qualité liées aux traitements) et d'enrichir (à partir d'autres informations issues de l'image) la base de données en fonction des fonctionnalités intégrées à notre outil ou des sources utilisées.

Un bref historique du cadastre napoléonien et récent

- ¹³ Le cadastre a été créé au lendemain de la Révolution française dans le but de mieux répartir l'impôt voté par l'Assemblée constituante en 1791 entre les « unités administratives » du territoire, c'est-à-dire les communes. L'objectif affiché était de couvrir les dépenses de l'État grâce à « une taxe foncière répartie par égalité proportionnelle sur toutes les propriétés bâties et non bâties » (Clergeot 2009).
- ¹⁴ Pour répondre au problème de « la méconnaissance des ressources économiques de chaque unité administrative », les communes ont été autorisées dès août 1791 à « lever à leurs frais » un cadastre. L'administration a encadré cette nouvelle prérogative des communes par le décret du 16-23 septembre 1791 qui définit les méthodes à appliquer pour dresser les plans cadastraux. Mais c'est à partir de la loi du 15 septembre 1807 suivie du règlement impérial du 27 janvier 1808 que fut décidée la mise en œuvre d'un cadastre parcellaire général entièrement financé par l'État. Les travaux qui se déroulèrent de 1808 jusqu'à la fin des années 1840 donnèrent lieu au cadastre connu sous le nom de cadastre napoléonien ou d'ancien cadastre qui constitue la base du cadastre français contemporain.
- ¹⁵ La solution retenue pour la réalisation du cadastre parcellaire fut celle de Hautier, géomètre en chef et ancien professeur de géométrie pratique du cadastre. Elle consistait « à faire d'abord la triangulation calculée de la commune, à lever le périmètre de cette commune, celui des sections, à lever en outre les chemins et ruisseaux, à calculer les contenances des masses ou polygones formés par ces divisions et subdivisions, et à mesurer ensuite les parcelles de propriétés contenues dans ces grandes masses. » Ce mode opératoire avait vocation à « tirer parti des divers degrés de talents des arpenteurs » en chargeant les « plus instruits de lever le plan routier ou périmétral, et les simples arpenteurs, de lever, sous la direction des premiers, le détail des propriétés » (ministère des Finances 1808).
- ¹⁶ Le *Recueil méthodique des lois, décret, règlements, instructions et décisions sur le cadastre de la France* publié en 1811, met au point un code cadastral de 1 144 articles et définit notamment dans le « Titre IV – Arpentage » les procédures réglementaires à suivre, du levé sur le terrain jusqu'au dressage du plan. On y apprend ainsi qu'un plan parcellaire peut être construit à l'échelle 1/5 000, 1/2 500 ou 1/1 250 en fonction du morcellement (appréhendé par le nombre de parcelles par arpents métriques). Ce plan parcellaire est divisé en feuilles, comprenant une section, complète ou partielle, voire deux sections selon l'étendue de celles-ci (articles 213, 214 et 216). Concrètement une section est définie dans la documentation cadastrale comme « une partie du territoire communal déterminée dans le but de faciliter l'établissement, la consultation et la tenue des documents cadastraux⁸ » (ministère des Finances 1811). Son périmètre doit être constitué autant que faire se peut par des limites lui conférant « un caractère suffisant de stabilité (ex : routes, chemins, cours d'eau, fossés, etc.) ». L'ensemble du territoire communal est représenté sur un document au format grand aigle (75 cm × 106 cm), appelé tableau d'assemblage. Établi généralement à l'échelle 1/10 000, ce

dernier indique, entre autres informations, le découpage de la commune en sections cadastrales (identifiées par une lettre simple : A, B, C, etc.) et en feuilles.

17 Les documents graphiques que sont le plan parcellaire et le tableau d'assemblage s'accompagnaient d'une documentation littéraire composée d'un état de section (tableau donnant les différentes propriétés représentées sur une section) et d'une matrice cadastrale (registre récapitulatif renseignant pour chaque propriétaire l'ensemble de ses biens et leurs valeurs⁹).

18 L'article 261 du *Recueil méthodique* de 1811 impose aux plans dressés le respect de tolérances afin d'assurer la qualité et donc la fiabilité des informations portées au plan cadastral. Dans le cadre de « l'arpentage des propriétés rurales », l'erreur de distance ne doit ainsi pas excéder 1 % (« un centième ») pour les éléments de détail et 0,5 % (« un deux centièmes ») pour les grandes dimensions. Par exemple, pour le côté d'un polygone, l'erreur ne devait pas être supérieure à 0,5 % de la valeur totale de sa longueur.

19 Sur la qualité de la confection du cadastre à cette époque, on peut dire que si les conditions de réalisation restèrent globalement bonnes grâce à l'organisation napoléonienne du cadastre, les sources d'erreurs étaient tout de même nombreuses (manque de formation de géomètres, erreurs d'étalonnage, reprise de documents anciens, erreurs de reports lors des copies, etc.) [Clergeot 2009].

20 À partir de la loi du 31 juillet 1821, les départements et communes ont été tenus de financer les opérations cadastrales en lieu et place de l'État.

21 Une amélioration de la qualité géométrique du cadastre a été rendue possible par le règlement du 25 mars 1827 : c'est désormais un spécialiste, le « triangulateur », et non plus un géomètre de première classe moins qualifié, qui prend en charge la réalisation de la triangulation (Clergeot 2009). Vers 1850, l'arrivée d'un nouvel instrument, le tachéomètre, permet aussi d'effectuer des mesures plus précises et plus rapides.

22 Initialement, le plan cadastral n'était pas destiné à être renouvelé. En effet, l'allivrement, revenu net imposable fixé pour un bien foncier, étant considéré immuable, il était prévu de mettre à jour uniquement les matrices cadastrales lors d'opérations foncières. En réponse au besoin croissant à la fois de « conserver » les plans et de renouveler le cadastre napoléonien, de nombreux projets furent élaborés dans les années 1830 à 1850. La loi du 7 août 1850 autorise notamment les communes cadastrées depuis au moins 30 ans à renouveler, à leurs charges, leur cadastre. La procédure de renouvellement consiste à réaliser un levé parcellaire entièrement nouveau : les plans ont une codification différente de ceux antérieurs à 1840, la numérotation des parcelles a été refondue et différentes catégories d'éléments bâtis sont représentées. C'est le cas des bâtiments légers qui sont représentés par un rectangle avec les deux diagonales.

23 Parmi les grandes étapes dans l'histoire du cadastre récent après 1850, nous pouvons évoquer :

- la loi du 16 avril 1930 qui pose les principes de la conservation, aux frais de l'État, de tous les cadastres renouvelés et de la réfection, également aux frais de l'État, du cadastre des communes pour les cas exceptionnels

- la loi du 17 décembre 1941 qui est à l'origine de la création d'un service unique du cadastre et de la rénovation par voie de réfection consistant à effectuer un nouveau levé parcellaire appuyé sur une triangulation cadastrale, accompagné d'une délimitation des propriétés publiques et privées sous l'égide d'une commission communale de délimitation créée à cet effet
- les décrets des 4 janvier et 30 avril 1955 sur la réforme de la publicité foncière et la réglementation de la rénovation et de la conservation du cadastre
- la loi du 18 juillet 1974 qui autorise une nouvelle « rénovation de la rénovation » appelée remaniement pour répondre notamment aux besoins de rénover à nouveau les plans de territoires, souvent situés en périphérie des villes, qui ont connu une forte urbanisation

24

Cet historique, marqué par une succession de mises à jour, explique notamment les changements qui ont parfois affecté les styles d'écriture ou de figuré dans les plans du cadastre. Notre chaîne de traitement devra prendre en compte ces évolutions pour assurer la qualité de nos résultats.

Présentation de notre chaîne de traitement semi-automatique

Cadre conceptuel de notre démarche

25

Pour produire des données vectorielles géoréférencées multilatées à partir des plans cadastraux napoléoniens et récents scannés, nous nous appuyons sur les informations disponibles sur ces derniers. La feuille cadastrale contient un certain nombre de renseignements sous la forme d'écritures qui servent à identifier notamment son auteur et son année de création pour le cadastre napoléonien. Quelle que soit sa période de confection, elle décrit des espaces cadastrés (les parcelles sur lesquelles sont éventuellement situés des bâtiments) et non cadastrés (la voirie, les cours d'eau). Ces derniers permettent de délimiter les îlots (définis comme des regroupements de parcelles contiguës). Notre objectif est ici d'obtenir les contours des parcelles sous forme de polygones qui sont recomposés à partir des segments que nous extrayons des images. Par ailleurs un autre type d'information contenue dans les feuilles cadastrales correspond à des écritures et des symboles portant sur les espaces cadastrés et non cadastrés (par exemple, les signes conventionnels définissant la nature et la mitoyenneté des limites parcellaires ou les numéros de parcelles). Ces informations peuvent être exploitées pour enrichir notre base de données cadastrales. Leur détection n'est pas notre objectif dans les travaux présentés ici. Néanmoins, nous envisageons d'appliquer une méthode de type apprentissage profond pour cela.

26

Comme nous pouvons le voir sur l'extrait du cadastre de 1813 (figure 1), les contours de parcelles, que nous cherchons à détecter, peuvent être confondus avec des traits de construction tandis que les segments des caractères (noms de lieux-dits, numéros de parcelles...) peuvent être identifiés à tort comme portions de limites de parcelles. De plus, ainsi

que nous le rappellerons dans la partie « Zone d'étude et données », nous travaillons sur des documents potentiellement très dégradés par le temps, ce qui complique notre tâche.

FIGURE 1. EXTRAIT DU CADASTRE DE 1813 DE LA COMMUNE D'AUBIGNÉ-RACAN (FEUILLE 09)



Archives départementales de la Sarthe

Structure de la chaîne de traitement

27 Notre chaîne de traitement est un programme développé en Python et appuyé sur des bibliothèques *open source*. Il se décompose en trois parties (voir figure 2) : la vectorisation, le géoréférencement et le mosaïquage.

28 L'enchaînement choisi diffère de celui utilisé généralement dans les projets similaires, où le traitement commence par le géoréférencement. En effet, les parcelles sont définies par des limites rectilignes, en particulier sur les plus anciennes planches car les lignes étaient tracées à la règle. Commencer par le géoréférencement des images conduirait à l'obtention de lignes courbes. En débutant par la vectorisation, nous produisons des polygones avec des arêtes droites, propriété géométrique qui ne sera pas altérée lors de l'étape suivante de géoréférencement pendant laquelle seule la transformation des sommets sera effectuée.

29 Pour la première étape, la vectorisation, deux méthodes sont présentées et évaluées ici : l'une, classique, basée sur la transformée de Hough dite probabiliste (THP) et l'autre, plus novatrice, appuyée sur la méthode Line Segment Detector (LSD). Notons que dans Follin et Simonetto (2018), une autre approche utilisant des fonctions de GRASS (squelettisation du raster prétraité puis vectorisation de l'image du squelette par une méthode de « suivi de chemin ») a aussi été mise en œuvre mais n'est pas présentée ici (voir partie « Approches adoptées pour l'étape de vectorisation »).

30 La deuxième étape, le géoréférencement, est quant à elle semi-automatique. Elle a pour objectif de rendre superposable, avec une qualité optimale, le parcellaire vectorisé ancien au parcellaire numérique ré-

cent. Elle est précédée par une procédure manuelle de sélection des points de liaison qui servent au calcul d'un modèle géométrique de transformation des coordonnées depuis le référentiel image vers le système de coordonnées de référence. En reprenant la méthodologie de Herrault *et al.* (2013), plusieurs fonctions mathématiques ont été testées pour ce recalage et les résultats ont été comparés par une méthode classique de validation croisée. La fonction « ridge par noyau gaussien » s'est avérée être la meilleure avec une erreur moyenne quadratique inférieure à celle des autres méthodes testées (telles que la transformée polynomiale de second ordre employée par exemple pour les plans anciens mis à disposition sur Géopal) pour les plans les plus anciens (Follin, Fährasmane et Simonetto 2016).

31 La troisième étape, le mosaïquage des feuilles cadastrales vectorisées et géoréférencées, a pour finalité de fournir un assemblage dont la topologie est cohérente. Cette étape est automatique. Comme nous l'avons vu, une commune est composée de plusieurs sections cadastrales, chacune dessinée partiellement ou en totalité sur une ou plusieurs feuilles cadastrales, qu'il est nécessaire d'assembler afin d'obtenir une continuité dans le plan. Or, en sortie du traitement de géoréférencement, des défauts topologiques (chevauchements, trous) entre les parcelles voisines des différentes feuilles cadastrales sont constatés. Actuellement, cette étape n'est pas opérationnelle dans notre chaîne de traitement en Python, mais est réalisée à l'aide d'outils inclus dans QGIS pour traiter les cas les plus simples (présence d'une superposition ou d'un trou entre deux parcelles voisines qui devraient être contiguës). Les cas plus complexes (parcelles qui devraient être séparées par des espaces non cadastrés et sont en intersection à l'issue de notre géoréférencement) ne sont pas encore pris en charge. Pour résoudre ces derniers cas, qui n'ont pas été rencontrés lors de l'application de notre méthode, nous pouvons envisager d'utiliser un processus de conflation géométrique tel que celui présenté dans Touya *et al.* (2013).

32 Cette chaîne de traitement fournit en sortie une base de données multitudes qui montre les changements d'état (division, fusion) au cours du temps et qui constitue le support des analyses menées sur l'évolution du parcellaire cadastral.

FIGURE 2. CHAÎNE DE TRAITEMENT

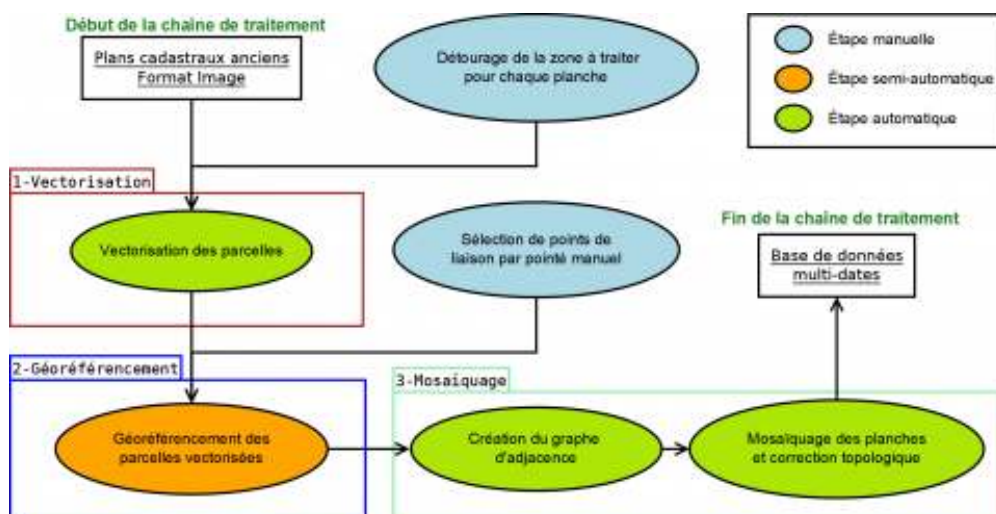


Image produite par les auteurs

33 Nous allons présenter plus en détail les deux approches évaluées pour l'étape de vectorisation.

Approches adoptées pour l'étape de vectorisation

34 De nombreuses approches ont été testées sur des images scannées de cartes, par exemple l'extraction des parcelles en utilisant des outils de squelettisation (Janssen, Duin et Vossepoel 1993), l'extraction des routes par transformée de Hough (Dhar et Chanda 2006) ou par correspondance de modèle (Chiang et Knoblock 2013). Ces approches s'accompagnent de prétraitements et de post-traitements plus ou moins complexes (segmentation, seuillages, opérations morphologiques, vectorisation...). Un état de l'art plus détaillé peut être lu dans Chiang, Leyk et Knoblock (2014).

35 Le LSD a été retenu dans Franken *et al.* (2020) concernant l'extraction de lignes sur des documents de cadastre scannés. Ce travail nous permet ainsi d'évaluer la pertinence du LSD sur le cadastre napoléonien comparé à une méthode plus classique, la THP.

36 Les deux méthodes évaluées, THP et LSD, visent à détecter les contours de parcelles à partir des images sous forme de segments qui seront convertis en polygones dans un deuxième temps. Nous commençons par un prétraitement dont les étapes sont sensiblement identiques pour les deux approches et terminons par les mêmes post-traitements.

Le prétraitement

37 La donnée d'entrée pour le LSD est une image de niveaux de gris, tandis que pour la THP, c'est une image binaire (c'est-à-dire à deux valeurs). Dans les deux cas, le prétraitement consiste en la création de deux images : l'une qui servira à la détection proprement dite, et l'autre, que l'on nomme dans la suite l'image de validation, qui sera utilisée dans l'étape de validation des segments détectés lors du post-traitement.

38 Ce prétraitement diffère selon l'époque des planches cadastrales. Pour générer l'image pour la détection à partir des planches les plus anciennes (1813, 1850), il comprend les étapes de conversions de l'image en niveaux de gris, en image binaire, puis l'application d'un masque détournant la zone à traiter.

39 Pour l'approche THP, le seuillage par hystérésis permet d'obtenir l'image binaire (0 ou 1). Ce seuillage a été adapté afin de tenir compte des variabilités radiométriques sur une même planche cadastrale. Pour cela, les seuils haut et bas sont estimés pour chaque pixel selon la moyenne locale des niveaux de gris calculée sur un voisinage de taille 3×3 . Nous avons nommé cette méthode le seuillage par hystérésis locale. Le prétraitement se termine par la suppression des petits objets isolés par des opérateurs de morphologie mathématique (figure 3).



Image produite par les auteurs

40

Pour la méthode LSD, nous avons fait le choix de rehausser le contraste car cet algorithme détecte plus facilement des segments à proximité de pixels dont le niveau de gris est élevé. Nous modifions les valeurs par le maximum local des niveaux de gris calculé sur un voisinage de taille 3×3 . La figure 4 illustre le résultat obtenu.

FIGURE 4. EXTRAIT DU CADASTRE DE 1813 DE LA COMMUNE D'AUBIGNÉ-RACAN (FEUILLE 09) APRÈS LE PRÉTRAITEMENT PAR L'APPROCHE LSD



Image produite par les auteurs

41

L'image de validation est à valeurs binaires et est obtenue selon des étapes similaires : conversion en niveaux de gris, seuillage par hystérésis locale (avec un voisinage 11×11), application du masque de détourage de

la zone à traiter.

42 Les planches scannées de 1972 sont de meilleure qualité que les plus anciennes et sont déjà codées sur 2 niveaux (noir et blanc). Leur prétraitement inclut deux étapes : une squelettisation (qui permet d'obtenir les contours d'épaisseur 1 pixel) et la suppression des petits objets isolés par des opérateurs de morphologie mathématique.

La détection par THP

43 Le principe général de la transformée de Hough est de passer de l'espace image 2D à l'espace des paramètres (correspondant aux paramètres d'une droite dans notre cas). La mise en œuvre consiste à comptabiliser le nombre de pixels contribuant à chaque droite décrite dans l'espace des paramètres. Cela constituera une nappe d'accumulation, où un pic indique la probable présence d'une droite. L'implémentation utilisée ici est celle proposée dans la bibliothèque Scikit-image pour Python et décrite dans Galamhos, Matas et Kittler (1999). Elle est locale et probabiliste car elle permet de détecter des segments et non des droites, et utilise des probabilités qui lui permettent d'accélérer son exécution. Cette fonction a été appliquée en fixant de manière empirique les valeurs pour ses quatre paramètres (seuil, longueur minimale d'un segment, longueur minimale d'un trou et liste des angles décrivant l'orientation des segments).

La détection par LSD

44 L'algorithme LSD permet de détecter les frontières sans passer par une détection des points de contours (Von Gioi *et al.* 2012). Pour cela, l'algorithme étudie les discontinuités de niveaux de gris et réalise plusieurs contrôles pour valider ou supprimer les points de segments. Il va permettre, notamment grâce à une fonction de croissance de région, de créer des segments là où sont présents des alignements de pixels avec des directions de gradients proches. L'implémentation du LSD a été réalisée en langage Python afin d'être intégrée dans notre outil. Elle n'est pas optimisée comme peut l'être celle de THP, codée en Cython.

45 Le LSD est pensé pour détecter les segments entre deux zones contrastées. Son objectif est donc sensiblement différent de celui que nous recherchons : détecter des traits foncés tracés sur un fond plus clair. L'application a montré qu'on obtenait ainsi deux traits de part et d'autre de la ligne à détecter. Pour atteindre notre objectif, une transformation est gérée dans le post-traitement.

Les post-traitements

46 Nous définissons deux post-traitements ayant pour objectif respectif de nettoyer les segments extraits à l'étape précédente puis de tenter de former les contours fermés des parcelles.

47 Le premier post-traitement vise à nettoyer l'ensemble des segments extraits à l'étape précédente en se basant sur les informations visibles sur la planche. En effet, des contrôles de présence de points de contours sont effectués avec l'image de validation construite lors du prétraitement. Les segments détectés sont rastérisés pour permettre ce contrôle et calculer un « taux de correspondance » (pourcentage de pixels du seg-

ment rastérisé correspondant à des points de contours). Ces contrôles servent à valider ou non un segment et sont intégrés dans les quatre fonctions principales (figure 5) du post-traitement décrites ci-dessous.

- La fonction « sous_segments » découpe les segments sécants, puis procède à la validation des sous-segments ainsi obtenus.
- La fonction « fusions » rassemble en un seul segment les segments parallèles (écartés d'une fois l'épaisseur moyenne d'une ligne sur l'image) ou les segments colinéaires et voisins.
- La fonction « prolongements » prolonge des segments pour les intersecter, si ce prolongement est validé avec le taux de correspondance (cas B de la figure 6). Un seul test de prolongement non valide est permis par extrémité, afin de limiter au maximum la détection de segments dans des espaces non cadastrés de type route (cas le plus fréquent des erreurs observées après un prolongement).
- La fonction « jointures » ajoute un segment entre deux extrémités de segments (cas C de la figure 6). Chaque jointure permet de connecter une extrémité isolée (nœud pendant) à une autre extrémité de segment (isolée ou non) à condition qu'elle soit validée par le taux de correspondance.

48

Il existe donc deux types de fonctions :

- celles de création de nouveaux segments (« prolongements » et « jointures ») qui servent à combler des lacunes ou des faux négatifs
- celles de correction (« sous_segments » et « fusions ») qui suppriment, découpent ou simplifient des segments *a priori* faux ou redondants

49

À l'issue de ce premier post-traitement, nous obtenons un ensemble de segments simplifiés et validés car on considère après contrôle qu'ils correspondent à une limite visible sur la planche. Néanmoins, certaines connexions entre segments sont manquantes et les contours de parcelles apparaissent incomplets.

FIGURE 5. POST-TRAITEMENT 1 : DÉTECTION DES SEGMENTS CORRESPONDANT À DES LIMITES

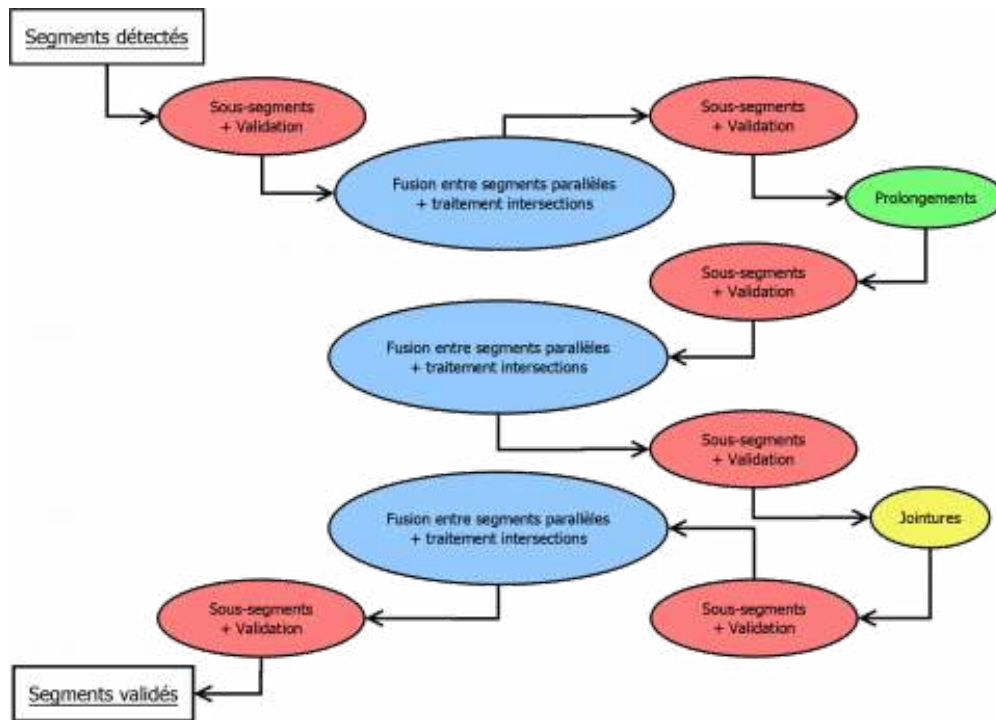
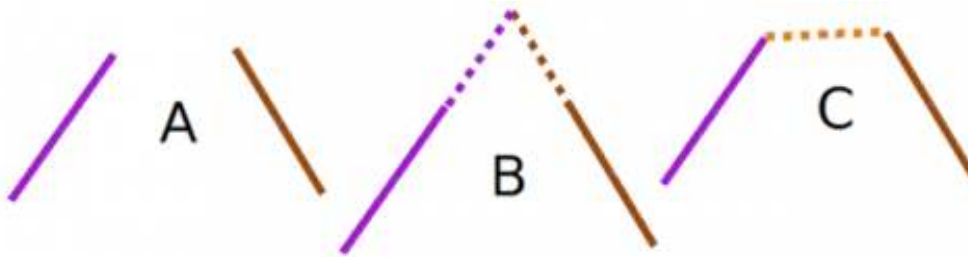


Image produite par les auteurs

FIGURE 6. FONCTIONS « PROLONGEMENTS » ET « JOINTURES »



État initial en A, morceaux de segments créés avec la fonction « prolongements » en B, et nouveau segment créé avec la fonction « jointures » en C.

Image produite par les auteurs

Le second post-traitement vise à corriger les représentations lacunaires de parcelles et à construire des entités polygonales à partir des entités linéaires sans opérer de validation comme dans l'étape précédente (figure 7).

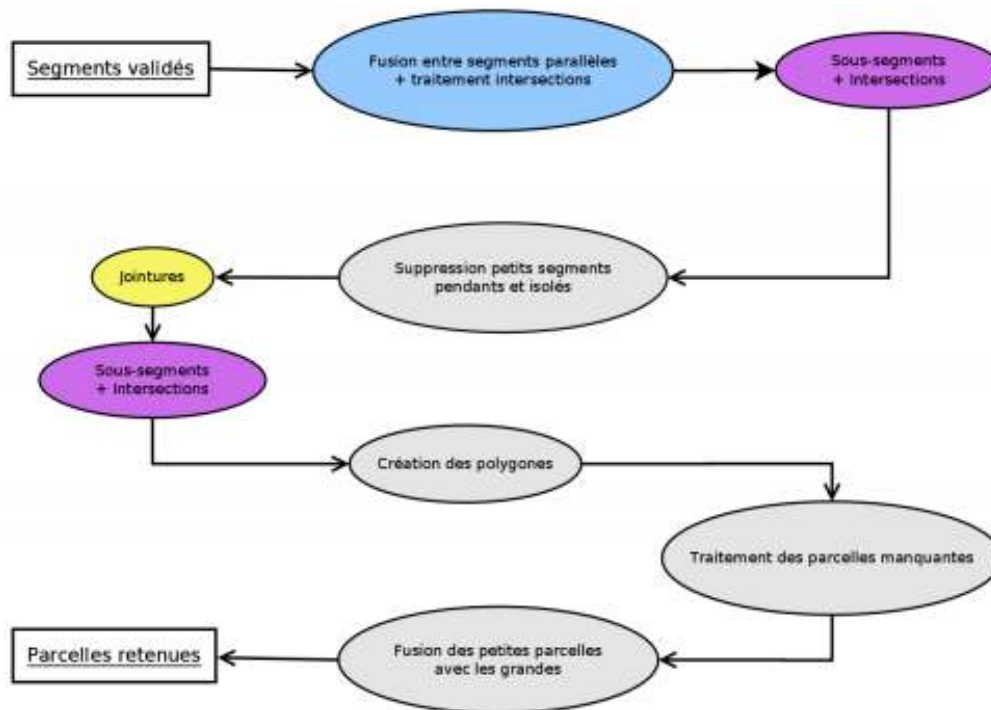


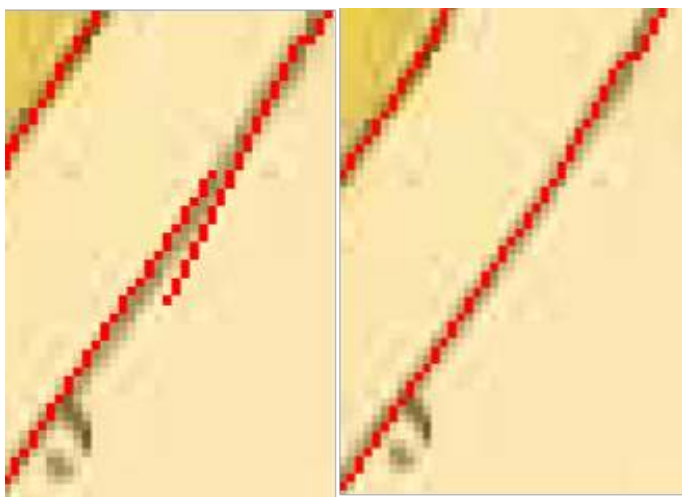
Image produite par les auteurs

51 Ainsi la première partie du second post-traitement utilise les mêmes fonctions que précédemment : « sous_segments » et « fusions », mais sans réaliser de contrôle par le taux de correspondance. Cela autorise le remaniement et la suppression de segments, même s'ils ont été validés comme les petits segments isolés, c'est-à-dire présentant un nœud pendent (figures 8 et 9).

52 Dans la seconde partie de ce post-traitement, les polygones sont générés en utilisant la fonction « polygonize_full » du module Shapely (Gillies 2021). Cette dernière fusionne les lignes concourantes. Ce traitement est imparfait et « oublie » un certain nombre de polygones. Deux types de parcelles manquantes apparaissent ainsi et sont traités :

- Les parcelles fusionnées sont des polygones simples coupés en plusieurs morceaux par des polygones. Elles auraient dû former plusieurs polygones distincts.
- Les parcelles localisées en bordure d'îlot qui forment une « encoche », alors qu'elles devraient être fermées par une ligne du côté extérieur à l'îlot. Dans la figure 10, un îlot dont trois parcelles ont été initialement détectées, comporte ainsi deux « encoches » A et B sur sa bordure droite dont l'une, B, est en fait un polygone omis par la fonction de polygonisation, le trait en rouge qui la clôt n'ayant pas été pris en compte. La correction effectuée par notre post-traitement permet d'obtenir la parcelle verte.

FIGURE 8. EXEMPLE DE RECONNEXION DE SEGMENTS



Segments avant reconnexion, à gauche, après reconnexion, à droite
Image produite par les auteurs

FIGURE 9. EXEMPLE DE FUSION



Segments avant fusion, en haut et après fusion, en bas
Image produite par les auteurs

FIGURE 10. CORRECTION D'UNE PARCELLE CONSIDÉRÉE À TORT COMME UNE « ENCOCHE » EN BORDURE D'ÎLOT

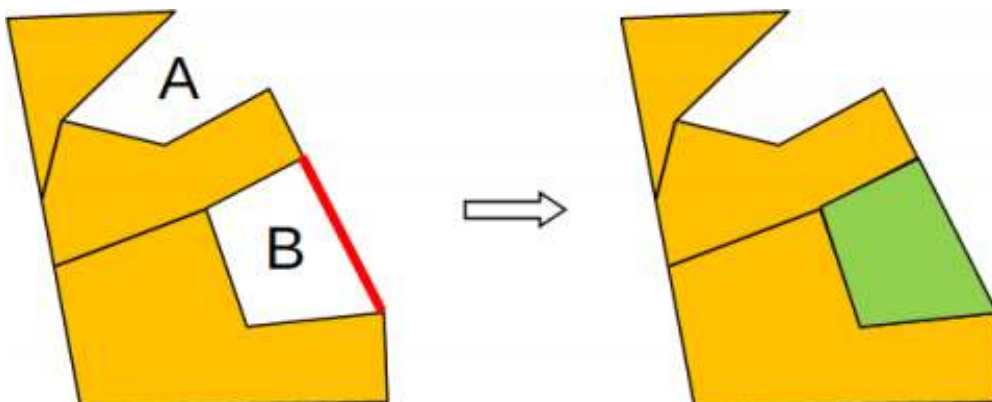


Image produite par les auteurs

53 Enfin les petits polygones (définis empiriquement en fonction des données que nous avons traitées, en l'occurrence ceux dont la surface est inférieure à 15 m²) sont fusionnés avec le polygone adjacent ayant l'enveloppe convexe présentant la plus grande surface d'intersection avec leurs propres géométries.

54 Ce second post-traitement permet d'obtenir les polygones qui décrivent les parcelles. Nous considérons que les quelques erreurs topologiques restantes peuvent être traitées manuellement. Pour cela, l'opérateur disposera de la planche de départ, des parcelles détectées et des segments validés (avec le taux de correspondance) à l'issue des post-traitements.

Application de l'étape de vectorisation

Zone d'étude et données

55 Nous disposons de planches scannées de différentes années d'une commune rurale de la vallée du Loir localisée dans le sud de la Sarthe (Aubigné-Racan). Ce territoire est caractérisé par un habitat regroupé dans les centres-bourgs et dispersé dans les campagnes où l'on trouve des parcelles de taille moyenne à grande comme dans la plupart des régions rurales françaises ayant été partiellement ou totalement remembrées. Sur l'ensemble de la zone traitée, environ 50 % des parcelles présentent ainsi une superficie comprise entre 350 et 2 300 m² pour les années 1813 et 1850. Ces planches ont été scannées par les archives départementales de la Sarthe.

56 Le jeu test utilisé ici se compose des planches 009 (1813 ; voir figure 11), 114 et 115 (1850 ; voir figure 12), et 151 et 152 (1972 ; voir figure 13). Leurs caractéristiques sont résumées dans le tableau 1.

TABLEAU 1. PROPRIÉTÉS DES FEUILLES CADASTRALES ANCIENNES SCANNÉES D'AUBIGNÉ-RACAN

Année	1813	1850	1972-1974
Pas du scanner (DPI)	200	200	400
Effectifs de planches	1	2	2
Échelle	1/2 500	1/2 000	1/2 000
Taille d'un pixel sur le terrain (m)	0,32	0,25	0,13
Dimensions des images en pixels	7 800 × 5 300	8 000 × 5 300	16 000 × 11 000

57 Les planches de 1813 surtout, et, dans une moindre mesure, celles de 1850 sont souvent de mauvaise qualité en raison du jaunissement du papier lié au temps et de nombreuses traces laissées par l'homme (marques de pliures, déchirures, taches d'humidité ou d'encre, traits de construction). Cela complique la vectorisation automatique d'autant plus que la reconnaissance même visuelle des limites de parcelles peut s'avérer difficile. D'autre part, selon la zone géographique, les planches peuvent présenter des caractéristiques variables : orientation et/ou échelle différente, présence d'agrandissements, voire couleur du papier variable (de très clair à très foncé). L'épaisseur des lignes reste néanmoins constante sur toutes les planches d'une même année.

58

Notons qu'en 1813, certaines parcelles sont fusionnées, mais les limites n'ont pas été effacées (juste barrées, avec un symbole assez reconnaissable). Et pour une même zone géographique, la transition entre 1813 et 1850 peut se traduire par la division d'une planche de 1813 en plusieurs planches de 1850.

59

Les planches de 1972-1974 sont, quant à elles, dans un état de conservation parfait. Nous distinguons beaucoup plus facilement les caractères et les limites, puisqu'il n'y a plus de marques de crayon ou d'écrits.

FIGURE 11. PLANCHE DE 1813 ÉTUDIÉE (PARTIE SUD DE LA PLANCHE 009, COMMUNE D'AUBIGNÉ-RACAN)



Archives départementales de la Sarthe

FIGURE 12. PLANCHE DE 1850 ÉTUDIÉE (PLANCHE 115, COMMUNE D'AUBIGNÉ-RACAN)



Archives départementales de la Sarthe

Afin d'évaluer la performance de la vectorisation automatique, avec LSD ou THP, et la qualité de nos résultats selon l'époque, nous avons réalisé une « vérité terrain » en vectorisant manuellement l'ensemble des limites des parcelles sur les 5 planches utilisées. La vectorisation manuelle a été chronométrée, ce qui tient compte des difficultés rencontrées lors de la reconnaissance visuelle des parcelles, liées à la qualité des planches (pour les années 1813 et 1850).

61 Cette « vérité terrain » est utilisée pour le calcul d'indicateurs statistiques et la comparaison des temps de vectorisation entre méthodes manuelle et automatique. Néanmoins, le temps d'exécution de nos méthodes automatiques n'est pas définitif car nos algorithmes ne sont ni optimisés, ni parallélisés. Soulignons que, même coûteuses en temps, les méthodes automatiques permettent de libérer du temps pour l'opérateur.

⁶² La « vérité terrain » est rastérisée pour évaluer la qualité des limites obtenues avant la formation des polygones.

⁶³ Nous nous intéressons à deux indicateurs statistiques :

- le rappel, nommé ici pourcentage de bonnes détections (PBD), qui est le pourcentage de pixels issus des segments détectés correspondant à une limite cadastrale dans la « vérité terrain » (vrais positifs) par rapport au nombre total de pixels appartenant à des limites cadastrales dans la « vérité terrain » (éléments pertinents à détecter)
- le complément à 100 de la précision, nommé ici pourcentage de mauvaises détections (PMD), qui est le pourcentage de pixels issus des segments détectés qui ne correspondent pas à une limite cadastrale dans la « vérité terrain » (faux positifs), par rapport au nombre total de pixels issus des segments détectés

Pour ces calculs, nous considérons une tolérance d'un pixel entre la référence et la détection.

À cela, nous ajouterons une estimation de la durée d'exécution réelle, du prétraitement à la fin de la détection. Ce temps dépend du matériel utilisé, qui est ici le même pour chacune des exécutions réalisées, à savoir un nœud de calcul composé de deux processeurs Intel Xeon cadencés à 2,20 GHz et 62 Go de RAM.

Pour le LSD, des expérimentations ont démontré l'intérêt, à la fois selon le temps d'exécution et selon la qualité des résultats, d'une approche de type « *divide and conquer* » consistant à traiter l'image par blocs plutôt que l'image entière. Une taille de bloc à traiter d'environ 2 000 × 1 300 pixels est apparue optimale.

TABLEAU 2. RÉSULTATS OBTENUS EN SORTIE DES ALGORITHMES DE DÉTECTION

Critère	Algorithme de détection	1813 Planche 009	1850 Planches 114 · 115	1972 Planches 151 · 152
Temps d'exécution	THP	0 h 08	0 h 09 · 0 h 08	0 h 03 · 0 h 03
	LSD	<u>2 h 21</u>	<u>1 h 12 · 0 h 58</u>	<u>1 h 56 · 1 h 48</u>
PBD (rappel)	THP	<u>84,6 %</u>	95,7 % · 98,9 %	99,4 % · 99,4 %
	LSD	87,2 %	97,5 % · 98,9 %	99,9 % · 100 %
PMD (1 – précision)	THP	18,2 %	38,2 % · 33,7 %	37,6 % · 23,2 %
	LSD	<u>31,9 %</u>	<u>43,9 % · 47,6 %</u>	37,9 % · <u>29,3 %</u>

Les meilleurs résultats sont indiqués en gras et les moins bons résultats sont soulignés.

D'après le tableau 2, la THP est systématiquement beaucoup plus rapide que le LSD. Cette mauvaise performance pour le LSD doit cependant être relativisée car, comme nous l'avons vu, l'implémentation de la méthode LSD pourrait être améliorée. Pour les années 1813 et 1850, la THP est plus précise, mais le LSD est légèrement plus performant en matière de bonnes détections pour l'année 1813.

À ce stade, la THP semble meilleure que le LSD mais elle présente le défaut de produire de multiples segments pour une seule limite à détecter (figure 14). Par ailleurs, elle est non déterministe, c'est-à-dire qu'à partir des mêmes données, elle peut produire des résultats différents à chaque exécution. Le LSD fournit en revanche des résultats plus « propres » en matière de linéaires détectés.

FIGURE 14. EXEMPLE DE RÉSULTAT DE DÉTECTION SUR LA PLANCHE 115 AVEC THP (A) ET LSD (B)



Image produite par les auteurs

Résultats à l'issue du second post-traitement

68

Rappelons que le second post-traitement aboutit à la création des polygones. Afin d'analyser les résultats issus de celui-ci, nous appellerons :

- parcelle « correcte », une parcelle de référence (issue de la « vérité terrain ») dont l'aire est représentée à 95 % par un unique polygone détecté (l'aire commune entre le polygone et la parcelle est supérieure à 95 % de l'aire totale de la parcelle et de l'aire totale du polygone)
- parcelle « divisée », une parcelle de référence dont l'aire est représentée à 95 % par plusieurs polygones détectés (l'aire commune entre ces polygones et la parcelle est supérieure à 95 % de l'aire totale de la parcelle)
- parcelle « englobée », une parcelle de référence dont l'aire est représentée à 95 % dans une portion de polygone, c'est-à-dire qu'elle est incluse avec d'autres parcelles dans ce polygone (95 % de l'aire totale de la parcelle est représentée dans moins de 95 % de l'aire du polygone)
- parcelle « partielle », une parcelle de référence dont l'aire n'est pas représentée à 95 % par un ou plusieurs polygones détectés (l'aire commune entre la parcelle et le ou les polygones détectés est inférieure à 95 % de l'aire totale de la parcelle)
- parcelle « non détectée », une parcelle de référence absente (c'est-à-dire qu'aucun polygone détecté n'a d'aire commune avec la parcelle de référence)

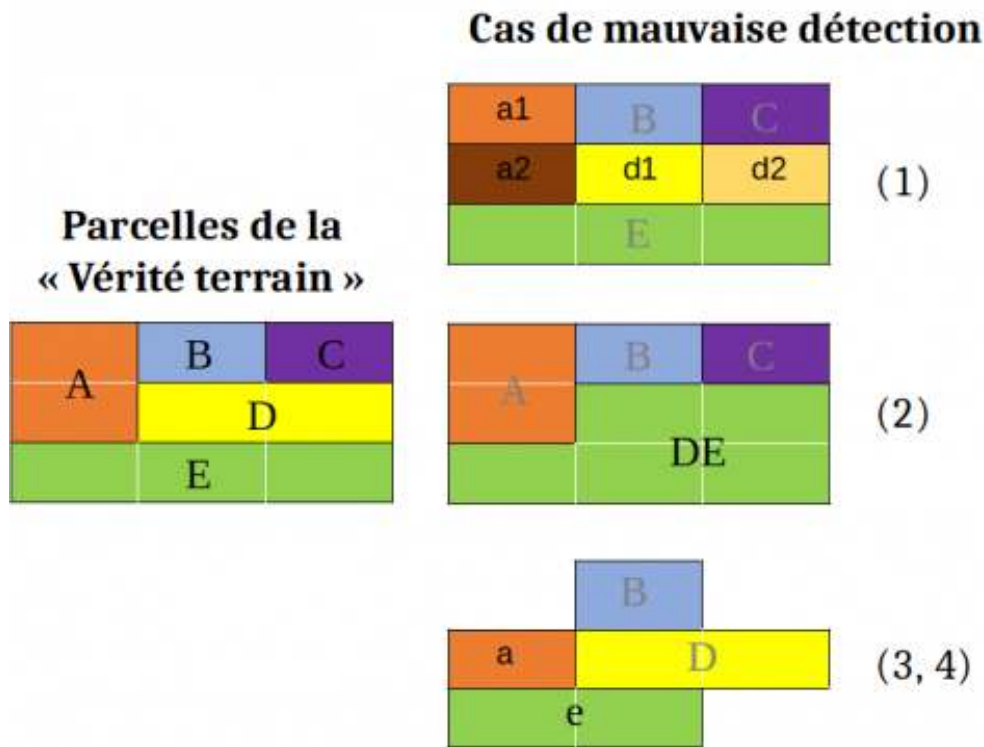
69

Les différents cas de mauvaise détection de parcelles sont présentés dans la figure 15.

La proportion de parcelles « correctes » et « divisées » obtenues mesure la qualité de l'algorithme en termes de vrais positifs et permet d'appréhender sa bonne reconnaissance de la structure du cadastre.

70

La figure 16 présente les statistiques obtenues quant aux cas de mauvaise détection ainsi que le nombre de parcelles détectées et les temps de traitement pour la saisie manuelle, la THP et le LSD. Plus la planche est récente, meilleur est le résultat.



(1) parcelles divisées (A et D), (2) parcelles englobées (D et E), (3) parcelles partielles (A et E) et (4) parcelle non détectée (C)

Image produite par les auteurs

71 Pour 1813, la THP apparaît plus rapide mais ses statistiques sont médiocres, la « partie verte » ne correspondant qu'à 23,8 % des parcelles à détecter. Ainsi, malgré presque cinq heures d'exécution supplémentaires, et une « partie verte » qui atteint difficilement les 50 %, le LSD apparaît plus performant.

72 Pour 1850, le LSD apparaît plus intéressant à la fois pour la qualité de ses résultats, sa « partie verte » atteignant 73,2 % contre 49,5 % pour la THP, et pour son temps d'exécution. Ce dernier varie selon le nombre de segments et parcelles détectés, leurs répartitions et leurs tailles. Par exemple, pour 1850, si la THP dure 34 heures, en réalité les deux tiers de ce temps sont consacrés à l'étude de la planche 114 (l'autre tiers pour la planche 115). Il est difficile de trouver l'origine précise de ces fluctuations.

73 Pour 1972, les résultats des deux méthodes sont nettement plus proches avec une « partie verte » atteignant 61 % pour la THP contre 58 % pour le LSD. Le temps d'exécution est plus déterminant puisque le LSD le fait quasiment doubler. C'est pourquoi, nous pensons que la THP est préférable pour 1972.

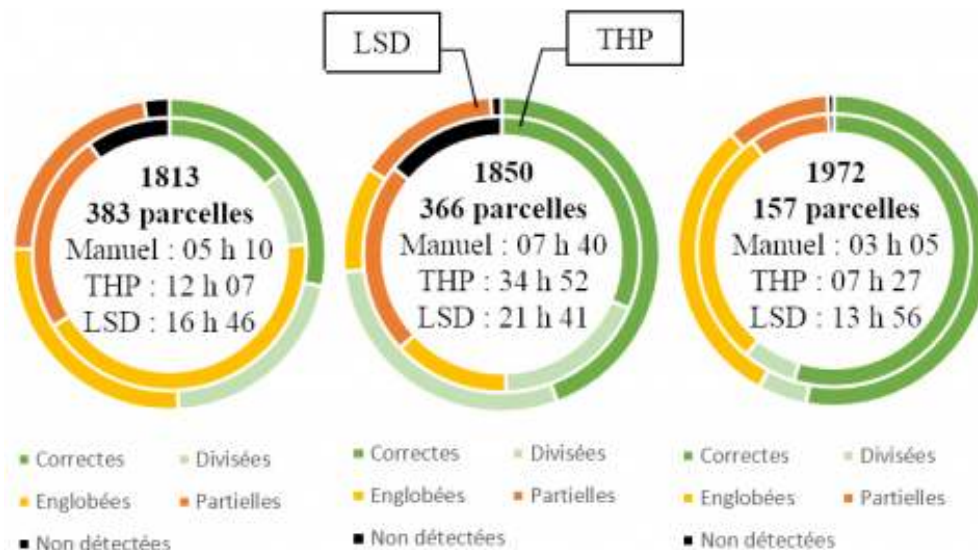


Image produite par les auteurs

74 Ces dernières observations ont permis de mettre en avant l'avantage de notre approche basée sur le LSD par rapport à la THP pour les planches napoléoniennes, ainsi que l'intérêt des post-traitements proposés. Mais, ces résultats étant calculés sur un échantillon très réduit (cinq planches sur plus de cent soixante, rien que sur Aubigné-Racan), ils sont difficilement généralisables. Notre outil, qui a été développé et organisé de façon à être aisément paramétrable et mis en œuvre, devra donc être employé sur d'autres corpus de données pour confirmer ces conclusions.

Conclusions et perspectives

75 Nos travaux portant sur un outil semi-automatique de traitement du cadastre ancien s'inscrivent dans un contexte général de prise de conscience de l'importance du patrimoine cartographique ancien. Le cadastre napoléonien qui est le fruit d'une histoire complexe au lendemain de la Révolution Française, a beaucoup de choses à nous apprendre sur l'évolution de nos territoires au cours du XIX^e siècle. Nous avons présenté ici l'étape clé de la vectorisation automatique des parcelles en comparant deux approches basées sur des algorithmes reconnus dans la détection de primitives géométriques à partir d'images : le Line Segment Detector et la transformée de Hough probabiliste. La mise en œuvre de ces approches a requis le développement d'étapes de prétraitement et post-traitement afin de les adapter au mieux aux caractéristiques de notre objet d'étude. Notre méthode, toujours en phase de développement, a pour l'instant été appliquée sur cinq planches scannées d'une commune rurale du sud de la Sarthe. Dans ce cadre, l'approche basée sur le LSD est apparue supérieure à celle basée sur la THP pour les planches les plus anciennes tandis que la THP semble préférable pour le cadastre récent.

76 Cependant, il reste des points à améliorer qui peuvent concerner l'automatisation du processus ou la qualité des résultats. Il apparaît d'abord nécessaire d'évaluer la robustesse de notre méthode en évaluant son efficacité sur un plus large éventail de feuilles cadastrales, avec des caracté-

ristiques différentes (paysages plus urbains, feuilles présentant des formes de dégradations différentes...). Des travaux sont actuellement menés en ce sens.

77 Concernant la vectorisation, nous envisageons d'évaluer d'autres méthodes d'extraction d'entités linéaires telles que « Edge Drawing » (Topal et Akinlar 2012). Des méthodes d'apprentissage profond pourraient également être mobilisées à partir de ces premières annotations pour améliorer et accélérer l'étape de vectorisation des parcelles mais aussi pour extraire des symboles comme les limites approximatives des biens non délimités (BND), représentées en pointillé sur les planches de 1850, les numéros de parcelles, les toponymes et les mentions des types d'occupation du sol. Nous pourrions nous inspirer des travaux menés dans Guyomard *et al.* (2012) sur l'extraction de symboles, dans Heitzler et Hurni (2020) sur l'extraction des bâtis ou encore dans Laumer *et al.* (2020) sur l'extraction des étiquettes. Les informations géométriques et sémantiques ainsi obtenues pourraient être intégrées dans une base de données multidates du cadastre.

78 Une modélisation par graphes pourrait améliorer l'ensemble du processus incluant la vectorisation et le mosaïquage. L'idée serait d'adopter et d'adapter une modélisation de type graphe granulaire qui permet de considérer différents niveaux de détail (Stell 1999). En effet des graphes d'adjacence peuvent être construits aux différentes échelles du cadastre, notamment celle du tableau d'assemblage de la commune, celle de la section et celle de la feuille avec les parcelles et les espaces non cadastrés. Ce type de modélisation permettrait, par exemple, de contrôler les résultats de la vectorisation en vérifiant que les règles topologiques entre éléments du plan sont valides (par exemple, un bâtiment ne peut pas chevaucher un espace non cadastré).

79 Ces graphes granulaires auraient vocation à intégrer l'aspect temporel pour décrire l'évolution des parcelles, en modélisant les changements d'état, et servir à la création de la base de données multidates.

Bibliographie

Baily, Brian. 2007. « The Extraction of Digital Vector Data from Historic Land Use Maps of Great Britain Using Image Processing Techniques ». *E-perimetron* 2 (4) : 209-223.

Boutoura, Chrysoula et Evangelos Livieratos. 2006. « Some Fundamentals for the Study of the Geometry of Early Maps by Comparative Methods ». *e-Perimetron* 1 (1) : 60-70.

Chiang, Yao-Yi et Craig A. Knoblock. 2013. « A General Approach for Extracting Road Vector Data from Raster Maps ». *International Journal on Document Analysis and Recognition (IJ DAR)* 16 (1) : 55-81. <https://doi.org/10.1007/s10032-011-0177-1>.

Chiang, Yao-Yi, Stefan Leyk et Craig A. Knoblock. 2014. « A Survey of Digital Map Processing Techniques ». *ACM Computing Surveys (CSUR)* 47 (1) : 1-44. <https://doi.org/10.1145/2557423>.

Clergeot, Pierre. 2009. « Du cadastre napoléonien au cadastre en ligne sur Internet – 1^{re} partie ». *XYZ* 119 (2^e trimestre) : 49-59.

Costes, Benoît, Éric Grosso et Christine Plumejeaud. 2012. « Géoréférencement et appariement de données issues des cartes de Cassini : intégration dans un référentiel topographique actuel ». Dans *Actes de la conférence internationale de géomatique et analyse spatiale 2012. Liège, 6-9 novembre*, édité par Roland Billen, Marc Binard, Pierre Hallot et Jean-Paul Donnay, 96-112. Liège : unité de géomatique, université de Liège.

Dhar, Deeptendu Bikash et Bhabatosh Chanda. 2006. « Extraction and Recognition of Geographical Features from Paper Maps ». *International Journal of Document Analysis and Recognition (IJ DAR)* 8 (4) : 232-245. <https://doi.org/10.1007/s10032-005-0010-9>.

Dupouey, Jean-Luc, Jean Bachacou, Régine Cosserat, Serge Aberdam, Daniel Vallauri, Gérard Chappart et Marie-Anne Corvisier de Villèle. 2007. « Vers la réalisation d'une carte géoréférencée des forêts anciennes de France ». *Le Monde des cartes : revue du Comité français de cartographie* 19 : 85-98.

Follin, Jean-Michel, Maïté Fahrasmane et Élisabeth Simonetto. 2016. « An Open-Source Based Toolchain for the Georeferencing of Old Cadastral Maps ». *PeerJ Preprints* 4 : e2222v3. <https://doi.org/10.7287/peerj.preprints.2222v3>.

Follin, Jean-Michel et Élisabeth Simonetto. 2018. « Vers une semi-automatisation du processus d'intégration de plan cadastral ancien dans une base de données multi-dates ». Dans *Humanités numériques spatialisées (HumaNS). Actes de l'atelier Humanités numériques spatialisées (HumaNS'2018), Montpellier, 6 novembre*, édité par Carmen Brando, Francesca Frontini et Mathieu Roche, 58-63, Montpellier : SAGEO 2018. <https://agritrop-cirad.fr/589755/>.

Franken, Jeroen, Wim Florijn, Maarten Hoekstra et Eric Hagemans. 2020. « Rebuilding the Cadastral Map of The Netherlands, the Artificial Intelligence Solution ». *Proceedings of FIG-2020, Amsterdam*. Copenhagen : International Federation of Surveyors.

Galamhos, Charles, Jose Matas et Josef Kittler. 1999. « Progressive Probabilistic Hough Transform for Line Detection ». Dans *Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)* 1 : 554-560. IEEE. <https://doi.org/10.1109/CVPR.1999.786993>.

Gillies, Sean. 2021. *The Shapely User Manual*. Version 1.7.0. <https://shapely.readthedocs.io/en/latest/manual.html> (consulté le 27/02/2021).

Guyomard, Jonathan, Nicolas Thome, Matthieu Cord et Thierry Artieres. 2012. « Contextual Detection of Drawn Symbols in Old Maps ». Dans *Proceedings of the 19th IEEE International Conference on Image Processing*, 837-840. IEEE.

Heitzler, Magnus et Lorenz Hurni. 2020. « Cartographic Reconstruction of Building Footprints from Historical Maps : a Study on the Swiss Siegfried Map ». *Transactions in GIS* 24 (2) : 442-461. <https://doi.org/10.1111/tgis.12610>.

Herrault, Pierre-Alexis, David Sheeren, Mathieu Fauvel, Claude Monteil et Martin Paege-low. 2013. « A Comparative Study of Geometric Transformation Models for the Historical "Map of France" Registration ». *Geographia Technica* 1 : 34-46.

Iosifescu, Ionut, Angeliki Tsorlini et Lorenz Hurni. 2016. « Towards a Comprehensive Methodology for Automatic Vectorization of Raster Historical Maps ». *e-Perimetre* 11 (2) : 57-76.

Janssen, Rik D.T., Robert P.W. Duin et Albert M. Vossepoel. 1993. « Evaluation Method for an Automatic Map Interpretation System for Cadastral Maps ». Dans *Proceedings of the 2nd International Conference on Document Analysis and Recognition (ICDAR'93), Tsukuba, Japan*, 125-128. IEEE. <https://doi.org/10.1109/ICDAR.1993.395767>.

Laumer, Daniel, Hasret Gümgümcü, Magnus Heitzler et Lorenz Hurni. 2020. « A Semi-Automatic Label Digitization Workflow for the Siegfried Map ». Dans *Proceedings of the International Workshop on Automatic Vectorisation of Historical Maps, Budapest, 13 March*, 57-64. ICA Commission on Cartographic Heritage into the Digital. <https://doi.org/10.21862/avhm2020.07>.

Le Couédic, Mélanie, Samuel Leturcq, Xavier Rodier, Florent Hautefeuille, Étienne Fieux et Bertrand Jouve. 2012. « Du cadastre ancien au graphe. Les dynamiques spatiales dans les sources fiscales médiévales et modernes ». *ArcheoSciences. Revue d'archéométrie* 36 : 71-84. <https://doi.org/10.4000/archeosciences.3758>.

Ministère des Finances. 1808. *Administration des finances de l'Empire français*. Paris : Imprimerie impériale.

Ministère des Finances. 1811. *Recueil méthodique des lois, décret, réglemens, instructions et décisions sur le cadastre de la France*. Paris : Imprimerie impériale.

Noizet, Hélène et Éric Grosso. 2012. « Mesurer la ville : Paris de l'actuel au Moyen Âge. Les apports du système d'information géographique d'ALPAGE. » *Le Monde des cartes : revue du Comité français de cartographie* 211 : 85-100.

Perret, Julien, Maurizio Gribaudi et Marc Barthelemy. 2015. « Roads and Cities of 18th Century France ». *Scientific Data* 2 (1) : 150048. <https://doi.org/10.1038/sdata.2015.48>.

Stell, John G. 1999. « Granulation for Graphs ». Dans *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*, édité par Christian Freksa et David M. Mark, 417-432. Berlin, Heidelberg : Springer.

Topal, Cihan et Cuneys Akinlar. 2012. « Edge Drawing : a Combined Real-Time Edge and Segment Detector ». *Journal of Visual Communication and Image Representation* 23 (6) : 862-872. <https://doi.org/10.1016/j.jvcir.2012.05.004>.

Touya, Guillaume, Adeline Coupé, Jérémie Le Jollec, Olivier Dorie et Frank Fuchs. 2013. « Conflation Optimized by Least Squares to Maintain Geographic Shapes ». *ISPRS International Journal of Geo-Information* 2 (3) : 621-644. <https://doi.org/10.3390/ijgi2030621>.

Von Gioi, Rafael Grompone, Jérémie Jakubowicz, Jean-Michel Morel et Gregory Randall. 2012. « LSD : a Line Segment Detector ». *Image Processing On Line* 2 : 35-55. <https://doi.org/10.5201/ipol.2012.gjmr-lsd>.

Notes

1 <https://www.geopal.org>.

2 <https://www.crige-paca.org>.

3 <http://geopeople.ign.fr>.

4 <https://geohistoricaldata.org>.

5 <https://alpage.huma-num.fr>.

6 <http://w3.geode.univ-tlse2.fr/mode-respyr/accueil.php>.

7 <http://modelespace.univ-tlse2.fr>.

8 Ministère de l'Action et des Comptes publics. 2019. « CAD – Descriptif, usage et diffusion de la documentation cadastrale – Documentation cadastrale ». *Bulletin officiel des finances publiques – Impôts*. 05/11/2019. <https://bofip.impots.gouv.fr/bofip/5359-PGP.html/identifiant=BOI-CAD-DIFF-10-20191105>.

9 <http://www.archives.cg19.fr/recherche/fiches/categorie/167>.

Auteurs

Jean-Michel Follin

EA 4630 Laboratoire Géomatique et foncier (GeF), CNAM, Paris, France

Jean-Michel Follin est docteur de l'université de La Rochelle et maître de conférences en systèmes d'information géographique. Ses travaux de recherche s'intéressent à la modélisation et à l'analyse de données spatiotemporelles ou dynamiques en lien avec différents contextes d'application.

ORCID 0000-0001-7787-6121

jean-michel.follin@lecnam.net

Élisabeth Simonetto

EA 4630 Laboratoire Géomatique et foncier (GeF), CNAM, Paris, France

Élisabeth Simonetto est docteur de l'université Rennes 1 et maître de conférences en traitement numérique des images et photogrammétrie. Ses travaux de recherche concernent l'exploitation des images pour la reconnaissance d'objets et la mesure 3D dans différents domaines applicatifs.

ORCID 0000-0001-7870-5156

elisabeth.simonetto@lecnam.net

Anthony Chalais

EA 4630 Laboratoire Géomatique et foncier (GeF), CNAM, Paris, France

Anthony Chalais est ingénieur de l'École supérieure des géomètres et topographes du Mans et technicien en systèmes d'information géographique. Son mémoire d'ingénieur a eu pour but d'améliorer la chaîne de traitement semi-automatique de détection de parcelles décrite dans cet article.

anthony.chalais@sieeen.fr

Droits d'auteur



Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la [Licence Creative Commons Attribution 4.0 International](#).