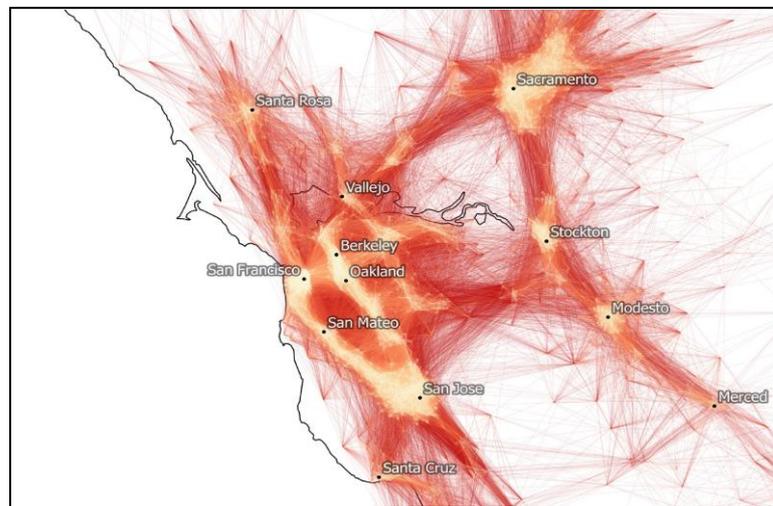


Comparaison du modèle gravitaire et radiatif dans la modélisation des flux de pendulaires

Christian Stadler

Sous la direction du Prof. François Bavaud



Source : Nelson, G.D., Rae, A. (2016). An Economic Geography of the United States : From Commuters to Megaregions. *PLOS ONE*, 11(11): e0166083. doi:10.1371/journal.pone.0166083

Abstract

The aim of this thesis is to compare the recently introduced radiation model with the traditional gravity model in the prediction of commuting flows at different spatial scales. Three different datasets are used to test the accuracy and robustness of the models, whose goodness of fit is assessed through different measures. It is shown that both the radiation and the gravity model fail to predict commuting flows in intra-urban areas such as the city of New York, whereas the overall performance of the gravity model is slightly better. The radiation model is indeed less flexible to changes in spatial scales, but gives competitive results in specific case studies.

Keywords : Gravity model, radiation model, commuting flows

Résumé

Le but de ce travail de mémoire est de comparer le récemment introduit modèle radiatif avec le traditionnel modèle gravitaire dans le cadre de la prédiction de flux pendulaires à différentes échelles spatiales. Trois jeux de données sont utilisés afin de tester la précision et la robustesse des modèles, dont l'ajustement aux données théoriques est évalué à travers différentes mesures. Il est montré que le modèle radiatif ainsi que le modèle gravitaire échouent dans la prédiction de flux pendulaires à une échelle intra-urbaine telle la ville de New York, alors que la performance globale du modèle gravitaire est légèrement meilleure. Le modèle radiatif est en effet moins flexible aux changements d'échelle d'analyse, mais donne des résultats compétitifs dans certains cas d'étude.

Mots clés : Modèle gravitaire, modèle radiatif, flux pendulaires

Table des matières

1	Introduction	5
2	Revue de littérature	6
2.1	Le modèle gravitaire	6
2.2	Le modèle radiatif	7
2.3	Les modèles contraints	8
2.4	Méthodes d'ajustement	8
2.4.1	L'indice de Sørensen-Dice	9
2.4.2	L'entropie relative	9
2.4.3	L'erreur quadratique moyenne normalisée	9
2.4.4	L' <i>Akaike Information Criterion</i>	10
2.5	Estimation des paramètres du modèle gravitaire basé sur une distribution de Poisson	10
2.5.1	Le modèle log-normal	10
2.5.2	Le modèle de Poisson	11
2.6	Estimation des paramètres dans le cadre de l'absence de données détaillées	12
3	Problématique	13
3.1	Hypothèses	13
4	Méthodologie	14
4.1	Présentation des données	14
4.2	Outils d'analyse	15
5	Résultats	16
5.1	États-Unis	16
5.1.1	Flux pendulaires sortant du comté de Queens - 2000	16
5.1.2	Flux pendulaires sortant du comté de Queens - 2013	24

5.1.3	Prédiction de flux pendulaires pour l'ensemble des États-Unis . . .	31
5.2	Suisse	32
5.2.1	Flux pendulaires sortant de la commune de Zurich	32
5.2.2	Prédiction de flux pendulaires pour l'ensemble de la Suisse	40
6	Discussion	41
7	Conclusion	43

Table des figures

1	Découpage spatial à l'échelle des communes	14
2	Découpage spatial à l'échelle des comtés	15
3	Flux pendulaires originaire du comté de Queens - 2000 (données effectives et prédictions)	17
4	Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle gravitaire (2000) - paramètres $[\alpha, \beta, \gamma] = [0.41, 0.85, 2.27]$	19
5	Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle radiatif classique (2000)	21
6	Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle radiatif paramétré (2000) - paramètre $[\alpha] = [0.96]$	22
7	Flux pendulaire originaires du comté de Queens - 2013 (données effectives et prédictions)	25
8	Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle gravitaire (2013) - paramètres $[\alpha, \beta, \gamma] = [0.41, 0.85, 2.27]$	27
9	Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle radiatif classique (2013)	29
10	Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle radiatif paramétré (2013) - paramètre $[\alpha] = [0.96]$	30
11	Comparaison des prédictions de flux pendulaires à l'échelle des Etats-Unis pour les données de 2000 (première rangée) et de 2013 (deuxième rangée)	33
12	Flux pendulaires originaires de la commune de Zurich (données effectives et prédictions)	34
13	Résidus de la prédiction de flux pendulaires originaires de la commune de Zurich par modèle gravitaire (2000) - paramètres $[\alpha, \beta, \gamma] = [0.67, 1.14, 1.85]$	36
14	Résidus de la prédiction de flux pendulaires originaires de la commune de Zurich par modèle radiatif (2000)	38
15	Résidus de la prédiction de flux pendulaires originaires de la commune de Zurich par modèle radiatif paramétré (2000) - paramètre $[\alpha] = [1.3]$	39
16	Comparaison des prédictions de flux de pendulaires à l'échelle de la Suisse	41

Liste des tableaux

1	Mesures d'ajustement - Flux pendulaires sortant du comté de Queens (2000)	23
2	Mesures d'ajustement - Flux pendulaires sortant du comté de Queens (2013)	31
3	Mesures d'ajustement - Flux pendulaires pour l'ensemble des États-Unis .	31
4	Mesures d'ajustement - Flux pendulaires sortant de la commune de Zurich (2000)	40
5	Mesures d'ajustement - Flux pendulaires pour l'ensemble de la Suisse . . .	41

1 Introduction

Un nombre important de travaux cherchant à prévoir les déplacements domicile-travail ont été réalisés durant ces dernières années, le phénomène de la pendularité occupant de plus en plus d'importance, notamment en Suisse¹. En effet, les temps ainsi que les durées de déplacements entre domicile et lieu de travail sont en hausse, générant d'intenses réseaux de relations socio-économiques. Ces derniers sont le vecteur de diverses dynamiques sociales et économiques, tels le trafic, les flux d'information ainsi que la propagation de maladies (Lenormand et al., 2012). L'estimation des flux ayant historiquement comme application principale l'ingénierie du trafic ainsi que la géographie et l'économie du transport (Commenges, 2016), il convient d'estimer au mieux ces patterns de mobilités humaines afin de faciliter le développement urbain, la planification des réseaux de transport et de distribution ainsi que l'aide à la décision politique (Liang et al., 2013).

Une multitude de modèles d'interactions spatiales ont été proposés dans le cadre de l'estimation des flux de pendulaires. Le modèle d'opportunités intervenantes de Stouffer, qui présuppose que le nombre de personnes se déplaçant sur une distance donnée est directement proportionnel à l'augmentation du pourcentage d'opportunités à cette distance (Stouffer, 1940), va poser les bases du récemment introduit modèle radiatif. Le modèle gravitaire de Zipf (1946), utilisé dans divers domaines autres que la géographie, telle la prédiction de flux commerciaux, est remis en question par le modèle radiatif de Simini et al. (2012). Effectivement, le modèle gravitaire et radiatif ont été le sujet de nombreuses comparaisons, comme par exemple dans les articles de Lenormand et al. (2012, 2016), Masucci et al. (2013), et Commenges (2016). Le but du présent travail est d'explicitier, en mettant avant différentes méthodes d'estimation et de calibration présentées dans la littérature (Flowerdew and Aitkin, 1982; Flowerdew and Lovett, 1988; Lenormand et al., 2012; Yang et al., 2014), en quoi le modèle radiatif diffère du modèle gravitaire dans le cadre de la modélisation de flux de pendulaires à l'échelle des USA et de la Suisse. Les résultats obtenus seront ensuite comparés à des jeux de données étasuniens et suisses, les performances du modèle gravitaire et radiatif n'ayant jamais été comparées dans le contexte helvétique.

1. La pendularité en Suisse en 2013 (2015, 23 juin). *Office fédéral de la statistique*. Repéré à : <http://www.bfs.admin.ch/bfs/portal/fr/index/themen/11/04/blank/04.html> (consulté le 15 avril 2016)

2 Revue de littérature

2.1 Le modèle gravitaire

Le principal modèle utilisé dans la modélisation de flux pendulaires est le traditionnel modèle gravitaire, popularisé par Zipf en 1946². Ce dernier, en analogie à la loi universelle de la gravitation de Newton, présuppose que le nombre de pendulaires T_{ij} entre une origine i et une destination j est proportionnel à la population à l'origine m_i et la population à destination n_j , et décroît avec la distance r_{ij} entre ces derniers, où α , β et γ sont les paramètres positifs du modèle et A un facteur de normalisation :

$$T_{ij} = A \frac{m_i^\alpha n_j^\beta}{r_{ij}^\gamma} \quad (1)$$

où

$$A = \frac{T}{\sum_{ij} \frac{m_i^\alpha n_j^\beta}{r_{ij}^\gamma}} \quad (2)$$

avec

$$T = T_{\bullet\bullet} \quad (3)$$

Malgré l'utilisation très fréquente du modèle susmentionné dans plusieurs domaines de recherche, Simini et al. (2012) mettent en avant un certain nombre de limitations liées au modèle gravitaire :

1. Il n'existe pas de dérivation rigoureuse de (1).
2. L'absence de lignes directrices théoriques entraîne l'utilisation de nombreux paramètres afin d'ajuster le modèle
3. La calibration des paramètres nécessite des données spécifiques (qui s'avèrent ne pas être disponibles dans certaines régions d'étude)
4. Le modèle gravitaire présente systématiquement des divergences prédictives
5. Selon (1), le nombre de pendulaires estimés augmente sans limite en fonction de la population de destination n_j , alors que le nombre de pendulaires ne peut dépasser la population à l'origine m_i
6. Le modèle étant déterministe, il ne peut expliquer des fluctuations dans le nombre de pendulaires entre deux emplacements

2. Zipf, G.K. (1946) The P1 P2/D Hypothesis : On the Intercity Movement of Persons. *American Sociological Review*, 11(6), 677-686

Néanmoins, il convient de remettre en question certaines de ces limitations. En effet, Wilson (1969) propose une dérivation du modèle gravitaire par le maximum d'entropie dans ses travaux. De plus, la limitation six ne se limite pas uniquement au modèle gravitaire, mais également aux modèles radiatifs (4) et (5), présentés par la suite.

Un autre élément problématique dans le cadre des modèles proposés dans ce travail est la prédiction des flux internes de chaque unité spatiale, soit des flux diagonaux. Comme évoqué dans Lenormand et al. (2016), il n'est pas possible de prédire les flux intra-unités avec le modèle radiatif. Effectivement, alors qu'il est possible d'approximer les distances internes dans le cadre du modèle gravitaire (par exemple à partir de la moitié de la racine carrée de la surface de l'unité spatiale) (Lenormand et al., 2016), le modèle radiatif quant à lui ne se prête pas à des estimations intra-unités. De ce fait, les flux diagonaux ne sont pas considérés, de sorte à ce que $T_{ii} = 0$.

2.2 Le modèle radiatif

Afin de dépasser ces limitations, Simini et al. (2012) introduisent un nouveau modèle sans paramètre capable de prédire des flux de mobilité, le modèle radiatif. Basé sur les processus de radiation et d'absorption et nécessitant uniquement des informations sur la distribution de la population, il s'inspire du modèle d'opportunités intervenantes de Stouffer (1940), qui suppose que le nombre de déplacements est relié au nombre d'opportunités entre origine et destination plutôt qu'à la distance les séparant (Yang et al., 2014).

Soit T_{ij} le nombre de pendulaires entre une origine i et une destination j avec une population respective m_i et n_j à distance r_{ij} l'une de l'autre, avec s_{ij} correspondant à la population totale dans le cercle de rayon r_{ij} centré en i et $T_{i\bullet} = \sum_{j \neq i} T_{ij}$. Ce modèle

supposant un nombre infini de flux, il convient d'ajouter un facteur de normalisation dans la cadre d'un système fini, où $M = \sum_i m_i$ correspond à la population totale de l'échantillon, comme présenté par Masucci et al. (2013) :

$$T_{ij} = \frac{T_{i\bullet}}{1 - \frac{m_i}{M}} \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \quad (4)$$

Une extension de ce modèle incluant un paramètre α afin de contrôler l'effet du nombre d'opportunités de travail entre la source et l'origine a également été proposé par Yang et al. (2014) et Lenormand et al. (2016) :

$$T_{ij} = \frac{[(m_i + m_j + s_{ij})^\alpha - (m_i + s_{ij})^\alpha](m_i^\alpha + 1)}{[(m_i + s_{ij})^\alpha + 1][(m_i + m_j + s_{ij})^\alpha + 1]} \quad (5)$$

N'ayant pas de temps de déplacement à disposition, les distances r_{ij} et s_{ij} seront mesurées à vol d'oiseau.

2.3 Les modèles contraints

Les modèles étudiés dans le cadre de ce travail peuvent être considérés à travers différents niveaux de contraintes. En effet, l'ajout d'un facteur de normalisation permet de spécifier une contrainte à l'origine et / ou à destination. Un modèle contraint à l'origine préservera donc le nombre de pendulaires sortant d'une région, alors qu'un modèle contraint à destination maintiendra le nombre de pendulaires attirés par une région. Le modèle doublement contraint, qui préserve le nombre de pendulaires générés et attirés par une région, nécessite une implémentation itérative afin de reconstruire une matrice OD empirique (Yang et al., 2014). Finalement, le modèle non-contraint préservera quant à lui uniquement le nombre total de déplacements.

Un niveau de contrainte plus élevé signifiant généralement de meilleures performances prédictives (Yang et al., 2014), il semble important de comparer des modèles ayant le même niveau de contrainte, comme souligné dans Lenormand et al. (2016). De ce fait, les modèles gravitaire et radiatif seront considérés comme étant contraints à l'origine, ce qui s'avère pertinent dans le contexte de la modélisation d'interactions spatiales telles les flux de pendulaires (Flowerdew and Lovett, 1988).

Le modèle gravitaire contraint à l'origine pourra prendre la forme suivante (où $f(r_{ij}) = r_{ij}^{-\gamma}$) :

$$T_{ij} = O_i A_i n_j f(r_{ij}) \quad (6)$$

où

$$A_i = \frac{1}{\sum_k n_k f(r_{ik})} \quad (7)$$

et

$$\sum_{j=1}^n T_{ij} = O_i \quad (8)$$

ce qui correspond à la relation suivante :

$$T_{ij} = O_i \frac{n_j f(r_{ij})}{\sum_k n_k f(r_{ik})} \quad (9)$$

avec O_i correspondant à la somme des flux sortant et A_i au facteur de normalisation spécifique à l'origine.

Le modèle radiatif contraint à l'origine peut quant à lui être écrit de la forme suivante :

$$T_{ij} = \frac{O_i}{\sum_k \frac{m_i n_k}{(m_i + s_{ik})(m_i + n_k + s_{ik})}} \cdot \frac{m_i n_k}{(m_i + s_{ik})(m_i + n_k + s_{ik})} \quad (10)$$

2.4 Méthodes d'ajustement

Afin d'apprécier la qualité prédictive des différents modèles étudiés, il convient de mesurer l'écart entre les données estimées et les données observées. Pour ce faire, des

mesures tels l'indice de Sørensen (1948), l'erreur quadratique moyenne normalisée ainsi que l'entropie relative (divergence de Kullback-Leibler) sont utilisées.

2.4.1 L'indice de Sørensen-Dice

Lenormand et al. (2012) proposent un indice de similarité construit à partir de l'indice de Sørensen, le *Common part of commuters (CPC)*. Ce dernier, qui a également été repris sous une autre forme par Masucci et al. (2013), permet de déterminer quelle part des flux observés est correctement reproduite par les flux estimés. Sa valeur est comprise entre 0 et 1, 0 correspondant à aucune relation entre les deux sets de données et 1 à deux sets de données identiques. Lorsque la somme des flux observés correspond à celle des flux estimés, tel que $\sum_{i,j}^n T_{ij} \simeq \sum_{i,j}^n \hat{T}_{ij}$, le CPC peut être interprété comme le pourcentage de connections de pendulaires correctement reproduit en termes d'origine et destination (Lenormand et al., 2012) :

$$CPC(T_{ij}, \hat{T}_{ij}) = \frac{2 \sum_{ij}^n \min(T_{ij}, \hat{T}_{ij})}{\sum_{ij}^n T_{ij} + \sum_{ij}^n \hat{T}_{ij}} \quad (11)$$

2.4.2 L'entropie relative

La divergence de Kullback-Leibler (Kullback and Leibler, 1951) est quant à elle une mesure de dissimilarité entre ici, deux tables de contingence T et \hat{T} de même effectif total $T_{\bullet\bullet}$, prenant la forme suivante :

$$D_{KL}(T||\hat{T}) = \frac{1}{T_{\bullet\bullet}} \sum_{ij}^n T_{ij} \ln \frac{T_{ij}}{\hat{T}_{ij}} \quad (12)$$

Egalement appelée *information gain statistic*, cette mesure non-symétrique prend une valeur minimale de 0 lorsque $T_{ij} = \hat{T}_{ij}$, et une valeur maximale positive infinie lorsque $T_{ij} > 0$ et $\hat{T}_{ij} = 0$ (Knudsen and Fotheringham, 1986). De plus, Knudsen and Fotheringham (1986) montrent qu'au vu de la transitivité de la relation, telle que

$$\sum_j^n \left(\sum_i^n T_{ij} \ln \frac{T_{ij}}{\hat{T}_{ij}} \right) = \sum_i^n \left(\sum_j^n T_{ij} \ln \frac{T_{ij}}{\hat{T}_{ij}} \right) \quad (13)$$

la mesure permet d'apprécier le degré d'ajustement pour des lignes et colonnes individuelles au sein d'une matrice.

2.4.3 L'erreur quadratique moyenne normalisée

Une autre mesure permettant d'évaluer le degré d'ajustement d'un modèle est l'erreur quadratique moyenne normalisée (*normalized mean square error*), telle que proposée par

Lenormand et al. (2016) :

$$\text{NMSE}(T, \hat{T}) = \frac{\sum_{i,j=1}^n (T_{ij} - \hat{T}_{ij})^2}{\sum_{i,j=1}^n T_{ij}} \quad (14)$$

Cette dernière a une valeur minimale de 0 dans le cas où les prédictions sont précisément identiques aux données observées, et une limite supérieure dépendant de la distribution des flux T_{ij} .

2.4.4 L' Akaike Information Criterion

Dans la cadre de l'appréciation de la qualité prédictive des différents modèles, il convient de prendre en compte le fait que certains auront des paramètres indépendants supplémentaires, ce qui se traduira généralement par de meilleures performances. De ce fait, il importe de pénaliser les modèles en fonction du nombre de paramètres ajustables à travers l'utilisation d'un critère tel l' *Akaike Information Criterion* :

$$\text{AIC}(k) = -2 \ln L(\hat{\theta}) + 2k \quad (15)$$

où k correspond au nombre de paramètres ajustables, ou de *dimensions* du modèle et $L(\hat{\theta})$ au maximum de la fonction de vraisemblance.

Le principe de la mesure est de trouver un compromis entre le nombre de paramètres libres, qui correspondent au critère de pénalité, et le log du maximum de la fonction de vraisemblance, qui correspond au critère d'insuffisance en termes d'ajustement (Bozdogan, 2000). Le modèle ayant l'AIC le plus faible sera donc privilégié.

2.5 Estimation des paramètres du modèle gravitaire basé sur une distribution de Poisson

2.5.1 Le modèle log-normal

Afin d'estimer les paramètres du modèle gravitaire proposé en (1), la méthode la plus traditionnelle (et la plus populaire) est de recourir au modèle log-normal. Ce dernier, présenté ci-dessous, permet une estimation des valeurs de α , β et γ à travers une régression linéaire multiple.

$$\ln T_{ij} = \ln k + \alpha \ln m_i + \beta \ln n_j - \gamma \ln r_{ij} + e_{ij} \quad (16)$$

Néanmoins, plusieurs auteurs tels Flowerdew and Aitkin (1982) et Arvis and Shepherd (2013) ont mis en avant un certain nombre de biais liés à l'utilisation des modèles log-linéaires dans le cadre de l'estimation des paramètres du modèle gravitaire.

Une première limitation, mise en avant par Flowerdew and Aitkin (1982), concerne la

non-conservation du total des flux. En effet, la régression produisant une estimation du logarithme des T_{ij} et non des T_{ij} en soi, la somme des flux estimés sera considérablement plus faible que la somme des flux observés (45467 vs 89101 dans l'exemple présenté dans Flowerdew and Aitkin (1982))³.

De plus, le processus de génération des flux T_{ij} implique une distribution des valeurs selon une loi log-normale, ce qui ne correspond pas forcément à la réalité, les flux de pendulaires pouvant uniquement être des nombres entiers non-négatifs (Flowerdew and Aitkin, 1982).

Le troisième problème, mise en avant dans Flowerdew and Aitkin (1982) et Arvis and Shepherd (2013), est que l'estimation des paramètres présente un biais dans le cas d'hétéroscédasticité du terme d'erreur e_{ij} .

Finalement, les flux nuls entre deux unités présentent également un problème majeur dans le cadre d'un loi log-normale, ces derniers ne pouvant pas être transformés en logarithme. Dans le cas des flux pendulaires suisses et étasuniens, ces derniers sont très nombreux (près de 97% des flux sont nuls dans la matrice des flux observés suisses, 98% dans le cas des USA). En effet, il semble tout à fait naturel qu'il n'existe pas de pendulaires pour chaque paire d'unité spatiale. Le fait d'ajouter une faible valeur à chaque 0, comme proposé dans Flowerdew and Aitkin (1982), peut donc avoir un impact conséquent sur le pouvoir prédictif du modèle.

2.5.2 Le modèle de Poisson

Afin de pallier à ces difficultés, le modèle de régression de Poisson a été mis en avant par Flowerdew and Aitkin (1982) ainsi que Silva and Tenreyro (2006). En effet, s'il l'on considère que le nombre de déplacements T_{ij} suit une loi de Poisson, avec comme moyenne λ_{ij} , alors la distribution de probabilité sera :

$$Pr(T_{ij}|\lambda_{ij}) = \frac{\exp(-\lambda_{ij})\lambda_{ij}^{T_{ij}}}{T_{ij}!} \quad (17)$$

où le paramètre λ_{ij} est lié de façon logarithmique à une combinaison linéaire du log des variables indépendantes (Flowerdew and Aitkin, 1982) :

$$\lambda_{ij} = \exp(\beta_0 + \alpha \ln m_i + \beta \ln n_j + \gamma \ln r_{ij}) \quad (18)$$

et la moyenne des déplacements est égale à la variance (phénomène d'équidispersion) :

$$E(T_{ij}|\lambda_{ij}) = \text{var}(T_{ij}|\lambda_{ij}) = \lambda_{ij} \quad (19)$$

Afin d'apprécier le degré d'ajustement du modèle, la mesure de la *deviance* a été proposée par Nelder and Wedderburn (1992) dans le cadre des modèles linéaires généralisés. Cette dernière permet de déterminer les variations dans les données qui ne sont pas perçues

3. Cela est être en rapport avec l'inégalité de Jensen, qui est due à la concavité de la fonction $\log(x)$.

par le modèle, et prend la forme suivante pour la distribution de Poisson (Flowerdew and Aitkin, 1982) :

$$deviance = 2 \sum_{i=1}^n \sum_{j=1}^n T_{ij} \ln\left(\frac{T_{ij}}{\hat{\lambda}_{ij}}\right) \quad (20)$$

ce qui correspond à

$$2 TD_{KL}(T||\hat{T}) \quad (21)$$

si

$$\hat{T} = \sum_{ij} \hat{T}_{ij} = T = \sum_{ij} T_{ij} \quad (22)$$

qui peut être mis en relation avec (15), où

$$-2 \ln L(\hat{\theta}) = 2 TD_{KL}(T||\hat{T}) \quad (23)$$

et donc

$$AIC = deviance + 2k \quad (24)$$

avec

$$AIC = 2 TD_{KL}(T||\hat{T}) + 2k \quad (25)$$

2.6 Estimation des paramètres dans le cadre de l'absence de données détaillées

Hormis l'estimation des paramètres par régression linéaire multiple, Lenormand et al. (2012) proposent une méthode globale afin de calibrer le paramètre β d'un modèle gravitaire exponentiel. Cette dernière, qui résulte d'une régression linéaire entre le β calibré et le logarithme de la surface moyenne $\langle S \rangle$ (en km^2) du maillage de la région d'étude pour 80 cas d'étude, prend la forme ci-dessous :

$$\beta = 0.315 \langle S \rangle^{-0.177} \quad (26)$$

Dans le cadre de l'extension du modèle radiatif, Yang et al. (2014) proposent quant à eux la relation suivante afin de déterminer le paramètre α :

$$\alpha = \frac{1}{36^{1.33}} \langle S \rangle^{0.665} \quad (27)$$

Ces méthodes de calibration semblent tout à fait pertinentes dans une optique de prédiction de flux, dans laquelle des données telles les flux effectifs ne sont pas disponibles (et donc dans laquelle une prédiction par régression linéaire multiple n'est pas possible).

3 Problématique

De nombreux travaux sont récemment parus dans le cadre de la comparaison de modèles d'interaction spatiale ainsi que dans l'optimisation de ces derniers. La plupart de ces derniers (Lenormand et al., 2012 ; Liang et al., 2013 ; Masucci et al., 2013 ; Yang et al., 2014 ; Kang et al., 2015 ; Lenormand et al., 2016 ; Commenges, 2016) mettent en lumière le modèle radiatif introduit par Simini et al. (2012), aboutissant à une multiplicité de résultats dont les divergences peuvent être expliquées par l'utilisation de méthodes et d'échelles d'étude différentes. En effet, alors que Simini et al. (2012) mettent en avant la supériorité du modèle radiatif, Masucci et al. (2013) et Lenormand et al. (2016) concluent à de meilleurs résultats avec le modèle gravitaire.

Les concepts théoriques décrits dans la revue de littérature vont permettre de comparer de manière rigoureuse les modèles gravitaires et radiatifs dans le cadre de la modélisation des flux de pendulaires, et ce à travers l'intégration de méthodes de calibration et d'ajustement différentes. Les résultats, qui seront établis sur trois jeux de données, seront présentés sous forme de cartes et complétés par les comparaisons d'indices (CPC, DKL, NMSE, AIC) ainsi que de quelques figures.

3.1 Hypothèses

1. Le modèle radiatif donnera une meilleure représentation des trajets à longue distance, alors que le modèle gravitaire sera plus précis dans la prédiction de flux à courte et moyenne distance, comme le montrent les travaux de Masucci et al. (2013) ; Yang et al. (2014) et Lenormand et al. (2016).
2. L'ajustement entre données estimées et observées sera meilleur pour le modèle gravitaire que pour le modèle radiatif au vue des résultats obtenus dans les études précédentes (Masucci et al., 2013 ; Yang et al., 2014 ; Commenges, 2016 ; Lenormand et al., 2016).
3. La sélection d'un modèle radiatif plus complexe en termes de paramètres réduit de manière importante les résidus et améliore l'ajustement aux données empiriques par rapport au modèle radiatif classique, comme mis en avant par (Lenormand et al., 2012 ; Yang et al., 2014 ; Kang et al., 2015)

4 Méthodologie

4.1 Présentation des données

Les estimations de flux de pendulaires seront appliqués à 3 jeux de données : les données pendulaires inter-comtés étasuniennes de 2000 et 2013 ⁴ ainsi que les données pendulaires suisses inter-communales de 2000. ⁵

Les données pendulaires inter-comtés étasuniennes de 2000 ont été utilisées dans un nombre important de travaux de recherche, et serviront de ce fait de set de référence permettant de vérifier l'exactitude des méthodes et outils utilisés afin de les appliquer aux deux autres jeux de données n'ayant pas encore été exploités dans la cadre de la prédiction de flux de pendulaires.

Les comtés étasuniens sont au nombre de 3108 (l'Alaska ainsi que Hawaï ne sont pas pris en compte) avec une superficie moyenne de 1225 km^2 , les communes suisses sont au nombre de 2324 avec une superficie moyenne de 15.15 km^2 . Le nombre de pendulaires est de 1'961'262 pour les données suisses, 34'097'929 pour les données étasuniennes de 2000 et 38'314'141 pour les données étasuniennes de 2013.

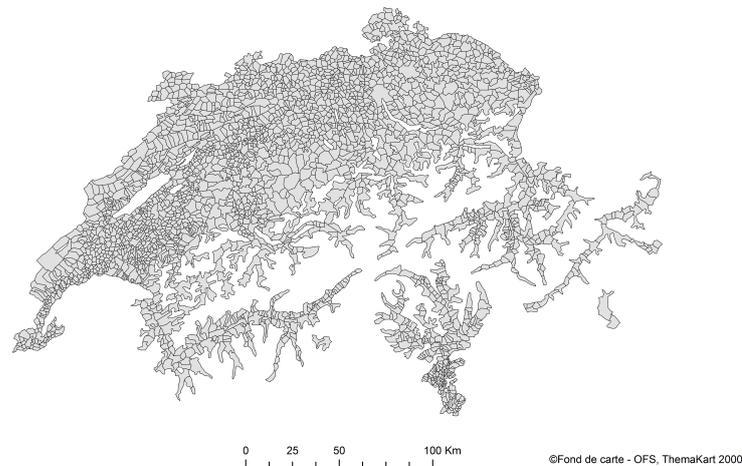


FIG. 1 : Découpage spatial à l'échelle des communes

4. Disponibles en ligne sur : <https://www.census.gov/population/www/cen2000/commuting/index.html> et <https://www.census.gov/hhes/commuting/>

5. Disponible en ligne sur : <https://www.bfs.admin.ch/bfs/fr/home/statistiques/catalogues-banques-donnees/tableaux.html>

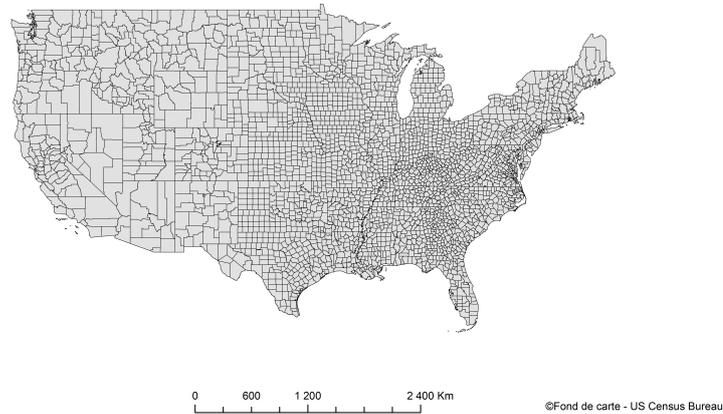


FIG. 2 : Découpage spatial à l'échelle des comtés

Les données obtenues incluent :

- Le nombre de pendulaires effectifs T_{ij} entre chaque unité i et j (y inclus les flux diagonaux)
- Le nombre d'habitants m_i de chaque unité i
- La distance r_{ij} entre les centroïdes de chaque unité i et j dans le cadre des données étasuniennes

4.2 Outils d'analyse

Une implémentation Java des modèles gravitaires et radiatifs a été mise à disposition par Lenormand et al. (2016), qui est disponible à l'adresse suivante : <https://github.com/maximelenormand/Trip-distribution-laws-and-models>. L'estimation des flux est effectuée à l'aide de ces scripts, alors que la calibration des paramètres est implémentée sur le programme R (à l'aide de la formule glm). Les méthodes d'ajustement sont effectuées en Java et en R. La matrice des distances r_{ij} entre les communes suisses n'étant pas fournie par le site de l'OFS, cette dernière a également été créée à l'aide du programme R , en calculant les distances entre les centroïdes de chaque commune. Les principaux package R utilisés sont *Igraph* et *Imap*. Les résultats des différents modèles sont cartographiés à l'aide du logiciel ArcGIS.

5 Résultats

Ce chapitre a pour objectif de comparer les performances prédictives des modèles gravitaires et radiatifs dans le cadre de données pendulaires suisses et étasuniennes. Cette comparaison est faite à travers différentes mesures d'ajustement (cf. chapitre 2.4), et illustrée par des représentations cartographiques. Afin de faciliter la visualisation de ces dernières, les flux cartographiés ont été sélectionnés en fonction d'une origine particulière pour la Suisse ainsi que les États-Unis, respectivement la commune de Zurich et le comté de Queens, qui représentent des unités spatiales majeures en termes du nombre de pendulaires sortants O_i .

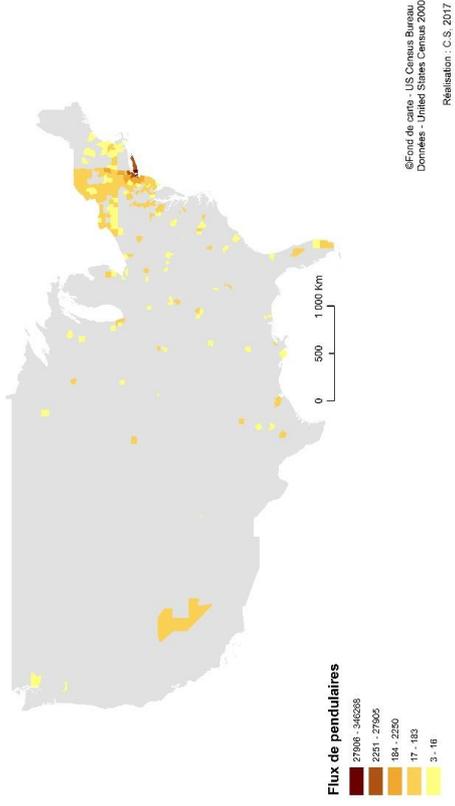
La présentation des résultats est divisée en deux sections, la première traitant des données étasuniennes, la deuxième des données suisses. Chaque section comprend ensuite des sous-sections traitant des différentes échelles d'analyse ainsi que des différents modèles. Les flux pendulaires étasuniens ayant été analysés à travers deux jeux de données, ces derniers sont traités de manière distincte.

5.1 États-Unis

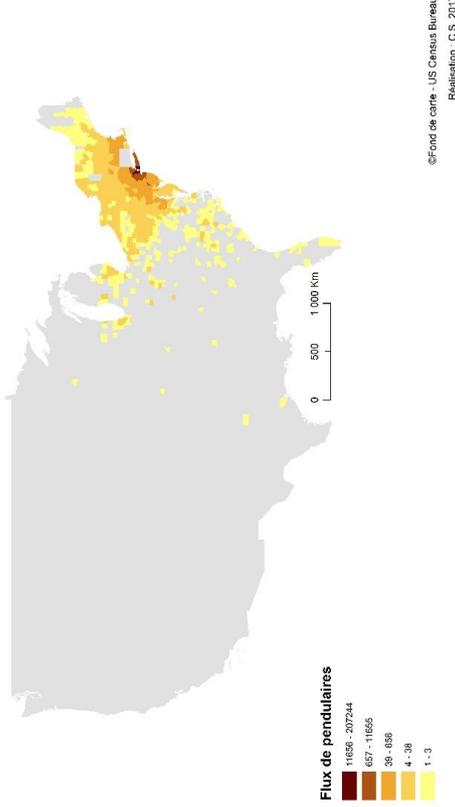
5.1.1 Flux pendulaires sortant du comté de Queens - 2000

Des cartes représentant les flux pendulaires originaires du comté de Queens ont été réalisées (Fig. 3 et 7) afin de compléter les cartes de résidus, sur lesquelles est basée l'analyse des divergences entre données effectives et prédictions des différents modèles. Le comté de Queens compte un nombre de pendulaires sortant O_i de 563'155, où les principaux comtés de destination sont New York (Manhattan), Kings (Brooklyn) et Nassau (Long Island) avec respectivement 346'268, 87'350 et 65'125 pendulaires venant de Queens (Fig. 3). En effet, ces 3 comtés captent 88% des flux pendulaires sortant de Queens, et sont de ce fait mis en évidence sur les cartes de résidus.

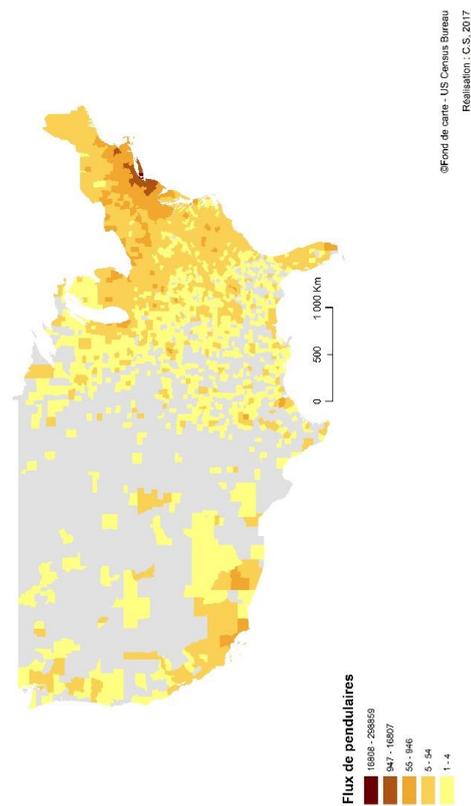
Flux pendulaires originaires du comté de Queens (2000)



Prédiction de flux pendulaires originaires du comté de Queens par modèle gravitaire (2000)



Prédiction de flux pendulaires originaires du comté de Queens par modèle radiatif (2000)



Prédiction de flux pendulaires originaires du comté de Queens par modèle radiatif paramétré (2000)

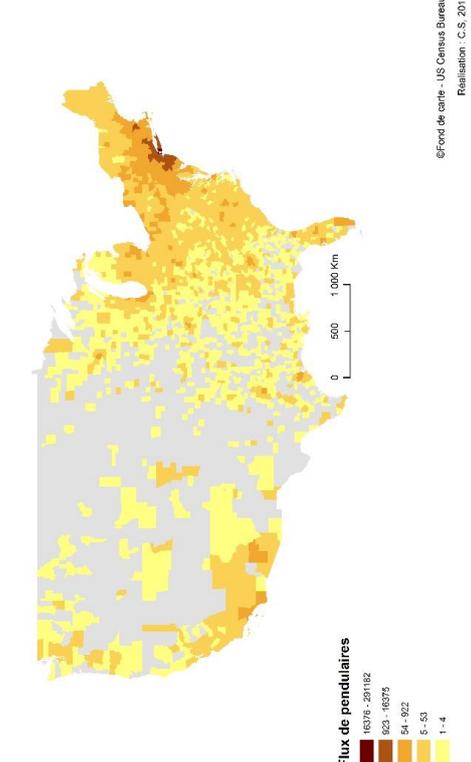


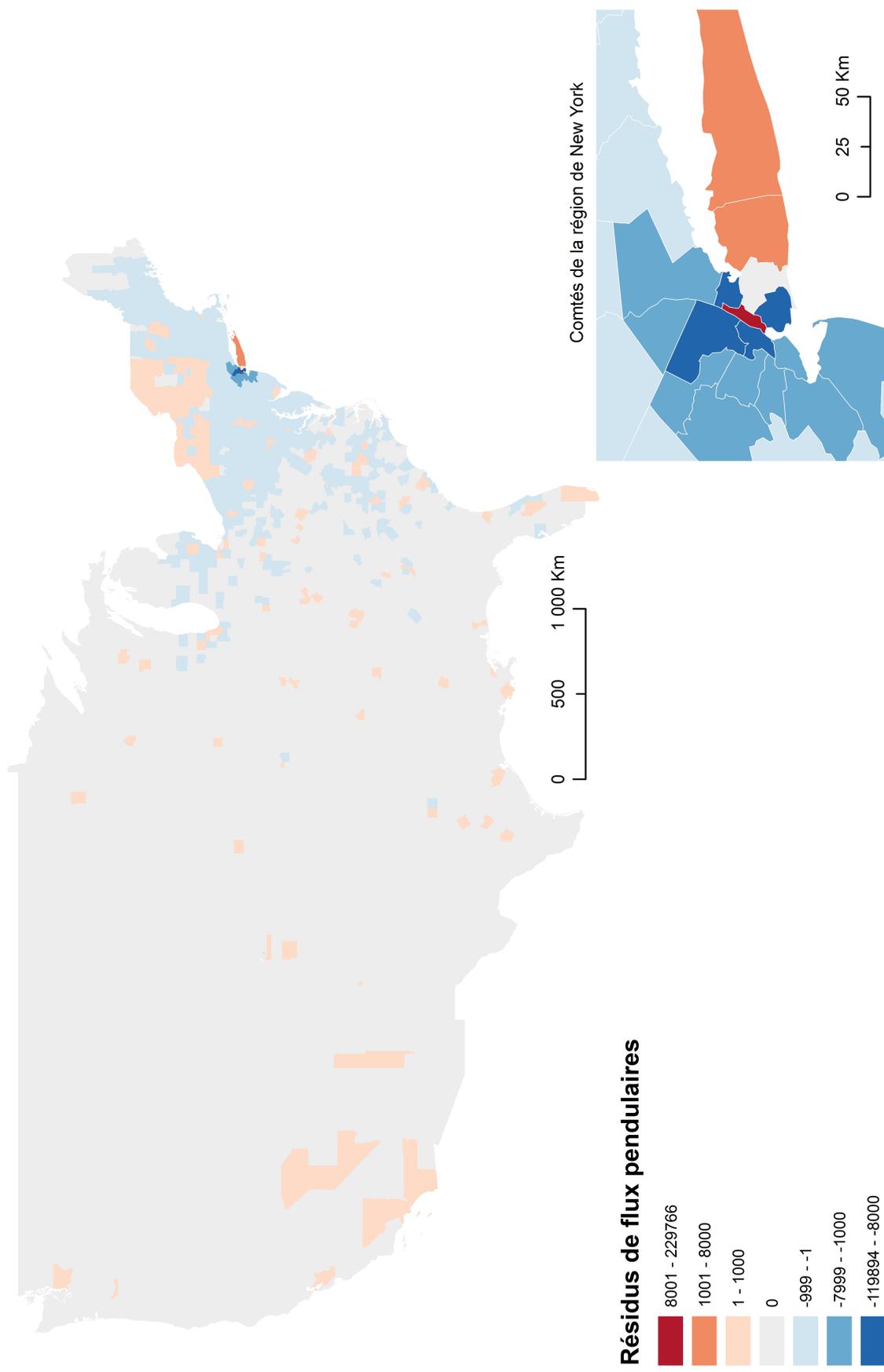
FIG. 3 : Flux pendulaires originaire du comté de Queens - 2000 (données effectives et prédictions)

Prédiction par modèle gravitaire La carte suivante (Fig. 4) représente les résidus de flux pendulaires, soit la différence entre données effectives et données simulées. Les flux positifs (dégradé rouge) signifient donc que les données simulées sont inférieures aux données effectives, ce qui correspond à une sous-estimation de flux de pendulaires. Les flux négatifs (dégradé bleu) quant à eux représentent une surestimation du modèle. Les résidus nuls, qui signifient que les données effectives correspondent à la prédiction, sont représentés en gris.

Le comté ayant les résidus positifs les plus importants est celui de New York, avec 229'766. Cette sous-estimation extrêmement marquée peut être en partie expliquée par le fait que la région de Manhattan compte "uniquement" 1'537'195 habitants en 2000, alors que des régions comme le comté de Kings en comptait 2'465'326. Les deux étant adjacents au comté de Queens (et donc à faible distance), le modèle gravitaire, qui repose uniquement sur les données de populations ainsi que les distances reliant les centroïdes des unités spatiales (comme explicité en (1)), aura tendance à prédire un nombre de pendulaires proportionnel à cette population, sachant que les comtés de Kings et New York sont plus ou moins à la même distance du comté de Queens (respectivement 14,28 et 13,22 km). En effet, les flux prédits entre Queens et Kings par le modèle gravitaire sont de 207'244, ceux entre Queens et New York uniquement 116'502. Néanmoins, la région de Manhattan est un centre économique majeur (1'458'766 de personnes venaient y travailler depuis d'autres comtés en 2000), facteur qui n'est pas pris en compte par le modèle gravitaire utilisé.

Cela permet donc également d'expliquer la surestimation très importante des flux allant dans le comté de Kings. En effet, le modèle surestime de 119'894 le nombre de pendulaires allant dans cette région au vu du nombre important d'habitants.

Un autre point important à relever est que le modèle a une tendance générale à légèrement surestimer les flux à courtes et moyenne distance (comme le montre la zone bleu claire sur la carte), et à sous-estimer les flux plus distants, notamment allant dans d'autres grandes villes étasuniennes comme Miami, Los Angeles, Las Vegas, Seattle, etc. Cette sous-estimation de flux distants par le modèle gravitaire est tout à fait normale vu la nature de la relation décrite en (1).



©Fond de carte - US Census Bureau
Réalisation : C.S., 2017

FIG. 4 : Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle gravitaire (2000) - paramètres $[\alpha, \beta, \gamma] = [0.41, 0.85, 2.27]$

Prédiction par modèle radiatif classique Dans le cadre des prédictions par modèle radiatif sans paramètre (Fig. 5), les comtés ayant les valeurs résiduelles extrêmes sont également New York et Kings. Effectivement, le comté ayant le plus de résidus positifs est à nouveau New York, cette fois avec la valeur de 279'612. Malgré le fait que le modèle radiatif ne soit pas basé sur la distance séparant les unités spatiales, mais plutôt sur le nombre d'opportunités dans un certain rayon (voir (2)), le modèle n'est pas à même de capter l'attrait important de la région de Manhattan. Cela est dû au fait que le nombre d'opportunités du modèle radiatif est proportionnel à la population de l'unité spatiale, d'où le même problème que dans le paragraphe précédent.

La surestimation du modèle radiatif est encore plus importante que celle du modèle gravitaire, avec une valeur extrême de 211'509 pendulaires dans le comté de Kings. De plus, le modèle a une tendance générale à surestimer les flux, qu'ils soient à courte, moyenne ou longue distance.

Prédiction par modèle radiatif paramétré Les comtés ayant les valeurs résiduelles extrêmes restent inchangés par rapport au modèle radiatif classique, même si ces dernières sont légèrement différentes (Fig. 6). En effet, le comté de New York compte 279'870 résidus positifs, alors que le comté de Kings compte 203'832 résidus négatifs. Même si la surestimation pour le comté de Kings est un peu plus faible que dans le cas précédent, les tendances prédictives générales sont très similaires. Malgré quelques sous-estimations de flux pendulaires à longues distances de la région de New York (tels les comtés de Nye au Nevada et Navajo en Arizona), la surestimation est prépondérante.

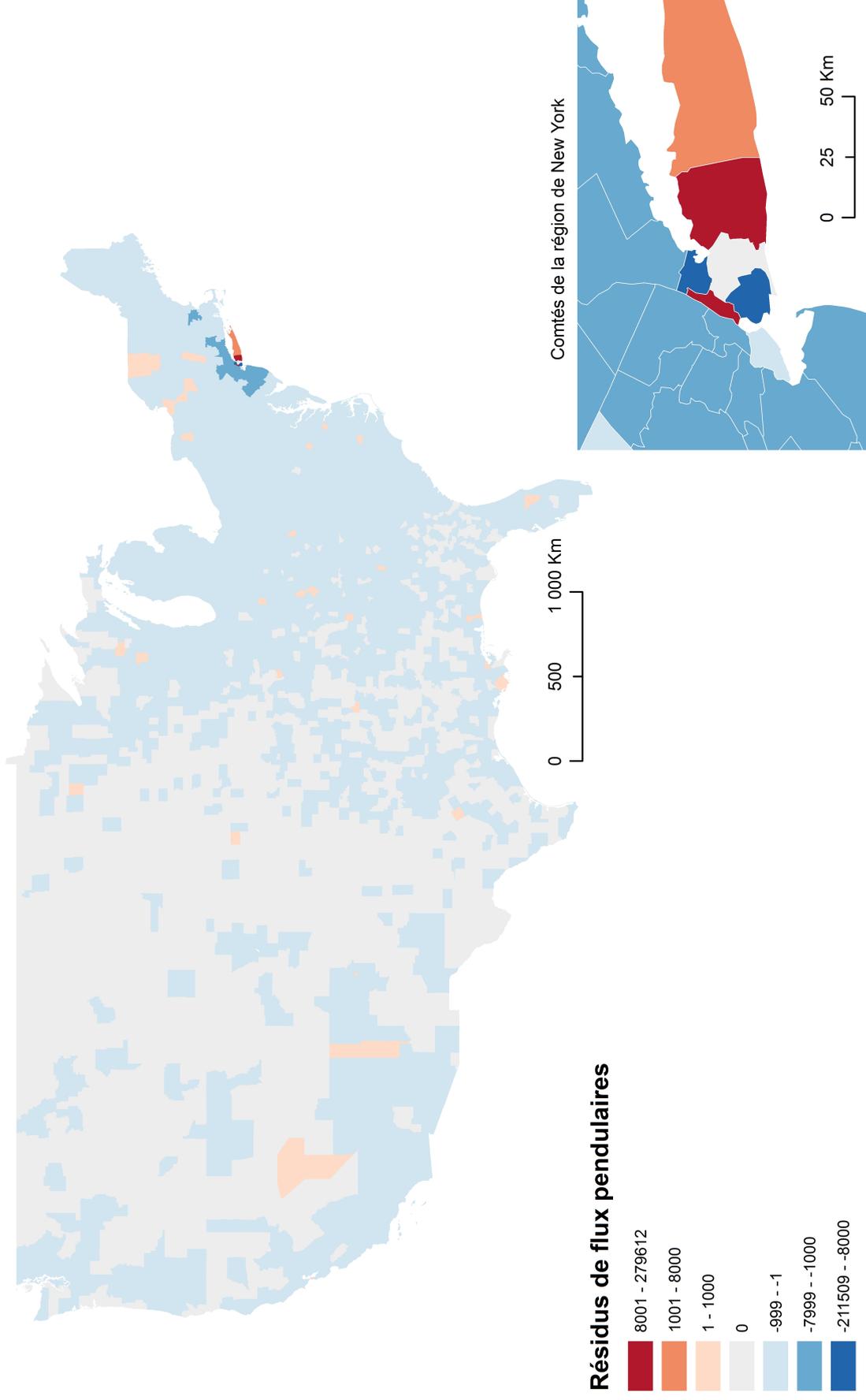
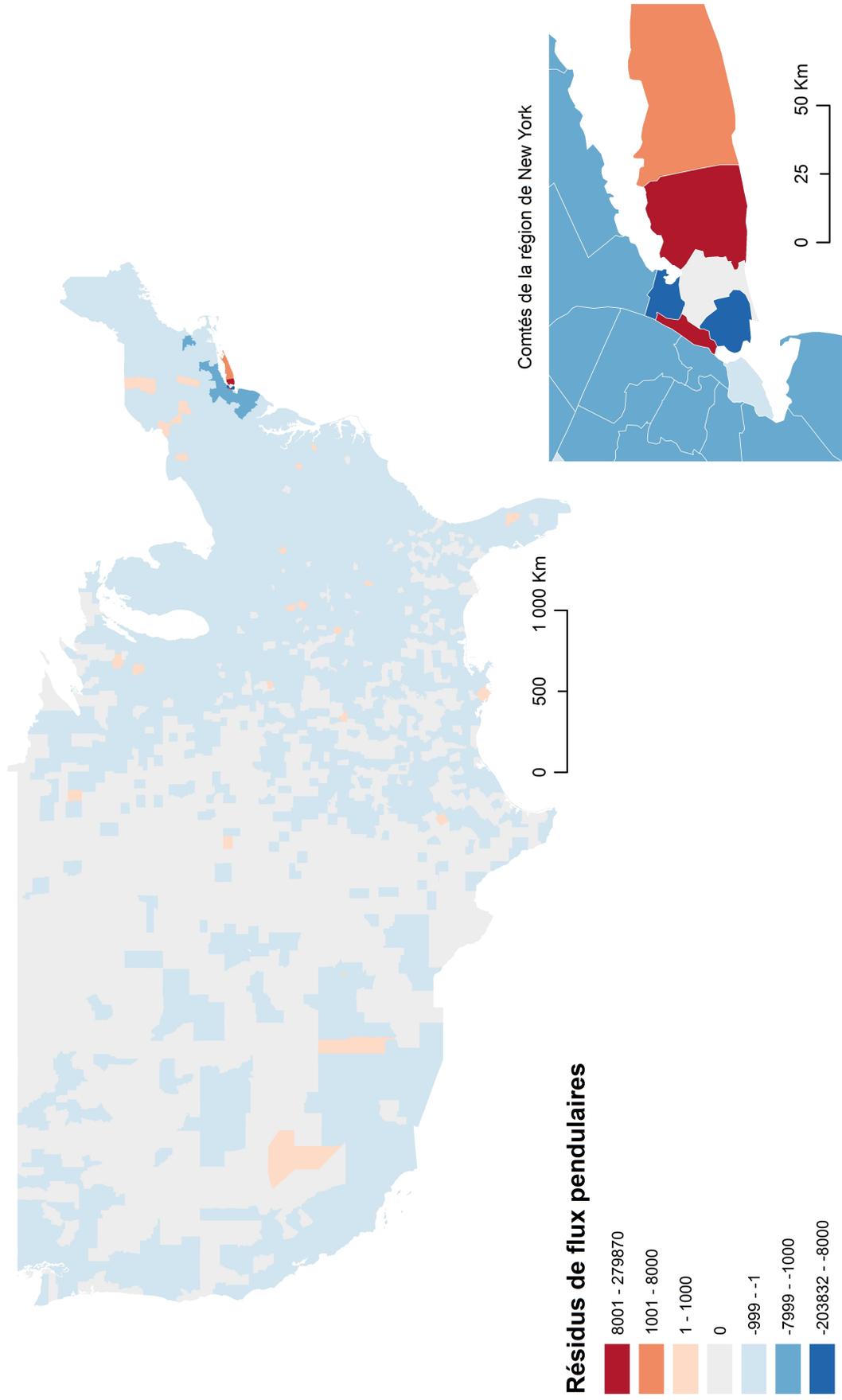


FIG. 5 : Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle radiatif classique (2000)



©Fond de carte - US Census Bureau
Réalisation : C.S, 2017

FIG. 6 : Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle radiatif paramétré (2000) - paramètre $[\alpha] = [0.96]$

Discussion Afin de déterminer quel modèle donne la meilleure prédiction de flux pendulaires, un certain nombre de mesures d’ajustement ont été effectuées (cf. Tab 1).

Le modèle gravitaire est celui ayant le meilleur ajustement à la vue des résultats ci-dessous. En effet, avec un CPC de 0.56, une NMSE de 74, une entropie relative de 0.522 et un AIC de 603’302.7, il surpasse de manière non-négligeable le modèle radiatif sans et avec paramètres. Cela va à l’encontre des résultats obtenus par Simini et al. (2012), qui analysent les flux pendulaires sortant du comté de New York (Manhattan) avec le même jeu de données. Effectivement, Simini et al. (2012) mettent en avant la supériorité du modèle radiatif sans paramètre par rapport au modèle gravitaire (paramétré). Néanmoins, la contrainte des modèles n’a pas été abordée dans leur article (cf. chapitre 2.3), et le comté d’origine n’est pas le même que dans ce travail.

Lenormand et al. (2016) ont également travaillé avec le jeu de données pendulaires états-unien, et ce à une échelle plus globale. Dans le cadre des flux pendulaires sur l’ensemble des États-Unis, le modèle gravitaire avait le meilleur ajustement par rapport aux données observées. De plus, Lenormand et al. (2016) soulignent que le modèle gravitaire prenant une fonction d’impédance de puissance tel que $f(r_{ij}) = r_{ij}^{-\gamma}$ fournit une bonne prédiction des flux à grandes distances.

Le modèle radiatif sans paramètre présente des performances prédictives relativement limitées, avec un CPC de 0.42, une NMSE de 97.2, une entropie de 0.879 et un AIC de 992’138.4. La surestimation importante observée dans (Fig 5 et Fig 6) résulte en un ajustement faible, ce qui est également le cas dans le modèle radiatif paramétré. Ce dernier a été paramétré grâce à la méthode proposée en (27), résultant en une valeur de $\alpha = 0.96$. Étant très proche de 1, cela explique les mesures d’ajustement très similaires entre le modèle radiatif sans paramètre et celui avec. De ce fait, l’AIC laisse suggérer que le modèle radiatif sans paramètre serait plus adapté que celui avec.

Malgré la supériorité du modèle gravitaire dans le cadre de la prédiction de flux pendulaire sortant de comté de Queens, il convient de souligner qu’aucun des modèles est à même de fournir des prédictions exactes. Avec des résidus allant dans l’ordre des centaines de milliers, des améliorations sont nécessaires dans un contexte urbain dense tel celui de la ville de New York, comme mis en avant par Liang et al. (2013). En effet, l’incapacité des modèles à prévoir des flux de pendulaires à une échelle intra-urbaine pourrait être dû au fait qu’il existe différentes habitudes et préférences de voyages en fonction des échelles considérées (Liang et al., 2013). Une solution à ce problème pourrait être de concevoir un modèle incluant un paramètre tenant compte du nombre d’emplois dans la région étudiée.

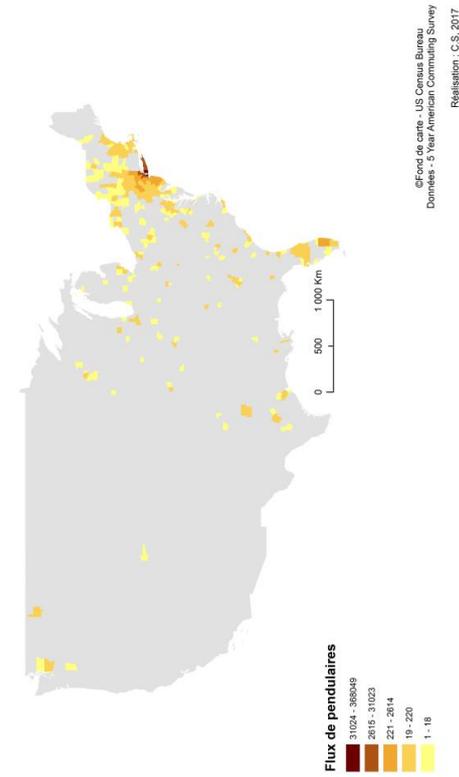
TAB. 1 : Mesures d’ajustement - Flux pendulaires sortant du comté de Queens (2000)

Modèle	NMSE	CPC	KLD	AIC
Gravitaire	74	0.56	0.535	603302.7
Radiatif classique	97.2	0.421	0.880	992138.4
Radiatif paramétré	96	0.421	0.882	994016.2

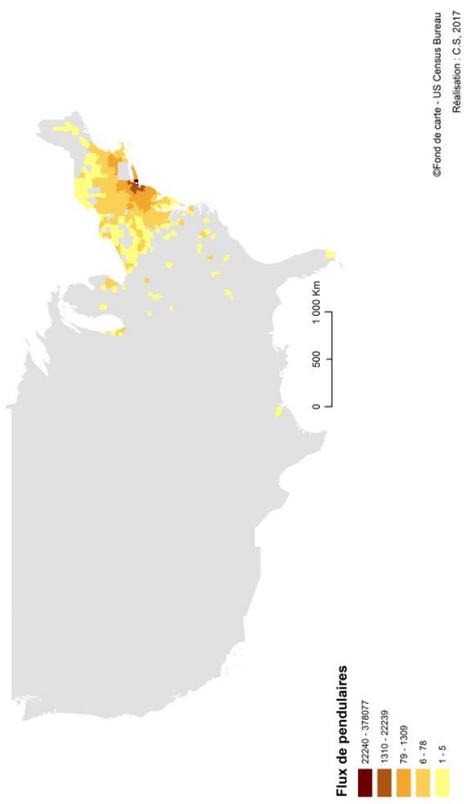
5.1.2 Flux pendulaires sortant du comté de Queens - 2013

Le jeu de données pendulaires étasunien de 2013 (Fig. 7) est sensiblement différent de celui de 2000, notamment dans le cadre de flux sortant du comté de Queens. Ce dernier compte un nombre de pendulaires sortant O_i de 598'961, où les principaux comtés de destination sont les comtés de New York (Manhattan), Nassau (Long Island) et Kings (Brooklyn), avec respectivement 368'049, 70'964 et 70'162 pendulaires venant de Queens.

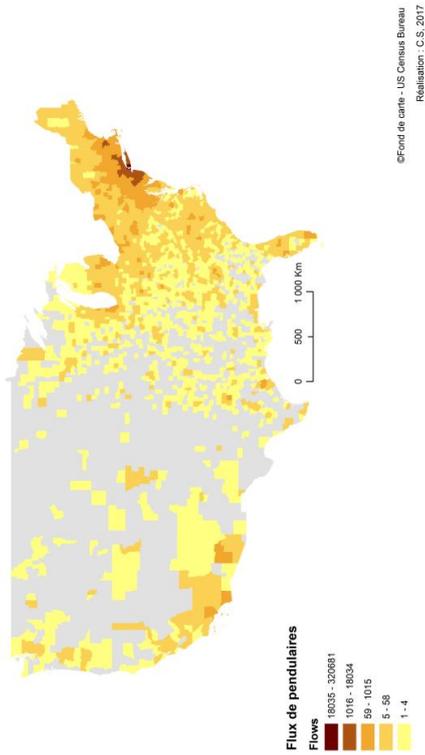
Flux pendulaires originaires du comté de Queens (2013)



Prédiction de flux pendulaires originaires du comté de Queens par modèle gravitaire (2013)



Prédiction de flux pendulaire originaires du comté de Queens par modèle radiatif (2013)



Prédiction de flux pendulaires originaires du comté de Queens par modèle radiatif paramétré (2013)

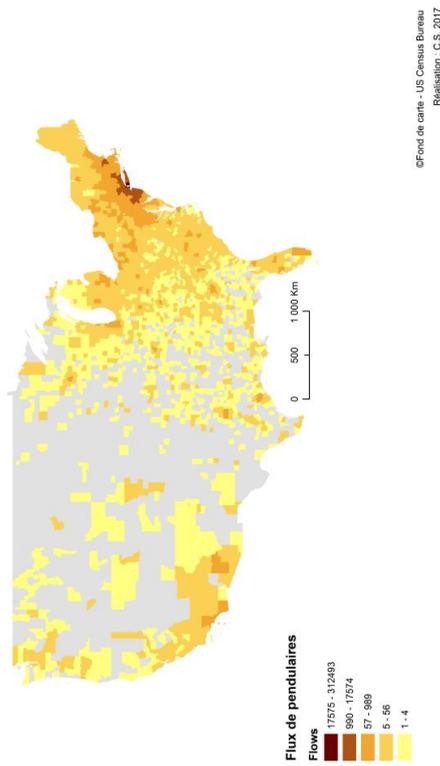
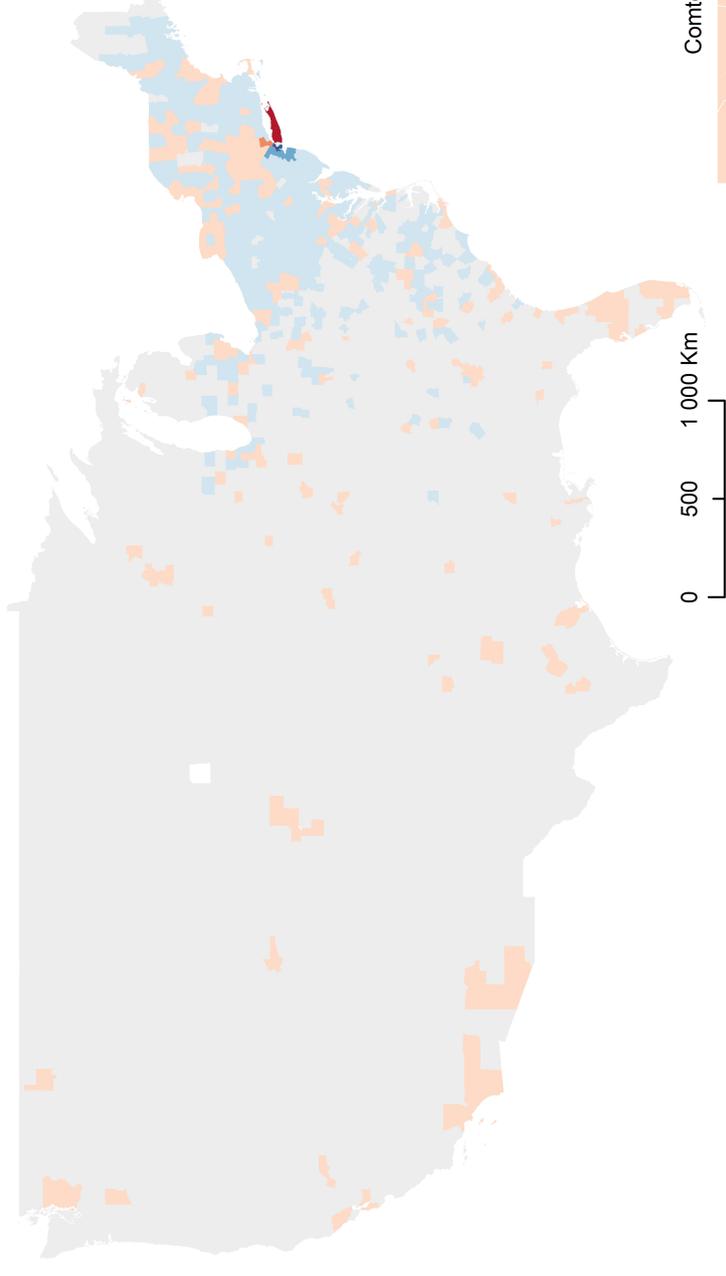


FIG. 7 : Flux pendulaire originaires du comté de Queens - 2013 (données effectives et prédictions)

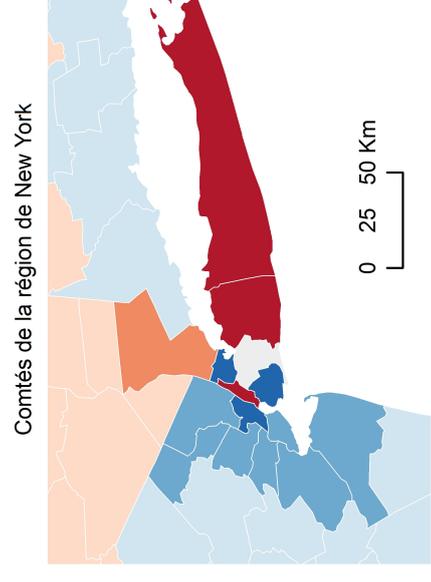
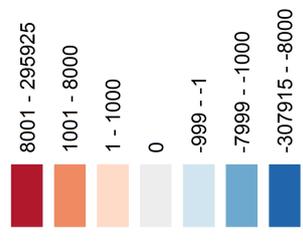
Prédiction par modèle gravitaire Les résidus positifs extrêmes se situent une fois de plus dans le comté New York, étant au nombre de 295'925, suivi par le comté de Nasau avec 37'009 résidus positifs (Fig. 8). Cette sous-estimation très marquée à New York peut être expliquée de manière similaire que pour le jeu de données de 2000.

Le comté ayant la surestimation la plus marquée est Kings, avec 307'915 résidus négatifs. La surestimation du comté de Kings peut à nouveau en partie être expliquée par son nombre d'habitants très élevé (2'592'149 en 2013).

Même s'il n'est pas possible de distinguer une tendance générale de sous ou surestimation des flux pendulaires à courtes et moyennes distances du comté de Queens (comme cela était le cas pour les données de 2000), le modèle gravitaire sous-estime en majeure partie les flux à longues distances.



Résidus de flux pendulaires

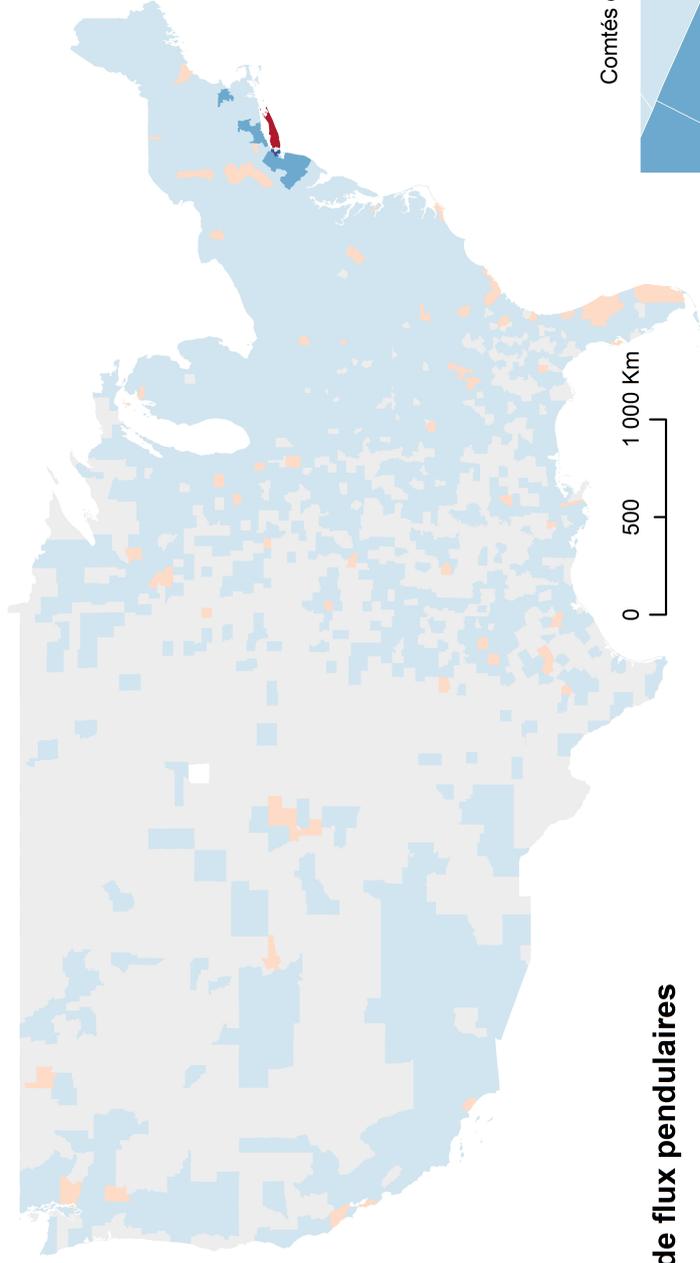


Comtés de la région de New York
 ©Fond de carte - US Census Bureau
 Réalisation : C.S., 2017

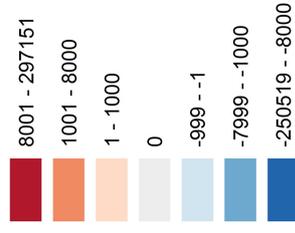
FIG. 8 : Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle gravitaire (2013) - paramètres $[\alpha, \beta, \gamma] = [0.41, 0.85, 2.27]$

Prédiction par modèle radiatif classique La prédiction par modèle radiatif non-paramétré (Fig. 9) sous-estime également de manière importante les comtés de New York et Nassau, avec des résidus positifs de respectivement 297'151 et 49'005. Les résidus négatifs extrêmes quant à eux sont situés dans le comtés de Kings, avec -250'519. La surestimation est prépondérante dans le cadre du modèle radiatif, même si elle est moins marquée que dans le cadre du jeu de données de 2000. En effet, le modèle va même sous-estimer certains flux distants, notamment en direction de grandes villes telles, Los Angeles, Miami, San Francisco, Seattle, etc.

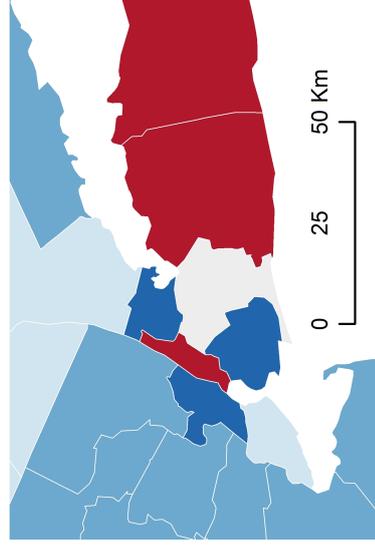
Prédiction par modèle radiatif paramétré Le modèle radiatif paramétré (Fig. 10) a des résidus très proches du modèle radiatif classique, sous-estimant également de manière importante les comtés de New York et Nassau avec respectivement 297'388 et 48'633 résidus positifs. La surestimation quant à elle est un peu moins marquée, avec une valeur extrême de -242'331 dans le comté de Kings. Comme dans le cadre du jeu de données de 2000, la prédiction du modèle paramétré est très proche du modèle radiatif classique vu la valeur du paramètre α de 0.96.



Résidus de flux pendulaires

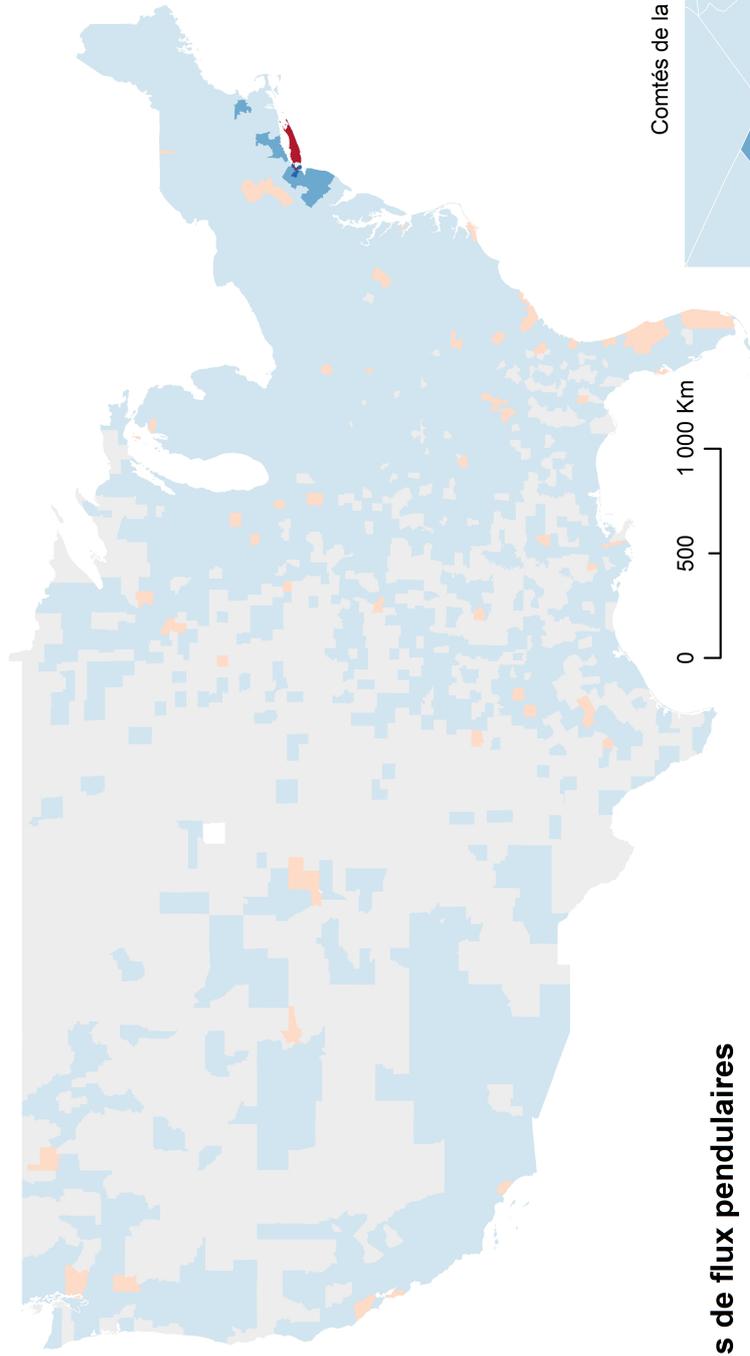


Comtés de la région de New York

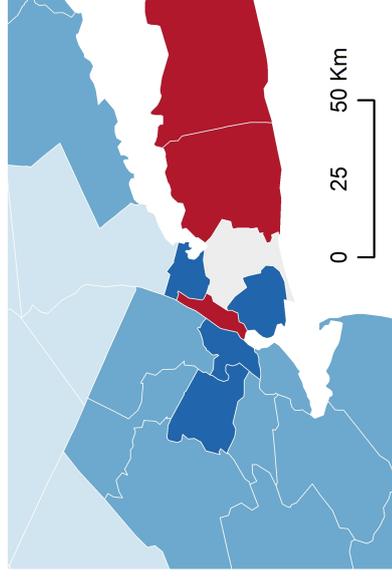


©Fond de carte - US Census Bureau
Réalisation : C.S., 2017

FIG. 9 : Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle radiatif classique (2013)



Comtés de la région de New York



©Fond de carte - US Census Bureau
Réalisation : C.S., 2017

FIG. 10 : Résidus de la prédiction de flux pendulaires originaires du comté de Queens par modèle radiatif paramétré (2013)
- paramètre $[\alpha] = [0.96]$

Discussion La table des mesures d’ajustement (Tab. 2) montre des résultats très proches en termes de CPC pour les 3 modèles étudiés, avec le modèle radiatif paramétré et le modèle gravitaire ayant la valeur la plus élevée. En effet, les mesures de NMSE, de KLD ainsi que d’AIC sont également relativement proches, néanmoins le modèle radiatif paramétré semble être le plus adapté dans le cadre de la prédiction des flux pendulaires sortant du comté de Queens pour l’année 2013. Il convient toutefois de souligner que les valeurs très faibles de CPC sont également liées au biais important lié au contexte intra-urbain dense de la ville de New York.

TAB. 2 : Mesures d’ajustement - Flux pendulaires sortant du comté de Queens (2013)

Modèle	NMSE	CPC	KLD	AIC
Gravitaire	112.3	0.386	1.13	1353658
Radiatif classique	102.6	0.385	1.128	1351256
Radiatif paramétré	101.3	0.386	1.126	1348862

5.1.3 Prédiction de flux pendulaires pour l’ensemble des États-Unis

La figure suivante (Fig. 11) permet de comparer les prédictions de flux pendulaires inter-comtés pour l’ensemble des États-Unis. En effet, une cartographie des flux n’étant pas possible en termes de visualisation, des nuages de points ont été réalisés afin de visualiser l’écart entre les flux observés non-nuls et les prédictions. La ligne verte représentant la droite de régression et la rouge la diagonale où $x = y$, plus l’écart entre ces dernières est élevé, plus les flux prédits sont éloignés des données effectives.

Dans le cadre du jeu de données de 2000, le modèle gravitaire donne la meilleure prédiction au vu des mesures d’ajustement de la table 3. Néanmoins, les modèles radiatifs produisent des prédictions de qualité très similaire en termes de CPC, mais avec une NMSE, une entropie relative et un AIC plus élevés. La supériorité du modèle gravitaire a également été mise en avant par Lenormand et al. (2016) dans le cadre du jeu de données étasuniens de 2000.

TAB. 3 : Mesures d’ajustement - Flux pendulaires pour l’ensemble des États-Unis

Modèle	NMSE	CPC	KLD	AIC
Gravitaire - 2000	72.1	0.630	0.722	49230603
Radiatif classique - 2000	154.2	0.621	1.242	84764655
Radiatif paramétré - 2000	150.5	0.623	1.198	81703823
Gravitaire - 2013	78	0.581	0.851	65720028
Radiatif classique - 2013	96.45	0.59	1.313	101398812
Radiatif paramétré - 2013	95.19	0.59	1.271	98155288

Les prédictions de flux de pendulaires pour le jeu de données de 2013 sont un peu moins précises, les modèles radiatifs ayant la meilleure qualité d'ajustement en termes de CPC. Néanmoins, les mesures de NMSE, KLD et AIC penchent clairement en faveur du modèle gravitaire, qui semble également être le modèle le plus adapté pour prédire les flux pendulaires à l'échelle des États-Unis pour les données de 2013. Le léger décalage en termes de performance prédictive entre le jeu de données de 2000 et celui de 2013 est perceptible sur la Fig. 11, l'écart entre la diagonale et la droite de régression étant plus marqué sur les nuages de point de la deuxième rangée (et donc du jeu de données de 2013).

5.2 Suisse

Suivant la logique d'analyse des données étasuniennes, des cartes représentant les flux pendulaires originaires de la commune de Zurich ont été réalisées (Fig. 12), accompagnées de cartes de résidus.

La commune de Zurich compte au total 38'032 pendulaire sortant O_i , où les communes limitrophes captent 32% des flux sortants (soit 11'858), les communes attirant le plus de personnes étant Kloten, Opfikon et Schlieren, avec respectivement 3'473, 2'814 et 2'064 pendulaires.

5.2.1 Flux pendulaires sortant de la commune de Zurich

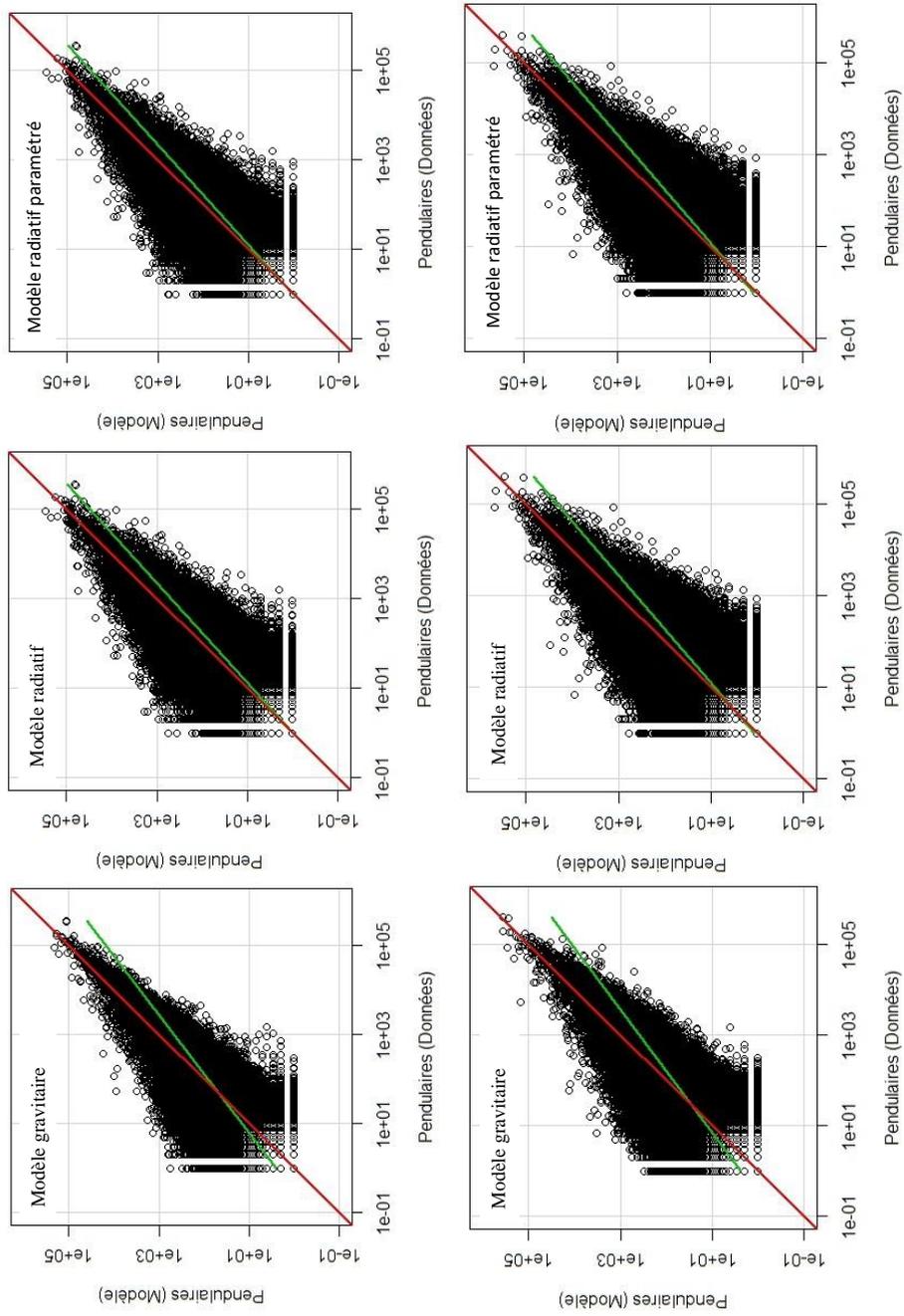
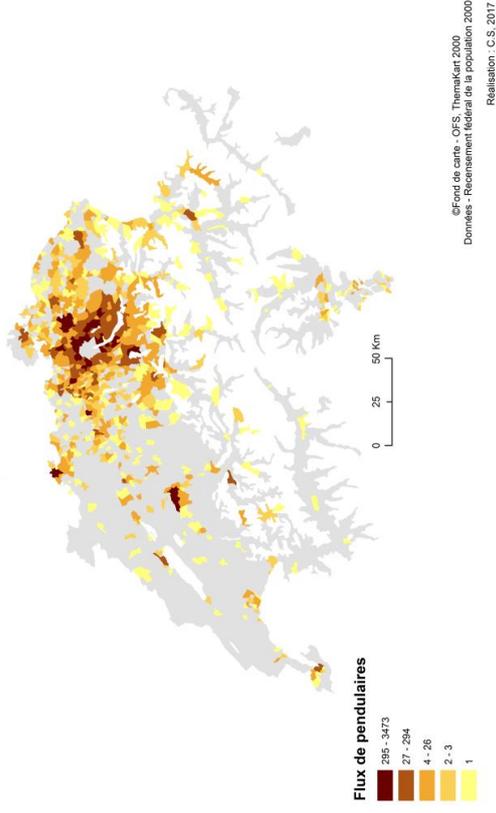
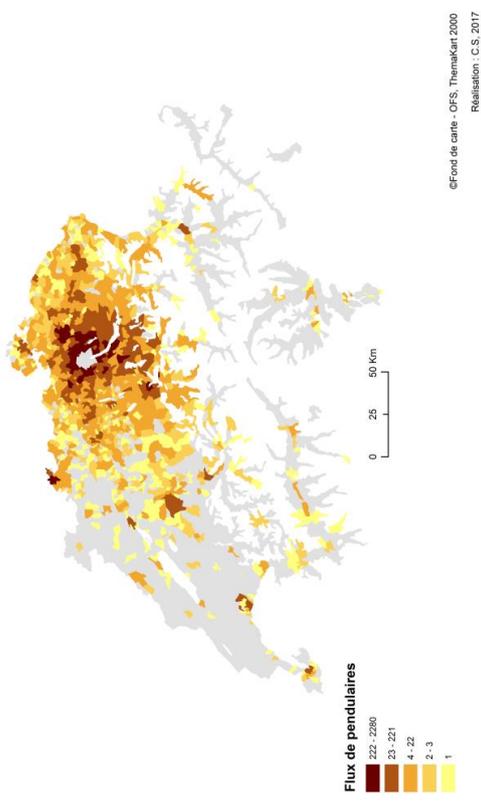


FIG. 11 : Comparaison des prédictions de flux pendulaires à l'échelle des Etats-Unis pour les données de 2000 (première rangée) et de 2013 (deuxième rangée)

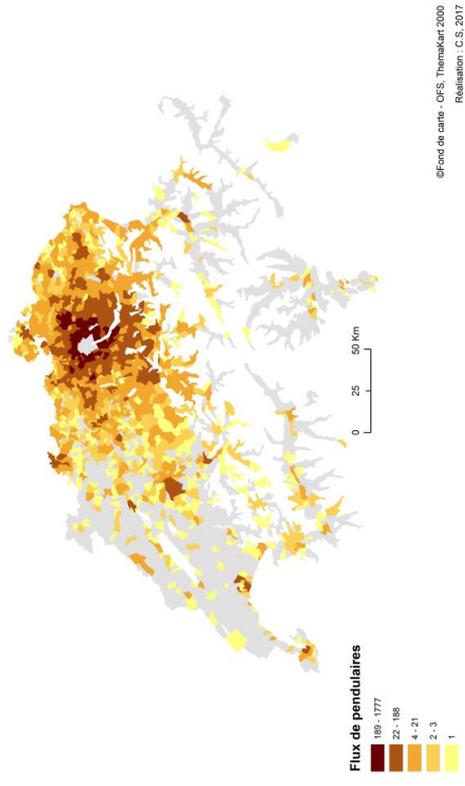
Flux pendulaires originaires de la commune de Zürich (2000)



Prédiction de flux pendulaires originaires de la commune de Zürich par modèle gravitaire



Prédiction de flux pendulaires originaires de la commune de Zürich par modèle radiatif classique



Prédiction de flux pendulaires originaires de la commune de Zürich par modèle radiatif paramétré

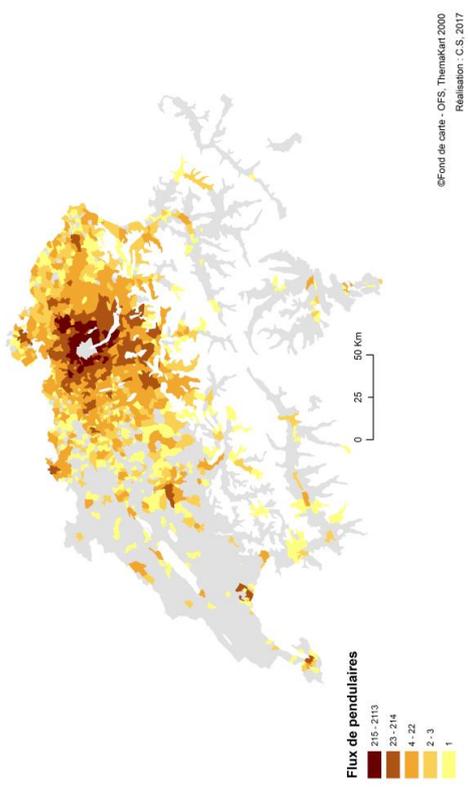


FIG. 12 : Flux pendulaires originaires de la commune de Zurich (données effectives et prédictions)

Prédiction par modèle gravitaire Les communes étant le plus sous-estimées par le modèle gravitaire sont Kloten, Opfikon et Schlieren avec respectivement 2579, 1478 et 708 résidus positifs (Fig. 13). La sous-estimation extrême pour la commune de Kloten peut être expliquée d'une manière similaire à la sous-estimation du comté de Manhattan pour la prédiction par modèle gravitaire pour les données étasuniennes de 2000. En effet, le modèle prédit uniquement 894 pendulaires allant de Zurich à Kloten au vu de sa population et la distance qui relie les deux communes. Néanmoins, la commune de Kloten abrite l'aéroport le plus important du pays, et ainsi un nombre très important d'emplois (23'026 personnes venaient travailler dans la commune de Kloten depuis d'autres communes en 2000), phénomène qui n'est pas capté par le modèle gravitaire.

Les valeurs les plus importantes en termes de surestimation sont quant à elles moins élevées, avec -506 et -424 résidus négatifs pour les communes de Dübendorf et Oberengstringen.

Le modèle gravitaire, qui surestime légèrement les flux pendulaires de la plupart des communes à courte et moyenne distance de Zurich, a tendance à sous-estimer les flux allant vers d'autres villes majeures telles Bern, Zug et Aarau.

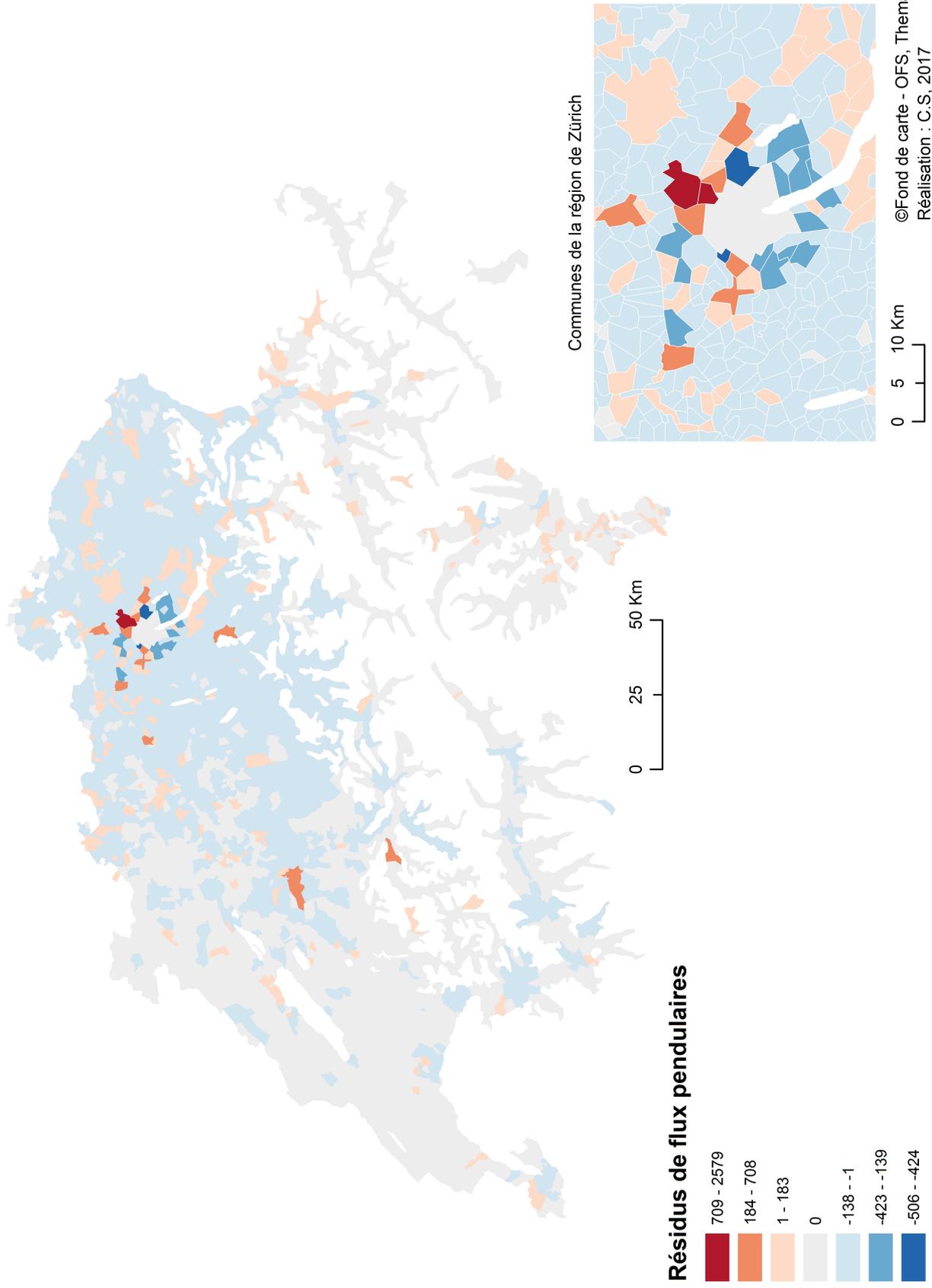


FIG. 13 : Résidus de la prédiction de flux pendulaires originaires de la commune de Zurich par modèle gravitaire (2000) - paramètres $[\alpha, \beta, \gamma] = [0.67, 1.14, 1.85]$

Prédiction par modèle radiatif classique La prédiction par modèle radiatif sans paramètre (Fig. 14) sous-estime également de manière importante les communes de Kloten, Opfikon et Schlieren avec respectivement 2665, 1556 et 755 résidus positifs. Les résidus négatifs extrêmes sont quant à eux situés à Oberengstringen, Uitikon et Buchs avec des valeurs respectives de -467, -260 et -246. Le modèle radiatif a également une tendance générale à sous-estimer les flux allant vers d'autres villes importantes, la surestimation étant néanmoins prépondérante à courte et moyenne distance de Zurich.

Prédiction par modèle radiatif paramétré Les résidus positifs les plus élevés dans le cadre du modèle radiatif paramétré (Fig. 15) se situent à Kloten, Opfikon et Winterthur, avec des valeurs respectives de 2572, 1250 et 607. Les résidus négatifs les plus importants sont une fois de plus à Oberengstringen et Uitikon, avec des valeurs respectives de -624 et -358. La tendance générale du modèle radiatif paramétré est proche du modèle radiatif classique, la surestimation à moyenne et longue distance étant cependant légèrement moins marquée, comme le montrent la supériorité de résidus nuls dans la région romande dans le cadre de la prédiction paramétrée.

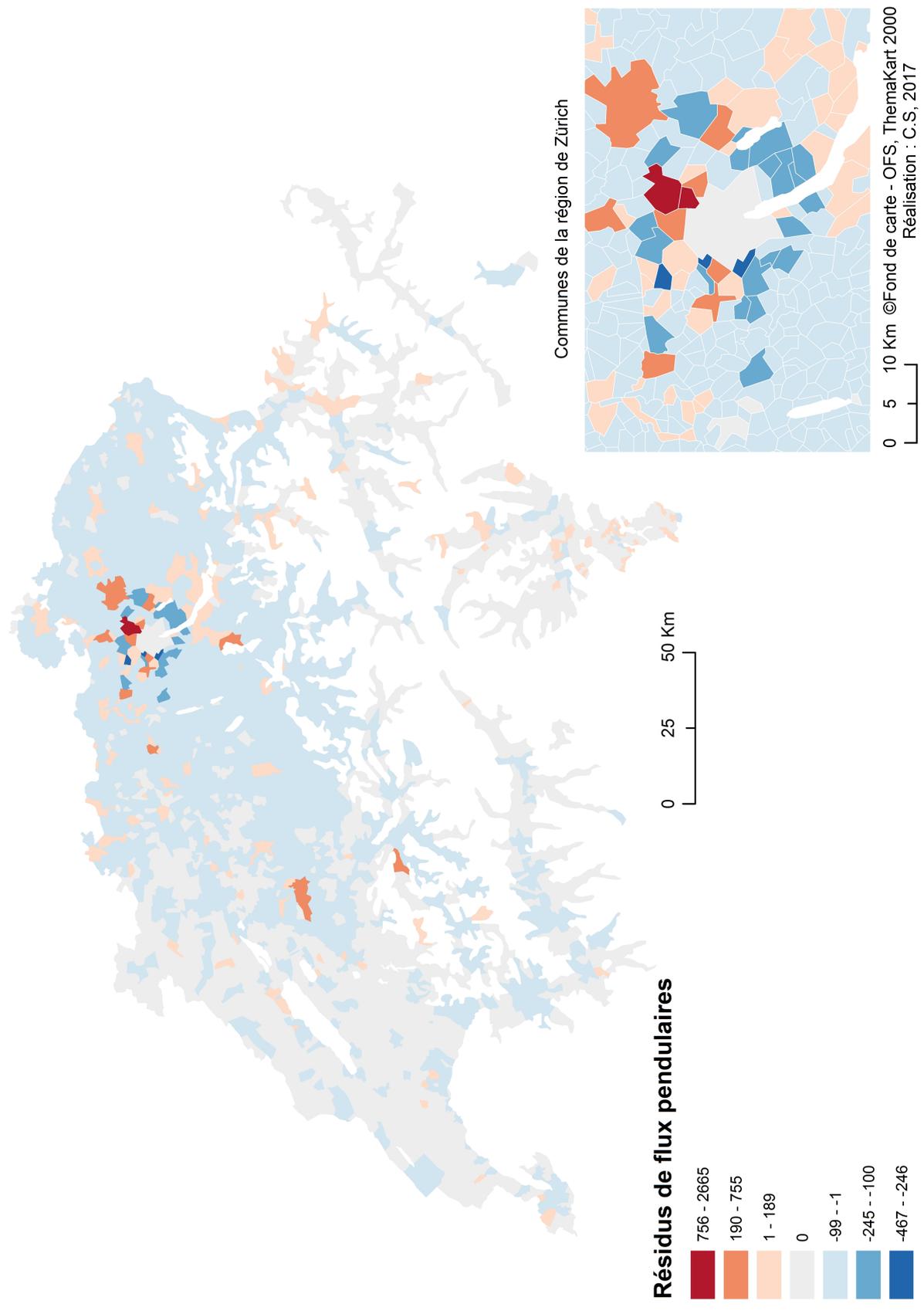


FIG. 14 : Résidus de la prédiction de flux pendulaires originaires de la commune de Zurich par modèle radiatif (2000)

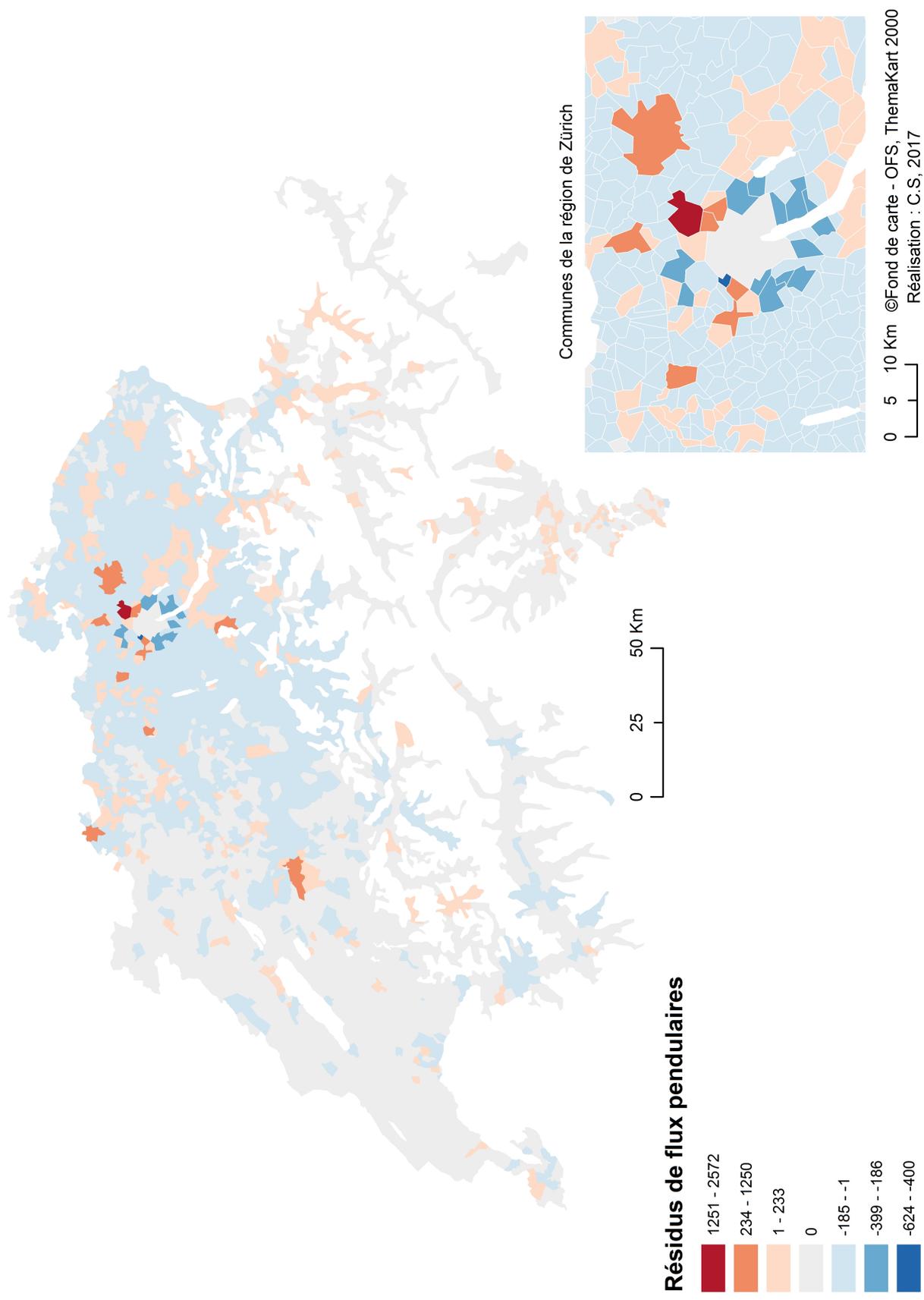


FIG. 15 : Résidus de la prédiction de flux pendulaires originaires de la commune de Zurich par modèle radiatif paramétré (2000) - paramètre $[\alpha] = [1.3]$

TAB. 4 : Mesures d'ajustement - Flux pendulaires sortant de la commune de Zurich (2000)

Modèle	NMSE	CPC	KLD	AIC
Gravitaire	53.4	0.716	0.784	59665.57
Radiatif classique	55.7	0.689	0.519	39504.55
Radiatif paramétré	53.1	0.7	0.821	62482.28

Discussion Les mesures d'ajustement (Tab. 4) montrent des résultats relativement proches pour les trois modèles étudiés, le modèle radiatif classique ayant les meilleures qualités prédictives en termes de KLD et d'AIC. Néanmoins, le modèle gravitaire présente un meilleur ajustement en termes de NMSE et CPC. Au vu de la différence relativement faible en termes de NMSE et CPC entre les différents modèles, mais aussi du fait que le modèle radiatif présente de bien meilleures valeurs du KLD et de l'AIC que les autres modèles, le modèle radiatif semble être le plus approprié pour prédire les flux pendulaires sortant de la commune de Zurich.

La différence non-négligeable entre les mesures d'ajustement suisses et étasuniennes peut être expliquée par le fait que le cadre intra-urbain est sensiblement différent dans les deux cas. En effet, le comté de Queens compte un nombre de pendulaires sortant O_i de 252'958 en 2013 et 563'155 en 2000, alors que la commune de Zurich en compte 38'032 dans la même année. Dans le cadre des données étasuniennes, une large partie des flux sortant de Queens était captée par les comtés à proximité immédiate (88% en 2000 et 65% en 2013), les flux pendulaires restant donc largement dans un contexte intra-urbain. Dans le contexte helvétique, la structure des mouvements pendulaires sortant de Zurich n'est pas autant limitée à la ville la plus peuplée du pays, mais implique également une grande partie de flux interurbains, comme le montre la carte en haut à gauche de la Fig. 12. De ce fait, les prédictions des flux pendulaires sortant de la commune de Zurich sont beaucoup moins biaisées par la difficulté à prédire des flux intra-urbains, comme mentionné auparavant.

5.2.2 Prédiction de flux pendulaires pour l'ensemble de la Suisse

Afin de comparer les performances prédictives des différents modèles dans le cadre des flux pendulaires intercommunaux pour l'ensemble de la Suisse, des nuages des points ont été réalisés afin d'accompagner les mesures d'ajustement, une cartographie étant difficilement réalisable.

Les nuages de points de la Fig. 16 montrent que le modèle radiatif est celui ayant le plus grand décalage entre données empiriques et prédictions (Fig. 16 (b)). En effet, cela est confirmé par les mesures d'ajustement (Tab. 5), le modèle radiatif classique ayant les valeurs les plus élevées (en donc les moins bonnes performances prédictives) en termes de NMSE, de KLD ainsi que d'AIC.

Néanmoins, la différence de qualité d'ajustement entre le modèle gravitaire et le modèle

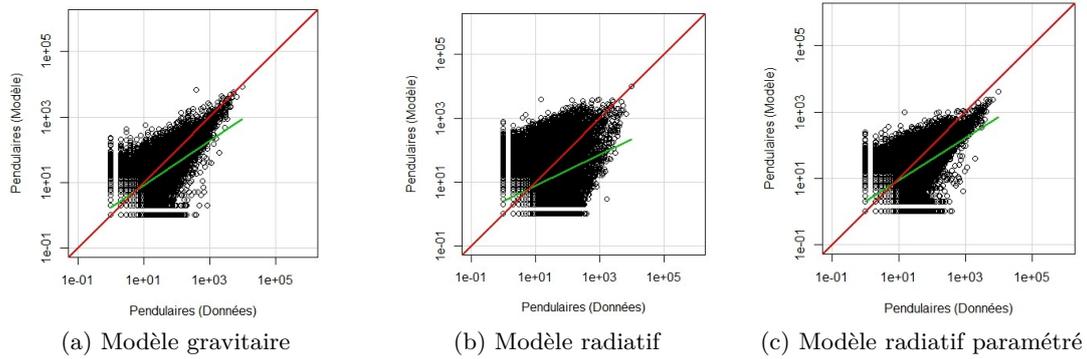


FIG. 16 : Comparaison des prédictions de flux de pendulaires à l'échelle de la Suisse

TAB. 5 : Mesures d'ajustement - Flux pendulaires pour l'ensemble de la Suisse

Modèle	NMSE	CPC	KLD	AIC
Gravitaire	175.3	0.601	3.719	14589104
Radiatif classique	1031.8	0.342	11.472	45000764
Radiatif paramétré	274.8	0.503	3.588	14075571

radiatif paramétré n'est pas autant perceptible sur les nuages de points de la Fig. 16. De plus, la sélection du meilleur modèle grâce aux mesures d'ajustements (Tab. 5) s'avère compliquée, le modèle gravitaire étant supérieur en termes de NMSE et CPC, le modèle radiatif quant à lui étant meilleur en termes de KLD et AIC. Cependant, la différence en termes de NMSE et CPC entre modèle gravitaire et modèle radiatif paramétré est plus importante que celle impliquant les mesures de KLD et d'AIC. De ce fait, le modèle à retenir dans ce cas d'étude semble être le modèle gravitaire.

6 Discussion

Après avoir mis en lumière les résultats des différents modèles de prédiction, il convient de revenir sur les hypothèses émises dans le chapitre 3.1.

Afin de déterminer si le modèle gravitaire est plus précis dans la prédiction de flux à courte et moyenne distance et les modèles radiatifs dans les trajets à longue distance (H1), il semble pertinent de revenir sur les différentes cartes de résidus présentées dans le chapitre 5. En effet, selon H1, les cartes de résidus des modèles radiatifs auraient plus de résidus nuls à longues distances de l'origine cartographiée que celles des modèles gravitaires. A l'inverse, les cartes de résidus des modèles gravitaires auraient plus de résidus nuls à faible et moyenne distance de l'origine cartographiée que celles des modèles radiatifs. Néanmoins, il n'est pas possible de distinguer de manière claire un intervalle de distance spécifique dans lequel les modèles gravitaire et radiatif auraient de meilleures

performances prédictives, les modèles radiatifs ayant même plus de résidus à longues distances que les modèles gravitaires (et donc moins de résidus nuls, ce qui correspond à moins de précision), notamment dans le cadre étasunien. Il n'est donc pas possible de valider H1 à travers l'analyse des cartes de résidus. Une manière beaucoup plus précise et moins approximative de répondre à cette hypothèse aurait été d'analyser les résidus de flux pendulaires en fonction de la distance, comme proposé dans Masucci et al. (2013).

Les résultats mis en avant dans le chapitre précédent permettent de répondre à H2, selon laquelle le modèle gravitaire aurait des meilleures performances prédictives que le modèle radiatif. En effet, cette affirmation peut être en partie validée, notamment pour 4 des 6 cas d'études. Effectivement, le modèle gravitaire présente les meilleures performances prédictives sauf dans le cas des flux pendulaires sortant du comté de Queens pour les données de 2013, où le modèle radiatif paramétré semble plus approprié, ainsi que le flux sortant de la commune de Zurich, qui sont le mieux estimés par le modèle radiatif classique.

Afin de déterminer si l'ajout d'un paramètre se justifie dans le cadre du modèle radiatif (H3), il convient de prendre en compte le critère d'information d'Akaike (15), et de comparer la valeur de ce dernier pour le modèle radiatif avec et sans paramètre. Les résultats obtenus dans le chapitre 5 montrent que le modèle radiatif paramétré présente un AIC plus faible (et donc de meilleurs résultats) dans 4 sur 6 cas d'études. En effet, l'ajout d'un paramètre supplémentaire ne se justifie pas en termes de gain de performances prédictives dans le cadre des données pendulaires sortant du comté de Queens (2000) ainsi que pour les flux pendulaires sortant de la commune de Zurich.

Néanmoins, il convient de remettre en question la pertinence du critère d'information d'Akaike au vu des résultats obtenus dans le chapitre précédent. Comme expliqué auparavant (chapitre 2.4.4), l'AIC contient un terme de pénalité dépendant du nombre de paramètres ajustables, sans lequel le meilleur modèle sera généralement le plus complexe. Travaillant dans le cas présent avec des modèles contenant respectivement, 3, 1 et aucun paramètre ajustable, le critère de pénalité $2K$ est très faible relativement à la *deviance* vu le nombre total d'observations important. De ce fait, l'AIC ne permet pas de mettre en avant une information concernant l'ajustement du modèle qui n'est pas déjà transmise par l'entropie relative au vu de la relation décrite en (25), comme le montre l'annexe A, qui met en avant le rang de chaque modèle en fonction de l'entropie relative et de l'AIC. Effectivement, le critère d'information d'Akaike ne pénalisant pas suffisamment les modèles paramétrés, les meilleurs modèles en termes de divergence de Kullback-Leibler sont les mêmes en termes d'AIC.

Une autre solution aurait donc été d'appliquer le critère d'information bayésien (BIC) proposé par Schwarz (1978), dont la pénalité dépend de la taille de l'échantillon.

7 Conclusion

Le présent travail a permis de comparer les performances prédictives des modèles gravitaires et radiatifs dans le cadre de la modélisation de flux de pendulaires. Les trois modèles retenus (modèle gravitaire à 3 paramètres, modèle radiatif sans paramètre et un modèle radiatif à un paramètre) ont été testés sur trois jeux de données, deux étasuniens et un suisse. Le modèle ayant les meilleures performances prédictives dans 4 des 6 cas d'étude est le modèle gravitaire. En effet, le modèle gravitaire s'avère relativement robuste et est à même de produire de bons résultats en termes de mesures d'ajustement, que cela soit pour une origine particulière ou l'ensemble d'un territoire, tels la Suisse ou les États-Unis.

Le modèle radiatif sans paramètre, tel que proposé par Simini et al. (2012), donne lieu à de très bons résultats dans le cadre des flux pendulaires sortant de la commune de Zurich. Néanmoins, le modèle est moins flexible en ce qui concerne les changements d'échelle d'analyse, comme par exemple dans le cadre des flux pour l'ensemble de la Suisse, où il est sujet à de fortes imprécisions.

Le modèle radiatif paramétré, proposé dans les travaux de Yang et al. (2014) et Lenormand et al. (2016), semble être le modèle le plus approprié dans le cadre de la prédiction de flux de pendulaires sortant du comté de Queens pour les données de 2013. Comme énoncé auparavant, l'ajout d'un paramètre au modèle radiatif classique est justifié dans une majorité des cas étudiés, notamment pour l'ensemble des flux pendulaires intercommunaux en Suisse, où l'ajout d'un paramètre entraîne une augmentation de performances prédictive non-négligeable (cf. Tab. 5).

Le choix de l'analyse des flux pendulaires sur deux pays très différents, que cela soit en taille, population ou structure de mouvements pendulaires, a permis non seulement de tester la "résistance" des modèles aux changements d'échelle, mais également de mettre en avant l'importance de la sélection réfléchie d'un modèle en fonction de la structure des entités spatiales étudiées. En effet, les résultats du chapitre 5 ont montré que l'analyse des flux pendulaires à travers une origine particulière, notamment le comté de Queens, est fortement influencée par des imprécisions liées au milieu urbain. Dans un contexte pareil, il serait alors intéressant de sélectionner un modèle capable de prédire des mouvements intra-urbain, comme présenté dans Liang et al. (2013).

La calibration des paramètres du modèle gravitaire ayant été réalisée par régression linéaire multiple, il est pertinent de se projeter dans une perspective prédictive future et s'interroger sur la valeur de ces paramètres en l'absence de données empiriques. Dans le cas où l'on voudrait, par exemple, prédire les flux pendulaires sortant de la commune de Zurich en 2020, il est possible de partir de l'hypothèse que les structures pendulaires générales restent inchangées. En effet, si l'on part du principe que les réseaux routiers, les réseaux de transport ainsi que les différentes voies de communication ne seront pas sujets à des modifications majeures, les paramètres estimés pour l'année 2000 pourraient être à nouveau utilisés. Effectivement, les paramètres utilisés dans le cadre des prédictions étasuniennes pour 2013 sont les mêmes que ceux ayant été estimés pour le jeu de données de 2000, résultant en des mesures d'ajustement tout à fait correctes (cf. Tab. 2). Cepen-

dant, le cas des flux sortant du comté de Queens en 2013 montre un changement majeur dans la structure des flux de pendulaires, phénomène qui influencerait très probablement la calibration des paramètres. De ce fait, les modèles ne sont pas à l'abri de changements structurels majeurs dont la prévisibilité peut s'avérer limitée, comme dans le cas de la crise financière de 2007-2008. De plus, il convient d'intégrer les phénomènes plus prévisibles tels des investissements majeurs dans certaines aires géographiques créant des opportunités d'emplois, et ayant de ce fait un impact considérable sur les modèles de prédiction.

Références

- Arvis, J.-F. and Shepherd, B. (2013). The Poisson quasi-maximum likelihood estimator : a solution to the ‘adding up’ problem in gravity models. *Applied Economics Letters*, 20(6) :515–519.
- Bozdogan, H. (2000). Akaike’s Information Criterion and Recent Developments in Information Complexity. *Journal of Mathematical Psychology*, 44(1) :62–91.
- Commenges, H. (2016). Modèle de radiation et modèle gravitaire : Du formalisme à l’usage. *Revue Internationale de Géomatique*, 26(1) :79–95.
- Flowerdew, R. and Aitkin, M. (1982). A METHOD OF FITTING THE GRAVITY MODEL BASED ON THE POISSON DISTRIBUTION*. *Journal of Regional Science*, 22(2) :191–202.
- Flowerdew, R. and Lovett, A. (1988). Fitting Constrained Poisson Regression Models to Interurban Migration Flows. *Geographical Analysis*, 20(4) :297–307.
- Kang, C., Liu, Y., Guo, D., and Qin, K. (2015). A Generalized Radiation Model for Human Mobility : Spatial Scale, Searching Direction and Trip Constraint. *PLOS ONE*, 10(11) :e0143500.
- Knudsen, D. C. and Fotheringham, A. S. (1986). Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling. *International Regional Science Review*, 10(2) :127–147.
- Kullback, S. and Leibler, R. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1) :79–86.
- Lenormand, M., Bassolas, A., and Ramasco, J. (2016). Systematic comparison of trip distribution laws and models. *Journal of Transport Geography*, 51 :158–169.
- Lenormand, M., Huet, S., Gargiulo, F., and Deffuant, G. (2012). A Universal Model of Commuting Networks. *PLoS ONE*, 7(10) :e45985.
- Liang, X., Zhao, J., Dong, L., and Xu, K. (2013). Unraveling the origin of exponential law in intra-urban human mobility. *Scientific Reports*, 3.
- Masucci, A. P., Serras, J., Johansson, A., and Batty, M. (2013). Gravity versus radiation models : On the importance of scale and heterogeneity in commuting flows. *Physical Review E*, 88(2).
- Nelder, J. A. and Wedderburn, R. W. M. (1992). Generalized Linear Models. In Kotz, S. and Johnson, N. L., editors, *Breakthroughs in Statistics*, pages 547–563. Springer New York, New York, NY. DOI : 10.1007/978-1-4612-4380-9_39.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2) :461–464.

- Silva, J. S. and Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, 88(4) :641–658.
- Simini, F., González, M. C., Maritan, A., and Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature*, 484(7392) :96–100.
- Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Biol. Skr.*, (5) :1–34.
- Stouffer, S. A. (1940). Intervening Opportunities : A Theory Relating Mobility and Distance. *American Sociological Review*, 5(6) :845.
- Wilson, A. G. (1969). The Use of Entropy Maximising Models, in the Theory of Trip Distribution, Mode Split and Route Split. *Journal of Transport Economics and Policy*, 3(1) :108–126.
- Yang, Y., Herrera, C., Eagle, N., and González, M. C. (2014). Limits of Predictability in Commuting Flows in the Absence of Data for Calibration. *Scientific Reports*, 4.
- Zipf, G. K. (1946). The P 1 P 2 D Hypothesis : On the Intercity Movement of Persons. *American Sociological Review*, 11(6) :677.

Modèles Données	Modèle gravitaire		Modèle radiatif		Modèle radiatif paramétré	
	Rang AIC	Rang DKL	Rang AIC	Rang DKL	Rang AIC	Rang DKL
Queens (2000)	1	1	2	2	3	3
Queens (2013)	3	3	2	2	1	1
USA (2000)	1	1	3	3	2	2
USA (2013)	1	1	3	3	2	2
Zurich (2000)	2	2	1	1	3	3
Suisse (2000)	2	2	3	3	1	1
	Rang AIC	Rang DKL	Rang AIC	Rang DKL	Rang AIC	Rang DKL