



ESPACE

LETG



ThéMA
UMR 6049 Théoriser & Modéliser pour Aménager

57^e
Colloque
ASRDLF

Association
de Science Régionale
de Langue Française

Estimation prospective des flux journaliers de navetteurs à l'aide de modèles gravitaires

Analyse de sensibilité dans un contexte de données massives

D. Josselin, UMR ESPACE, F. Audard, LETG, S. Carpentier-Postel, ThéMA



Introduction

Mission Prospective de la Région PACA (scénarios)

Comparaison systématique de différents modèles gravitaires cherchant à concilier :

⇒ *la facilité d'accès au données*

⇒ *la simplicité de mise en œuvre*

⇒ *la qualité de la prévision*

afin d'assurer la transparence et la répliquabilité de la démarche

⇒ *Geomatica (Josselin et al., 2021)*

Différents modèles gravitaires pour estimer les flux

Auteur	Objectif	Formule	Masse	Distance ($d_{ij} > 0$)	Commentaire
Tobler 1970	Évaluation de l'attractivité selon la distance	$F_{ij} = \frac{1}{d_{ij}^\gamma}$	Non prise en considération	Effet négatif	Relation uniquement liée à la distance; problème de division par 0
Reilly 1931; Converse 1951	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 1687	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 log-linéaire; problème de division par 0; pas d'unité
Wilson 1975	Estimation de flux migratoires	$F_{ij} = aM_i \frac{M_j f(C_{ij})}{\sum_{k=1}^n M_k f(C_{ik})}$	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 $f(\Delta)$ (Δ = 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

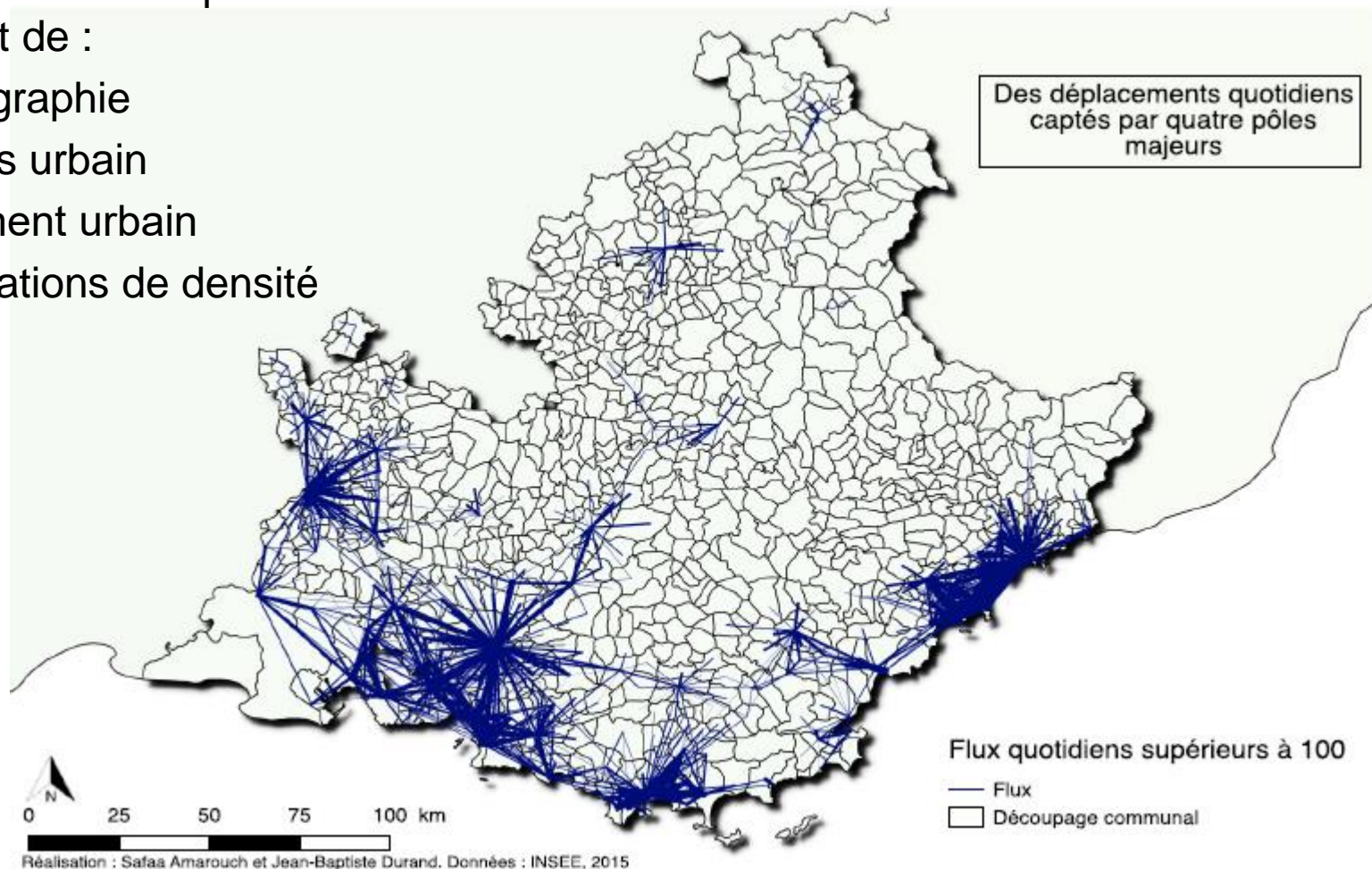
Nota : M_i , masse origine du flux; M_j , masse destination du flux; α , exposant de la masse d'origine; β , exposant de la masse de destination; γ , exposant de la distance; a , coefficient multiplicateur; C_{ij} , coût de déplacement de i vers j ; c , constante.

Aire d'étude : région PACA

Région très contrastée.

Les déplacements dépendent
notamment de :

- la topographie
- le semis urbain
- l'étalement urbain
- les variations de densité



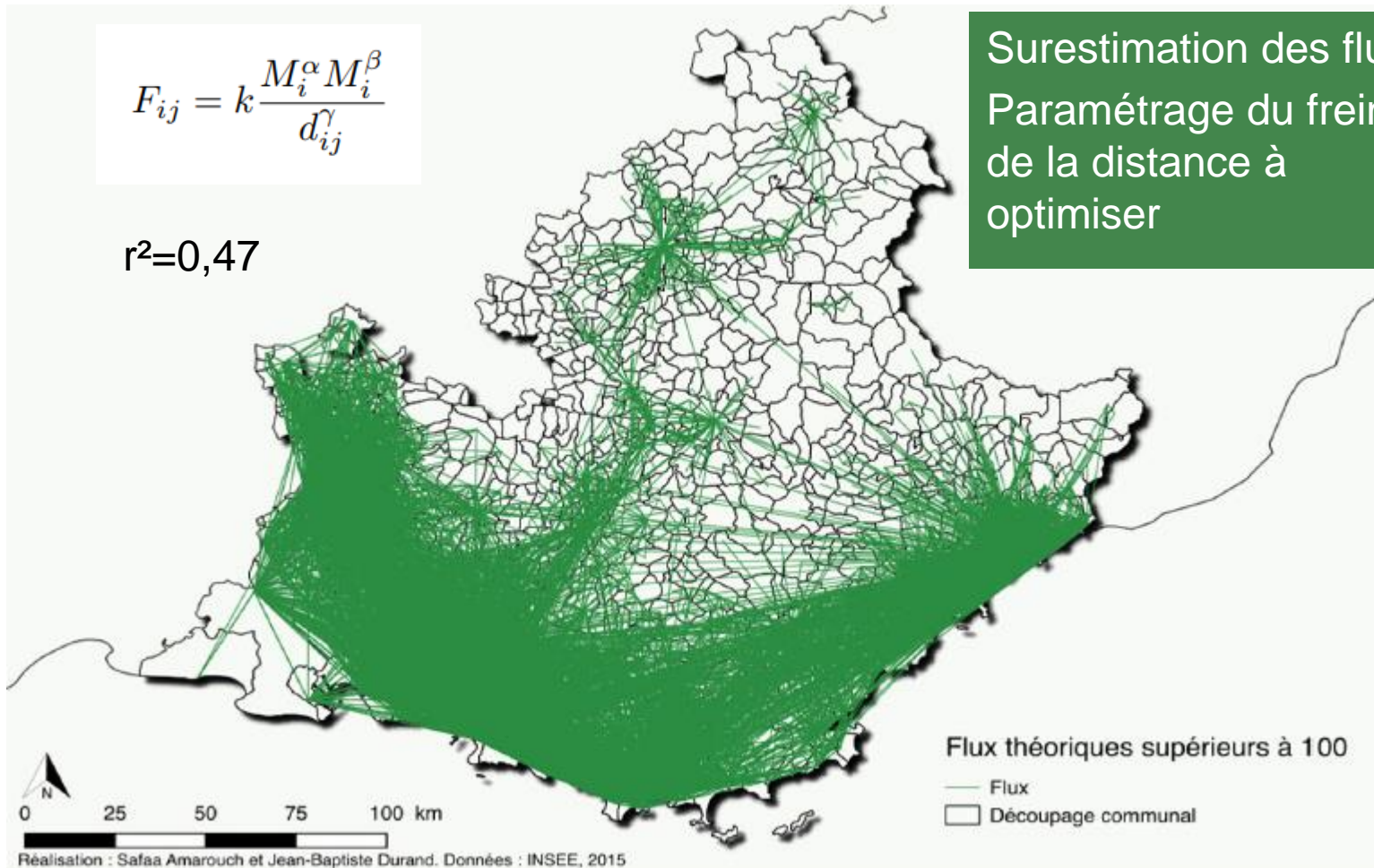
Approche exploratoire : application du modèle de Newton

$\alpha = 1$
 $\beta = 1$
 $\gamma = 2$

$$F_{ij} = k \frac{M_i^\alpha M_j^\beta}{d_{ij}^\gamma}$$

$r^2=0,47$

Surestimation des flux
Paramétrage du frein
de la distance à
optimiser



Les navettes intracommunales sont ignorées ($d=0$)

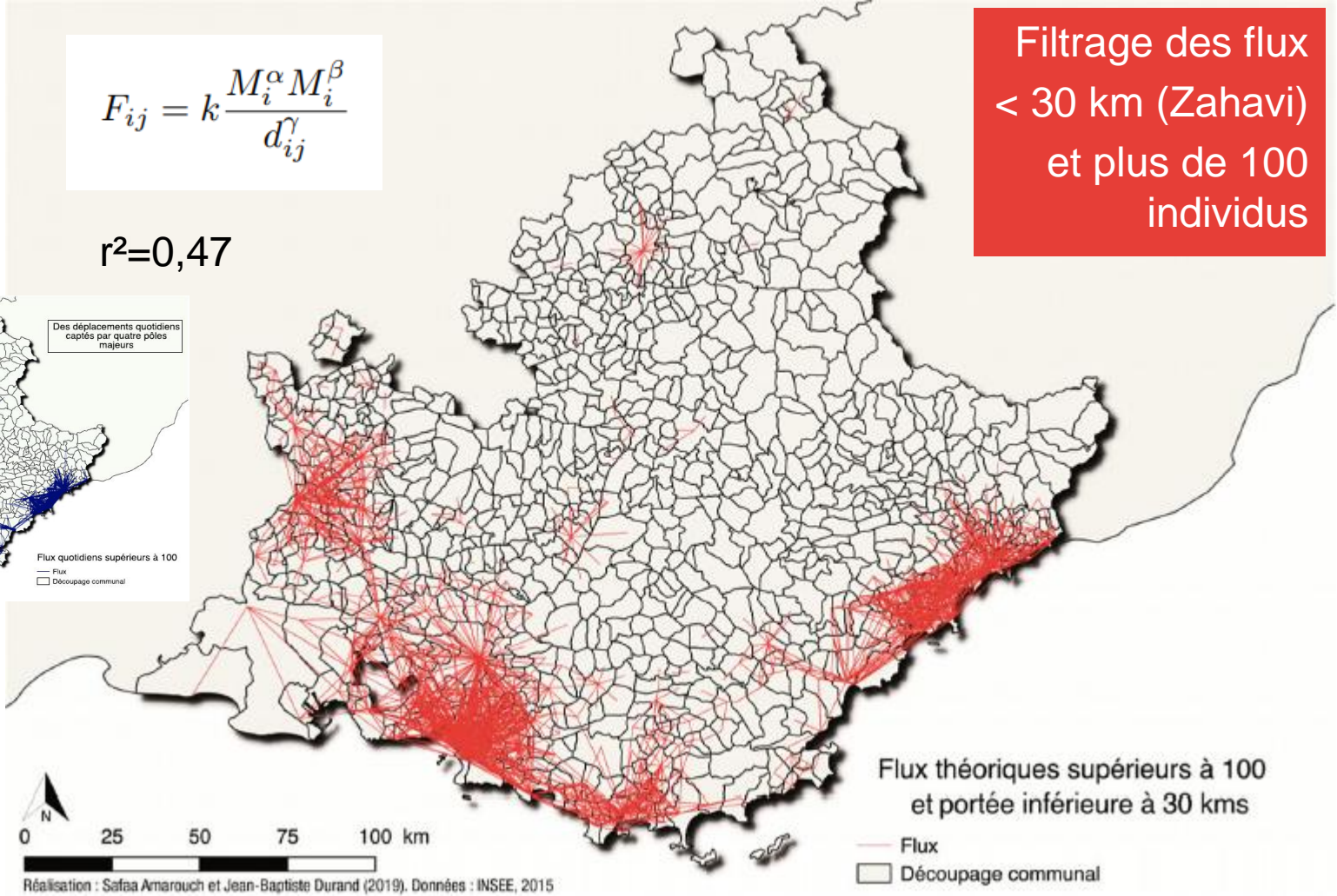
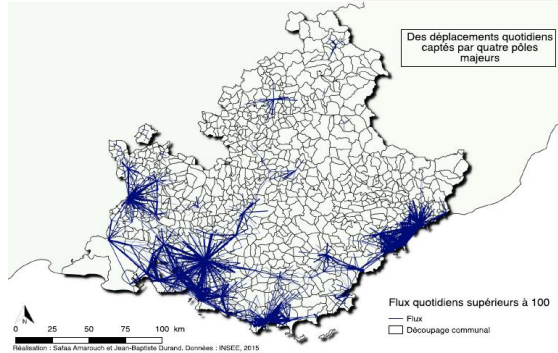
Approche exploratoire : application du modèle de Newton

$\alpha = 1$
 $\beta = 1$
 $\gamma = 2$

$$F_{ij} = k \frac{M_i^\alpha M_j^\beta}{d_{ij}^\gamma}$$

$r^2=0,47$

Filtrage des flux
< 30 km (Zahavi)
et plus de 100
individus



Les navettes intracommunales sont ignorées ($d=0$)

Démarche

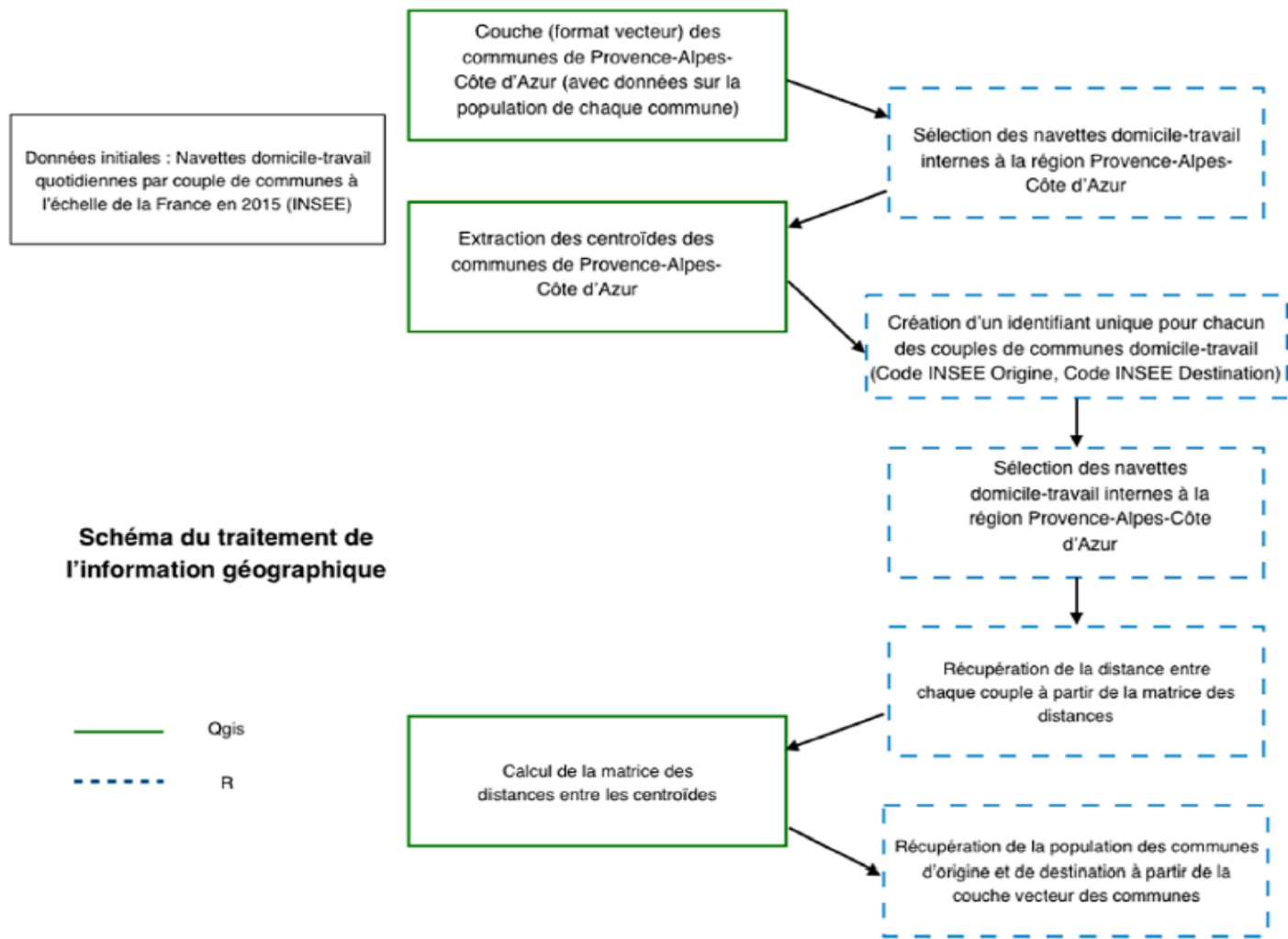
Au-delà d'une comparaison visuelle, une analyse approfondie de la qualité des modèles est nécessaire.

⇒ Focalisation sur les déplacements domicile–travail internes des communes et des arrondissements des métropoles de la région Provence-Alpes-Côte d'Azur

⇒ Données RGP INSEE 2016

18 modèles produits et comparés

Traitement de l'information géographique



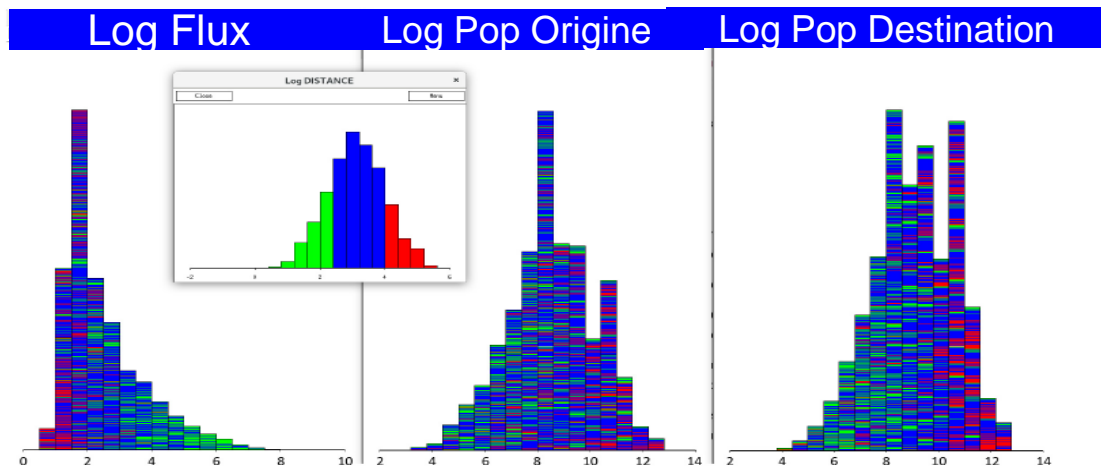
Réalisation : Safaa Amarouch et Jean-Baptiste Durand

Modélisation

Application d'un modèle de Newton généralisé et linéarisé par une fonction log :

$$(1) \quad \log(F_{ij}) = \alpha \log(M)_i + \beta \log(M)_j - \gamma \log(d_{ij}) + \log(k)$$

- ⇒ les exposants deviennent alors des « élasticités » qui identifient les poids respectifs des critères dans l'estimation des flux (F_{ij})
- ⇒ on peut alors appliquer des régressions log-linéaires multiples pour estimer ces élasticités



Variation des paramètres (1/2)

Distance (D_{ij}) :

- remplacer les valeurs nulles par 1 m
- calculer un cercle de rayon R de portée théorique basée sur la surface S
- générer aléatoirement par *jittering* des distances allant de 0 (exclu) à ce rayon théorique R

Masse des origines (M_i)

- population de l'entité
- nombre d'actifs (de 15 à 64 ans)

Masse des destinations (M_j)

- population de l'entité
- nombre d'emploi

Variation des paramètres (2/2)

Taille de l'échantillon

- tirage aléatoire dans l'échantillon de 1%, 2%, 5%, 10%, 25%, à 50% des individus ($n=250$ à $n=12000$)

Seuil maximal de portée spatiale de flux

- enlever les flux de longue distance avec des seuils variables pour tester la contrainte de budget-temps de Zahavi (*ces choix impactent aussi la taille de l'échantillon*)

Seuil minimal de quantité de flux par entité

- enlever les communes possédant un nombre de flux inférieur à un certain seuil, afin de réduire si possible le bruit dans le modèle dû aux *outliers* potentiels (*ces modifications réduisent plus ou moins la taille de l'échantillon*)

Résultats (1/2)

DONNÉES de FLUX			FORMULE de la RÉGRESSION LOG-LINEAIRE				QUALITÉ du MODÈLE							
O : origine, D : destination			Élasticités (exposants des variables)				Intervalles de confiance à 1% [min ; max]				Ajustement		Variabilité des résidus	
ID	Description	Effectif N ≈	Intercept log(k)	DISTANCE γ	POP_ORIG α	POP_DEST β	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	24 000	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	24 000	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	24 000	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	24 000	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	24 000	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	24 000	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 %	12 000	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	4,703	-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

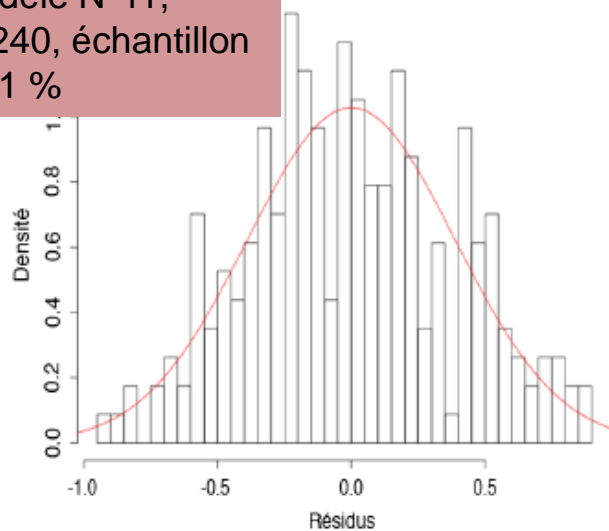
Résultats (2/2)

DONNÉES de FLUX		FORMULE de la RÉGRESSION LOG-LINEAIRE					QUALITÉ du MODÈLE							
O : origine, D : destination			Élasticités (exposants des variables)			Intervalle de confiance à 1 % [min ; max]				Ajustement	Variabilité des résidus			
ID	Description	Effectif $N \approx$	Intercept log(k)	DISTANCE γ	POP_ORIG α	POP_DEST β	Intercept	DISTANCE	POP_ORIG	POP_DEST	R^2 ajusté	RMSE	SRMSE	Entropie
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	23 000	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	19 500	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	15 000	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	23 000	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	16 500	3,98	-1,183	0,267	0,415	[3,896; 4,064]	[-1,206; -1,160]	[0,255; 0,278]	[0,404; 0,417]	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	10 000	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

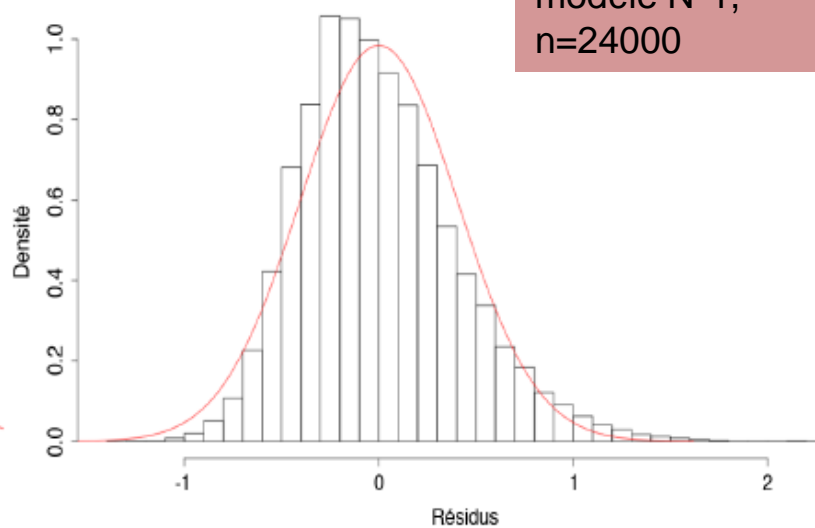
Nota : α , exposant de la masse d'origine; β , exposant de la masse de destination; γ , exposant de la distance; RMSE, somme des carrés des résidus du modèle, erreur quadratique moyenne; SRMSE, erreur quadratique moyenne standardisée.

Analyse des résidus

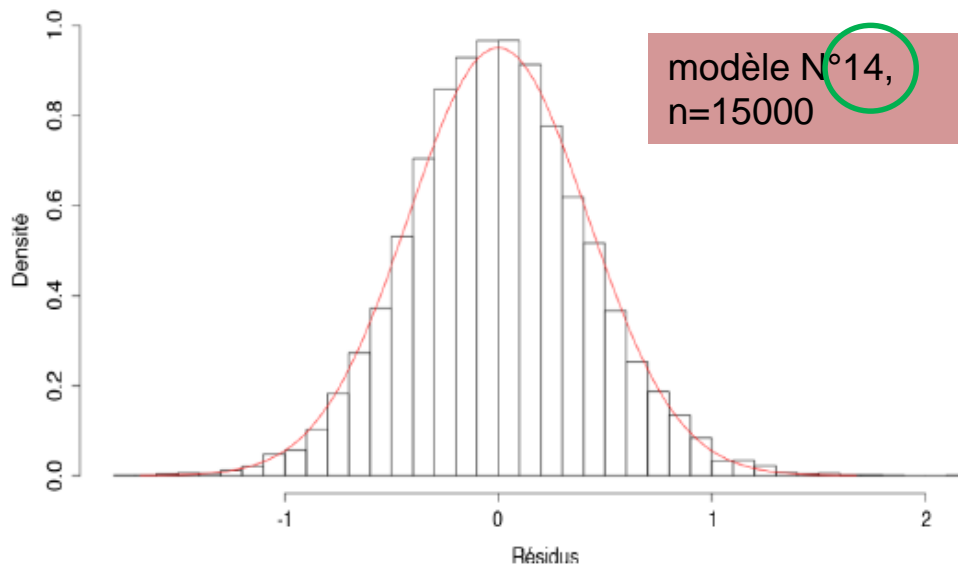
modèle N°11,
n=240, échantillon
de 1 %



modèle N°1,
n=24000



modèle N°14,
n=15000

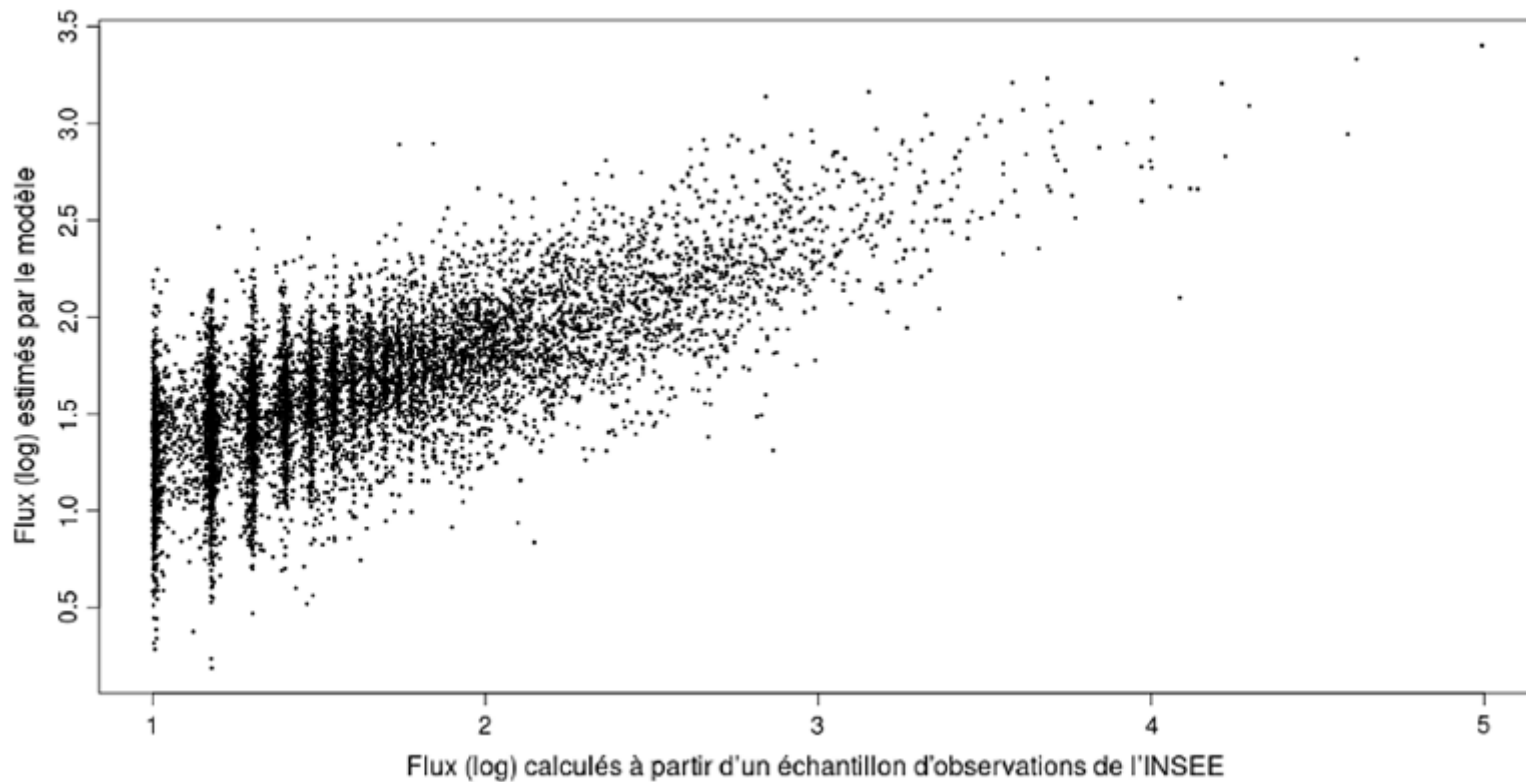


Validation statistique

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

Analyse des résidus

Comparaison des valeurs log des flux observés et théoriques pour le modèle N°18



Discussion / perspectives (1/2)

Principaux résultats:

- ⇒ élasticité négative de la distance et élasticités positives pour les masses
- ⇒ l'effet de la distance entre les entités sur la quantité de migrations alternantes est en moyenne 4 à 5 fois supérieur à celui de la masse de l'entité à l'origine du flux et 2 à 3 fois supérieur à celui de la masse de l'entité destination.
- ⇒ les caractéristiques de l'entité de destination interviennent jusqu'à 2,5 fois plus que l'entité d'origine dans la prévision du flux. C'est donc l'attractivité de la commune de travail qui est déterminante par rapport à son émissivité, très fortement régulée par la distance.

On s'éloigne sensiblement des modèles de la loi universelle de gravitation de Newton ou d'approches de type Reilly, où la distance est affectée d'un exposant 2

Discussion / perspectives (2/2)

Si les valeurs des élasticités varient d'un modèle à l'autre, les rangs et les ordres de grandeur sont conservés, montrant une grande cohérence dans les résultats, et ce en dépit parfois de grandes variations d'effectifs dans les échantillons.

On constate qu'il est possible de descendre jusqu'à 10 % de l'échantillon (soit près de 1200 individus sur 24 000), sans réellement perdre en qualité d'ajustement.

Un échantillon de 10% ne prenant en compte que les flux de moins de 30 km, les populations actives des communes d'émission et le nombre d'emplois des communes réceptrices peut réduire la massivité des données tout en produisant des résultats relativement corrects.



ThéMA
UMR 6049 Théoriser & Modéliser pour Aménager



Estimation prospective des flux journaliers de navetteurs à l'aide de modèles gravitaires

Analyse de sensibilité dans un contexte de données massives

D. Josselin, UMR ESPACE, CNRS

didier.josselin@univ-avignon.fr

F. Audard, LETG, S. Carpentier-Postel, ThéMA



Merci de votre attention !