



HAL
open science

Estimer des flux de navetteurs avec un modèle gravitaire. Application géomatique en Région Provence-Alpes-Côte d'Azur (France)

Didier Josselin, Samuel Carpentier-Postel, Frédéric Audard, Safaâ Amarouch, Jean-Baptiste Durand, Nathalie Brachet, Mathieu Coulon, Louissette Garcin

► To cite this version:

Didier Josselin, Samuel Carpentier-Postel, Frédéric Audard, Safaâ Amarouch, Jean-Baptiste Durand, et al.. Estimer des flux de navetteurs avec un modèle gravitaire. Application géomatique en Région Provence-Alpes-Côte d'Azur (France). *Geomatica*, 2020, 10.1139/geomat-2020-0009 . hal-02974718

HAL Id: hal-02974718

<https://hal.science/hal-02974718v1>

Submitted on 22 Oct 2020

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

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

1 Estimer des flux de navetteurs avec un modèle gravitaire

2 Application géomatique en Région Provence-Alpes-Côte d'Azur (France)

3
4 Didier Josselin¹, Samuel Carpentier-Postel³, Frédéric Audard², Safaâ Amarouch⁴, Jean-Baptiste
5 Durand⁴, Nathalie Brachet¹, Mathieu Coulon¹, Louissette Garcin⁴

6 1. UMR ESPACE, CNRS, Avignon Université

7 2. UMR ESPACE, CNRS, Aix Marseille Université

8 3. UMR ThéMA, CNRS, Université Bourgogne Franche-Comté

9 4. Master GEOTER - Géomatique et conduite de projet, Avignon Université

10 Correspondance : didier.josselin@univ-avignon.fr

11 *RÉSUMÉ.* Cet article réinvestit la modélisation des flux de déplacement domicile-travail au moyen des
12 modèles gravitaires de flux. À partir d'une demande émanant de la région Provence-Alpes Côte d'Azur,
13 un modèle répondant au principe de parcimonie, basé sur des données publiques ouvertes et facilement
14 mobilisables, a été établi afin de répondre au besoin de prévision tout en assurant la transparence et la
15 répliquabilité de la démarche. À partir de ce modèle, différentes hypothèses (type de variable sur les
16 masses des entités géographiques, variations des calculs des courtes distances à vol d'oiseau, contrainte
17 de portée spatiale due au budget-temps de déplacement, ré-échantillonnage aléatoire) ont été modélisées
18 dans une optique de prospective et d'aide à la décision. Les résultats montrent qu'il est possible d'obtenir
19 un modèle assez robuste avec les données disponibles de l'INSEE et un modèle gravitaire log-linéaire,
20 tout en réduisant sensiblement la taille de l'échantillon traité.

21 *ABSTRACT.* This paper investigates the modeling of journeys to work commuting flows using gravity
22 models. Based on a study for the Provence-Alpes Côte d'Azur region, a model that meets the principle of
23 parsimony, based on publicly available data, was designed to meet the need for forecasting, while
24 ensuring transparency and reproducibility of the method. Different models are processed, tuning several
25 input variables such as the masses of geographical entities, the way to compute short distances as the
26 crow flies, the time budget constraint from Zahavi, and successive resampling. The results show that it is
27 possible to build a robust model of correct quality to estimate the commuting flows, using INSEE data and
28 a log-linear gravity computation, while reducing the sample size.

29 *MOTS-CLÉS :* migrations alternantes intercommunales, modèle gravitaire, modèle de Newton
30 généralisé, aménagement régional, région Provence-Alpes-Côte d'Azur

31 *KEYWORDS:* intercity commuting, gravity model, Newton generalized modeling, regional planning,
32 Provence-Alpes-Côte d'Azur region
33

34 1. Introduction

35 Parce qu'elles sont fondamentalement l'expression de l'espace et du temps, les mobilités
36 spatiales intéressent de très près la géographie. La compréhension des modes d'organisation de
37 l'espace suppose, au-delà de la description des arrangements des lieux, d'analyser en détail les
38 échanges entre ces lieux, considérant que ces deux aspects – la structure, le processus – ne sont
39 que les deux faces d'une même pièce. Sans espacement, sans distance, pas d'échange et sans
40 échange pas de distanciation, de différenciation possible. Depuis les travaux sur les migrations de
41 Ernst Ravenstein (1885)¹, jusqu'à la formulation de la première loi de la géographie de Waldo
42 Tobler (1970)², les géographes se sont largement penchés sur la relation entre distances et flux.
43 Pour ce faire, de nombreux travaux se sont appuyés sur un transfert de modèle à partir de la
44 théorie de la gravitation universelle de Isaac Newton. Ces modèles d'interaction spatiale – parmi
45 lesquels on peut citer ceux de William Reilly (1931), John Stewart (1947), ou encore David Huff
46 (1963) – s'intéressent tour à tour à la question du volume des échanges entre les lieux ou de
47 l'attraction potentielle des lieux. Ces travaux, dont les apports et les limites sont aujourd'hui
48 largement documentés dans la littérature (voir par exemple : Wilson, 1975 ; Ubøe, 2004),
49 demeurent d'un apport précieux pour cerner les dynamiques macro-géographiques.

50 Dans le domaine des transports et des mobilités quotidiennes, cette question de la modélisation
51 des flux, bien qu'ancienne, reste d'actualité eu égard aux nombreux enjeux sociétaux et
52 environnementaux sous-jacents. L'épuisement progressif des ressources naturelles et
53 l'irréversibilité de l'impact de nombreuses activités humaines sur l'environnement sont en effet
54 largement liés à ce secteur. Selon l'International Energy Agency, ils contribuent en 2016, au
55 niveau mondial, pour 30% des émissions de CO₂ et le volume de ces émissions a augmenté de
56 28% depuis l'an 2000. Longtemps appréhendée sous l'angle principal de la congestion routière et
57 de la planification des projets d'infrastructures, la question de la gestion des flux de déplacement
58 est désormais abordée sous l'angle du report modal de la voiture vers les modes collectifs et
59 actifs. Elle constitue ainsi une question majeure pour la géographie et l'aménagement. Les
60 évolutions contemporaines des systèmes métropolitains marqués par l'étalement urbain et son
61 corolaire, la dépendance automobile (Newman et Kenworthy, 1989) ont en effet entraîné une
62 remise en cause de l'hégémonie de la ville motorisée, issue de la transition urbaine. D'autres
63 phénomènes majeurs sont directement liés à cet étalement : les dynamiques de ségrégation socio-
64 spatiale, ou encore la performance économique des espaces urbains (Cervero, 2001) sont
65 aujourd'hui pointés comme des facteurs limitant leur durabilité et justifiant l'étude des relations
66 entre formes urbaines et déplacements.

67 Ces évolutions et enjeux sont largement pris en compte par les acteurs de l'aménagement du
68 territoire. En France, les Régions ont, depuis la loi portant Nouvelle Organisation Territoriale de
69 la République (loi NOTRE, 2015), l'obligation³ d'élaborer un Schéma Régional d'Aménagement,
70 de Développement Durable et d'Égalité des Territoires (SRADDET). Ce schéma, qui vise
71 notamment à rompre avec les logiques sectorielles pour développer une vision intégrée de
72 l'aménagement régional, s'appuie sur un certain nombre de scénarios et prévisions qui supposent
73 un effort important de modélisation spatiale, notamment dans le domaine des transports. Dans le
74 contexte de l'élaboration de travaux de prospective territoriale régionale des mobilités à l'horizon
75 2030 et 2050, la région Provence-Alpes-Côte d'Azur nous a ainsi sollicité pour mener un projet

2 1 "1. Most migration is over a short distance [...] 3. Long-range migrants usually move to urban areas; 4. Each
3 migration produces a movement in the opposite direction (although not necessarily of the same volume)".

4 2 "I invoked the first law of geography: everything is related to everything else but near things are more related
5 than distant things."

6 3 À l'exception de la région d'Ile-de-France, les régions d'Outre-mer et la Corse.

76 de recherche action visant notamment à analyser et modéliser les flux régionaux. Face à cet
77 exercice prospectif, des choix ont été réalisés parmi un large panel de données et de modèles. Cet
78 article vise à restituer notre démarche de modélisation en considérant tant les enjeux scientifiques
79 (la modélisation des flux à l'échelle régionale), qu'opérationnels (l'appropriation et la
80 répliquabilité de la démarche par les acteurs).

81 Cet article se divise en quatre parties. La première propose de mettre en perspective les données
82 et grandes familles de modèles existants dans la prévision de la mobilité quotidienne et de les
83 situer quant à leurs avantages respectifs. La partie suivante aborde plus spécifiquement la
84 modélisation par les modèles gravitaires et leurs différences d'applications. Enfin, nous
85 présenterons les choix de modèles effectués et leurs résultats quant à la modélisation au sein de la
86 recherche-action menée dans ce travail.

87 **2. Données disponibles en France pour réaliser les estimations des** 88 **mobilités**

89 Les récentes évolutions en termes de capacités de collecte, de traitement et de stockage de
90 données modifient en profondeur la recherche en sciences sociales, et avec elle l'étude des
91 mobilités quotidiennes. Ces avancées, illustrées notamment par l'avènement des *Big Data*,
92 incitent bien souvent les chercheurs à complexifier leurs modèles de prévision des mobilités en
93 mobilisant un nombre important de variables. Pourtant, l'une des caractéristiques de ces grands
94 jeux de données est certainement leur hétérogénéité et la variabilité de leur disponibilité en
95 fonction des aires d'études choisies. Ainsi, il existe un risque de perte de reproductibilité, critère
96 pourtant essentiel de la démarche de modélisation. Sans nier que le foisonnement actuel des
97 sources de données constitue un cadre favorable à l'élaboration de nouvelles méthodes, notre
98 réflexion, qui s'inscrit notamment dans un transfert opérationnel vers les questions
99 d'aménagement régional, s'attache ici à la parcimonie en termes de données, ainsi qu'à la
100 reproductibilité des modèles choisis (Tableau 1).

101 Le choix d'une méthode d'analyse ou de prévision des flux, s'il s'opère en fonction d'objectifs
102 spécifiques, est également fortement dépendant des ressources disponibles en termes
103 d'informations sur les mobilités elles-mêmes, ainsi que sur la population qui se déplace. Ces
104 données peuvent alors être catégorisées en fonction de l'objet qu'elles renseignent, à savoir : les
105 individus, les ménages ou les véhicules.

106 **2.1. Données centrées sur les individus**

107 Les données centrées sur les individus relèvent elles-mêmes de deux sous catégories, déterminant
108 leur champ d'application. Tout d'abord, certaines bases de données renseignent à la fois les
109 mobilités individuelles et les données socio-démographiques liées à l'individu. Ainsi, les données
110 de migrations alternantes issues de recensement général de la population (INSEE⁴, 2015, 2016)
111 décrivent les mobilités liées aux motifs travail et études, tout en fournissant pour chaque individu
112 un certain nombre de caractéristiques socio-démographiques. Les fichiers dits « détails » des «
113 *Mobilités professionnelles* » fournissent pour chaque individu interrogé : la commune de
114 résidence (supposée être le lieu origine des navettes), le lieu de travail (supposé être le lieu de
115 destination), le mode de transport principal déclaré, ainsi que des données d'âge, de catégorie

8 4. Institut national de la statistique et des études économiques

116 socio-professionnelle, de condition d'emplois etc... Cette base de données est mise à jour de
117 manière régulière depuis 2006 ; à ce jour, les données les plus récentes datent de 2016. Bien que
118 l'établissement de séries temporelles soit faisable, notons qu'il n'est pas possible de suivre un
119 même individu dans le temps, car il n'existe pas d'identifiant commun entre les différentes
120 années. Ces données géolocalisées permettent également un certain nombre d'appariements avec
121 des données contextuelles utiles à l'analyse des flux (par exemple des taux de chômage localisé,
122 des données d'accessibilité, etc.). Elles permettent la segmentation des matrices de flux par mode
123 de transport, ce qui est utile pour la calibration des modèles (notamment des modèles gravitaires,
124 inclus ou non dans une démarche de modélisation en 4 étapes). L'ensemble de ces données est
125 disponible par simple téléchargement, sans autre formalité. Si leur accès est simple, leur
126 manipulation reste délicate, notamment en raison de la taille des fichiers, et suppose une bonne
127 expertise dans le domaine de l'analyse de données géographiques.

128 Plus récemment, de nouvelles sources de données qui renseignent les déplacements des personnes
129 ont fait leur apparition. Il s'agit notamment des données de localisation des smartphones que
130 détiennent les opérateurs ou fournisseurs de services. Dans le domaine de l'étude des mobilités,
131 ces *big data* suscitent un grand intérêt de la communauté scientifique, dans la mesure où elles
132 pourraient notamment permettre de documenter certains phénomènes mal pris en compte par les
133 sources de données traditionnelles (variations temporelles, diversité des motifs, etc.). Une autre
134 perspective consiste à vouloir substituer aux données classiques ces nouvelles sources,
135 notamment pour des questions de coût de collecte et de répétition des mesures. Si des tests
136 récents pour la constitution de matrices origines-destinations se sont montrés encourageants
137 (Bonnell et al., 2017), cette démarche comporte quelques inconvénients potentiels, parmi
138 lesquels : la dépendance aux opérateurs privés, l'absence de contrôle de la représentativité, ainsi
139 que l'absence de données socio-démographiques associées aux traces de géolocalisation.

140 2.2. Données centrées sur les ménages

141 En France, plusieurs grandes enquêtes permettent de mesurer de manière récurrente la mobilité à
142 l'échelle des ménages. Depuis 1967, plusieurs enquêtes, dénommées « Enquêtes Nationales
143 Transport et Déplacement » (ENTD), portant sur les comportements de mobilité des français ont
144 été réalisées pour le ministère français des transports par l'INSEE et l'IFSTTAR⁵ (1967, 1974,
145 1982, 1994, 2008, 2019). Il s'agit d'enquêtes représentatives au niveau national, effectuées sur un
146 échantillon important de répondants : par exemple pour l'enquête 2008, plus de 20 000 ménages
147 ont été interrogés. Bien que cet échantillon soit conséquent, il ne permet pas d'envisager des
148 analyses géographiques à des échelles fines (cette réserve varie selon les régions, car
149 l'échantillon a été augmenté pour certaines d'entre elles). L'intérêt de ces enquêtes repose
150 notamment sur le grand nombre de variables disponibles, la représentativité de l'échantillon au
151 niveau national, ainsi que l'existence d'une série temporelle longue. Si elles ne constituent pas
152 nécessairement la meilleure base pour une analyse des flux, elles représentent une source
153 d'information précieuse en ce qui concerne les comportements (choix modal, budgets-temps,
154 etc.) et le parc automobile et sont en cela très complémentaires des matrices origines-destinations
155 issues du recensement de l'INSEE (2015, 2016).

156 À l'échelle des grandes agglomérations, les comportements de mobilité sont renseignés, en
157 France, par les Enquêtes Ménages Déplacements (EMD), qui, comme leur nom l'indique,
158 considèrent le ménage comme entité de base de l'acquisition de données. Elles sont produites, sur
159 des périmètres variables, depuis le milieu des années 70. Elles se basent sur un protocole

10 ⁵ Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux

160 reproductible, appelé « *standard Certu* ». Elles sont aujourd’hui commanditées et conduites par
161 les collectivités territoriales avec l’appui méthodologique du Centre d’Études et d’Expertise sur
162 les Risques, l’Environnement, la Mobilité et l’Aménagement (CEREMA). Elles sont
163 généralement réalisées selon une périodicité d’environ dix ans, mais ne présentent pas de
164 caractère synchronisé sur l’ensemble du territoire national et sont rarement menées
165 conjointement, même aux échelles régionales. Ces enquêtes ont des périmètres qui peuvent
166 évoluer dans le temps. Ils s’étendent au minimum sur le périmètre des transports urbains, mais
167 sont souvent plus étendus. Ils correspondent parfois à des limites fonctionnelles (aires urbaines,
168 bassins de vie), parfois à des limites administratives (départementales). Ces enquêtes renseignent
169 les déplacements des membres d’un même ménage lors d’un jour de semaine et peuvent être
170 accompagnées d’informations optionnelles sur les déplacements du weekend. Elles sont
171 également composées de plusieurs fichiers, renseignant les caractéristiques du ménage, les
172 caractéristiques des individus, un descriptif détaillé de chaque déplacement, ainsi que différents
173 modules complémentaires. En France, l’ensemble de ces données est disponible, sur demande
174 écrite, auprès de l’ADISP (Archives de Données Issues de la Statistique Publique). Leur
175 utilisation nécessite cependant un important travail de prétraitement et implique des
176 manipulations spécifiques pour l’appariement avec d’autres données spatialisées, dépendant de
177 périmètres souvent différents.

178 **2.3. Données centrées sur les véhicules**

179 Les données de comptage routier de flux et de trafic correspondent à des informations recueillies
180 sur des durées et des étendues variables par des capteurs localisés sur la voirie. Elles permettent
181 de connaître la charge (quantité de passages) des tronçons de route et d’estimer des modèles
182 probabilistes de flux. On connaît le nombre et les grands types de véhicules qui passent à un
183 endroit donné, mais on ne sait pas d’où ils viennent, ni où ils vont. En lien avec la théorie des
184 graphes, ce type de mesures peut être utile, si elles sont comparées à des calculs théoriques de
185 *centralité intermédiaire*, c’est-à-dire permettant d’identifier quel tronçon de route est
186 particulièrement attracteur, de par sa topologie dans l’ensemble du réseau routier. Ces tronçons
187 peuvent d’ailleurs être identifiés comme autant de points de sensibilité structurelle du réseau
188 impactant sa résilience. Dans le cas de l’estimation des flux, on peut utiliser ces informations
189 pour savoir *a priori*, à l’aide de calculs de plus courts chemins, quels seraient, en principe, les
190 cheminements les plus usités.

191 Notons par ailleurs l’existence de données de trafic en temps réel, construites à partir des
192 observations dynamiques des flux et des historiques statistiques de la charge sur le réseau
193 (congestion). Des entreprises comme TomTom, Google, ou Géoconcept les utilisent, voire les
194 commercialisent. Leur utilité est plutôt le suivi et l’optimisation dynamique des transports.

195 **2.4. Données centrées sur les services**

196 On notera enfin diverses sources de données issues de l’analyse de la qualité des services, via les
197 opérateurs de transport (observation à travers les rapports de fonctionnement réalisés auprès de la
198 clientèle). Ces analyses peuvent découler également d’enquêtes qualitatives (entretiens) ou
199 quantitatives (non exhaustives, avec échantillonnage) sur des services de transports particuliers.

200

201 L'ensemble des sources de données présentées ici (Tableau 1) peuvent donc être différenciées en
 202 fonction de critères (exhaustivité, coût, attributs, propriété, granularité, usage principal, qualité),
 203 que ce soit sur la caractérisation de ces données ou sur l'utilisation qu'elles permettent pour la
 204 description, la compréhension ou la prévision des mobilités.

205 Parmi ces données, pour des raisons de complétude de couverture de la région Provence Alpes
 206 Côte d'Azur, nous choisissons d'utiliser les données échantillonnées de l'INSEE sur les
 207 migrations alternantes intercommunales.

208

Intitulé	Exhaustivité	Coût	Attribut	Propriété	Granularité	Usage principal	Qualité
Insee « mobilités professionnelles »	Exhaustif jusqu'en 1999, puis échantillon au quart	Elevé pour la collecte, gratuit à l'usage	Préférence révélée, Origine destination travail études	Libre	Individus, IRIS, communes	Régional, national (uniformité)	Bonne représentativité
Enquêtes Nationales sur les Transports et les Déplacements (ENTD)	Non, échantillon	Elevé pour la collecte, gratuit à l'usage	Individus, ménages, équipements, trajets	Accès sur demande	Individus, ménages	National	Représentativité nationale, pas d'exploitation possible à des échelles fines
Enquête Ménage Déplacements	Non exhaustif, mais représentatif sur les aires concernées	Elevé	Ménage, individus, déplacement	Accès sur demande	Individus, zone de tirage	Local	Bonne représentativité
Comptage flux et trafic	Oui, sur périodes et lieux donnés	Faible (matériel)	Mesures, fréquences	Publique non libre, partenariat	Tronçon routier	Local (département)	Exhaustif mais locale et peu généralisable
Traces mobiles et réseaux sociaux	Partielle, big data non validées	Intégré au service, élevé pour l'usager et en traitement	Pas d'attributs sociaux, trace géométrique observée	Privée, accès variable sur demande	Série de points, couverture des antennes ou précision du GPS	Potentiel national (données massives)	En quantité, mais peu précis et peu représentatif
Observation de service	Normalement oui (opérateur)	Faible, suivi marketing	Possible profils utilisateurs, préférences révélées	Privée RGPD-CNIL	Variable selon service et clients	Local, évaluation et suivi	Non généralisable mais détaillé
Enquête quantitative	Non mais on vise échantillon représentatif	Assez élevé	Possible, profils utilisateurs, préférences déclarées	Variable RGPD-CNIL	Variable (possible à l'adresse)	Local, recherche de tendances et lois	Potentiellement généralisable, non exhaustif
Entretiens qualitatifs	Non	Variable	Possible profils-types, préférences déclarées	Variable RGPD-CNIL	Variable (possible à l'adresse)	Local, recherche de signaux faibles, innovations	Non généralisable mais très détaillé

209 TABLEAU 1. Principales sources de données disponibles en France pour analyser les mobilités à
 210 une échelle régionale (RGPD : Règlement Général sur la Protection des Données, CNIL :
 211 Commission Nationale de l'Information et des Libertés)

212 3. Typologie des modèles gravitaires pour l'estimation spatialisée des 213 flux de personnes

214 Depuis les premiers modèles gravitaires, dès les années 1930, en passant par les modèles à quatre
 215 étapes issus de la *Chicago Area Transport Study* dans les années 1950, jusqu'aux modèles
 216 d'interaction transport-urbanisme qui ont émergé dans les années 1990, de nombreuses approches
 217 ont été développées pour modéliser et simuler les déplacements des individus. Ces modèles se
 218 distinguent par leur niveau d'analyse (les individus ou des unités spatiales), leurs échelles
 219 (urbaine, régionale, nationale...) ou encore leur niveau de complexité (en intégrant ou non la

220 dynamique des réseaux et de l'espace bâti). Si l'absence de prise en compte de la rétroaction de la
221 distribution des flux et des accessibilités afférentes sur l'agencement des activités et des réseaux
222 est un point critique des modèles gravitaires, ces derniers demeurent un outil privilégié de la
223 simulation de la distribution des flux, encore souvent mobilisés comme composante de modèles
224 plus sophistiqués. Cette longévité de l'approche gravitaire est due à la qualité de l'ajustement
225 qu'ils permettent, ainsi qu'à la relative simplicité de leur mise en place (tant en termes de
226 données d'entrée que de traitements), notamment dans une perspective opérationnelle.

227 3.1. Principes généraux

228 Le modèle gravitaire découle de la propriété très générale selon laquelle deux masses s'attirent ou
229 entretiennent des échanges privilégiés d'autant plus fortement qu'elles sont importantes et
230 proches. Waldo Tobler la définit en 1970 comme la première loi de la géographie à savoir « *tout*
231 *interagit avec tous, mais deux objets qui sont proches ont davantage de chances d'interagir que*
232 *deux objets qui sont éloignés* ». Cette propriété, étayée par des observations anciennes,
233 notamment dans le domaine des migrations, est analogue à la loi de la gravitation universelle
234 d'Isaac Newton. Elle impacte directement les flux entre les masses, et ainsi, à l'échelle
235 géographique qui nous occupe, les mobilités entre les villes. Évidemment, les formes de cette
236 interaction et de cette attractivité diffèrent selon le contexte considéré.

237 Plus particulièrement, dans l'espace géographique, notamment les géographes, apparaissent
238 plusieurs éléments : la *friction*, c'est à dire la difficulté à se déplacer, dans un espace qui est
239 différencié à cause de l'*anisotropie spatiale* (propriété reposant sur le constat que les territoires
240 ne sont pas homogènes, notamment du point de vue de leur topographie et leur population
241 résidente), mais également à cause de la *non symétrie des échanges* entre deux masses
242 équivalentes (des villes ou des stations de service de transport, par exemple) et de l'effet de la
243 topologie d'un ensemble de masses en interactions concurrentielles, constituant un réseau
244 particulièrement complexe à appréhender. Autrement dit, le modèle gravitaire peut être utilisé
245 pour estimer les mobilités, via l'analyse des couples de masses en interaction, ou comme un
246 système d'interactions entre un ensemble de masses. Dans le cas des flux terrestres de personnes
247 et des migrations individuelles à l'échelle d'une région, le support de la mobilité est le réseau
248 viaire et les lieux d'attractivité sont représentés par les villes. L'espace est découpé en communes
249 ou en unités spatiales infra-communales (en France les IRIS⁶ de l'INSEE), dont le poids est
250 quantifié par leur population.

251 En géographie, deux grands types de modèles existent pour les approches gravitaires.

252 3.2. Modèles basés sur les interactions par couple d'entités

253 Le premier regroupe les modèles qui considèrent les interactions entre deux masses (par couple).
254 A la base, la probabilité de partir d'une commune d'origine vers une commune de destination est
255 directement liée aux deux masses en présence et à la distance qui les sépare. S'agissant du
256 modèle d'attraction universelle de Newton, le flux établi entre la ville d'origine et la ville de
257 destination est le produit de leurs populations respectives, divisé par la distance au carré, facteur
258 important de réduction de l'interaction, selon la loi de Tobler. Mais on voit rapidement un
259 premier défaut d'un tel modèle. La relation entre les deux villes est symétrique, ce qui ne se
260 vérifie pas dans l'espace géographique : par exemple des grandes villes offrent des quantités et
261 des qualités élevées de formations, qui sont plus attractives pour les étudiants des petites villes

14 ⁶ Îlots Regroupés pour l'Information Statistique

262 n'en possédant pas. Un second défaut apparaît également dans la littérature (Josselin, Nicot,
263 2003) : l'impact de la distance sur les flux est variable. L'exposant de 2 affecté à la distance peut
264 donc changer selon les flux qui sont considérés.

265 Suite à ces constats, les scientifiques ont adapté le modèle de Newton en le généralisant et en
266 affectant aux masses (c'est-à-dire les populations urbaines), à l'instar de la distance, un exposant
267 variable. Par une transformation des variables en log de la formule généralisée de Newton pour
268 les estimer, on obtient une relation log-linéaire dans laquelle les poids de chaque critère considéré
269 (distance, population de la commune d'origine, population de la commune de destination, etc.)
270 représentent une élasticité et identifient le rôle de chaque critère dans l'explication des flux. Afin
271 de réduire les résidus du modèle, ces coefficients peuvent être estimés par une régression
272 multiple. Une autre difficulté réside dans la nécessité de déterminer un facteur d'échelle, noté k ,
273 permettant d'ajuster les flux estimés par le modèle aux unités de référence des observations.
274 Enfin, on notera la difficulté de prendre compte les courtes distances proches de zéro, qui souvent
275 correspondent à des flux relativement importants et pourtant mathématiquement incalculables
276 (soit une division par zéro est impossible, soit le chiffre au dénominateur est tellement petit que la
277 valeur du flux estimé atteint des ordres de grandeurs irréalistes). Pour évaluer si un modèle
278 gravitaire de type Newton généralisé a correctement estimé des flux, il convient de réaliser une
279 régression linéaire entre les flux théoriques et les flux réels. L'apparition d'un nuage allongé,
280 pouvant être correctement résumé par une droite avec un coefficient de détermination élevé,
281 indique un pouvoir prédictif relativement bon.

282 En complément, le modèle de Reilly (1931), se basant également sur un couple de masses,
283 permet de déterminer l'aire d'influence d'une ville par rapport à une autre (Converse, 1951). Pour
284 tous ces modèles, la distance est un facteur explicite clé et joue un rôle de frein ou de délimitation
285 de frontière, lié aux différences d'attractivité, réciproques ou relatives, des villes en présence.

286 **3.3. Modèle intégrant les occasions interposées**

287 Le second groupe de modèles estime la relation entre des villes prises dans leur contexte. Ils
288 diffèrent des précédents dans le sens où, d'une part, leur formule mathématique intègre les autres
289 masses environnantes (concrètement, les villes qui se trouveraient dans le périmètre d'étude et
290 d'influence), d'autre part, la distance est considérée comme une fonction clé dans la construction
291 même du modèle. De fait, les modèles de Stouffer (1940) et de Wilson (1971) intègrent
292 respectivement une forme de friction liée aux interactions topologiques des villes avoisinantes et
293 à la distance, en prenant en compte l'environnement et les *occasions interposées* (c'est à dire les
294 entités spatiales disposées dans un cercle d'un rayon d'influence donné autour de la ville
295 d'origine). Ces villes situées dans un cercle d'influence, constituent potentiellement une source
296 de captation de flux théoriques non négligeable, en fonction de l'organisation spatiale des masses
297 et des populations en présence, qui déterminent leur attractivité. Plus précisément, on retrouve
298 par exemple dans le modèle de Wilson une forte analogie avec le modèle gravitaire de Newton, à
299 la différence près que le flux est centré sur la ville d'origine et redressé (normalisé) par
300 l'ensemble des flux que cette même ville peut avoir avec les autres destinations possibles. De
301 cette façon, plus les échanges d'une ville avec des villes environnantes sont grands, plus
302 l'échange de celle-ci avec une ville en particulier est faible. Par ailleurs, une fonction de friction,
303 riche et modulable, permet de calculer les contraintes de déplacement en fonction de la distance.

304 Comme nous le verrons, à partir du fichier des migrations alternantes intercommunales de
 305 l'INSEE de 2016 en région Provence Alpes Côte d'Azur, nous avons réalisé un modèle gravitaire
 306 d'estimation des migrations alternantes, en utilisant le modèle de Newton généralisé.

307 3.4. Tableau comparatif des modèles gravitaires

308 Dans la Tableau 2, nous présentons et comparons 5 approches complémentaires courantes du
 309 modèle gravitaire, selon leur objectif et les effets négatifs ou positifs des masses et de la distance
 310 les séparant :

- 311 • la première loi de la géographie de Tobler (1970) ;
- 312 • la fameuse loi d'attraction universelle de Newton (1964) qui est généralisée pour cet
 313 usage ;
- 314 • la loi de Reilly identifiant les bassins d'attractivité gravitaires (Reilly, 1931) ;
- 315 • les occasions interposées mises en évidence par Stouffer (1940) ;
- 316 • l'approche de Wilson (1961) qui fait intervenir une fonction de friction dans l'espace.

317
 318

Auteur	Objectif	Formule	Masse	Distance ($d_{ij} > 0$)	Commentaire
Tobler (1970)	Evaluation de l'attractivité selon la distance	$F_{ij} = \frac{1}{d_{ij}^\gamma}$	Non prise en compte	Effet négatif	Relation uniquement liée à la distance ; problème de division par 0
Converse Reilly (1931)	Détermination de bassins d'attraction	$B_{ij} = \frac{d_{ij}}{1 + \sqrt{\frac{M_i}{M_j}}}$	Rapport de masse j/i ; effet négatif	Au numérateur ; effet positif ; calcul de bassin et frontière	Pas directement de calcul de flux, mais de bassin
Newton (1684)	Estimation de flux pendulaires	$F_{ij} = \frac{M_i^\alpha M_j^\beta}{d_{ij}^\gamma}$	Produit de masses ; asymétrique si exposants différents ; effet positif	Effet négatif	Calcul des élasticités avec modèle loglinéaire ; problème de division par 0 ; pas d'unité
Wilson (1971)	Estimation de flux migratoires	$F_{ij} = a M_i \frac{M_j f(C_{ij})}{\sum_{k=1}^k M_k f(C_{ik})}$	Produit de masses normalisé ; effet positif	Effet négatif indirect (à travers une fonction de coût ou de friction f)	Introduction indirecte des opportunités interposées
Stouffer (1940)	Estimation de flux par comptage des occasions interposées dans un rayon R	$F_{ij} = -a \cdot \log(f(\Delta)) + c$	à travers $f(\Delta)$ (s = nombre cumulé d'occasions interposées X dans le cercle de rayon R)	Effet négatif	Pas d'explicitation de la distance ; formalisation continue des occasions interposées

319 **TABLEAU 2. Différents modèles gravitaires pour estimer les flux**

320 Comme le tableau le montre, les différentes formulations des principaux types de modèles
 321 gravitaires sont anciennes. Pour autant, un certain nombre de travaux récents portent sur ces
 322 outils et explorent différentes perspectives, témoignant ainsi de l'intérêt persistant de ces modèles
 323 pour répondre aux problématiques contemporaines. Par exemple, Lenormand *et al.* (2016) ont
 324 réalisé une comparaison systématique de différents modèles gravitaires pour estimer les flux
 325 domicile-travail dans six pays, montrant ainsi la pertinence de ces modèles dans des aires
 326 géographiques différentes. En 2018, Li *et al.* ont élaboré un modèle gravitaire permettant de
 327 générer la distribution de flux en estimant les lieux d'emplois à partir d'une base de point
 328 d'intérêts (POI). Plus récemment, Liu *et al.* ont proposé en 2020 une extension des modèles
 329 gravitaires prenant en compte les caractéristiques de l'occupation du sol et utilisant des données
 330 de flux issus des traces de téléphonie mobile.

331 Cette liste des développements récents n'est pas exhaustive, et l'examen de la littérature montre
 332 que la plupart des travaux récents tendent vers une sophistication accrue des modèles originaux,
 333 par l'inclusion de nouvelles variables permettant de mieux appréhender la diversité des types de
 334 lieux et des types d'individus, reposant parfois sur des sources de données nouvelles. À l'inverse

335 de ces démarches, notre positionnement consiste au contraire à proposer ici des modèles
336 parcimonieux en termes de données, qui puissent être implémentés simplement sur différents
337 terrains, afin de favoriser leur appropriation par les acteurs territoriaux devant réaliser des
338 prévisions, dans le cadre de l'élaboration des différents schémas d'aménagement.

339 4. Migrations alternantes en région Provence Alpes Côte d'Azur

340 4.1. Site d'étude

341 Nous étudions les migrations alternantes de la région Provence-Alpes-Côte d'Azur en France.
342 Cette région est caractérisée par sa grande diversité. Elle est composée de 6 départements, un peu
343 moins de 1 000 communes et compte un peu plus de 5 millions d'habitants. Dans les zones
344 urbaines, au premier rang desquelles Marseille, Nice et Aix en Provence, et, dans une moindre
345 mesure, Toulon et Avignon, la population peut être importante dans les cœurs de ville. Mais
346 l'étalement en zone périurbaine, voire au-delà, y est aussi légion du fait de l'attrait du modèle
347 pavillonnaire et de politiques publiques peu contraignantes jusqu'à une période récente. L'accès à
348 la région est favorisé par l'autoroute A7 qui perce littéralement la région du Nord vers le Sud et
349 draine d'importants flux de tous ordres. La région possède aussi un grand espace rural et
350 montagnoux, dans sa partie (Nord-)Est, où la densité de population est très faible. Par ailleurs,
351 une grande partie de son périmètre touche la côte méditerranéenne (Figure 1).

352

353

354

355

356

357

358

359

360



361 FIGURE 1. La région Provence-Alpes-Côte d'Azur dans le Sud de la France (crédits : ESPACE,
362 toutes illustrations libres de droits, source du modèle numérique de terrain : IGN, logiciels
363 utilisés pour la conception : ArcGIS et Illustrator)

364 La région Provence-Alpes-Côte d'Azur est donc une région très contrastée. La nature des
365 déplacements de ses populations dépend de sa morphologie, en termes de topographie comme de

366 semis urbain. Les migrations alternantes, domicile-travail ou domicile-études sont certes
367 largement polarisées par quelques métropoles ou villes, mais elles présentent aussi une grande
368 variété due à l'étalement urbain et aux variations de densité. Pour les appréhender, nous avons
369 recours aux données de mobilités issues du recensement de la population de l'INSEE, organisme
370 officiel de production de la statistique nationale en France (service statistique public).

371 **4.2. Données utilisées**

372 **4.2.1. Recensement de la population et migrations alternantes de l'INSEE**

373 Le recensement de la population française (INSEE, 2019) a connu en 2004 un changement
374 radical de méthode de production (Faivre, 2017). En effet, l'INSEE a décidé de passer du
375 principe de recensement exhaustif de la population française effectué à intervalles de temps
376 réguliers (7 à 9 ans) à un mode de recensement rotatif continu (Brilhault & Caron, 2016). Avec
377 cette méthode, les communes de moins de 10000 habitants sont recensées exhaustivement tous
378 les cinq ans par roulement, selon 5 groupes de rotation disjoints définis aléatoirement. Pour les
379 communes de 10 000 habitants ou plus, elles font l'objet chaque année d'une enquête de
380 recensement au taux moyen de 8 %.

381 Selon l'INSEE⁷, le nouveau dispositif présente deux avantages par rapport aux collectes
382 exhaustives antérieures : il permet la publication annuelle de résultats statistiques récents
383 (antérieurs de 3 ans à la date de publication) sur différents thèmes et pour tout le territoire et il
384 permet le calcul d'évolutions robustes, pour tout échelon du territoire, sur des périodes plus
385 courtes que par le passé (cinq ans ou plus). Ce mode de recensement assure ainsi une meilleure
386 qualité du suivi des structures et de leur évolution en termes de population. Au total, sur un cycle
387 de 5 ans, environ 40 % des logements de la commune sont recensés. Les résultats sont affectés
388 d'une imprécision liée au sondage, estimée par un coefficient de variation, rapport entre la
389 variance des valeurs des échantillons et leur moyenne. Plus l'effectif est réduit, plus le coefficient
390 de variation est élevé et plus l'imprécision est forte. Par exemple, pour un effectif de 4 900 établi
391 au recensement de la population 2011, basé sur les enquêtes annuelles de 2009 à 2013, le
392 coefficient de variation vaut 3 %. Globalement, on peut souligner la grande robustesse des
393 échantillons du recensement de la population produits par l'INSEE.

394 Les données concernant les migrations alternantes sont extraites du recensement de la population.
395 Il s'agit d'une quantification des mobilités journalières pendulaires. Elles fournissent donc une
396 estimation fiable des flux liés à l'activité professionnelle. Ces migrations pendulaires peuvent être
397 justifiées par des motifs de travail ou d'étude, ou les deux associés. Elles peuvent avoir également
398 deux granularités : l'individu auquel est affecté un poids statistique dans l'échantillon ou la
399 commune qui agrège les individus se déplaçant. Ces données sont riches car elles intègrent bon
400 nombre de variables sur l'emploi, le travail, les modes de transport, les caractéristiques de
401 l'individu (sexe, catégorie socio-professionnelle) et du ménage, etc.

402 Dans notre étude, nous analysons les informations disponibles sur les migrations alternantes de
403 2016, et, dans une moindre mesure, de 2015. Dans le paragraphe suivant, des informations de
404 cadrage sont présentées sous forme de cartes avec les fichiers agrégés des flux
405 domicile-travail/études en 2015. Puis, dans la section traitant de la modélisation, nous étudions
406 exclusivement les migrations alternantes journalières domicile-travail à l'échelle de l'individu,
407 extraites du dernier fichier disponible de 2016. Si l'information recueillie est bien à l'échelle de

19 ⁷ https://www.insee.fr/fr/statistiques/fichier/2383177/fiche-conseils_synthese_2019-06-25.pdf

408 l'individu (pondéré statistiquement), le grain de traitement du flux, c'est à dire son origine et sa
409 destination, demeure à l'échelle de la commune. L'intérêt est de maîtriser la méthode
410 d'agrégation des données individuelles au niveau de la commune, afin de produire une
411 quantification des migrations alternantes domicile-travail, intercommunale et journalière, selon
412 différents critères.

413 Outre les flux observés, issus des estimations par l'échantillon de l'INSEE, nous exploitons
414 d'autres informations au niveau de la commune, selon qu'elle soit origine ou destination du flux.
415 Les masses M_i et M_j peuvent être quantifiées par la population, le nombre d'actifs, l'emploi. La
416 distance, quant à elle, est simplement calculée à vol d'oiseau entre les centroïdes des communes.
417 Non que nous défendions sans discernement la justesse de ce mode d'estimation du coût
418 d'accessibilité spatiale (cf. la discussion d'Héran, 2009), mais parce que le calcul du temps sur le
419 réseau est très coûteux. De plus, le recours au coût sur le réseau viaire complexifie fortement la
420 construction de la matrice origine-destination : comment agréger les vitesses de déplacement
421 pour assigner des valuations justes aux arêtes du graphe selon les moments de la journée et les
422 encombrements des routes ? Que signifie, *in fine*, l'accès à un polygone que constitue la
423 commune ? Doit-on considérer la commune comme un point localisé à la mairie ? Doit-on
424 agréger les quartiers ou les villages ou considérer plusieurs destinations internes à la commune ?
425 De fait, l'affinage de la variable distance et la complexité qu'il induirait s'oppose à une mise en
426 œuvre générique de modèles gravitaires parcimonieux d'estimation des flux pour une collectivité
427 territoriale.

428 Le projet étant réalisé dans le cadre d'un programme de recherche-action avec la région
429 Provence-Alpes-Côte d'Azur, nous ne considérons que les flux ayant lieu en son sein. Flux
430 entrant et sortant dans/de la région sont donc exclus du traitement. La portée potentielle
431 maximale des flux est donc de l'ordre de 300 km, ce qui offre déjà une grande amplitude spatiale
432 et une large quantité de flux à explorer, étant entendu qu'il existe une relation inverse entre la
433 distance et la fréquence des flux, relation que nous quantifions par la suite dans les modèles
434 gravitaires testés. En définitive, le fait de ne considérer que les flux à l'intérieur du périmètre
435 régional ne biaise pas trop statistiquement l'échantillon, car l'effet de la distance, en lien avec la
436 contrainte quotidienne du budget-temps de transport, réduit la fréquence des flux journaliers
437 domicile-travail à très longue portée. Toutefois, cela n'évite pas les effets de bordure : des flux de
438 portée variable (entrants ou sortants) avec les autres régions (et le pays) limitrophes ne font pas
439 partie de notre échantillon. Mais ces flux sont de tous ordres et de toutes portées et leur absence a
440 pour principale conséquence de diminuer la taille de l'échantillon.

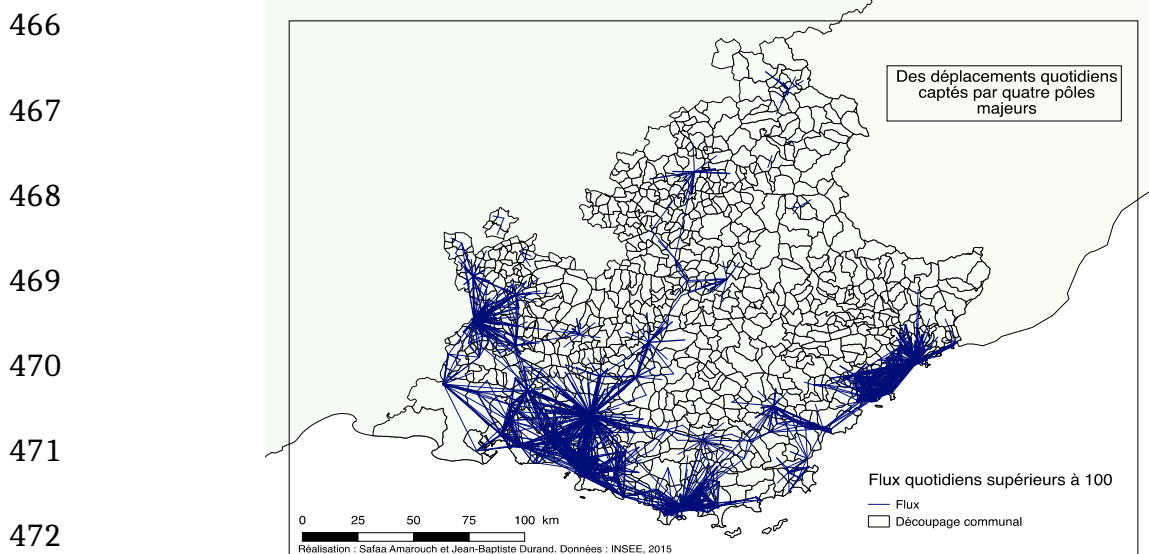
441 Enfin, notons que la région est considérée comme l'objet à analyser globalement en termes de
442 flux. Nous ne distinguons pas de bassins de mobilité dues aux polarisations des métropoles
443 (exemple : Marseille). Les modèles que nous réalisons visent donc, sur la base du principe de
444 gravité universel, à englober tous les cas de figures observés en région Provence-Alpes-Côte
445 d'Azur, pour un maximum de généralisation.

446 **4.2.2. Structures des flux intercommunaux régionaux**

447 À partir de données agrégeant les flux domicile-travail et domicile-études de 2015, nous
448 présentons ici quelques cartes donnant une représentation synthétique de la densité et de la portée
449 des déplacements des navetteurs. Elles sont le fruit d'un travail réalisé en 2019 (Amarouch *et al.*,
450 2019).

451 La figure 2 dessine les flux intercommunaux pour les communes ayant plus de 100 mobilités
452 quotidiennes. Les communes de petite taille (moins de 100 flux par jour) ont été enlevées pour
453 une meilleure lisibilité de la structure des flux. On visualise en conséquence la polarisation
454 engendrée par les métropoles marseillaise, aixoise et niçoise, mais également l'attractivité des
455 villes littorales, comme Toulon ou Canne au Sud, ou intermédiaires, comme Grasse, Avignon à
456 l'Ouest, ainsi que Digne et Gap en zone de montagne.

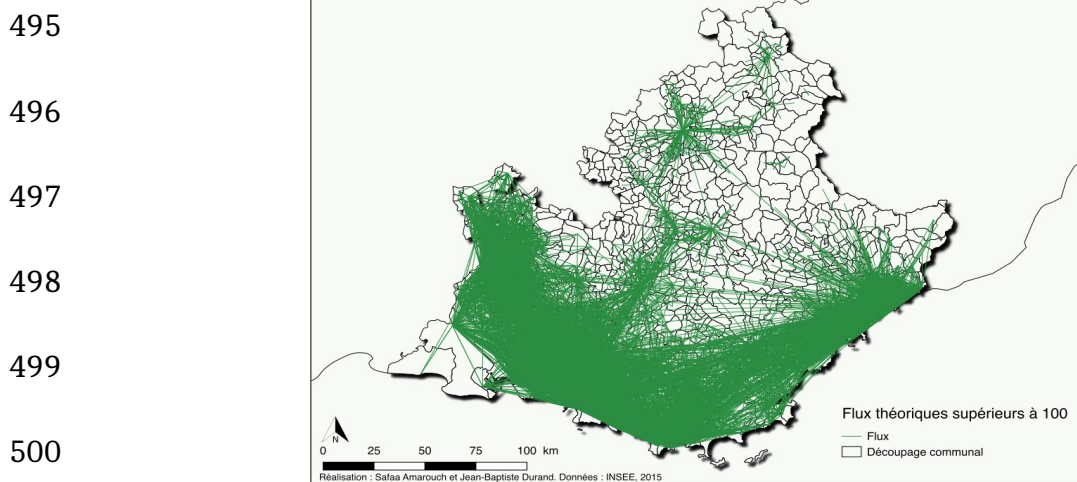
457 En appliquant un modèle gravitaire standard de Newton avec la population comme masses des
458 communes d'origine et de destination (cf tableau 2, avec respectivement $\alpha = 1$ et $\beta = 1$) et la
459 distance au carré ($\gamma = 2$) dans le modèle gravitaire (Amarouch *et al.*, 2019, Josselin *et al.*, 2020),
460 on obtient les flux présentés en figure 3. À noter que les flux internes aux communes, de distance
461 nulle, ont été ici enlevés. On constate que les flux théoriques calculés par le modèle dans les
462 mêmes conditions que l'observation produite par l'INSEE surestiment globalement la quantité de
463 flux réels. On devine toutefois le squelette des corridors des déplacements, mais celui-ci est noyé
464 dans un écheveau inextricable de flux, rendant la carte peu lisible. Pour information, le R^2 du
465 modèle obtenu était de 0.47, ce qui reste correct, mais relativement faible.



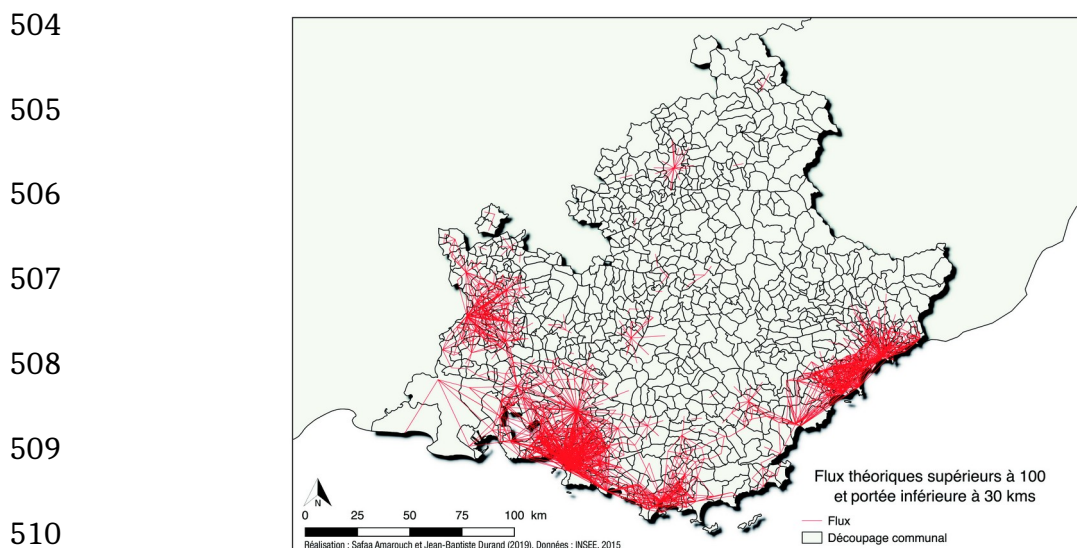
473 FIGURE 2. *Migrations alternantes journalières (motifs : travail et études) observées en Provence-*
474 *Alpes-Côte d'Azur (échantillon rotatif de 2015, INSEE « Mobilités professionnelles »)*

475 Ces résultats amènent à discuter le modèle de Newton plus en profondeur, afin d'en améliorer
476 son usage, d'évaluer son pouvoir de prédictibilité des migrations pendulaires, avec différents
477 paramètres. D'une part, il s'agit de rechercher les meilleurs coefficients (exposants) du modèle de
478 Newton, dans sa version généralisée, pour mieux l'ajuster, et déterminer dans l'équation la valeur
479 optimale du coefficient k . En effet, une des raisons de la différence flagrante entre les cartes 2 et
480 3 réside dans la précision de ces paramètres, et notamment la valeur de k . D'autre part, le modèle
481 peut se baser sur différentes sources d'informations pour quantifier les masses et les distances, en
482 testant, par exemple, différentes variables en entrée (population, emploi, actifs, notamment). Le
483 calcul de la distance peut également varier et prendre en compte, ou non, les flux intra-
484 communaux.

485 Du point de vue thématique, on peut analyser si la sélection de certains flux améliore la qualité
486 du modèle. Sous contrainte du budget-temps (Zahavi & Talvitie, 1980 ; Joly, 2005), on peut, par
487 exemple, enlever les flux de trop grande portée. C'est ce qu'illustre la figure 4, où l'on n'a
488 cartographié que les flux de moins de 30 km pour les communes comptant plus de 100 migrations
489 alternantes journalières. Visuellement, la structure qui en résulte semble relativement proche de
490 l'observation (carte de la figure 2). Mais au-delà d'une comparaison visuelle, une analyse
491 approfondie de la qualité des modèles produits est nécessaire. C'est ce que nous proposons de
492 réaliser dans cet article, cette fois-ci sur les toutes dernières données disponibles en France sur le
493 sujet et datant de 2016, en réalisant un focus sur les déplacements domicile-travail internes des
494 communes et des arrondissements des métropoles de la région Provence-Alpes-Côte d'Azur.



501 *FIGURE 3. Flux théoriques supérieurs à 100 : ils paraissent largement surestimés, au moins*
502 *spatialement, pour les longues distances, correspondant souvent à des flux de faibles valeurs*
503 *(échantillon rotatif de 2015, INSEE « Mobilités professionnelles »)*



511 *FIGURE 4. Flux théoriques intercommunaux supérieurs à 100 (en quantité) et selon une*
512 *contrainte de budget-temps de 30 min (soit 30 km en moyenne) (échantillon rotatif de 2015,*
513 *INSEE « Mobilités professionnelles »)*

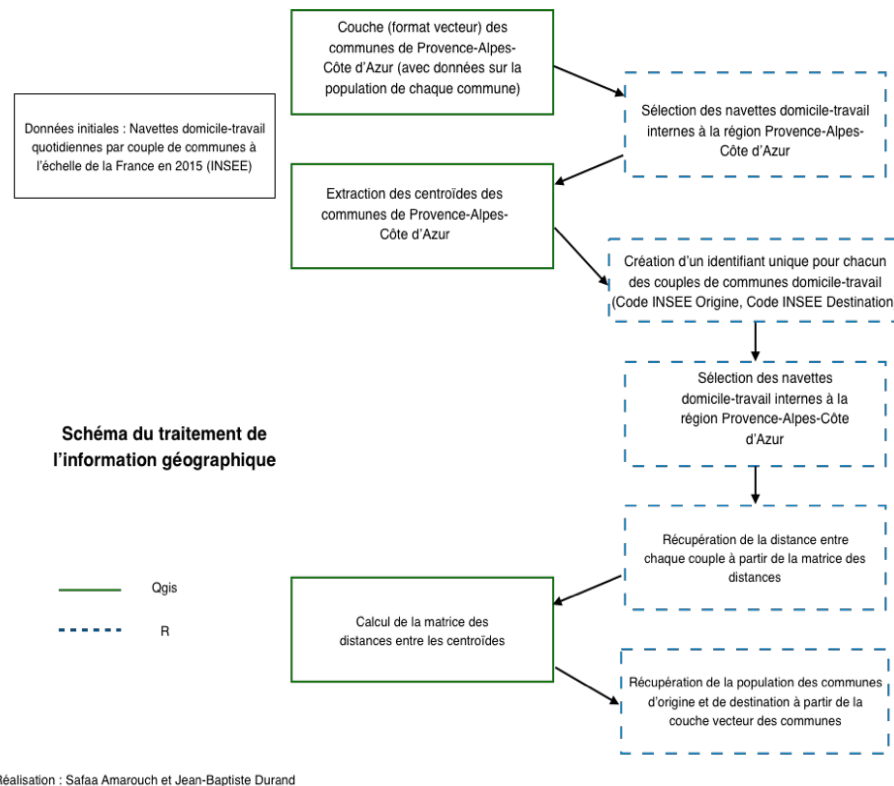
514 4.3. Application d'un modèle gravitaire sur les migrations 515 alternantes quotidiennes intercommunales

516 Dans cette section, nous présentons la méthodologie : le traitement de l'information
517 géographique, le modèle de Newton log-linéarisé et l'analyse exploratoire détaillée du modèle,
518 amenant à comparer la qualité de 18 modèles, avec une batterie de tests statistiques associés.

519 4.3.1. Traitement de l'information géographique

520 La figure 5 expose la méthodologie de traitement. Celle-ci a nécessité deux grandes phases. Dans
521 un premier temps, nous constituons une base de données spatiales avec les communes et leurs
522 centroïdes pour la région Provence Alpe Côte d'Azur. Puis, nous extrayons les flux observés des
523 fichiers des navettes domicile-travail, ainsi que les distances intercommunales. Les traitements
524 ont été réalisés au sein d'un tableur, d'un Système d'Information Géographique (QuantumGIS) et
525 d'un logiciel de traitement statistique (R). Sous R, les flux sont agrégés à la commune selon
526 différents critères (cf. section suivante des résultats). Dans un second temps, nous calculons les
527 flux théoriques en région Provence-Alpes-Côte d'Azur à partir des populations communales, en
528 utilisant un modèle gravitaire généralisé de Newton, via une régression loglinéaire multiple
529 (Josselin et Nicot, 2003). Nous comparons les estimations du modèle aux observations, ainsi que
530 les structures spatiales induites par les flux.

531



532 FIGURE 5. *Traitement de l'Information Géographique pour la modélisation des migrations*
533 *alternantes en région Provence Alpe Côte d'Azur (exemple : domicile-travail/études en 2015 ;*
534 *Amarouch et al., 2019)*

535 4.3.2. Linéarisation du modèle de Newton généralisé

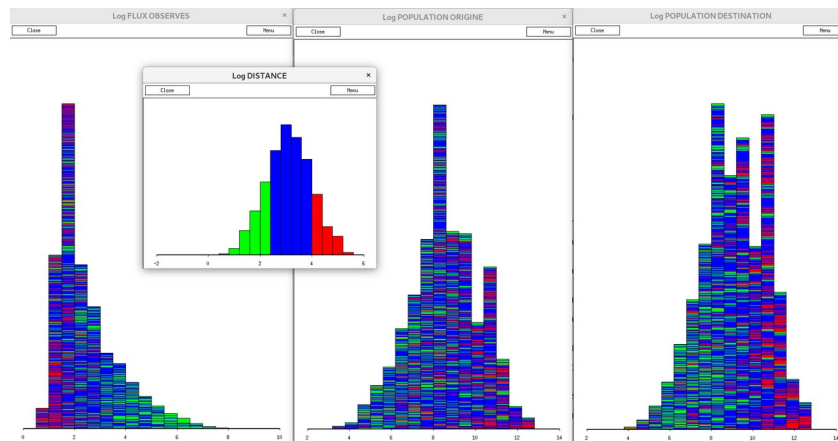
536 Le modèle de Newton généralisé (présenté dans le tableau 2) peut être linéarisé par une
537 transformation globale en \log (Burger et al., 2019). Les exposants deviennent alors des *élasticités*
538 qui identifient les poids respectifs des critères dans l'estimation des flux. On peut alors appliquer
539 des régressions log-linéaires multiples pour estimer ces élasticités. Par ailleurs, les distributions
540 de la population communale et des distances étant hyperboliques, une transformation en \log leur
541 permet de devenir des distributions la plupart du temps unimodales, relativement symétriques (cf.
542 Figure 6). La formule du modèle log-linéaire généralisé gravitaire (en spécification $\log\text{-log}$)
543 devient alors (i) :

$$544 \log(F_{ij}) = \alpha \log(M)_i + \beta \log(M)_j - \gamma \log(d_{ij}) + \log(k) \quad (i)$$

545 Dans notre modèle, les flux observés fournis par l'INSEE sont la variable à prédire, les masses
546 des entités d'origine et de destination des flux et la distance les séparant sont les variables
547 explicatives. Le retour aux données initiales est réalisé via l'équation du modèle de flux par
548 l'usage de la fonction exponentiel. On revient alors sur la formule de référence du modèle de
549 Newton généralisé (ii), avec ses flux F_{ij} , k une constante, M_i et M_j les populations des
550 communes (avec leurs exposants respectifs α et β) et d_{ij} la distance (avec son exposant γ)
551 séparant les deux communes.

$$552 F_{ij} = k \frac{M_i^\alpha M_j^\beta}{d_{ij}^\gamma} \quad (ii)$$

553



554 FIGURE 6. Analyse exploratoire des distributions statistiques croisées (en log népérien) des flux
555 observés, de la distance et des populations des communes d'origine et de destination (source :
556 INSEE, « Mobilités professionnelles », 2015).

557 4.3.3. Critères de calibrage et stratégie d'exploration du modèle

558 Plusieurs types de critères ont été considérés selon l'aspect du modèle que nous souhaitons faire
559 varier. Le principe est de rechercher le meilleur modèle, c'est à dire celui qui prédit le mieux les
560 flux, en testant à la fois différentes sources de données et/ou façon de les calculer. L'idée sous-
561 jacente est d'analyser si des sources d'information facilement accessibles suffisent à produire un
562 modèle de qualité correcte. En complément, on s'attachera à évaluer l'importance des gains
563 éventuels de qualité du modèle, en cas de choix de données plus spécifiques et/ou *a priori* mieux

564 adaptées au modèle du point de vue thématique. Par exemple, est-ce que la variable *emploi* pour
565 définir la masse des entités de destination améliore substantiellement le modèle, par rapport à
566 l'usage de la variable *population*, plus facile à obtenir, mais moins pertinente si l'on considère
567 l'estimation des mobilités domicile-travail ? Les critères testés touchent notamment les quatre
568 types de variables engagées dans la modélisation.

569 Concernant la méthode d'agrégation des flux observés, deux cas ont été considérés :

- 570 • un individu dans le fichier de l'INSEE est un flux et l'opération consiste alors à
571 simplement en réaliser la somme par entité spatiale ;
- 572 • un individu est affecté de son poids statistique dans l'échantillon et l'on calcule alors la
573 somme pondérée par entité, donnant la plupart du temps des nombres réels pour les flux.

574 Pour la distance, nous avons précédemment justifié l'usage de la distance à vol d'oiseau dans le
575 présent article. Ce choix ne résout toutefois pas la question de la division par zéro imposée pour
576 les flux réalisés au sein des entités (et donc de distance nulle) (Commenges & Commenges,
577 2017). Nous proposons ici plusieurs approches à comparer :

- 578 • remplacer les valeurs nulles par 1 mètre (méthode fréquemment utilisée) ;
- 579 • calculer un cercle de rayon R de portée théorique basée sur la surface S l'entité ($R =$
580 $\sqrt{S/\pi}$) ;
- 581 • générer aléatoirement par jittering des distances allant de 0 (exclu) à ce rayon théorique R .

582 Pour la masse M_i des entités à l'origine des flux, nous pouvons choisir deux sources de données :

- 583 • la population de l'entité ;
- 584 • le nombre d'actifs de 15 à 64 ans, *a priori* plus adapté pour les déplacements domicile-
585 travail.

586 Enfin, pour la masse M_j des entités de destination des flux, l'alternative était la suivante :

- 587 • la population de l'entité ;
- 588 • le nombre d'emplois sur la commune, *a priori* plus adapté pour les déplacements
589 domicile-travail.

590 Au-delà de ces critères concernant les variables, nous testons trois autres dimensions du modèle :
591 la taille de l'échantillon permettant de calibrer le modèle gravitaire log-linéaire, un seuil maximal
592 de portée spatiale de flux et un seuil minimal de quantité de flux par entité. Ainsi, on peut
593 appliquer les modifications suivantes sur l'échantillon de référence :

- 594 • réaliser un tirage aléatoire dans l'échantillon de référence, en prélevant 1, 2, 5, 10, 25,
595 50 % des individus, faisant varier d'environ 250 à 12 000 l'effectif de flux traités par le
596 modèle ;
- 597 • enlever les flux de longue distance avec des seuils variables pour tester la contrainte de
598 budget temps de Zahavi illustrée précédemment ; ces choix impactent aussi la taille de
599 l'échantillon ;
- 600 • enlever les communes possédant un nombre de flux inférieur à un certain seuil, afin de
601 réduire si possible le bruit dans le modèle dû aux *outliers* potentiels ; ces modifications
602 réduisent plus ou moins la taille de l'échantillon.

603 Devant la forte combinatoire de ces critères, nous ne présenterons pas la suite qu'une sélection de
604 modèles typés en fonction de leurs résultats.

605 4.3.4. Évaluation du modèle

606 Les modèles log-linéaires (*log-log*) renvoient la fonction minimisant les résidus, avec les
607 élasticités des variables explicatives et l'ordonnée à l'origine, l'effectif de l'échantillon et les
608 degrés de liberté, ainsi qu'une batterie de tests statistiques.

609 Comme toute régression linéaire, nous pouvons ainsi avoir recours à un certain nombre de
610 métriques statistiques pour évaluer l'ajustement du modèle, que nous avons observées dans les
611 résultats comparés sous le logiciel R, mais que nous ne retiendrons pas toutes dans le tableau 3
612 synthétique des modèles :

- 613 • pour le modèle dans sa globalité : R^2 et R^2 ajusté, *erreur type*, *F-statistique*, *p-value*,
614 critères d'information d'Akaike (*AIC*, également dans sa forme bayésienne, *BIC*), nuage
615 de points croisant flux observés et flux théoriques ; il s'agit ici de mesurer la qualité du
616 modèle de régression ;
- 617 • pour les variables explicatives et l'intercept : *erreur type*, *t-statistique*, *p-value*,
618 *intervalles de confiance* à 1 %, *moyenne*, *médiane*, *matrice de covariance*, test de
619 multicolinéarité par le facteur d'inflation de la variance (*VIF*) ; il s'agit ici de fournir des
620 indicateurs statistiques sur les variables et leurs relations ;
- 621 • pour les résidus du modèle : somme des carrés des résidus du modèle (*SSE*), erreur
622 quadratique moyenne (*RMSE*), erreur quadratique moyenne standardisée (*SRMSE* =
623 *RMSE*, moyenne des résidus), *entropie* de Shannon, *quantiles* des résidus, *distribution* des
624 résidus du modèle avec distribution théorique gaussienne de mêmes moyenne et écart-
625 type pour évaluer l'homoscédasticité des résidus ; il s'agit ici d'estimer la variabilité de la
626 prédiction, via différents indicateurs statistiques appliqués sur les résidus.

627 5. Résultats statistiques

628 5.1. Synthèse des résultats statistiques

629 Au total, 18 modèles ont été retenus en fonction de leurs résultats (tableau 3). Ils peuvent
630 maintenant être comparés à l'aide d'une sélection de métriques.

631 L'ensemble des modèles partage les éléments suivants, montrant déjà une grande robustesse dans
632 leur mise en oeuvre :

- 633 • les *p-values* sont toujours extrêmement faibles ($< 10^{-16}$)
- 634 • toutes les variables sont significatives au regard des t-statistiques
- 635 • les tests *F* rejettent systématiquement l'hypothèse *H0* d'indépendance statistique
- 636 • les tests de multicolinéarité sont toujours nuls pour toutes les variables

637

638 Les différents modèles sont comparés en détails par différents tests dans le tableau 3 :

- 639 • *identification du modèle* : ID, description textuelle et effectifs de l'échantillon ;
- 640 • *formule de la régression log-linéaire* : élasticités des trois variables explicatives (valeurs
641 des exposants du modèle de Newton indiquant les poids des variables), ordonnée à
642 l'origine (*intercept*) ;
- 643 • *qualité du modèle de régression* :
 - 644 ◦ intervalles de confiance à 1 % de l'*intercept* et des *élasticités* (montrant une
 - 645 incertitude très limitée quant aux résultats obtenus) ;
 - 646 ◦ qualité de l'ajustement avec R^2 ajusté (critère standard) ;

647 ○ variabilité des résidus estimée par trois critères complémentaires : erreur quadratique
648 moyenne (approche moyenne sensible aux valeurs extrêmes), erreur quadratique
649 moyenne standardisée (pouvant être assimilée à un coefficient de variation des
650 résidus), et entropie de Shannon calculée sur les classes des distributions (plus
651 robuste, portant ici sur un nombre équivalent de classes entre les modèles).

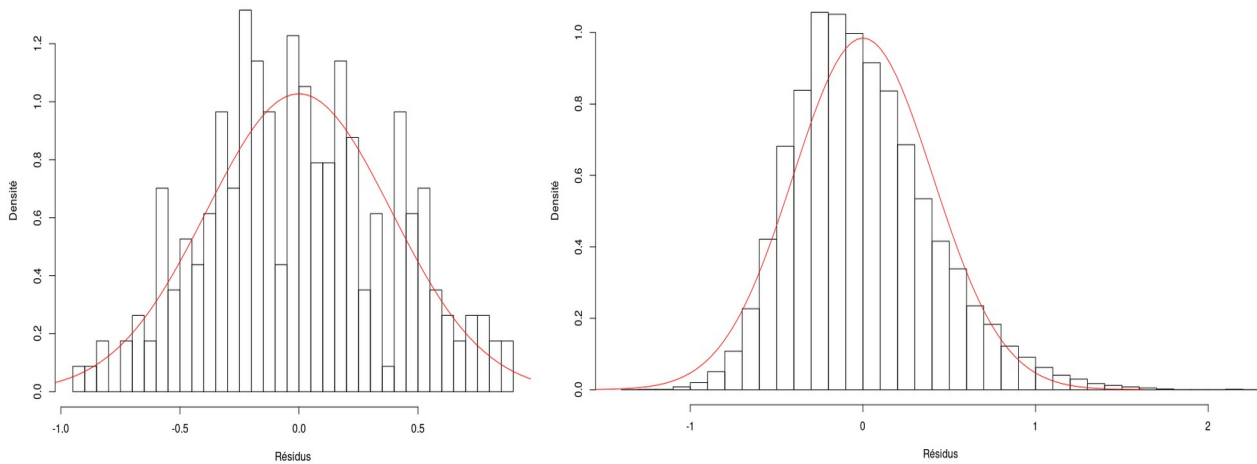
652

653

654

655

656



658

659

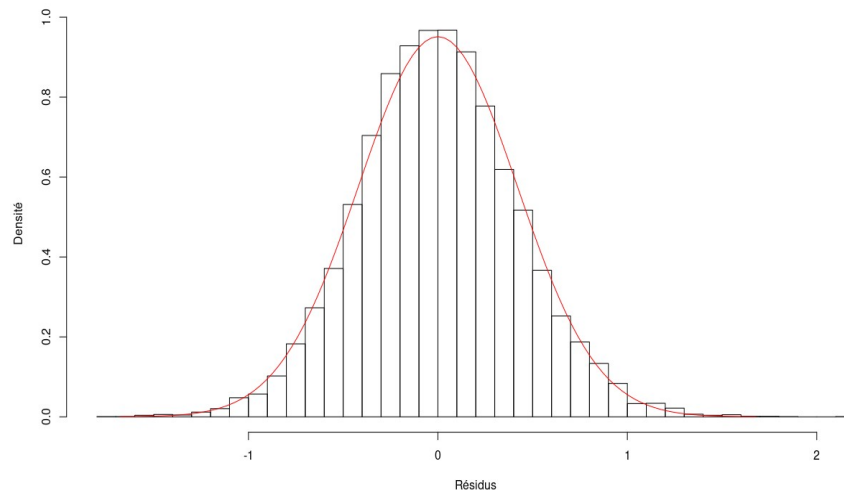
660

661

662

663

664



665

666 *FIGURE 7. Normalité des résidus des modèles ; en haut à gauche : modèle #11, effectif de 240,*
667 *échantillon de 1 % ; en haut à droite : modèle #1, effectif de 24000 ; en bas : modèle #14,*
668 *effectif de 15000 ; la courbe correspond à la distribution théorique gaussienne*

669 Concernant l'homoscédasticité des résidus, considérant que les tests de normalité ne sont pas
670 toujours robustes, nous observons les distributions de l'erreur et présentons 3 cas, parmi les 18
671 modèles de régression calculés (figure 7), dont la moins bonne (à gauche, modèle #11, effectif de
672 240, échantillon de 1 %) et la meilleure (à droite, modèle #14, effectif de 15000).

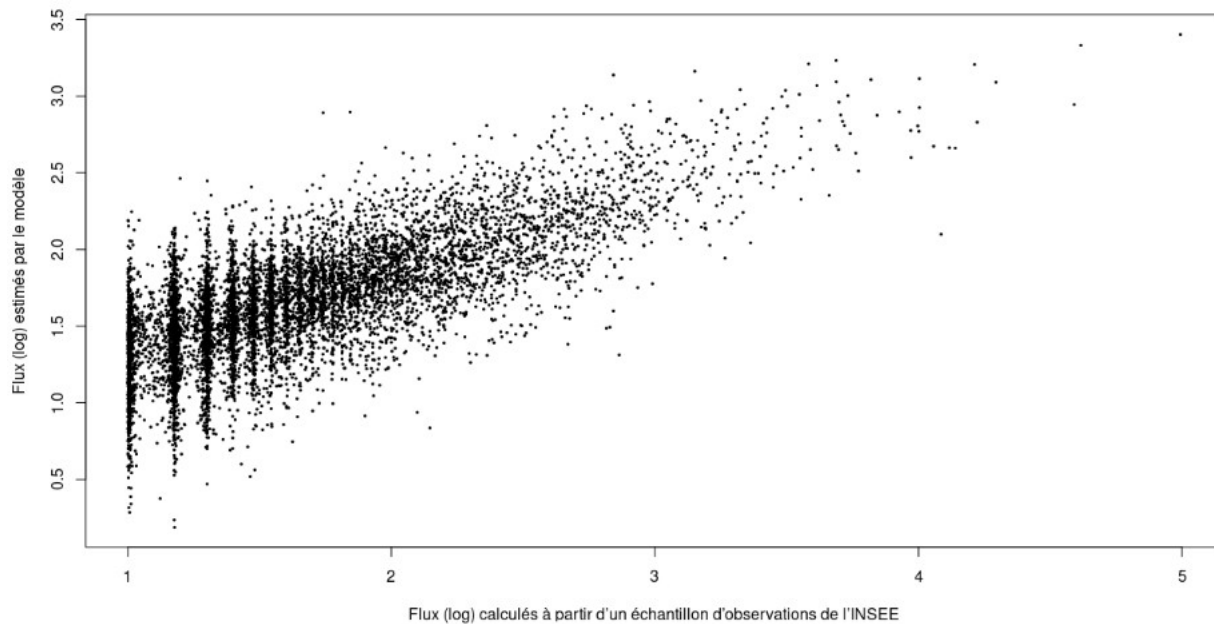
673

674

DONNÉES de FLUX			FORMULE de la RÉGRESSION LOGLINEAIRE				QUALITÉ du MODÈLE							
O : origine, D : destination			Elasticités				Intervalles de confiance à 1% [min,max]				Ajustement		Variabilité des résidus	
ID	Description	Effectif N =	Intercept log(k)	DISTANCE α	POP_ORIG β	POP_DEST γ	Intercept	DISTANCE	POP_ORIG	POP_DEST	R ² ajusté	RMSE	SRMSE	Entropie
1	Flux comptés par commune (sauf internes) ; population communale en O et D	24000	1.755	-0.908	0.311	0.381	[1.673;1.836]	[-0.928;-0.887]	[0.300;0.321]	[0.370;0.391]	0.45	0.4	0.92	2.23
2	Flux pondérés par commune (sauf internes) ; population communale en O et D	24000	3.027	-0.953	0.153	0.405	[2.943;3.110]	[-0.973;-0.932]	[0.142;0.163]	[0.393;0.415]	0.42	0.41	0.4	2.02
3	Flux pondérés par commune dont internes (calcul par rayon) ; population communale en O et D	24000	3.572	-1.085	0.160	0.411	[3.494;3.649]	[-1.104;-1.064]	[0.148;0.170]	[0.399;0.422]	0.46	0.43	0.4	2.06
4	Flux pondérés par commune (internes par calcul aléatoire dans rayon) ; population communale en O et D	24000	3.098	-0.985	0.164	0.413	[3.028;3.168]	[-1.002;-0.966]	[0.153;0.175]	[0.401;0.424]	0.47	0.42	0.4	1.98
5	Flux pondérés par commune (internes par calcul aléatoire dans rayon) ; actifs 15-64 ans en O et population communale en D	24000	3.152	-0.982	0.163	0.413	[3.083;3.221]	[-1.000;-0.964]	[0.152;0.174]	[0.401;0.423]	0.47	0.42	0.4	1.99
6	Flux pondérés par commune (internes par calcul aléatoire dans rayon) ; actifs 15-64 ans en O et nombre d'emplois en D	24000	3.626	-1.070	0.181	0.423	[3.561;3.690]	[-1.087;-1.052]	[0.170;0.190]	[0.413;0.432]	0.53	0.4	0.37	1.93
7	Idem ci-dessus ; échantillon aléatoire de 50 %	12000	3.583	-1.059	0.181	0.421	[3.493;3.673]	[-1.083;-1.035]	[0.166;0.195]	[0.408;0.434]	0.53	0.4	0.37	2.07
8	Idem ci-dessus ; échantillon aléatoire de 25 %	6000	3.564	-1.065	0.186	0.432	[3.422;3.706]	[-1.102;-1.026]	[0.162;0.208]	[0.410;0.452]	0.53	0.4	0.37	2.08
9	Idem ci-dessus ; échantillon aléatoire de 10 %	2400	3.589	-1.073	0.186	0.435	[3.394;3.783]	[-1.125;-1.020]	[0.153;0.218]	[0.404;0.465]	0.55	0.39	0.36	2.06
10	Idem ci-dessus ; échantillon aléatoire de 5 %	1200	3.956	-1.162	0.220	0.413	[3.645;4.267]	[-1.242;-1.081]	[0.174;0.264]	[0.371;0.454]	0.56	0.4	0.37	2.26
11	Idem ci-dessus ; échantillon aléatoire de 1 %	240	4.763	-1.445	0.267	0.52	[4.026;5.380]	[-1.628;-1.261]	[0.175;0.359]	[0.424;0.615]	0.66	0.38	0.33	2.79
12	Flux pondérés par commune dont internes (calcul par rayon) ; actifs 15-64 ans en O et nombre d'emplois en D ; seulement flux de moins de 100 km	23000	4.355	-1.260	0.187	0.446	[4.274;4.435]	[-1.281;-1.238]	[0.176;0.197]	[0.435;0.455]	0.53	0.4	0.37	2.03
13	Idem ci-dessus, mais seulement flux de moins de 50 km	19500	4.511	-1.345	0.208	0.485	[4.413;4.608]	[-1.371;-1.319]	[0.196;0.219]	[0.473;0.495]	0.54	0.41	0.35	2.06
14	Flux pondérés par commune (internes par calcul aléatoire dans rayon) ; actifs 15-64 ans en O et nombre d'emplois en D ; seulement flux de moins de 30 km	15000	4.593	-1.402	0.222	0.516	[4.469;4.716]	[-1.435;-1.369]	[0.209;0.236]	[0.503;0.528]	0.55	0.41	0.33	2.18
15	Flux pondérés par commune dont internes (calcul par rayon) ; actifs 15-64 ans en O et nombre d'emplois en D ; seulement communes ayant plus de 1 flux	23000	3.952	-1.141	0.207	0.405	[3.881;4.023]	[-1.160;-1.122]	[0.196;0.216]	[0.395;0.414]	0.53	0.39	0.35	2.15
16	Flux pondérés par commune (internes par calcul aléatoire dans rayon) ; actifs 15-64 ans en O et nombre d'emplois en D ; seulement communes ayant plus de 5 flux	16500	3.98	-1.183	0.267	0.415	[3.896;4.064]	[-1.206;-1.160]	[0.255;0.278]	[0.404;0.427]	0.55	0.37	0.28	2.13
17	Flux pondérés par commune dont internes (calcul par rayon) ; actifs 15-64 ans en O et nombre d'emplois en D ; seulement communes ayant plus de 10 flux	10000	4.146	-1.214	0.277	0.428	[4.045;4.247]	[-1.244;-1.184]	[0.263;0.292]	[0.414;0.443]	0.53	0.35	0.22	2.19
18	Flux pondérés par commune (internes par calcul aléatoire dans rayon) ; actifs 15-64 ans en O et nombre d'emplois en D ; seulement communes ayant plus de 10 flux et à moins de 30 km	8000	4.41	-1.31	0.275	0.464	[4.273;4.548]	[-1.350;-1.270]	[0.258;0.292]	[0.448;0.481]	0.54	0.36	0.21	2.22

TABLEAU 3. Résultats synthétiques des modèles

679 À titre d'exemple, la figure 8 dessine le nuage comparant les flux observés et les flux théoriques
680 obtenus avec le modèle #18 (flux des communes possédant au moins 10 flux de moins de 30 km
681 à vol d'oiseau).



682 *FIGURE 8. Comparaison des log des flux observés et théoriques pour le modèle #18*

683 5.2. Comparaison des modèles

684 L'interprétation des résultats peut être réalisée à plusieurs niveaux (Tableau 3) : les élasticités des
685 variables du modèle, la qualité des modèles en fonction des effectifs, l'impact du choix des
686 données pour les variables explicatives, l'effet des seuils géographiques dans les résultats obtenus
687 et l'adéquation de la modélisation proposée à la recherche-action partenariale.

688 Quels que soient les modèles, on observe une élasticité négative de la distance et des élasticités
689 positives pour les masses, validant logiquement d'une part la loi de Tobler, d'autre part, l'usage
690 d'un modèle d'attraction de type Newton pour l'analyse et l'estimation de flux domicile-travail.
691 Le rapport entre les élasticités, donnant les poids respectifs des variables dans l'estimation des
692 flux s'avère particulièrement intéressant : l'effet de la distance entre les entités sur la quantité de
693 migrations alternantes est en moyenne 4 à 5 fois supérieur à celui de la masse de l'entité à
694 l'origine du flux et 2 à 3 fois supérieur à celui de la masse de l'entité destination. Par conséquent,
695 les caractéristiques de l'entité de destination interviennent jusqu'à 2,5 fois plus que l'entité
696 d'origine dans la prévision du flux. C'est donc l'attractivité de la commune de travail qui est
697 déterminante par rapport à son émissivité, très fortement régulée par la distance. En ce sens, on
698 s'éloigne sensiblement des modèles de la loi universelle de gravitation de Newton ou d'approches
699 de type Reilly, où la distance est affectée d'un exposant 2 (soit un impact seulement double de la
700 distance par rapport aux masses). Si les valeurs des élasticités varient d'un modèle à l'autre, les
701 rangs et les ordres de grandeur sont conservés, montrant une grande cohérence dans les résultats,
702 et ce en dépit parfois de grandes variations d'effectifs dans les échantillons.

703 Plusieurs variations de critères ont permis de modifier les effectifs des modèles :
704 l'échantillonnage aléatoire, les seuils de quantité de flux par entité et de distance des flux, et, dans
705 une moindre mesure, la conservation ou l'adaptation des flux internes des entités (définition d'un
706 cercle théorique de déplacement avec, selon les cas, positionnement aléatoire de la destination).
707 Les modèles #6 à #11 décrivent une réduction progressive des effectifs des échantillons de 100 %
708 à 1 %. Le bon R^2 obtenu par le modèle #11 est un artefact statistique (effectif de 240 et faible
709 homoscedasticité, cf. Figure 7), contrecarré par l'entropie des résidus. Il est toutefois très
710 intéressant de constater que l'on peut descendre jusqu'à 5 %, soit près de 1 200 individus choisis
711 au hasard dans l'échantillon de référence (de 24000 individus), sans réellement perdre en qualité
712 d'ajustement ($R^2 > 0,5$) même si l'on voit s'élargir progressivement l'intervalle de confiance,
713 restant somme toute très réduit (modèle #10 : valeur estimée de l'intercept de 3,95 ; intervalle de
714 confiance de [3,64-4,26]). Il est donc possible, dans la plupart des situations observées, de réduire
715 l'échantillon de manière assez drastique, sans grande perte de qualité (notamment avec une bonne
716 conservation de la normalité des résidus). Cela est conforme à la loi des grands nombres
717 appliquée à ce type de données géographiques, qui stipule qu'assez rapidement, si un échantillon
718 est représentatif, les statistiques de l'échantillon convergent avec celles de la population. Notons
719 enfin que, parmi les variables explicatives, le test VIF a la plus petite valeur pour M_i (la masse de
720 l'entité d'origine), alors que c'est précisément cette même variable qui a l'élasticité la plus faible.
721 Par ricochet, cela met en valeur la consistance de la relation des deux autres variables
722 explicatives (distance entre les deux entités et masse de l'entité de destination) avec la variable à
723 expliquer (les flux).

724 Observons maintenant l'impact du choix des variables explicatives. Tout d'abord, il est étonnant
725 de constater une baisse du R^2 lorsqu'on utilise les flux pondérés (modèle #2). En y regardant de
726 plus près, on note une plus grande variabilité des indices RMSE, SRMSE et de l'entropie du
727 modèle #1 (faisant le choix du décompte simple des individus par entité pour estimer les flux
728 observés). De plus, ce premier modèle diverge singulièrement des autres en termes d'élasticité,
729 en équilibrant le poids respectif des entités en origine vs en destination. La méthode d'agrégation
730 de données, ne faisant pas l'objet d'un redressement par pondération, peut potentiellement éluder
731 les différences d'attractivités entre les communes rurales (ou de petites tailles) et les plus grandes
732 villes, sans compter les problèmes de baisse de représentativité statistique.

733 De son côté, la population active prise comme masse en origine et en destination grève
734 partiellement la qualité du modèle (surtout sur les masses de destination, moins pour les origines,
735 comme dans le modèle #5), puisque son remplacement par la capacité d'emplois pour les entités
736 de destination (modèle #6) suffit à ajouter environ 8 % d'explication de la variance du modèle.
737 Cela reste toutefois en deçà de ce à quoi l'on aurait pu s'attendre. La conservation et l'adaptation
738 des distances à vol d'oiseau montre un léger gain dans l'ajustement du modèle (de 3 % pour le
739 rayon simple du modèle #3, à 4% pour le jittering dans le rayon du modèle #4) : le traitement du
740 zéro dans les flux doit être réalisé, ne serait-ce que pour prendre en compte ces flux
741 statistiquement peu nombreux (diagonale de la matrice des flux), mais massifs en valeurs.

742 Discutons maintenant de l'effet des seuils sur les quantités de flux et les distances, en lien avec
743 les cartes présentées dans la section §3. La mise en œuvre d'un seuil du nombre minimal de flux,
744 respectivement 1 pour le modèle #15, 5 pour le modèle #16 et 10 pour le modèle #17, ne joue que
745 très peu sur les résultats mais réduit l'échantillon. De la même façon, que l'on enlève les flux de
746 plus de 100 km (modèle #12), de 50 km (modèle #13) ou de 30 km (modèle #14), cela ne change
747 pas beaucoup la donne, sauf à l'amélioration de l'homoscedasticité, en dépit de la réduction des
748 effectifs des échantillons. De même, la combinaison des deux seuils (effectif minimal de flux

749 communal fixé à 10 et distance maximale de 30 km, modèle #18), conserve de bons résultats en
750 réduisant la taille de l'échantillon à 8 000 (1/3), mais en élargissant en conséquence les
751 intervalles de confiance. Dans tous ces cas de figure, on peut toutefois se demander si cela
752 constitue une piste d'amélioration. D'un côté, on économise certes de l'énergie en ciblant et en
753 réduisant les flux. D'un autre côté, la continuité et la complétude de l'information statistique ne
754 sont pas conservées.

755 Conclusion

756 Les résultats obtenus dans cette recherche-action dévoilent quelques éléments méthodologiques
757 intéressants, pour ce qui concerne les migrations alternantes régionales, c'est-à-dire les flux dont
758 les motifs sont le travail, et potentiellement les études, et sont réalisés au sein même des
759 communes et des villes de la région Provence-Alpes-Côte d'Azur.

760 Les modèles de type gravitaire, et, notamment le modèle généralisé de Newton, sont en effet
761 utilisables, avec une qualité globale de prédiction tout à fait correcte, pour autant que les modèles
762 soient ajustés avec des coefficients pondérant les effets des différents critères, et appliqués sur
763 des variables appropriées. Cela valide les effets croisés de l'émissivité et de l'attractivité des
764 communes et de la friction affectant les déplacements pendulaires. Cette méthode peut d'ailleurs
765 potentiellement s'adapter à tout type d'analyse de migration et de mobilité.

766 Les analyses statistiques des modèles et exploratoires des flux révèlent que plus la distance est
767 grande et moins les flux sont nombreux, ce qui semble relativement évident d'un point de vue
768 géographique. Plus précisément, les modèles montrent que l'effet de la distance est jusqu'à 5 fois
769 supérieur à celui du poids des communes ou arrondissements municipaux de la région Provence-
770 Alpes-Côte d'Azur. La distance demeure donc un paramètre majeur ! Peut-être le développement
771 extensif du télétravail amènera dans quelques années à un changement marqué des pratiques de
772 mobilité pendulaire... que des modèles gravitaires adaptés permettront toujours d'estimer ?

773 Les flux d'effectifs intermédiaires, entre 10 et 100 par jour et par commune, structurent la forme
774 statistique de l'analyse, car ils sont suffisamment nombreux pour rendre le modèle robuste. Ils
775 représentent en quelque sorte une forme de rationalité spatio-temporelle d'*homo mobilis*. Ils
776 expliquent en grande partie la relative stabilité du modèle, pour lequel, nous l'avons vu, les
777 indicateurs de qualité n'évoluent que de 10 % environ. C'est dire le poids de la structure
778 topologique des communes et de leurs attractivités relatives sur le déploiement et l'organisation
779 spatiale des flux.

780 Contrairement à d'autres modèles (LUTI, SMA), le modèle gravitaire utilisé ici ne prend pas en
781 compte les modes de transport disponibles, ni l'état du trafic, facteurs pourtant très contraignants
782 sur les mobilités. Dans des travaux ultérieurs à objectif essentiellement scientifique et
783 méthodologique, il serait par ailleurs intéressant de considérer la distance réseau, ou bien
784 d'utiliser des régressions plus robustes (régression de Poisson, régressions robustes, régressions
785 par quantiles, etc.).

786 Posons-nous maintenant la question de l'adéquation de la modélisation proposée à la recherche-
787 action conduite avec le partenaire régional. Si l'on vise la parcimonie et la généralisation, le
788 niveau minimal d'un modèle « économe » pourrait être d'utiliser la distance à vol d'oiseau,

789 complétée par la méthode simple des rayons pour prendre en compte les flux internes, la
790 population comme masse d'entité d'origine et la quantité d'emplois comme masse d'entité de
791 destination. En complément, une réduction jusqu'à 10 % de l'échantillon des flux observés par
792 tirage aléatoire suffirait à obtenir de bons estimateurs, en réduisant de façon importante le coût
793 des traitements, à la condition de pouvoir appliquer une pondération aux populations enquêtées.
794 De plus, si ces modèles très économes en termes de données nécessaires montrent de bons
795 résultats du point de vue global, ils induisent certainement des erreurs locales dans les prévisions
796 qui peuvent parfois être importantes.

797 Enfin, notons que les résultats ne portent que sur les migrations alternantes, ne représentant
798 qu'une relativement faible part de l'ensemble des mobilités dans leur ensemble, et que les
799 pérégrinations (déplacements à arrêts et motifs multiples) ne sont pas appréhendées. Une prise en
800 compte de l'ensemble des flux en PACA, via l'acquisition de données massives de suivi de
801 véhicules par exemple, constituerait un travail de plus grande envergure, pouvant
802 avantageusement compléter la modélisation parcimonieuse prônée dans le contexte de cette
803 étude.

804 **Remerciements**

805 Les auteurs de cet article remercient vivement :

806 - la Région Provence-Alpes-Côte d'Azur pour son partenariat et son soutien financier dans ce
807 projet de recherche-action pédagogique impliquant notamment les étudiant.e.s du Coursus Master
808 Ingénierie « Modélisation Géographique des Territoires » d'Avignon Université ;

809 - Laetitia Faure pour le travail de mise en forme de l'article dans le cadre du transfert de LaTeX à
810 Word.

811 **Bibliographie**

812 Amarouch S., Durand J.-B., Josselin D., Garcin L., Brachet N., Audard F., Carpentier-Postel S..
813 (2019). Estimer des flux de navetteurs avec un modèle gravitaire. SAGEO 2019 : Spatial
814 Analysis and Geomatics, Nov 2019, Clermont-Ferrand, France. (halshs-02528166), résumé 4
815 pages et poster.

816 Bonnel P., Hombourger É., Olteanu-Raimond A.-M., & Smoreda, Z. (2017). Apports et limites
817 des données passives de la téléphonie mobile pour la construction de matrices origine-
818 destination. *Revue d'Economie Regionale et Urbaine*, Octobre(4), 647-672.

819 Brilhault G. et Caron N., (2016). Le passage à une collecte par sondage : quel impact sur la
820 précision du recensement ? *Économie et statistique*. N°483-484-485, 23-40.

821 Burger M., van Oort F., Meijers E., (2019). Examining Spatial Structure Using Gravity Model, in
822 *The Mathematics of Urban Morphology* (Ed. Luca d'Acci), Modeling and Simulation in
823 Science, Engineering and Technology, Birkhäuser, pp. 471-479.

- 824 Cervero R. (2001). Efficient Urbanisation: Economic Performance and the Shape of the
825 Metropolis. *Urban Studies*, 38(10): 1651-1671. <https://doi.org/10.1080/00420980120084804>
- 826 Commenges H., Commenges D. (2017). La question du zéro dans le modèle gravitaire. 13^e Actes
827 des Rencontres de Théo Quant, Besançon, 17-19 Mai 2017. 294-298. [http://thema.univ-
828 fcomte.fr/theoq/pdf/2017/communications/2017_Commenges_question_zero_modele_gravitai
829 re.pdf](http://thema.univ-fcomte.fr/theoq/pdf/2017/communications/2017_Commenges_question_zero_modele_gravitai)
- 830 Converse P. D. (1951). Development of marketing: Fifty years of progress. In *H. g. wales,
831 changing perspectives in marketing*, University of Illinois.
- 832 Faivre S. (2017). L'échantillonnage des enquêtes auprès des ménages dans le recensement annuel
833 de la population. Département des méthodes statistiques. INSEE. Mesurer pour comprendre.
834 Version n°1. 10 octobre 2017. 4 pages.
- 835 Huff D. L. (1963). A Probabilistic Analysis of Shopping Center Trade Areas. *Land Economics*,
836 39(1): 81-90. <http://doi.org/10.2307/3144521>
- 837 INSEE. (2015, 2016). Fichiers des « mobilités professionnelles ». <https://www.insee.fr/fr/information/2383243>
838
- 839 INSEE. (2019). Recensement de la population. Conseils d'utilisation – Synthèse. 19 pages.
- 840 Joly I. (2005). *L'allocation du temps au transport : de l'observation internationale des budgets-
841 temps de transport aux modèles de durées*. Laboratoire d'Economie des Transports.
- 842 Josselin D., Marcon A., Audard F., Carpentier-Postel S. (2020). Modélisation et analyse spatiales
843 des pratiques de mobilité partagée et de transport flexible en région PACA. Rapport de
844 recherche. Région Provence-Alpes-Côte d'Azur. (halshs-02526409).
- 845 Josselin D., Nicot B. (2003). Un modèle gravitaire géoéconomique des échanges commerciaux
846 entre les pays de l'U.E., les PECO et les PTM. *Cybergeo : European Journal of Geography*.
- 847 Lenormand M., Bassolas A., & Ramasco J. J. (2016). Systematic comparison of trip distribution
848 laws and models. *Journal of Transport Geography*, 51, 158-169.
849 <https://doi.org/10.1016/j.jtrangeo.2015.12.008>
- 850 Li M., Kwan M.-P., Wang F., & Wang J. (2018). Using points-of-interest data to estimate
851 commuting patterns in central Shanghai, China. *Journal of Transport Geography*, 72,
852 201-210. <https://doi.org/10.1016/j.jtrangeo.2018.09.004>
- 853 Liu, Y. Fang, F. & Jing Y. (2020). How urban land use influences commuting flows in Wuhan,
854 Central China : A mobile phone signaling data perspective. *Sustainable Cities and Society*, 53,
855 101914. <https://doi.org/10.1016/j.scs.2019.101914>
- 856 Newman P.G.W., Kenworthy J.R. (1989). *Cities and Automobile Dependence: An International*

- 857 *Sourcebook*. Aldershot, UK: Gower.
- 858 Ravenstein E. G. (1885). The Laws of Migration. *Journal of the Statistical Society of London*,
859 48(2): 167-235. <http://doi.org/10.2307/2979181>
- 860 Reilly W. (1931). *The law of retail gravitation*. Knickerbocker Press.
- 861 Stouffer S. A. (1940). Intervening opportunities: A theory relating mobility and distance.
862 *American Sociological Review*, vol. 5, no 6, p. 845–867.
- 863 Stewart J. Q. (1947). Empirical Mathematical Rules concerning the Distribution and Equilibrium
864 of Population. *Geographical Review*, 37(3): 461-485. <http://doi.org/10.2307/211132>
- 865 Tobler W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region.
866 *Economic Geography*, 46: 234-240. <http://doi.org/10.2307/143141>.
- 867 Ubøe J. (2004). Aggregation of Gravity Models for Journeys to Work. *Environment and*
868 *Planning A: Economy and Space*, 36(4): 715-729. <https://doi.org/10.1068/a3635>
- 869 Wilson A. G. (1975). Some new forms of spatial interaction model: A review. *Transportation*
870 *Research*, 9(2): 167-179. [http://doi.org/10.1016/0041-1647\(75\)90054-4](http://doi.org/10.1016/0041-1647(75)90054-4)
- 871 Zahavi Y. et Talvitie A., (1980). Regularities in travel time and money expenditures,
872 *Transportation Research Record*, 750 :13-19.