View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Hal-Diderot

Modèle de radiation et modèle gravitaire : du formalisme à l'usage

Hadrien Commenges

To cite this version:

Hadrien Commenges. Modèle de radiation et modèle gravitaire : du formalisme à l'usage. Revue Internationale de Géomatique, Lavoisier, 2016, 26 (1), pp.79-95. <halshs-01319588v2>

HAL Id: halshs-01319588

https://halshs.archives-ouvertes.fr/halshs-01319588v2

Submitted on 23 May 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.

COMMENGES H. (2016) « Modèle de radiation et modèle gravitaire : du formalisme à l'usage », *Revue Internationale de Géomatique*, vol.26, n°1, pp.79-95.

Modèle de radiation et modèle gravitaire : du formalisme à l'usage

Hadrien Commenges
Laboratoire Eau, Environnement et Systèmes Urbains
UPEC, ENPC, AgroParisTech
LABEX Futurs Urbains
hadrien.commenges@leesu.enpc.fr

Résumé Simini, González, Maritan et Barabási ont récemment proposé un « modèle universel » appliqué à la mobilité et aux migrations, le modèle de radiation, proposition qui s'appuie sur une critique sévère du modèle gravitaire. En confrontant les deux modèles les auteurs posent plusieurs questions cruciales concernant les formalismes et les usages des modèles de prévision des flux : questions théoriques et méthodologiques sur la qualité d'un modèle de prévision mais aussi questions sur le succès d'une approche de modélisation, c'est-à-dire sur sa diffusion et son maintien. L'objectif de cet article est d'examiner le lien entre ces deux aspects : la qualité du modèle, qui tient *en première analyse* à son formalisme, et le succès du modèle, qui tient à son usage.

Abstract Simini, González, Maritan et Barabási recently proposed a "universal model for mobility and migration patterns" : the radiation model. They compare the radiation model to the gravity model and conclude that the latter suffers important theoretical inconsistencies and systematic predictive discrepancies. By confronting both models, they address crucial issues about formalisms and usages of spatial interaction models : theoretical and methodological issues about the quality of a predictive model, but also issues about the success of a modelling approach, that is, its diffusion and stability. This paper aims to highlight the link between both aspects : the quality of the model, which at first sight relies on its formalism, and the success of the model, which relies on its usage.

Mots-clés Modèle de radiation, modèle gravitaire, interaction spatiale, navettes

Keywords Radiation model, Gravity model, Spatial interaction, Commuting

1 Introduction

Dans un article récemment publié dans la revue *Nature*, Simini, González, Maritan et Barabási proposent un « modèle universel » appliqué à la mobilité et aux migrations : le modèle de radiation (Simini *et al.* 2012b). Les auteurs comparent ce modèle au modèle de gravité en termes de fondements théoriques et de résultats empiriques. Ils affirment que le modèle de gravité souffre de faiblesses théoriques importantes et qu'il produit des résultats empiriques médiocres. Ils concluent que « toute forme de loi gravitaire est intrinsèquement inappropriée pour décrire de façon satisfaisante les navettes domicile-travail » (Simini *et al.* 2012a, p. 9). Leur article présente le formalisme mathématique du modèle de radiation et le compare à celui du modèle gravitaire, il présente ensuite une comparaison de résultats obtenus avec les deux modèles sur des jeux de données variés : flux domicile-travail, migrations résidentielles, flux de marchandises et appels téléphoniques. Le jugement semble sans appel : le modèle de radiation est plus élégant dans son formalisme, il est systématiquement meilleur dans ses prévisions.

Curieusement, cet article a suscité peu de réactions chez les géographes, et cela malgré l'importance du modèle gravitaire dans leur pratique ; malgré le fait que le modèle de radiation, comme le modèle gravitaire, est importé de la physique ; et enfin, malgré le prestige des auteurs et de la revue qui l'a publié. En

annonçant un modèle « universel », non dépendant des contextes d'application (*context-free*), ces auteurs posent pourtant des questions fondamentales pour le géographe. En effet, le débat entre les approches nomothétiques et les approches idiographiques, présent dans plusieurs disciplines de sciences humaines et sociales, est particulièrement marqué en géographie. La géographie théorique et quantitative en diffusant son programme a amené l'idée que des phénomènes sociaux pouvaient faire l'objet d'une modélisation quantitative (Pumain et Robic 2002), à condition d'être calibrée pour bien reproduire l'observation. L'article de Simini *et al.* va plus loin puisqu'il propose un modèle sans paramètre, donc indépendant du contexte d'application.

Le domaine d'utilisation du modèle gravitaire s'est considérablement étendu au cours du 20e siècle, il fait aujourd'hui partie de la boîte à outils de plusieurs courants de la géographie, de la démographie et de l'économie. Au fur et à mesure de cette extension, il a subi des critiques nombreuses et variées : critiques externes, issues de chercheurs opposés aux approches modélisatrices (Fichelet *et al.* 1970) ; critiques internes, issues des modélisateurs eux-mêmes (Duffus *et al.* 1987). En l'attaquant une nouvelle fois, Simini *et al.* posent plusieurs questions cruciales concernant les formalismes et les usages des modèles de prévision des flux : questions théoriques et méthodologiques sur la qualité d'un modèle de prévision mais aussi questions sur le succès d'une approche de modélisation, c'est-à-dire sur sa diffusion et son maintien. L'objectif de cet article est d'examiner le lien entre ces deux aspects : la qualité du modèle, qui tient *en première analyse* à son formalisme mathématique, et le succès du modèle, qui tient à son usage. Ce lien ne va pas de soi, comment expliquer sinon le succès du modèle gravitaire malgré son apparente médiocrité.

Cet article avance en deux temps : il questionne d'abord le caractère universel du modèle de radiation en le confrontant à des données empiriques, à savoir les flux de navettes domicile-travail en Île-de-France de 1968 à 2010. Il aborde ensuite, dans une perspective historique, la question de l'usage des modèles en revenant aux origines des modèles de gravité et d'opportunités interposées. Les réflexions sur ces trois approches de modélisation – gravité, radiation et opportunités – conduisent à examiner le lien entre le formalisme d'un modèle et son usage. À travers un examen détaillé de l'usage des modèles d'interaction spatiale en contexte urbain, l'article pose des bases pour une compréhension plus générale de l'usage des modèles en sciences sociales.

2 Les modèles au prisme de leur formalisme et de leurs résultats

2.1 Face-à-face entre gravité et radiation

L'article de Simini *et al.* compare le modèle gravitaire et le modèle de radiation sur des jeux de données variés mais, pour illustrer le raisonnement, il prend toujours l'exemple des navetteurs. Le même choix est fait ici : le raisonnement est illustré par des exemples portant sur les navettes domicile-travail. Après une présentation de la grande variété des domaines d'application du modèle gravitaire, les auteurs présentent son formalisme de la façon suivante :

$$T_{ij} = \frac{m_i^{\alpha} n_j^{\beta}}{f(r_{ij})} \tag{1}$$

Le nombre de déplacements entre i et j (T_{ij}) dépend de la population à l'origine (m_i), de la population à destination (n_j) et d'une fonction d'impédance ($f(r_{ij})$) qui représente la friction entre i et j et peut prendre en compte la distance, le temps et/ou le coût. Ils énumèrent ensuite ce qu'ils considèrent être les six limitations principales du modèle gravitaire :

1. La loi gravitaire ne peut pas être dérivée.
2. Il n'y a pas de fondements théoriques pour paramétrer le modèle.
3. Le modèle doit être alimenté par des données variées.

4. La qualité des prévisions est médiocre.
5. Le nombre estimé de navetteurs n'est pas limité : si la population de j est grande, le nombre de navetteurs partant de la zone i peut être supérieur à sa population.
6. La loi gravitaire est déterministe.

Le modèle proposé est formulé en termes de radiation et d'absorption, il ne nécessite que quatre variables : le total de navetteurs à l'origine (T_i), la population de i (m_i), la population de j (n_j) et enfin la population totale dans le cercle de centre i et de rayon égal à la distance entre i et j (s_{ij}).

$$T_{ij} = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \tag{2}$$

Ce modèle résoudrait les six limitations et incohérences du modèle gravitaire mais les auteurs insistent particulièrement sur deux d'entre elles : la limitation (2) et la limitation (3). À la différence du modèle gravitaire, le modèle de radiation ne nécessite que peu de données en entrée, la population des zones et les flux à l'origine. Le modèle de radiation n'a pas de paramètres à calibrer, il est *parameter-free*. Du point de vue des résultats, le modèle de radiation produit des estimations qui sont systématiquement meilleures que le modèle gravitaire, ce que montre la suite de l'article sur quatre jeux de données différents.

2.2 Le modèle de radiation en contexte urbain

La proposition de Simini *et al.* a provoqué peu de réactions critiques, les deux principales étant un article de Masucci *et al.* (2013) et un article de Lenormand *et al.* (2012). Masucci *et al.* confrontent le modèle de radiation et le modèle gravitaire sur différents jeux de donnés. En travaillant sur le périmètre du Grand Londres, les auteurs mettent en évidence un biais de comparaison dans le travail initial : le résultat de la comparaison dépend étroitement de l'échelle spatiale de modélisation.

À l'échelle d'une aire urbaine, les deux modèles peinent à estimer correctement les flux. Le modèle de radiation est meilleur que le modèle gravitaire pour estimer des flux lointains vers des zones peu peuplées alors que le modèle gravitaire est meilleur pour estimer des flux proches vers des zones très peuplées. Ce premier point explique en partie pourquoi les jeux de données testés par Simini *et al.* donnent toujours l'avantage au modèle de radiation : il s'agit de flux à l'échelle d'un pays, les États-Unis pour les navettes, pour les migrations résidentielles et pour les flux de marchandises et un « pays d'Europe occidentale » pour les appels téléphoniques.

Masucci *et al.* discutent ainsi de l'apport que représente le modèle de radiation en le testant à des échelles diverses. Le présent travail teste le modèle de radiation sur un seul espace d'étude mais à différentes époques. Il s'appuie sur les navettes domicile-travail entre les communes de la région Île-de-France [1] de 1968 à 2010 et mobilise six éditions du recensement : 1968, 1975, 1982, 1990, 1999 et 2010. Le modèle de radiation est calculé pour chacune de ces dates et six matrices de flux estimés sont produites.

Les résultats obtenus sont confrontés à deux autres modèles : un modèle aléatoire et un modèle gravitaire. Le modèle aléatoire est construit de façon à respecter la distribution des flux observés mais en les redistribuant aléatoirement dans la matrice origine-destination. Le modèle gravitaire utilisé est de la forme suivante :

$$T_{ij} = k \frac{E_i A_j}{d_{ij}^2} \tag{3}$$

Le flux entre i et j est proportionnel à la masse de flux émis par i (E_i) et à la masse de flux attirés par j (A_j) et inversement proportionnel à la distance qui sépare i et j (d_{ij}). Le modèle est calibré par un simple

1. Plus précisément, il s'agit des flux ayant leur origine et leur destination en Île-de-France. Les données sont issues des exploitations complémentaires du recensement (sondage au quart) mises à disposition par le Centre Maurice Halbwachs (Réseau Quetelet). Les modifications de périmètres de certaines communes sont prises en compte en restituant les flux dans les périmètres communaux de 2010.

paramètre k, c'est-à-dire qu'il n'y a pas de contrainte aux marges. La fonction d'impédance n'est pas non plus calibrée, c'est une fonction puissance avec un exposant égal à 2. Ce modèle de gravité est donc simple et probablement mauvais puisqu'il ne bénéficie de calibrage ni pour les sommes marginales de flux, ni pour le type de fonction de friction (loi puissance), ni pour le paramétrage de cette fonction (exposant égal à 2).

Pour mesurer la qualité de l'estimation des flux par ces trois modèles - radiation, gravité, aléatoire - la matrice de flux prédite est comparée à la matrice de flux observée. La mesure de dissimilarité la plus simple est le pourcentage d'erreur absolue (APE - *absolute percentage error*), calculé entre les valeurs observées (O_ij) et les valeurs espérées (E_{ij}) :

$$APE_{ij} = \left| \frac{O_{ij} - E_{ij}}{O_{ij}} \right| \times 100 \tag{4}$$

Cette mesure présente un inconvénient, elle n'est bornée que d'un côté : si l'observé (O_{ij}) est largement supérieur à l'espéré (E_{ij}), l'erreur maximale est de 100 %, mais si c'est l'inverse l'erreur maximale n'est pas bornée. Cette mesure a fait l'objet de plusieurs reformulations pour pallier ce problème (Hyndman et Koehler 2006). La mesure utilisée ici est le pourcentage d'erreur absolue bornée (CAPE - *Censored Absolute Percentage Error*) proposé par Cools *et al.* (2010).

$$CAPE_{ij} = \min(100, \left| \frac{O_{ij} - E_{ij}}{O_{ij}} \right| \times 100) \tag{5}$$

Le CAPE est une mesure locale de l'erreur, calculée pour tous les couples origine-destination, mais il est possible de calculer à partir de la matrice de pourcentages d'erreur des résumés numériques (cf. figure 1). Deux types de résumés numériques sont utilisés par la suite : la moyenne des pourcentages d'erreur, mesure globale caractérisant l'ensemble de la matrice ; la moyenne marginale des pourcentages d'erreur, mesure caractérisant l'erreur à l'origine (moyenne pour tous les j) ou l'erreur à destination (moyenne pour tous les i).

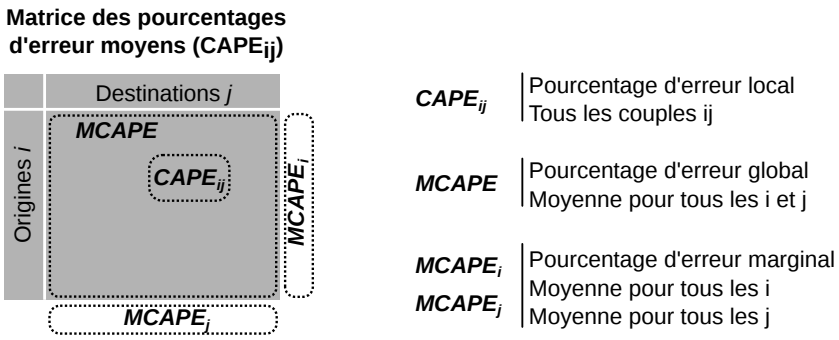

FIGURE 1 – Mesures locales et globales des pourcentages d'erreur

La moyenne des pourcentages d'erreur absolue bornée (MCAPE), mesure assez proche du cœfficient de Sørensen-Dice utilisé par Masucci *et al.* (2013), quantifie bien la qualité globale de l'estimation des modèles d'interaction spatiale. Cependant, lorsqu'elle est calculée sur l'ensemble de la matrice de flux, elle surestime largement cette qualité du fait que la majorité des cellules sont vides. En effet, cette matrice est constituée d'environ 1,7 million de couples de communes (1 300 par 1 300) et il n'y a aucun navetteur dans 90 à 95 % des cas. Par exemple, sur la matrice de flux du recensement 2010, on observe 128 000 flux non nuls sur les 1,7 million de couples possibles, ce qui représente 7,6 % du total. La figure 2 montre ainsi que les trois modèles, y compris le modèle aléatoire, ont un pourcentage d'erreur très faible lorsque celui-ci est calculé avec les cellules vides. L'évolution de cette courbe au cours du temps, en particulier le pic de 1999, n'a rien à voir avec la qualité des modèles. Elle dépend uniquement de la masse de cellules non vides, masse bien supérieure en 1999 du fait que ce tableau est issu de l'exploitation principale du recensement (exhaustif) et non de l'exploitation complémentaire (sondage au quart) comme le reste des tableaux.

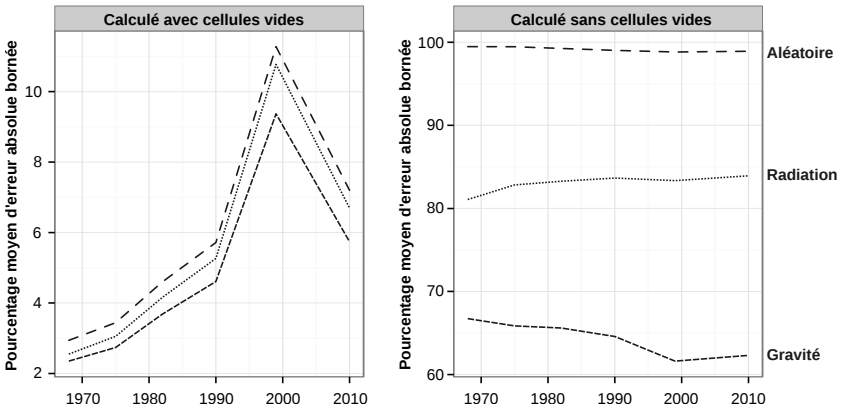

FIGURE 2 – Évolution de la qualité d'estimation des trois modèles (1968-2010)

Ainsi, l'augmentation globale, entre 1968 et 2010, du pourcentage d'erreur calculé avec cellules vides ne signifie pas que les modèles utilisés sont de plus en plus mauvais pour estimer les flux. Cette augmentation indique un remplissage progressif de la matrice de flux intercommunaux : on observe au cours du temps un nombre croissant de couples origine-destination distincts, c'est-à-dire une diversification globale des flux observés en Île-de-France.

Le MCAPE calculé sans cellules vides élimine ce biais [2], c'est donc cet indicateur qui doit être pris en compte. Le premier constat concernant la qualité de l'estimation des flux par les trois modèles touche à son importance : le modèle aléatoire est logiquement très mauvais, avec un pourcentage d'erreur proche de 100 % ; le modèle de radiation affiche un pourcentage qui oscille entre 82 et 85 % ; le modèle gravitaire est nettement meilleur, avec des valeurs comprises entre 63 et 67 %. Le second constat porte sur l'évolution du pourcentage d'erreur du modèle de radiation au cours du temps, pourcentage qui augmente continuellement entre 1968 et 2010. Pour expliquer cette tendance, il faut spatialiser la mesure de qualité : où les estimations du modèle sont-elles les plus mauvaises ?

Le différentiel de qualité des estimations du modèle de radiation est cartographié sur les données du recensement de 1975. Pour chaque commune, la moyenne du MCAPE est calculée sur les flux à destination (pourcentage d'erreur marginal, cf. figure 1). Cette variable, dont la distribution est fortement dissymétrique à droite, est discrétisée en quartiles. Pour mieux comprendre la répartition de ce pourcentage d'erreur, cette carte est accompagnée d'une régression linéaire simple entre la population résidente dans les communes et le nombre d'emplois localisés (cf. figure 3). Le nuage de points reprend la même variation de couleur que la carte issue de la discrétisation en quartiles du MCAPE.

La lecture de la carte et du graphique montre à quel point le modèle de radiation peine à estimer les flux en direction des arrondissements parisiens. Plus généralement, ses estimations sont particulièrement mauvaises lorsque la relation entre population résidente et emplois localisés est déséquilibrée. C'est le cas pour les 1^{r}, 2e, 8e et 9e arrondissements. C'est également le cas pour les communes peu peuplées qui accueillent des équipements exceptionnels, et donc un très grand nombre d'emplois : Orly, Roissy-en-France et Rungis par exemple. Ce résultat n'est pas surprenant puisque le modèle de radiation ne prend en compte que les populations résidentes. Il ne permet pas de saisir le rapport, particulièrement complexe dans une grande aire urbaine, entre la répartition des résidences et la répartition des emplois.

2. Dans ce calcul, la moyenne des pourcentages d'erreur est calculée après suppression des cellules où le flux observé est égal à 0, mais plus largement lorsque le flux est strictement inférieur à 40. Sur un tableau issu d'un sondage au quart, ceci revient à considérer comme non représentatifs les flux qui concernent moins de 10 individus.

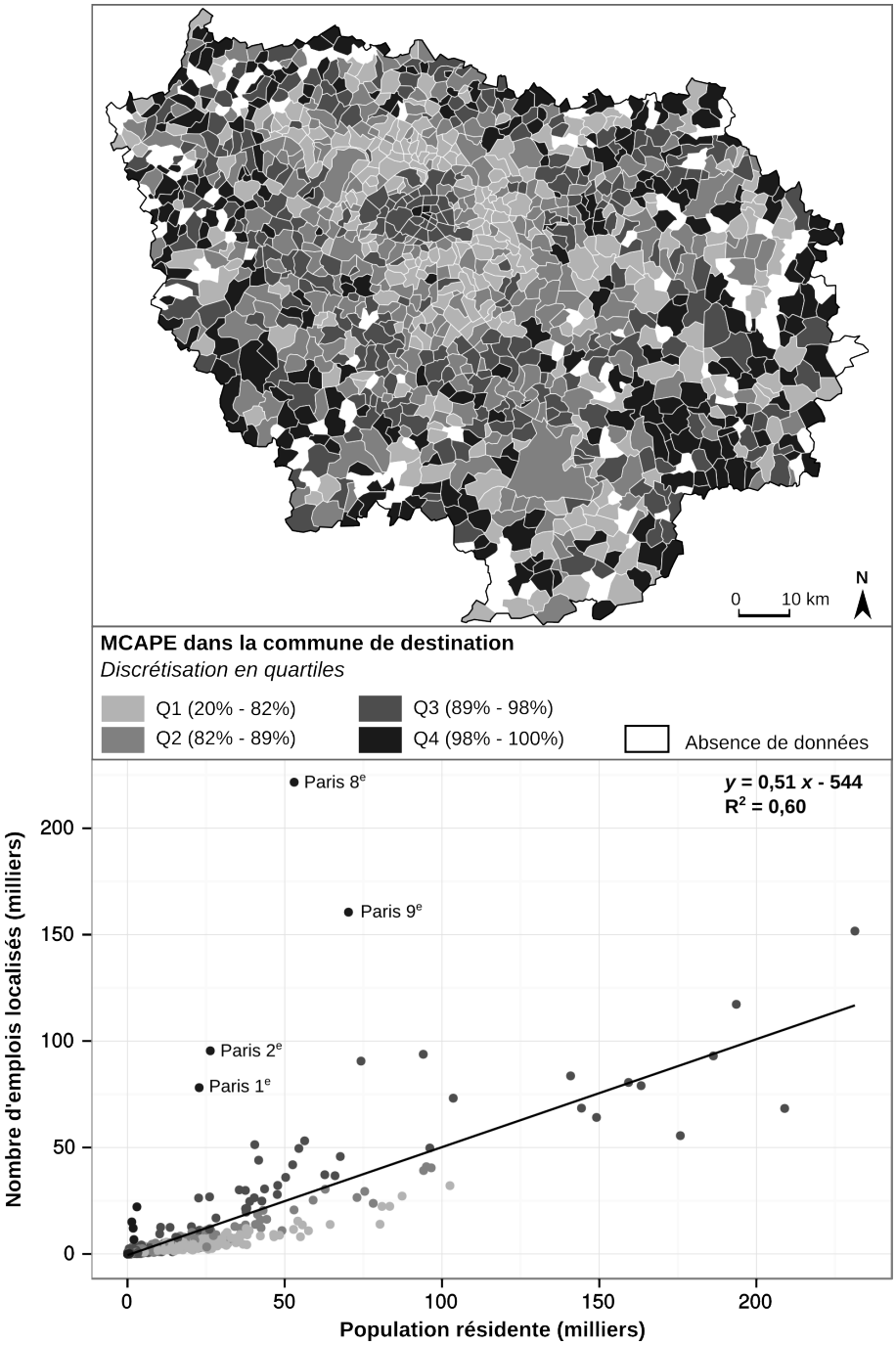

FIGURE 3 – Spatialisation de la mesure de qualité d'estimation (RP 1975)

Il est important de noter que la corrélation entre le nombre de résidents et le nombre d'emplois dépend étroitement de la finesse du maillage. À titre d'exemple, une régression linéaire sur ces deux variables au niveau départemental affiche un cœfficient de détermination supérieur à 0,9 (contre 0,6 sur la figure 3) avec très peu de valeurs exceptionnelles. C'est en partie pour cette raison que les résultats obtenus par Simini *et al.* sur les *counties* américains sont meilleurs que ceux obtenus ici sur un maillage communal. Les auteurs de l'article insistent beaucoup sur la diversité des données auxquelles s'applique leur modèle – flux de personnes, de marchandises, appels téléphoniques – mais ils ne font en aucun cas varier la taille du maillage. Le modèle de radiation est présenté comme un modèle universel et *context-free* mais il est en fait très affecté par la variation de taille du maillage (*modifiable areal unit problem*). Cette variation de taille peut affecter à la fois la corrélation entre deux variables mesurées dans les mailles (Openshaw et Taylor 1979) mais aussi la distribution des flux entre ces mailles (Courgeau 1973).

La qualité de l'estimation du modèle de radiation dépend de cette corrélation résidents-emplois, or celle-ci varie au cours du temps. Dans le contexte francilien, la seconde moitié du 20e siècle est caractérisée par une dispersion croissante des couples domicile-travail. Ce phénomène peut être quantifié par plusieurs indicateurs caractérisant les communes, par exemple le pourcentage d'actifs occupés travaillant dans leur commune de résidence (qualifié ici d'*autocontention*) et la part des emplois communaux occupés par des résidents de la commune (*autosuffisance*). En Île-de-France, ces deux indicateurs décroissent fortement entre 1968 et 2010, passant respectivement de 50 à 18 % et de 70 à 42 %.

Un autre aspect attire l'attention sur la figure 3 : les communes en bordure de l'espace d'étude affichent une erreur élevée, nombreuses sont celles qui se situent dans le quatrième quartile du pourcentage d'erreur. En effet, le modèle a été calculé ici en ne prenant en compte que les communes de l'espace d'étude. Or, pour estimer un flux entre deux communes i et j sans être affecté par un effet de bord, le modèle de radiation devrait prendre en compte toutes les unités spatiales comprises dans le cercle de rayon ij. L'exemple de la figure 4 montre que, dans le cas d'un flux entre deux communes situées aux limites opposées de la région Île-de-France, il faudrait disposer d'une information sur les communes des 8 départements de la région mais aussi sur les communes de 14 départements supplémentaires dont certains ne sont pas même limitrophes de l'espace d'étude. De façon plus générale, pour appliquer le modèle de radiation sur un espace d'étude circulaire de rayon r, il faut disposer de données sur un espace circulaire de même centre et de rayon égal à $3r$, espace qui aurait donc une surface neuf fois supérieure à celle de l'espace d'étude.

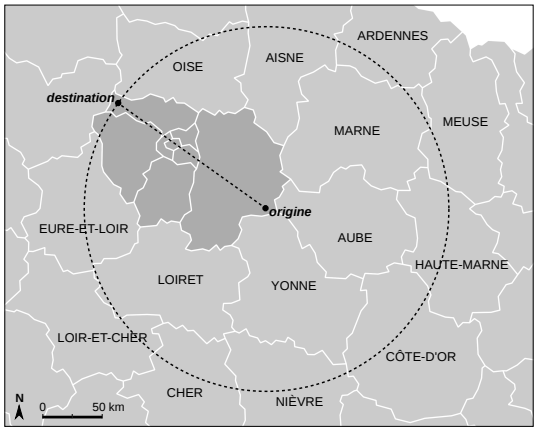

FIGURE 4 – Exemple de périmètre nécessaire pour l'estimation du modèle de radiation

En guise de conclusion provisoire, la qualité des prédictions du modèle de radiation dépend étroitement de la corrélation entre population et emplois, corrélation variable dans l'espace, ce qu'ont montré Masucci *et al.* (2013), corrélation variable selon la finesse du maillage des unités spatiales, corrélation variable dans le temps pour un même espace d'étude, ce qui vient d'être montré. En outre, l'avantage annoncé par Simini *et al.* concernant le peu de données nécessaires pour le modèle de radiation est contrebalancé par le besoin de disposer de ces données sur un espace beaucoup plus vaste que l'espace d'étude lui-même.

Le modèle gravitaire ne souffre pas des mêmes inconvénients : d'une part, il prend en compte la distribution spatiale des résidents et des emplois et ne dépend donc pas d'une corrélation hypothétique entre ces deux

variables ; d'autre part, il ne nécessite pas d'information sur des unités spatiales hors de l'espace d'étude. La comparaison qui vient d'être effectuée doit maintenant faire l'objet d'une mise en perspective historique mettant l'accent sur l'usage qui est fait des modèles de prévision des flux en milieu urbain.

3 Les modèles au prisme de leur usage

3.1 À l'origine des modèles d'interaction spatiale

Le modèle de radiation rappelle fortement le modèle d'opportunités interposées (*intervening opportunities*) proposé par Stouffer en 1940 pour expliquer la distribution des déplacements inter-étatiques aux États-Unis (Stouffer 1940; 1960). Le principe des opportunités interposées est exposé de la façon suivante (1940, p. 846) : « le nombre d'individus se déplaçant à une distance donnée est proportionnel au nombre d'opportunités offertes à cette distance et inversement proportionnel au nombre d'opportunités interposées », ce qui s'écrit :

$$T_{ij} = k_i O_i \left[\exp(-\alpha x_{j-1}) - \exp(-\alpha x_j)\right] \tag{6}$$

Le flux entre i et j (T_{ij}) est proportionnel à la masse de migrants originaires de i (O_i) ainsi qu'au différentiel entre les opportunités offertes à destination (x_j) et les opportunités qui s'interposent entre i et j (x_{j-1}). Le paramètre α est une probabilité constante qu'une opportunité soit satisfaisante et k_i est un paramètre d'ajustement qui assure l'égalité des flux émis et attirés [3]. Si la formulation mathématique du modèle de Stouffer diffère de celle du modèle de radiation, le principe est semblable : entre deux unités spatiales i et j, le flux dépend de la masse de migrants originaires de i (*radiation process* chez Simini *et al.*) et du degré d'attraction de l'espace qu'il faut traverser pour se rendre de i à j (*absorption process*). Il y a également une différence entre les deux modèles dans la façon de quantifier la capacité d'absorption de l'espace interposé : chez Stouffer (1960), le nombre d'opportunités interposées entre i et j se comptabilise à l'intérieur d'un cercle de diamètre ij alors que, dans le modèle de radiation, ce nombre d'opportunités est comptabilisé à l'intérieur d'un cercle de centre i et de rayon ij.

L'apparition du modèle de Stouffer doit être replacée dans le contexte américain des années 1940, qui voit fleurir un grand nombre de modèles d'interaction spatiale. Dès le début des années 1940, Stewart (1942) pose les bases des modèles de potentiel gravitaire, plusieurs auteurs formulent des modèles d'interaction spatiale de type gravitaire : Zipf (1946), Carroll (1949, 1952), Dodd (1950), Ikle (1954) ou encore Casey (1955). Le fameux principe de Tobler (1970, p. 236), « everything is related to everything else, but near things are more related than distant things », est issu de ce foisonnement de modèles d'interaction spatiale. Tous ces modèles portent ce principe, malgré la diversité de leurs formalismes mathématiques.

C'est au sein de ce milieu « techno-scientifique » (Hottois 1984) que les modèles commencent à être appliqués massivement sur des données empiriques, mobilisant les premiers ordinateurs scientifiques commercialisés par IBM. Les articles fondateurs des deux modèles, celui de Ravenstein (1885) pour le modèle gravitaire et celui de Stouffer pour le modèle d'opportunités, portent sur les migrations résidentielles à une échelle interurbaine. Cependant, leur application massive et informatisée a lieu dans le domaine de l'ingeniérie du trafic américaine, qui vise à estimer des flux dans l'espace urbain pour planifier des infrastructures de transport. C'est donc dans ce domaine, ingéniérie du trafic puis géographie et économie des transports qu'il faut rechercher les modalités de compétition entre modèles.

Dans ce domaine, les deux modèles, d'opportunités et de gravité, s'intègrent dans une *chaîne de modélisation* plus vaste, appelée « modèle à quatre étapes » (*four-step model*) (Bates 2000, McNally 2000). Le premier module, dit de génération (*trip generation*), consiste à évaluer une quantité de déplacements attirés et émis par les zones constituant l'espace d'étude. Le second module, dit de distribution (*trip distribution*),

3. La formulation du modèle a changé entre les deux articles fondateurs de Stouffer et les reformulations postérieures. L'équation présentée ici est issue de la généralisation proposée par Wilson (1970).

consiste à évaluer la répartition géographique de ces flux, par un modèle de gravité ou d'opportunités interposées. À ce bloc originel de deux modules se sont ajoutés en aval deux modules supplémentaires : l'un de choix modal, l'autre d'affectation au réseau. Cet ensemble constitue le modèle à quatre étapes, utilisé aux États-Unis depuis le milieu des années 1950 et en France depuis le milieu des années 1960 (Dupuy 1975, Chatzis 2009, Commenges 2013a). Malgré les nombreuses innovations dans le domaine de la modélisation des déplacements, le modèle à quatre étapes est encore utilisé aujourd'hui dans de nombreux pays, dont la France et les États-Unis (McNally 2000, Nguyen-Luong 2001).

Le modèle gravitaire et le modèle d'opportunités interposées sont utilisés pour distribuer les flux dès les premiers modèles de transport mis en place aux États-Unis et, par extension, dans le monde. Les deux premières études globales de transport américaines, toutes deux dirigées par Carroll, ont lieu à Détroit et Chicago au milieu des années 1950. L'étude de Détroit utilise un modèle gravitaire, celle de Chicago utilise un modèle d'opportunités interposées. À la suite de ces deux études fondatrices, le modèle gravitaire a essaimé dans toutes les études de transport et le modèle d'opportunités interposées n'a presque plus été utilisé dans ce cadre.

Comment expliquer que l'un des deux modèles ait complètement effacé l'autre alors que, du point de vue de la qualité des estimations, il n'y a pas de supériorité marquée de l'un ni de l'autre ? Willumsen et Ortúzar (2011) invoquent quatre raisons pour expliquer ce fait : (1) ses bases théoriques sont moins bien connues que celles du modèle gravitaire et il est plus difficile à comprendre pour les techniciens. (2) Il est plus difficile de mettre en place des matrices dont les destinations sont triées en fonction de la distance pour chaque origine. (3) Les estimation du modèle ne sont pas meilleures que celles du modèle gravitaire. (4) Il n'y a pas de logiciels implémentant le modèle d'opportunités.

L'argument (1) est contestable. De fait, c'est précisément l'argument contraire que donne Stouffer en 1940 : il considère que peu d'efforts ont été faits pour expliquer comment opère la distance et qu'il ne suffit pas de constater que cette dernière exerce un frein aux déplacements. Le modèle gravitaire, et plus généralement le principe d'interaction de Tobler, ne disent pas en quoi la probabilité qu'un individu se déplace entre deux lieux dépend de la distance qui les séparent. Le principe de Stouffer est beaucoup plus concret de ce point de vue : plus la distance est grande entre une origine et une destination, plus l'individu a de raisons (opportunités) de s'arrêter avant.

L'argument (2) est exact, le modèle d'opportunités est plus complexe à mettre en place, surtout dans un contexte où la puissance de calcul informatique est limitée. En effet, pour estimer un flux entre deux lieux i et j, le modèle gravitaire nécessite seulement trois valeurs (masse de i, masse de j, distance ij). Le modèle d'opportunités interposées nécessite les masses des unités spatiales d'origine et de destination, mais aussi l'ensemble des masses de toutes les unités spatiales situées dans le cercle de rayon ij, masses qui doivent de plus être ordonnées par la distance. Cependant, le modèle de Stouffer a pu être implémenté dès la fin des années 1950, cette difficulté technique n'est donc pas insurmontable.

L'argument (3) n'explique rien, les estimations du modèle d'opportunités ne sont pas meilleures que celles du modèle gravitaire, mais elles ne sont pas plus mauvaises. L'argument (4) est à la fois une cause et une conséquence, il n'y a pas de logiciels parce qu'il n'y a pas d'utilisateurs. Ce dernier point mériterait une discussion plus approfondie sur l'usage des modèles et sur les liens entre modèle et logiciel, mais cette discussion dépasse l'ambition de cet article.

La conclusion pour le moment est que la raison du succès du modèle gravitaire et de l'oubli du modèle d'opportunités doit être recherché ailleurs. La comparaison de deux modèles concurrents ne peut se focaliser, comme dans l'article de Simini *et al.*, que sur les formalismes, faisant fi des pratiques des modélisateurs.

3.2 Le modèle gravitaire et la production de connaissances

Revenons aux six limitations du modèle gravitaire annoncées par Simini *et al.* : à l'exclusion de la limitation (4) qui est un jugement de valeur contestable et contesté, toutes sont présentées par ces auteurs comme touchant au formalisme mathématique du modèle. En réalité, seules les limitations (1) et (6) se réfèrent strictement au formalisme mathématique de la loi de gravité, les limitations (2), (3) et (5) concernent tout autant l'usage du modèle que son formalisme. Le jugement sans appel qui émane de l'article à l'encontre

du modèle gravitaire vient principalement du fait que les auteurs le considèrent comme une méthode isolée, déconnectée de l'ensemble des pratiques des modélisateurs dans lequel elle s'intègre. Or, le succès du modèle gravitaire comme outil de prévision des flux dans un contexte urbain s'explique précisément par cette intégration au sein d'une « matrice technique » composée d'un ensemble de dispositifs et d'usages (Commenges 2013b).

La modélisation à quatre étapes est une *chaîne* de modélisation : chaque étape produit une sortie qui alimente l'étape suivante. La première étape produit une estimation des flux émis et attirés par chaque zone de l'espace d'étude, elle produit les marges de la matrice de flux. La deuxième étape distribue ces flux entre les zones, elle remplit la matrice de flux.

La génération des déplacements est modélisée le plus souvent par une régression linéaire qui prend en compte les caractéristiques de la population et des activités. Le nombre de déplacements émis par une zone dépend des caractéristiques démographiques des ménages, de leur revenu et du nombre de véhicules disponibles. Le nombre de déplacements attirés par une zone est fonction des activités qui s'y trouvent et de leurs caractéristiques. Le zonage de l'espace d'étude est établi dans le but de maximiser l'homogénéité des zones en termes de génération et d'attraction.

Les variables du modèle gravitaire ne sont donc plus les populations des zones, mais les variables d'émission et d'attraction calculées dans l'étape de génération :

$$F_{ij} = \alpha_i \beta_j O_i D_j f(C_{ij}) \tag{7}$$

Où α_i et β_j sont les facteurs de balancement qui assurent la contrainte aux marges, O_i et D_j sont les valeurs d'émission (total à l'origine) et d'attraction (total à destination) calculées à l'étape de génération, et $f(C_{ij})$ la fonction d'impédance ou de friction entre les unités spatiales i et j.

Voici le modèle gravitaire, tel qu'il s'insère dans les pratiques de modélisation en contexte urbain. Ce modèle a très peu été utilisé en utilisant comme masses les populations des unités spatiales (Ortúzar et Willumsen 2011). Dans l'usage qu'en font les modélisateurs, l'étape de génération condense une grande variété d'information sur l'usage du sol, sur la population et ses pratiques de mobilité, puis l'étape de distribution capitalise cette information pour estimer les flux.

C'est précisément ce lien qui explique le succès du modèle gravitaire face au modèle d'opportunités interposées. Le modèle gravitaire permet d'abord de ré-utiliser facilement les sorties de l'étape de génération des déplacements, ce qui serait également possible dans le modèle d'opportunités mais non dans le modèle de radiation. Il permet aussi de jouer sur la friction de la distance (du temps ou du coût) pour ajuster le modèle. La différence principale entre le modèle gravitaire et le modèle d'opportunités interposées est que le premier travaille avec une distance cardinale, qui peut être quantifiée, alors que l'autre travaille avec une distance ordinale, qui ne fait qu'ordonner les destinations possibles. Le modèle gravitaire permet ainsi de choisir une fonction d'impédance, fonction puissance ou exponentielle, mais aussi de calibrer cette fonction, par exemple en faisant varier l'exposant pour une fonction puissance.

Ce paramétrage, considéré comme une faiblesse par Simini *et al.*, est une façon de nourrir le modèle avec une information spécifique qui rend compte des pratiques de mobilité. En s'appuyant sur des données empiriques, il est facile de montrer que tous les motifs de déplacements ne sont pas égaux vis-à-vis de la distance. Par exemple, l'effort consenti pour se rendre vers un emploi est en moyenne supérieur à l'effort consenti pour faire des achats quotidiens. En d'autres termes, la friction de la distance est plus faible dans le premier cas que dans le second. Cette friction varie en fonction de plusieurs variables, comme le motif de déplacement mais aussi l'espace d'étude. C'est pourquoi les paramètres des modèles américains n'ont pas pu être ré-utilisés directement dans les premiers modèles réalisés en France (SERC 1964).

En outre, le fait de travailler avec une distance cardinale et de pouvoir calibrer le frein de la distance donne un cadre consistant avec les connaissances acquises sur l'organisation des flux dans l'espace urbain. En effet, le phénomène de l'allongement des distances parcourues en milieu urbain est documenté depuis le début des années 1970 (Koenig 1974, Zahavi 1974) : l'augmentation de la vitesse sur les réseaux de transport, c'est-à-dire la réduction du frein de la distance, se traduit non par un gain de temps mais par un

allongement des distances. Le modèle gravitaire prévoit cet allongement. En revanche, la formalisation du rôle de la distance dans le modèle de radiation, comme dans le modèle d'opportunités interposées, interdit de penser la distance comme un frein qui s'estompe avec l'amélioration des réseaux de transport.

Ainsi, calibrer un modèle gravitaire n'est pas un simple problème d'ajustement : l'objectif du calibrage n'est pas seulement d'ajuster au mieux la prévision du modèle mais aussi d'enrichir l'information obtenue par les données empiriques sur les pratiques de mobilité. Simini *et al.* insistent particulièrement sur ces deux limitations du modèle gravitaire : il a besoin d'être alimenté par une grande variété de données socio-économiques et il comprend plusieurs paramètres qui doivent être calibrés pour chaque contexte d'application. En proposant un modèle « universel » comprenant peu de données et aucun paramètre, ces auteurs brisent le lien qui fait la force de la chaîne de modélisation à quatre étapes : lien entre l'information produite par le modèle et l'information produite par ailleurs pour le nourrir. Le modèle ne peut être considéré comme un objet isolé dont l'unique objectif est de produire une estimation, il doit être considéré au sein d'un ensemble de dispositifs, en tant qu'il participe à produire de la connaissance sur les pratiques de mobilité et sur l'organisation urbaine.

Conclusion

À partir d'une critique du modèle gravitaire et d'une comparaison des résultats obtenus avec ce modèle et avec le modèle de radiation, Simini *et al.* formulent un jugement sans appel : « toute forme de loi gravitaire est intrinsèquement inappropriée pour décrire de façon satisfaisante les navettes domicile-travail » (Simini *et al.* 2012a, p. 9). Masucci *et al.* montrent que la comparaison est biaisée par le choix de jeux de données à une échelle nationale et que le modèle de radiation n'est pas apte à estimer des flux au sein d'une aire urbaine. Le présent travail poursuit cette comparaison sur une série historique de navettes et appuie les résultats obtenus par Masucci *et al.*.

Les deux articles cités comparent les modèles en termes de formalisme mathématique et de résultats empiriques, mais ils font peu de cas de l'usage qui en est fait. Considérant le modèle comme un dispositif ou objet technique, il est possible de lui appliquer un principe fondateur de l'anthropologie des techniques, « l'objet technique ne peut pas plus être confondu avec un dispositif matériel qu'avec l'ensemble des usages remplis par ce dispositif : il se définit très exactement comme le rapport construit entre ces deux termes » (Akrich 1987, p. 206). Un modèle ne peut être considéré comme un objet isolé dont le formalisme mathématique est suffisant pour le décrire. Le présent travail montre que le rapport entre le dispositif technique et son usage doit être pris en compte pour comprendre un modèle et pour juger de sa qualité.

Le modèle de radiation peut estimer des flux sans information préalable sur le contexte d'application du modèle, c'est-à-dire avec peu de données d'entrée et sans calibrage de paramètres. Le modèle gravitaire permet, par son usage au sein d'une approche de modélisation plus globale, de capitaliser une information existante sur la mobilité, information spécifique au contexte d'application, et d'enrichir les connaissances sur l'organisation urbaine.

Références

Akrich, M. (1987). Comment décrire les objets techniques ? *Techniques et cultures*, 9:49–64.

Bates, J. (2000). History of demand modelling. *In* Hensher, D. A. et Button, K. J., éditeurs : *Handbook of transport modelling*. Pergamon, Amsterdam.

Carroll, J. D. (1949). Some aspects of the home : work relationships of industrial worker. *Land Economics*, 25(4):414–422.

Carroll, J. D. (1952). The relation of home to work places and the spatial pattern of cities. *Social Forces*, 30(3):271–282.

Casey, H. J. (1955). Applications to traffic engineering of the law of retail gravitation. *Traffic Quarterly*, 9(1):23–35.

Chatzis, K. (2009). De l'importation des savoirs américains à la création d'une expertise nationale : la modélisation des déplacements urbains en france, 1950-1975. *In* Flonneau, M. et Guigueno, V., éditeurs : *De l'histoire des transports à l'histoire de la mobilité ?* Presses Universitaires de Rennes, Rennes.

Commenges, H. (2013a). *L'invention de la mobilité quotidienne. Aspects performatifs des instruments de la socio-économie des transports.* Thèse de doctorat, Paris.

Commenges, H. (2013b). Socio-économie des transports : une lecture conjointe des instruments et des concepts. *Cybergeo : European Journal of Geography*, Article 633:1–20.

Cools, M., Moons, E. et Wets, G. (2010). Assessing the quality of origin-destination matrices derived from activity travel surveys. *Transportation Research Record*, 2183:49–59.

Courgeau, D. (1973). Migrations et découpages du territoire. *Population*, 28:511–537.

Dodd, S. C. (1950). The interactance hypothesis : a gravity model fitting physical masses and human groups. *American Sociological Review*, 15(2):245–256.

Duffus, L. N., Alfa, A. S. et Soliman, A. H. (1987). The reliability of using the gravity model for forecasting trip distribution. *Transportation*, 14:175–192.

Dupuy, G. (1975). *Une technique de planification au service de l'automobile. Les modèles de trafic urbain.* Ministère de l'équipement, Paris.

Fichelet, R., Fichelet, M. et May, N. (1970). *Pour une approche écologique de l'utilisation des moyens de transport.* SERES, Paris.

Hottois, G. (1984). *Le signe et la technique : la philosophie à l'épreuve de la technique.* Aubier, Paris.

Hyndman, R. J. et Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.

Iklé, F. C. (1954). Sociological relationships of traffic to population and distance. *Traffic Quarterly*, 8(2):124–136.

Koenig, J.-G. (1974). Théorie économique de l'accessibilité urbaine. *Revue économique*, 25(2):275–297.

Lenormand, M., Huet, S., Gargiulo, F. et Deffuant, G. (2012). A universal model of commuting networks. *PLOS ONE*, 7:1–7.

Masucci, A. P., Serras, J., Johansson, A. et Batty, M. (2013). Gravity versus radiation models : on the importance of scale and heterogeneity in commuting flows. *Physical Review E*, 88(2):022812.

McNally, M. G. (2000). The four-step model. *In* Hensher, D. A. et Button, K. J., éditeurs : *Handbook of transport modelling.* Pergamon, Amsterdam.

Nguyen-Luong, D. (2001). *Modèles de prévision de trafic aux Etats-Unis.* IAURIF, Paris.

Openshaw, S. et Taylor, P. J. (1979). A million or so correlation coefficients : three experiments on the modifiable areal unit problem. *In* Wrigley, N., éditeur : *Statistical methods in the spatial sciences*, pages 127–144. Pion, London.

Ortúzar, J. d. D. et Willumsen, L. G. (2011). *Modelling transport.* John Wiley and Sons, Chichester, 4e édition.

Pumain, D. et Robic, M.-C. (2002). Le rôle des mathématiques dans une révolution théorique et quantitative : la géographie française depuis les années 1970. *Revue d'Histoire des Sciences Humaines*, 1:123–144.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Statistical Society of London*, 48(2):167–235.

SERC (1964). *Etudes de transports urbains. Analyse des méthodes américaines.* SERC.

Simini, F., González, M. C., Maritan, A. et Barabási, A.-L. (2012a). A universal model for mobility and migration patterns. Supplementary information. *Nature*, 484:1–27.

Simini, F., González, M. C., Maritan, A. et Barabási, A.-L. (2012b). A universal model for mobility and migration patterns. *Nature*, 484:96–100.

Stewart, J. Q. (1942). A measure of the influence of a population at a distance. *Sociometry*, 5(1):63–71.

Stouffer, S. A. (1940). Intervening opportunities : a theory relating mobility and distance. *American Sociological Review*, 5(6):845–867.

Stouffer, S. A. (1960). Intervening opportunities and competing migrants. *Journal of Regional Science*, 2(1):1–26.

Tobler, W. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240.

Wilson, A. G. (1970). *Entropy in urban and regional modelling.* Pion, London.

Zahavi, Y. (1974). Traveltime budgets and mobility in urban areas. Rapport technique, U.S. Department of Transportation.

Zipf, G. K. (1946). The P1 P2/D hypothesis : on the intercity movement of persons. *American Sociological Review*, 11(6):677–686.