

Éléments pour une représentation (carto)graphique des matrices de flux

Françoise Bahoken

► To cite this version:

Françoise Bahoken. Éléments pour une représentation (carto)graphique des matrices de flux. M@ppemonde, Maison de la géographie, 2014, 3 (115), 16p. hal-01262025

HAL Id: hal-01262025

<https://hal.archives-ouvertes.fr/hal-01262025>

Submitted on 26 Jan 2016

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.

Éléments pour une représentation (carto)graphique des matrices de flux

Françoise Bahoken

Résumé.— La cartographie des matrices de flux denses, formée de nombreux lieux et *a fortiori* de liens, est caractérisée par une complexité graphique intrinsèque, l'*effet spaghetti*. Il s'agit de la manifestation conjointe d'une *surcharge graphique* et d'une *couverture spatiale* trop importante. Les principaux modèles de représentation (carto)graphique des matrices apportent des solutions complémentaires qui se distinguent avant tout par la manière dont l'espace des flux est considéré.

Matrice de flux • origine-destination (OD) • complexité graphique • effet spaghetti • matrice ordonnée • graphe • carte de flux

Abstract.— Flow mapping of large matrices, formed by many places, is characterized by a *spaghetti effect* due to a *significant overlap* of links and a too large *spatial coverage*. The main graphic representation of such matrices provides complementary solutions that differ primarily in how the space of flow is taken into account in the graphic process.

Flow matrix • origin-destination (OD) • graphic complexity • spaghetti effect • ordered matrix • graph • flowmap

Resumen.— Los mapas de flujos, cuando las relaciones entre elementos son muy elevadas, plantean importantes problemas de complejidad y de representación gráfica, y generan efectos *tipo espagueti* que sobrecargan el mapa. Los modelos de representación (carto)gráficos mediante matrices ofrecen soluciones eficaces, sobre todo, por su forma de considerar los flujos en el espacio.

Matrices de flujos • origen-destino • gráficos complejos • gráficos tipo espagueti • matrices de grafos ordenadas • mapas de flujos

1. Introduction

Les méthodes de la cartographie choroplèthe sont dans l'ensemble bien connues (Cauvin *et al.*, 2008a, 2008b). Leur particularité est leur focalisation sur l'ensemble des lieux qui forment la zone d'étude, l'analyse ayant pour objectif d'illustrer un processus de différenciation spatiale (Pumain, Saint-Julien, 2001). Cette représentation s'appuie sur un fond de carte et une matrice d'information géographique de type (individu*variable) qu'il n'est pas nécessaire de simplifier *a priori* pour assurer la lisibilité de la figure et son efficacité dans la transmission du message. Autrement dit, l'ensemble des lieux de la zone d'étude peut être cartographié directement et de manière efficace sur une carte choroplèthe, ce qui n'est pas nécessairement le cas de la carte de flux origine-destination (OD).

La carte de flux OD est issue d'une matrice éponyme : elle correspond à un tableau croisé (individu*individu) où les lignes et les colonnes correspondent à des lieux, respectivement d'origine (i) et de destination (j). La valeur portée au croisement (i,j)

décrit le transfert d'une quantité observée, donnée : un flux (F_{ij}), pendant une période de temps (t) ; ses marges renseignent sur l'effet des flux sur la distribution des lieux d'origine ou de destination. La carte de flux est ainsi supposée illustrer un processus d'interaction spatiale qui peut être interprété soit du point de vue des flux, lorsque le cœur de la matrice est mobilisé, soit de celui des lieux, comme dans le cas d'une carte choroplèthe.

Le fait que l'objet géographique impliqué dans la cartographie d'une matrice de flux repose sur un vecteur : le lien entre (i) et (j) et non seulement sur un scalaire, le lieu (i) et/ou (j), engendre un certain nombre de difficultés théoriques et méthodologiques qui se répercutent au niveau graphique, quel que soit le support de la représentation. La matrice de flux présente en effet une structure multidimensionnelle comptant au moins quatre dimensions à représenter : les positions des lieux (i), (j), le lien (i,j) et la valeur du flux F_{ij} , ainsi que la temporalité (t), soit : (x_i, y_i) ; (x_j, y_j) ; F_{ij} ; t . Leur représentation graphique passe ainsi par la mise en œuvre d'un processus d'abstraction qui entraîne nécessairement la transformation des valeurs de la matrice. Si plusieurs transformations sont envisageables pour passer d'un tableau à une représentation graphique, elles peuvent toutefois être regroupées en trois familles.

La première famille résulte de la transformation des valeurs de (F_{ij}) en un objet coloré, pour aboutir à une matrice ordonnée. La seconde famille est issue de la transformation de (F_{ij}) en objets graphiques, généralement des liens (L) et des points (i), pour former un diagramme liens-nœuds surtout connu sous le terme de graphe. La troisième famille précise la précédente, en projetant ces objets graphiques dans un espace géographique ; elle conduit à inscrire la carte de flux dans un territoire.

La mise en œuvre de ces trois modèles ne pose pas de problème particulier pour les petites matrices. Elle s'apparente à un défi lorsque la matrice est dense, composée de plus d'une dizaine de lieux. C'est pourquoi nous allons mobiliser une matrice de flux décrivant plus de 12 000 valeurs de manière à ce qu'elle souffre de la complexité graphique caractéristique des matrices denses : l'*effet spaghetti*. Outre la présentation des données, nous allons définir cet effet, avant de présenter l'approche usuelle de sa résolution.

2. Présentation des données

La matrice de flux qui nous sert d'illustration fait partie d'une série de matrices de flux agrégés préparées par C. Grasland, à partir de données issues de l'annuaire démographique tchécoslovaque (*Pohyb Obyvatelstva*). Elles décrivent les migrations résidentielles réalisées à différentes dates entre 1965 et 1990, entre les 114 districts qui composent l'ex-Tchécoslovaquie (Grasland, 1994, p. 439). Cependant, seules les données relatives à l'année 1990 sont mobilisées dans le cadre de cet article.

Le choix d'utiliser cette matrice est lié à sa richesse méthodologique et thématique. Elle décrit effectivement des données harmonisées et stabilisées qui ont fait l'objet de nombreux travaux de recherche. Sur le plan thématique, l'ex-Tchécoslovaquie des années 1990 résulte d'une histoire politique marquée par l'évolution de ses frontières extérieures, qui donne lieu à de nombreux brassages de populations. Le cas d'étude est dès lors très intéressant, tant sur le plan des migrations de population que ces brassages engendrent que des reconfigurations sociospatiales qui en découlent. Parmi les résultats des travaux réalisés sur cet espace, nous pouvons mentionner ceux qui ont conduit à la mise en évidence d'effets spatiaux spécifiques

cartographiés : des effets de barrières ou de voisinage qui permettent de démontrer le rôle des limites de zones comme variables d'ajustement des valeurs de flux (Cattan, Grasland, 1994 ; Grasland, 1994 ; Bopda *et al.*, 2000).

En 1990, l'ex-Tchécoslovaquie est, en effet, formée de plusieurs échelons territoriaux parfaitement emboîtés qui résultent de reconfigurations spatiales historiques (figure 1) et influent sur les migrations internes.

L'échelon le plus fin auquel nous ayons eu accès est celui des 114 districts (*okresj*), qui date de 1961. Il peut être agrégé en dix régions administratives (les *kraj*) dont les limites suivent à peu près celles des trois anciennes républiques historiques qui sont, d'ouest en est : la République tchèque (52 districts), la Moravie (24 districts) et la Slovaquie (38 districts). Le troisième échelon est celui des frontières des Républiques socialistes tchèque et slovaque en vigueur jusqu'en 1989 et le quatrième échelon est celui du périmètre de l'ex-Tchécoslovaquie de 1990. À ces découpages administratifs correspondent « [...] des limites historiques fossiles qui avaient suivi le même tracé au cours du temps » (Bopda *et al.*, 2000, p. 108). Les limites linguistiques apparaissent toutefois plus perméables que les limites administratives, elles sont, en



1. Le découpage en 114 districts de l'ex-Tchécoslovaquie.

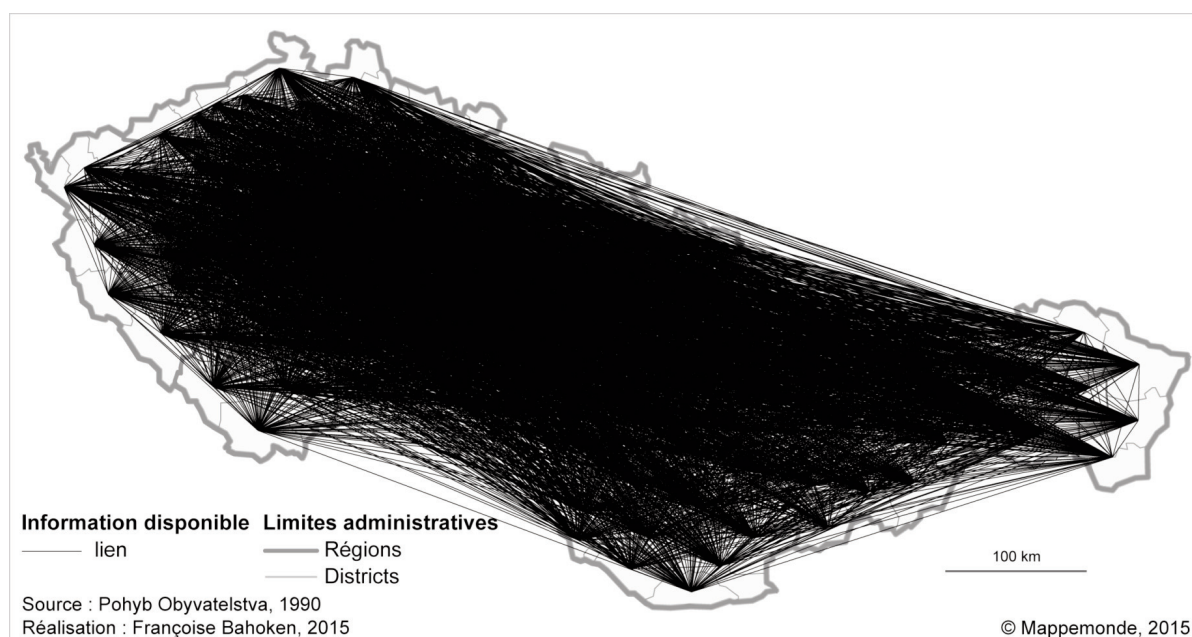
effet, susceptibles de se modifier « [...] au cours du temps à la faveur des migrations, des mariages et du renouvellement des générations » (*ibid.*).

Outre son intérêt thématique, cette matrice peut être considérée comme digne d'intérêt pour l'enseignement. Sur le plan pratique, elle est suffisamment dense pour pouvoir être intéressante, tout en restant aisément manipulable. Une version de ces flux, agrégée à l'échelle des dix régions, sert d'ailleurs régulièrement de support pédagogique pour l'analyse des interactions spatiales sous une forme cartographique ou non.

3. L'effet spaghetti : présentation du problème général de la cartographie de flux

La cartographie de la matrice de flux interdistricts de l'ex-Tchécoslovaquie (1990) consiste à représenter théoriquement 12 882 figurés (sachant que le nombre N de districts est égal à 114 et que l'on ne considère pas les flux internes : $[N*(N-1)]=(114*113)$. La matrice n'étant pas complète (seulement 84% de ses cases sont renseignées), 10 356 liens sont à représenter sur la carte. Les données étant spatialisées, nous représentons dans un premier temps les cases renseignées (*figure 2*).

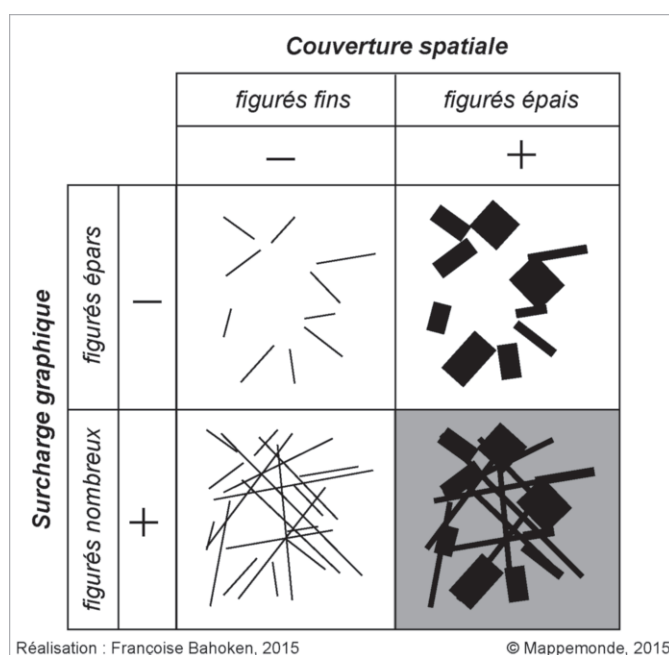
La *figure 2* est typique de la complexité graphique inhérente à la représentation des matrices de flux denses. Elle conduit à une figure « illisible, donc inutile » (Torricelli, 1997, p. 9), à l'image d'un « chaos » pour W.R. Tobler (1987), d'une « toile d'araignée » pour D.F.L. Dorling (1991). Nous leur préférons l'expression d'effet spaghetti que nous avons empruntée à J. Breukelmann *et al.* (2009) : « S'il y a beaucoup de flux, sur une carte, entre les origines et les destinations, leur tracé peut entraîner un motif 'spaghetti' ; tous les types de flux étant dessinés les uns sur les autres, brouillant les motifs généraux. Pour être en mesure de produire une carte lisible, seuls les principaux flux seront par exemple représentés » (Breukelmann *et al.*, 2009, p. 43). En faisant nôtre ce terme, nous proposons de le définir.



2. Cartographie des liaisons disponibles dans la matrice des flux de l'ex-Tchécoslovaquie (1990).

3.1. Définition de l'effet spaghetti

L'effet spaghetti est plus complexe qu'il n'y paraît sur la figure 2. Si l'on peut démontrer que son origine est avant tout d'ordre statistique, elle peut également l'être d'un point de vue graphique. Par rapport à l'approche de D. Fairbairn (2006), nous appréhendons cette complexité graphique de la carte de flux comme une difficulté liée à la compréhension de l'information représentée, en raison du nombre et de la densité de figurés représentés. Elle résulte en fait de la manifestation conjointe de deux sous-effets que nous qualifions, pour le premier : d'effet de surcharge graphique et pour le second : d'effet de couverture spatiale (figure 3).



3. Définition des composantes de l'effet spaghetti.

La *surcharge graphique* est liée à la quantité de figurés présents sur la carte. Elle correspond à un chevauchement ou à une superposition trop importante, souvent renforcée par la présence de longs figurés présentant des orientations variées. De fait, lorsque la charge graphique, c'est-à-dire le nombre de figurés, est trop grande, le message est brouillé, illisible.

La *couverture spatiale* est liée non pas au nombre de figurés, mais à la surface visuelle qu'ils occupent, c'est-à-dire à la quantité de teinte (généralement noire) visible sur un fond de carte (souvent blanc). On peut caractériser cette quantité de teinte par ses dimensions géométriques, étant donné qu'elle correspond au produit de la *longueur* par la *largeur* des liens qui décrivent les flux. Leurs dimensions représentent alors, respectivement, la cartographie de la distance géographique parcourue (la *distance cartographique*) entre les OD et la valeur du flux : la *largeur* étant proportionnelle à la valeur du flux, en raison de l'application du principe de sémiologie graphique correspondant (Bertin, 1967).

Ainsi, si la *surcharge graphique* peut être interprétée comme une densité de figurés, la *couverture spatiale* correspond à une densité de teinte ; elle est directement liée à la taille des figurés, au degré de recouvrement du fond de carte par des liens qui, lorsque ceux-ci sont trop épais (ou trop larges), conduit à assombrir la figure et cela, indépendamment du fait qu'ils soient trop nombreux ou non. C'est donc la combinaison de ces deux composantes qui définit, de notre point de vue, l'effet spaghetti.

Cette *complexité graphique* avérée de la cartographie des matrices denses constitue un réel problème, car l'efficacité d'une carte passe par la perception instantanée de l'information qu'elle est supposée transmettre. « L'excellence graphique correspond à une présentation bien conçue des données d'intérêt [...] [elle] vise à communiquer des idées complexes avec clarté, précision et efficacité ; [elle] est ce qui donne au lecteur le plus grand nombre d'idées dans le temps le plus court, tout en utilisant le moins d'encre sur un minimum d'espace » (Tufte, 1986, p. 51). Une transmission adéquate du message réside, par conséquent, dans l'articulation d'éléments d'ordres variés qui visent à réduire sa complexité visuelle (Mac Eachren, 1992).

Le raisonnement qui est traditionnellement adopté pour réduire cette *complexité graphique* consiste à agir au niveau analytique, lors du traitement des données de flux et non au niveau graphique.

3.2. Les principales familles de solutions envisageables

Les solutions apportées peuvent être classées en quatre grandes familles. La première est suggérée par W. Bunge (1973). L'auteur a proposé de résoudre le problème spécifique de la quantité d'information à représenter en la simplifiant, c'est-à-dire en la sélectionnant avec « parcimonie ». Dans ce registre, l'approche la plus courante consiste à éliminer brutalement une partie de l'information, afin de ne représenter que les flux qui sont considérés comme significatifs. C'est le cas de l'essentiel des cartes de flux qui ne représentent que les n flux les plus importants (en valeur, en tonnage, en volume...).

La seconde famille de solutions conduit à une réduction de la résolution des données, en utilisant des méthodes d'agrégation qui diminuent le nombre de lieux et, *a fortiori*, de liens. Elles sont cependant à manier avec précaution, car elles conduisent à un changement d'échelle géographique qui s'apparente, dans certains cas, à un changement d'objet d'étude.

La troisième famille est liée aux transformations cartographiques de positions dont l'effet varie, en fonction de l'échelle géographique. Elles conduisent à un changement de projection, à l'échelle mondiale ; à une modification de la position des lieux, à l'échelle locale. Si l'on prend l'exemple du logiciel CARTOD (Bonin *et al.*, 2013), le changement de la position des communes conduit à leur redistribution sur des cercles qui, lorsqu'elle est associée à une fusion graphique des liens, autorise une vision cardinale des principaux motifs des flux représentés.

Alors que les solutions apportées étaient, pendant de nombreuses années, exclusivement mises en œuvre lors du traitement des données, on observe ces dernières années l'émergence d'une quatrième famille de propositions purement graphiques, liées au développement des potentialités du support-écran. Les avancées réalisées dans le champ de l'informatique graphique (*InfoViz*) permettent désormais l'exploration interactive de l'image des flux, ainsi que leur sélection empirique. Elles rendent aussi opérationnelles d'anciennes méthodes de cartographie, en même temps qu'elles les popularisent. C'est le cas, par exemple, des possibilités très intéressantes de fusion graphique des liens (Holten, 2006) ou des actions menées sur la morphologie de la teinte (Mac Eachren, 1995) qui conduisent, dans certains cas, à renouveler l'image de la carte de flux (Bahoken, 2014).

Si ces solutions récentes étendent les possibilités graphiques de manipulation des matrices, il convient toutefois de prendre garde de ne pas tomber dans le piège de la

« curiosité cartographique » (Baron *et al.*, 2011) en réalisant des figures esthétiques, très lisibles, mais aussi très peu utiles sur le plan analytique. La réalisation d'une carte de flux est, en réalité, d'abord liée à des considérations d'ordres théoriques et méthodologiques indépendantes du support et des outils ; nous nous en tiendrons ici à sa version statique. L'objectif est de représenter les données disponibles dans la matrice de flux de l'ex-Tchécoslovaquie (1990), dont nous avons vu qu'elle était fortement sujette à l'*effet spaghetti*.

4. L'approche graphique de la représentation des matrices

La représentation graphique de données statistiques consiste à les présenter de manière simplifiée, voire résumée, sous une forme visuelle, c'est-à-dire par un graphique. Elle participe de la chaîne d'analyse d'un phénomène et son intérêt pour la compréhension de l'information n'est pas négligeable (Tukey, 1977). Elle est donc envisagée ici comme une étape d'une chaîne d'analyse globale des interactions spatiales, mais l'étape de modélisation n'est pas considérée ici.

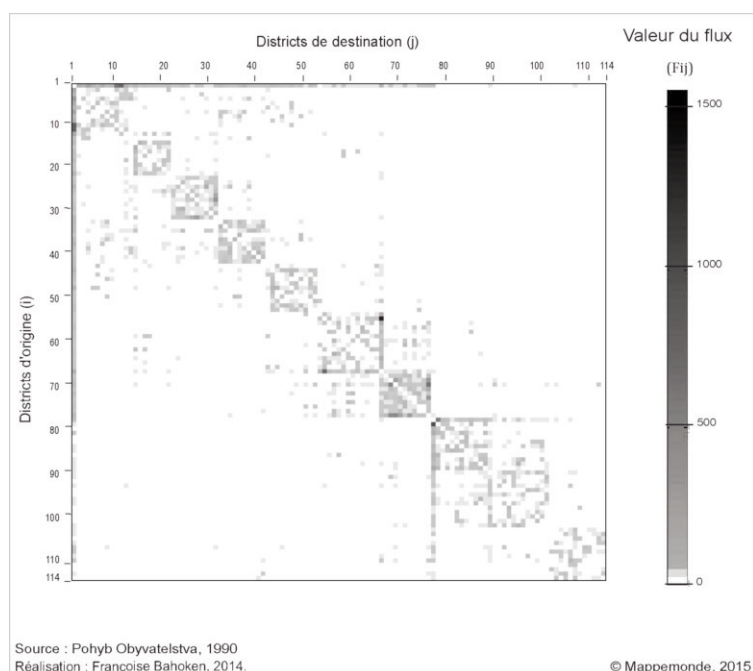
Les données étant localisées, nous explorons la figure des flux de l'ex-Tchécoslovaquie (1990) sous la forme de matrice ordonnée, de graphe et de carte de flux.

4.1. La matrice ordonnée

Le mode de représentation en matrice ordonnée, ou matrice-Bertin, a été mis au point d'abord manuellement par Jacques Bertin, à la fin des années 1960. Il correspond à une méthode de traitement graphique de données quantitatives, développée à l'heure où les outils informatiques permettant la mise en œuvre de traitements statistiques multivariés n'étaient pas généralisés. La méthode reste néanmoins d'actualité, il est désormais possible de la mettre en œuvre facilement, grâce à une application récente de l'équipe Aviz (<http://aviz.fr/bertifier>) de l'INRIA, qui en a fait un objet de recherche particulier.

L'objectif général de la matrice ordonnée est de recoder les valeurs de flux en niveaux colorés, puis à les ordonner de manière à aboutir à une typologie visuelle. La figure obtenue garde la forme initiale de la matrice : elle compte 114 lignes et 114 colonnes au croisement desquelles la valeur de flux est transformée en une teinte correspondante (figure 4).

Sur la figure 4, les cases blanches correspondent à l'absence de valeurs et les cases teintées à une valeur de flux, ces dernières s'étalent sur un intervalle compris entre 1 et 1567. Le nuancier de gris utilisé correspond à un recodage équivalent des valeurs selon la variable visuelle valeur. Le fait que ce recodage soit réalisé sur des effectifs ne s'inscrit pas en faux par rapport aux principes de sémiologie graphique définis par Jacques Bertin. L'objectif de l'ordonnancement est, en effet, de rechercher un ordre dans les valeurs de la matrice, « [...] des relations internes, c'est-à-dire des relations contenues dans les données inventoriées [...] » (Bertin, 2005, p. 70). En proposant ce modèle, l'auteur se place en explorateur de données, en *data scientist* comme on dit aujourd'hui. Il agit à la manière du statisticien qui cherche à classer ses données sans tenir compte du fait qu'elles pourraient être représentées graphiquement, il ne raisonne pas en cartographe qui tente de donner à voir ces données. Aussi élude-t-il les aspects visuels de la représentation : les idées de quantité et de proportionnalité entre les valeurs des cellules qui sont, en l'état, illisibles.



4. La matrice ordonnée des flux interdistricts de l'ex-Tchécoslovaquie.

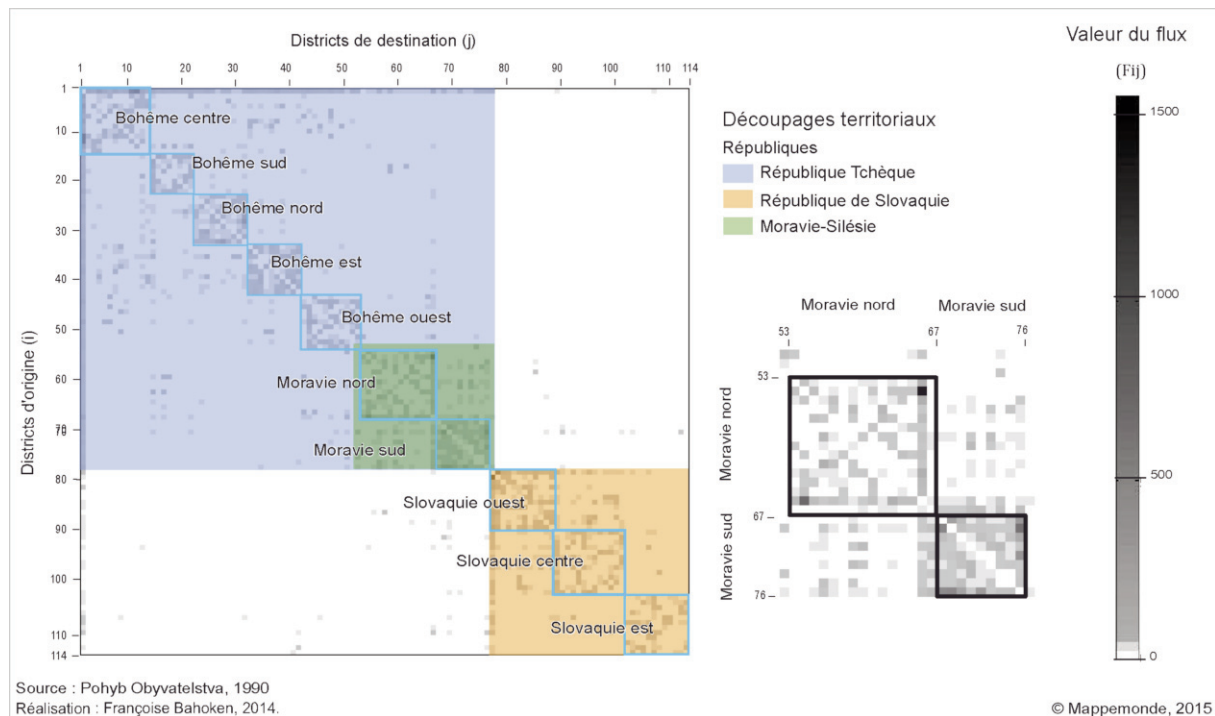
La matrice ordonnée présente de nombreux avantages qui sont liés aux potentialités analytiques de la figure, lorsqu'elle est utilisée dans un contexte exploratoire. Sur la [figure 4](#), les districts ont été ordonnés selon leur numérotation administrative : de 1 à 114 ([figure 1](#)). La typologie spatiale obtenue est robuste tout en étant purement visuelle et ce résultat est comparable à celui qui aurait pu être produit par un traitement mathématique (Bertin, 1980). La réduction des unités spatiales fait, en effet, apparaître la trace des dix ensembles régionaux et des trois républiques historiques, qu'il est possible de visualiser en interagissant avec la [figure 4](#).

Les données étant localisées, le résultat de cette typologie visuelle peut ensuite être interprété sur le plan géographique : on observe ainsi une forte concentration des flux à l'intérieur de chaque région correspondant au chef-lieu. La [figure 4](#) met en évidence la structure spatiale régionale des flux de l'ex-Tchécoslovaquie, ainsi que celle des trois anciennes républiques. Elle souligne naturellement le rôle joué par les frontières des différents échelons et l'antériorité de leurs relations qui se déploient au-delà du découpage administratif de 1990.

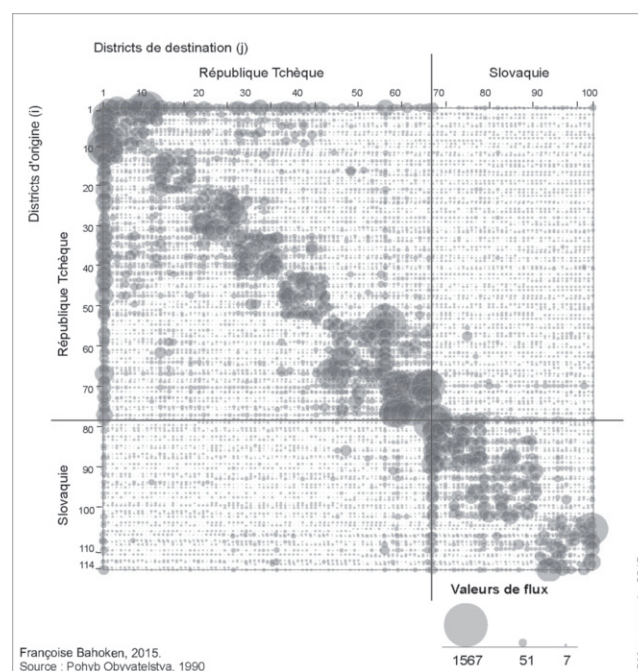
L'intérêt est également d'autoriser une analyse locale de ces flux. Pour cela, observons à titre d'exemple les flux internes à l'ex-République de Moravie (voir [figure 4bis](#)). Ils se produisent au sein des deux régions de Moravie Nord (MN) et Moravie Sud (MS), mettant en évidence la permanence des relations intra-régionales (MN, MS), mais aussi interrégionales, dans la limite du périmètre de l'ancienne Moravie-Silésie. On observe également la porosité de cette limite historique, visible par l'intensité de teinte des cellules correspondant aux flux impliquant des districts à l'époque « frontaliers », qui appartiennent à la République tchèque de 1990.

La bonne délimitation de ces entités administratives suggère plusieurs hypothèses déjà mentionnées, sur l'existence de relations préférentielles entre les districts appartenant à une même région (Bopda, Grasland, 1994 ; Grasland, 1994). Si la figure est efficace pour détecter la structure spatiale de ces flux, le réordonnancement des lignes et des colonnes entraîne une réduction des données qui est apparue discutable

à certains auteurs, car elle conduit à « [...] gommer les détails de la représentation des mouvements » (Cauvin, Reymond, 1991, p. 480). Sur la [figure 4](#), l'analyse est davantage menée au niveau régional qu'à celui des districts, ce qui pose le problème de la perte de l'information élémentaire fine (mesurée au niveau des districts). Pour y remédier, il est possible d'utiliser une variante du modèle qui, tout en conservant le même formalisme matriciel, représente les flux du point de vue des lieux. Dans la matrice en « semis de points » (Beguin, Pumain, 1994), la surface du figuré (de chacun des lieux) est rendue proportionnelle à la valeur du flux ([figure 4](#)). À l'inverse,



4bis. La matrice ordonnée des flux interdistricts de l'ex-Tchécoslovaquie (1990) avec un zoom sur la Moravie (matrice en noir et blanc) distinguant les parties nord et sud.



4ter. La matrice ordonnée en cercles proportionnels des flux interdistricts de l'ex-Tchécoslovaquie (1990).

dans les matrices en étoiles (Ward, 2002) souvent mobilisées en *InfoViz*, la transformation est double : la matrice demeure carrée, mais la valeur de la cellule correspond désormais à un lieu et non à un lien.

La représentation de flux, sous la forme de matrices ordonnées, présente l'inconvénient de ne pas permettre de visualiser leur géographie et leur utilité est surtout exploratoire. Ce constat explique probablement le faible intérêt porté à ce type de représentation en géographie, alors qu'il constitue un domaine de recherche foisonnant en *InfoViz*. Il s'agissait pourtant, pendant de nombreuses années, du seul cas de figuration efficace des flux sans simplification ni sélection *a priori* de l'information. C'est d'ailleurs la raison pour laquelle ses possibilités ont pu être maximisées, notamment grâce à son utilisation combinée avec le graphe.

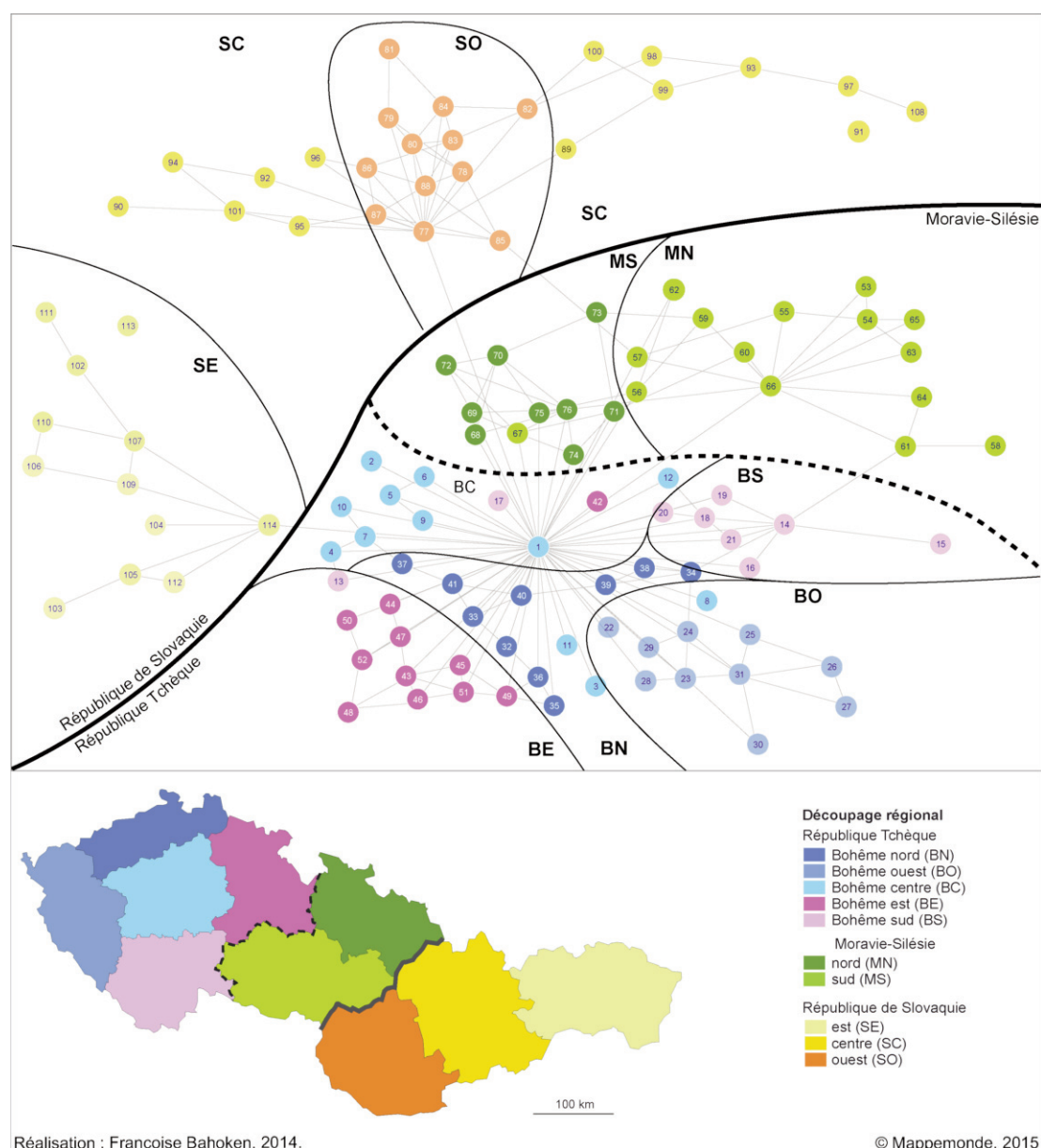
4.2. Le graphe de flux

La représentation d'une matrice de flux sous la forme d'un graphe non-planaire consiste à transformer ses valeurs en objets graphiques : souvent des lignes et des cercles, puis à les placer dans un plan ordinaire, généralement en deux dimensions. L'objectif est double : il s'agit soit de visualiser les liens qui existent entre les lieux, soit la formation de communautés. Dans le premier cas, la figure obtenue présente une complexité graphique similaire à celle de la carte de la [figure 2](#) qu'il n'est pas nécessaire de présenter. Effectivement, dans un graphe, les liens se croisent et les lieux se chevauchent (Henry, Fekete, 2008, p. 62). Dans le second cas, le graphe a pour objectif de décrire la formation de communautés d'individus interagissant plus eux qu'avec d'autres. Lorsque ces individus sont des lieux, ces communautés correspondent souvent à des « régions », formées d'un ensemble d'unités spatiales présentant une ou plusieurs caractéristiques d'appartenance régionale commune.

La [figure 5](#) présente un graphe des valeurs de flux de l'ex-Tchécoslovaquie (1990) supérieures à 100, soit 373 liaisons qui correspondent à 0,03% de l'ensemble. La position des nœuds n'y est pas aléatoire, elle résulte de l'application de l'algorithme de T. Fruchterman et E. Reingold (1991), supposé adapté à la visualisation des communautés spécifiques aux grands graphes. Pour résoudre les problèmes se posant à leur dessin, la solution générale consiste en effet à repositionner les lieux de manière à minimiser le chevauchement des liens, ou bien à ne représenter qu'une sélection d'entre eux. La couleur des points correspond à la région d'appartenance du district dans le découpage territorial ([figure 1](#)).

L'efficacité du graphe de la [figure 5](#) est avérée, dans la mesure où la structure régionale des flux est perceptible dans les regroupements des lieux en communautés. En effet, si on compare la [figure 6](#) et la [figure 1](#), on s'aperçoit que les communautés formées correspondent globalement aux limites régionales, elles-mêmes faisant écho à celles des anciennes républiques tchèque et slovaque.

Cependant, en l'absence de contrainte exercée sur la position des lieux, la représentation de la matrice de flux sous la forme de graphe n'apporte pas d'amélioration notable quant à la compréhension de la structure géographique des flux. Les solutions liées au repositionnement des objets permettent d'améliorer la lisibilité et l'esthétique du graphe, mais elles entraînent la perte des repères géographiques, à la différence de la carte de flux.



5. Un graphe de flux de l'ex-Tchécoslovaquie (1990).

4.3. La carte de flux

Dans une carte de flux, la conception de la position des objets (géographiques) diffère fondamentalement de la matrice ordonnée et du graphe. Elle opère, en effet, une transformation des valeurs (F_{ij}) en objets graphiques qui sont, en outre, projetés dans un espace métrique, muni d'un système de projection cartographique. La spécificité de la carte de flux réside par conséquent dans la référence à un espace qui peut être géographique, fonctionnel ou même social. Le caractère absolu de la localisation des objets sur une carte est d'autant plus important qu'il confère aux couples de lieux un certain nombre de propriétés spatiales liées à leur localisation et à leur espacement, c'est-à-dire à une acception d'éloignement (théorique, empirique, géographique ou social), pouvant être appréhendée sous diverses formes : la mesure peut être continue ou discrète, binaire ou ordinale.

Ainsi, « Puisqu'il n'est pas possible de représenter tous les flux, l'approche consiste à les sélectionner, ce qui signifie qu'un certain nombre d'entre eux – souvent la majorité – seront réputés moins saillants, quel que soit le critère utilisé pour sélectionner les flux. » (Holland, Plane, 2001, p. 90).

Lorsque l'objectif de la carte n'est pas de représenter toute l'information disponible dans la matrice de flux (Bahoken, 2014), l'une des solutions consiste donc à sélectionner les données à représenter. Plusieurs approches sont envisageables en fonction des objectifs fixés à la figure. L'une d'entre elles consiste à ne retenir que les valeurs qui répondent à un objectif de *significativité* de la figure. Cela passe par l'application d'un seuil (α) de *représentativité* de l'*information disponible* dans la matrice. Le résultat conduit à la cartographie des seules valeurs de flux supérieures au seuil ($F_{ij} > \alpha$).

La question est alors de savoir comment s'assurer que l'on cartographie l'*information significative*, en appliquant un *critère unique* de sélection à l'ensemble de la matrice, un critère qui agit au niveau global de la matrice. Une seconde question porte sur la comparabilité du résultat obtenu, puisque le raisonnement sous-jacent est identique à celui qui conduit à la réalisation de discrétisations statistiques : l'application d'un seuil unique correspondant à une dichotomie. Existe-t-il une méthode permettant d'estimer la valeur de ce seuil de manière que le résultat cartographique obtenu puisse être comparable à d'autres ?

Les méthodes de sélection étant, en effet, d'une robustesse variable, une manière simple d'y parvenir consiste à examiner la courbe de distribution des valeurs de flux (Tobler, 1982). Elle mobilise, en pratique, plusieurs outils graphiques : du diagramme de distribution à la courbe de Lorenz auxquels on propose de recourir. La raison tient au fait que le critère (α) agit en réalité de deux manières : d'une part, sur la part d'*information significative* de flux (F_{ij}) à représenter et, d'autre part, sur la part de figurés à représenter, de liens (L_{ij}), qui se doit d'être réduite pour garantir la lisibilité de la carte. Pour illustrer l'action de ce critère, nous avons réparti les flux de l'ex-Tchécoslovaquie (1990) en 4 classes de valeurs ([tableau 1](#)).

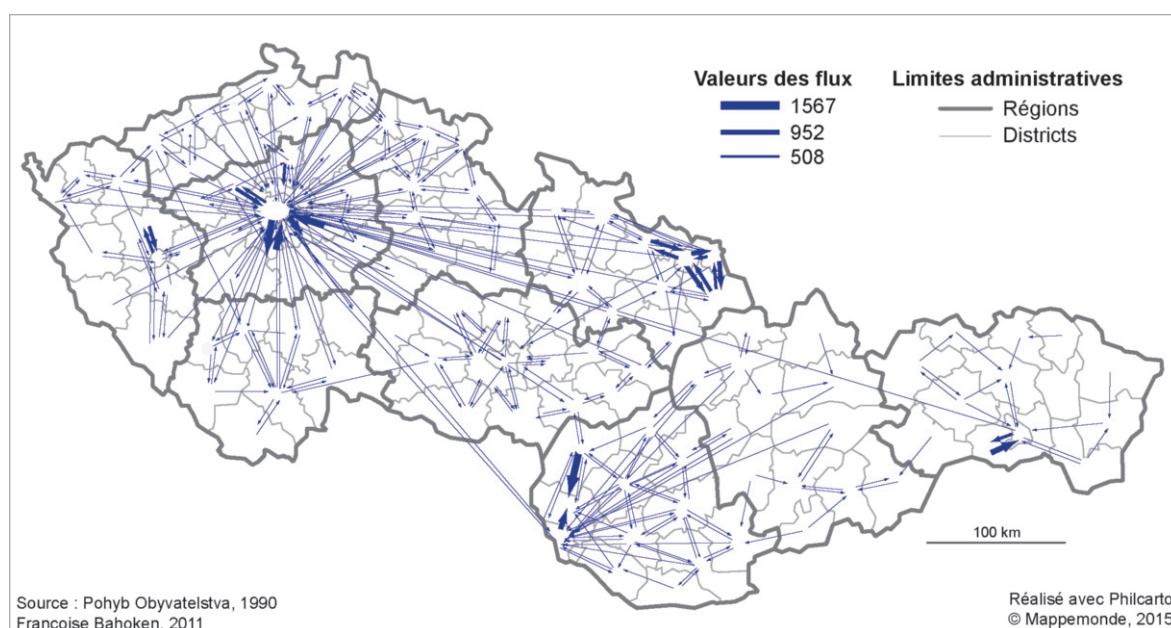
Nous observons sur le [tableau 1](#) que les plus gros flux, qui présentent les valeurs (F_{ij}) les plus fortes correspondent au nombre de liens (L_{ij}) le plus faible et inversement. La première classe des valeurs de flux comprises entre 1 et 100 correspond à 96,40 % du total des liens (L_{ij}) : ces liens, au nombre de 9 983, ne contribuent qu'à 8,06 % de l'information totale. Ils correspondent à la masse de petits flux diffus qui provoque en général un effet de *couverture spatiale* important. Tout en étant très nombreux, ces petits liens n'apportent pas d'information *significative* et leur

1. Part de flux (F_{ij}) versus part de liens (L_{ij}) de l'ex-Tchécoslovaquie

Numéro de classe	Classes de valeur des flux (F_{ij})	Total cumulé des valeurs de flux (F_{ij})	Part de flux (F_{ij}) en % du total	Nombre de liens (L_{ij})	Part de liens (L_{ij}) en % du total	Valeur de flux moyenne d'un lien
1	1-100	4 853	8,06	9 983	96,40	0,49
2	100-500	34 900	57,97	349	3,37	100
3	500-1 000	10 929	18,15	16	0,15	683
4	> 1 000	9 521	15,81	8	0,08	1 190
Total		60 203	100,00	10 356	100,00	6

représentation conduit à surcharger fortement la figure. Tout au plus masquent-ils l'information apportée par les autres classes qui sont plus intéressantes sur le plan graphique (moins de figurés sont représentés) et analytique (plus d'information est représentée). Nous en déduisons que cette première classe n'apporte pas grand-chose, voire rien, à la compréhension du système de flux. Aussi, si l'on ne considère pas d'autres variables susceptibles d'intervenir, les flux inférieurs à 100 peuvent être éliminés de l'analyse, *a fortiori* de la représentation. À l'inverse, dès la seconde classe des valeurs de flux supérieures à 100, la cartographie devient intéressante, que ce soit sous la forme de graphe (figure 5) ou de carte (figure 6).

Les flux de la figure 6 sont répartis en quatre classes de valeurs qui correspondent à celles du tableau 1, dans le cas d'une sélection des valeurs selon un critère global. Ils sont figurés par deux flèches, car les données correspondantes sont orientées. Cette précision n'est pas anodine, car le choix du figuré dépend du type de matrice de flux. Dans le cas de l'ex-Tchécoslovaquie (1990), la matrice est asymétrique et positive, ce qui signifie que les relations observées entre les districts sont bilatérales et les valeurs (F_{ij}) positives. Il s'agit d'un cas particulier de carte de flux puisque toutes les matrices sous-jacentes ne sont pas représentables par des flèches.



6. Cartographie des valeurs de flux interdistricts supérieures à 100.

Conclusion

La cartographie des matrices de flux correspond généralement à la représentation directe des données initiales, telles qu'elles ont été fournies par un organisme de collecte. Le résultat (figure 6) correspond à un premier niveau de cartographie descriptive, dont l'apparence est susceptible de varier en fonction de la manière dont les données ont été sélectionnées et représentées. Dans le cas de l'ex-Tchécoslovaquie (1990), le résultat peut aussi varier en raison du regroupement brutal des flux en classes de valeurs que nous avons réalisé, pour les besoins de la démonstration.

Cette carte, réalisée à partir des données observées, mélange des informations de nature différente qu'elle décrit sans les dissocier. C'est la raison pour laquelle elle présente une apparence compacte liée à un *effet d'amalgame*. Sur le plan sémiologique, cela se traduit par une figure peu lisible, même si elle est d'apparence claire. Pour y remédier, il convient d'appliquer un traitement aux données de flux au préalable à leur cartographie, a minima de sélectionner de manière plus robuste l'information, mais ce sont d'autres questions. Dans l'approche descriptive, la réponse à cet *effet d'amalgame* passe souvent par la production d'une multitude d'indicateurs variés (Holland, Plane, 2001), dont le choix n'est pas toujours aisé. Nous suggérons plutôt de décomposer la matrice de données initiales de manière à ne représenter, par exemple, que la partie dynamique du système de flux.

Bibliographie

- BAHOKEN F. (2014). « Pour un renouvellement de la sémiologie de la carte de flux ». *Revue Cartes et Géomatique*, Comité français de cartographie, vol. 4, n°222, p. 109-121.
- BARON M., ECKERT D., JEGOU L. (2011). « Peut-on démêler l'écheveau mondial des collaborations scientifiques ? » *M@ppemonde*, n°102 (2011-2), rubrique Internet. [En ligne : <http://mappemonde.mgm.fr/num30/internet/int11201.html>]
- BEGUIN M., PUMAIN D. (1994). *La Représentation des données géographiques. Statistique et cartographie*. Paris : Armand Colin, coll. « Cours Géographie », 192 p. ISBN : 2-200-21539-8
- BERTIN J. (1967). « La représentation du mouvement dans le plan ». In BERTIN J., *Sémiologie graphique*. Paris : Gauthier-Villars, p. 342-355.
- BERTIN J. (2005). *Sémiologie graphique : les diagrammes, les réseaux, les cartes*. Paris : Les réimpressions des éditions de l'EHESS, 4^e édition (1^{ère} éd. 1967), 452 p. ISBN : 978-2-713-22027-2
- BERTIN J. (1980). « Traitements graphiques et mathématiques. Différence fondamentale et complémentarité ». *Mathématiques et sciences humaines*, t. 72, p. 60-71.
- BONIN O., BAHOKEN F., MERMET S. (2013). « CARTOD/MAPOD : Visualiser la demande locale de transport ». *M@ppemonde*, n° 112 (2011-3), Rubrique Saint-Dié. [En ligne : <http://mappemonde.mgm.fr/num40/fig13/fig13401.html>]
- BOPDA A., GRASLAND C. (1994). « Migrations, régionalisation et régionalismes au Cameroun ». *Espace, Populations, Sociétés*, n°1, p. 109-129.
- BOPDA A., GRASLAND C., POULAIN M. (2000). « Évaluation comparative de l'influence des limites linguistiques sur les comportements migratoires : applications aux cas de la Belgique, du Cameroun et de la Tchécoslovaquie ». Dans *Régimes démographiques et territoires : les frontières en question*, Actes du colloque de l'AIDELF Les frontières en question (1988), La Rochelle, AIDELF, 20 p.
- BUNGE W. (1973). *Theoretical geography*. Lund: Gleerup, coll. « Lund Studies in Geography, Serie C: General and Mathematical Geography », n°1, 290 p. (2^e édition). ISBN : 91-40-02456-3
- BREUKELMAN J., BRINK G., DE JONG T., FLOOR H. (2009). « Selecting Flow Data ». In *Flowmap software v.3.2*. Utrecht : Utrecht University, coll. « Manual, Faculty of Geographical Sciences », p. 42-44.

- CATTAN N., GRASLAND C. (1994). « Migrations et effets de barrière en Tchécoslovaquie ». In REY V. (éd.), *La Tchécoslovaquie en 1992: transition, fragmentation, recomposition*. Fontenay-aux-Roses : Presses de l'ENS Fontenay-Saint-Cloud, coll. « Sociétés, Espaces, Temps », p. 97-120. ISBN : 2-902126-07-7
- CAUVIN C., ESCOBAR F., SERRADJ A. (2008a). *Cartographie thématique - Tome 1, Une nouvelle démarche*. Paris : Hermès Sciences Publications, coll. « Information Géographique, Traité IGAT », 284 p. ISBN : 978-2-7462-1535-1
- CAUVIN C., ESCOBAR F., SERRADJ A. (2008b). *Cartographie thématique - Tome 3, Méthodes quantitatives et transformations attributaires*. Paris : Hermès Sciences Publications, coll. « Information Géographique, Traité IGAT, Aspects fondamentaux de l'analyse spatiale », 282 p. ISBN : 978-2-7462-1884-0
- CAUVIN C., REYMOND H. (1991). « Interaction spatiale et cartographie : les solutions de W. Tobler ». *Espace, Populations, Sociétés*, n°3, p. 467-485.
- DORLING D.F.L. (1991). *The Visualization of Spatial Social Structure*, chap. 2 « People, Spaces and Places » et chap. 6 « Cobweb of Flows ». Newcastle upon Tyne : University of Newcastle upon Tyne, Phd. Thesis, 577 p.
- FAIRBAIRN D. (2006). « Measuring Map Complexity ». *Cartographic Journal*, vol. 43, n°3, p. 224 - 238.
- FRUCHTERMAN T., REINGOLD E. (1991). « Graph drawing by force-directed placement ». *Software Practice Experience*, vol. 21, n°11, p. 1129-1164.
- GRASLAND C. (1994). « Limites politiques et barrières migratoires : l'exemple de l'ex-Tchécoslovaquie (1965-1989) ». In GALLUSER W. (éd.), *Political Boundaries and Coexistence*. IGU Symposium, Basel, p. 425-440.
- HENRY N., FEKETE J.-D. (2008). « Représentations alternatives pour les réseaux sociaux ». *Réseaux*, n°152, p. 59-92.
- HOLLAND S.C., PLANE D.A. (2001). « Methods of mapping migration flow patterns ». *Southeastern Geographer*, vol. 41, n°1, p. 89-104.
- HOLTEN D. (2006). « Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data », *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, vol. 12, n°5, pp. 741-748.
- MAC EACHREN A.M. (1992). « Visualizing uncertain information ». *Cartographic Perspectives*, n°13, p. 10-19.
- MAC EACHREN A.M. (1995). « How Maps Work: Representation », *Visualization and Design*. New York, Londres : The Guilford Press, 513 p. ISBN : 0-89862-589-0
- PUMAIN D., SAINT-JULIEN T. (2001). « Les flux : déplacements et échanges, Chapitre 1 | Le changement dans l'espace, chapitre 4. In *Les Interactions spatiales*, Paris : Armand Colin, coll. « Cours Géographie », p. 134-181.
- TOBLER W.R. (1982). Cartographic study of movement tables, Presentation in the *National Computer Graphic Assembly - Session on Statistical Graphics - Mapping*, Anaheim, 17-06-1982.
- TOBLER W.R. (1987). « Experiments in migration mapping by computer », *American Cartographer*, n°14, pp. 155-163.
- TORRICELLI G.-P. (1997). « Navetteurs en Suisse et en Italie du Nord : image des flux et de leurs changements ». *Mappemonde*, n°4, p. 9-13.
- TUFTE E. (1986, 2001). *The Visual Display of Quantitative Information*. Cheshire : Graphics Press, 197 p. ISBN : 0-961-39214-2

TUKEY J.W. (1977). *Exploratory data analysis*. Reading : Addison-Wesley, 688 p.
ISBN : 0-201-07616-0
WARD M. (2002). « A Taxonomy of Glyph Placement Strategies for Multidimensional Data Visualization ». *Information Visualization*, p. 194-210.

Adresse de l'auteur

Françoise Bahoken, IFSTTAR/Département AME/SPLOTT, UMR Géographie-Cités/Équipe PARIS.
Courriel : francoise.bahoken@ifsttar.fr