

Axe transverse Milieux Urbains

"Statistical reliability of the modified areal weighted by control zones method to spatially downscale individual social data "

« Fiabilité statistique de la méthode de désagrégation avec pondération par zone de contrôle pour désagréger les données sociales »

Najla Touati, Wilda Jean Baptiste, Julia Hidalgo
Laboratoire Interdisciplinaire
Solidarités, Sociétés, Territoires
LISST

Objectifs

- Cette étude évalue l'erreur impliquée dans la réduction d'échelle des données sociales d'une grande maille spatiale à une maille plus petite.
- La méthode de réduction d'échelle évaluée est la méthode Modified Areal Weighting by Control Zones* (Langford, 1992 ; Goodchild, 1993 ; Plumejeaud, 2009). → utilisé dans MapUCE
- Partir d'une maille spatiale large (données INSEE ou FIDELI à l'échelle IRIS), et descendre à une maille plus fine (échelle RSU).
- Les évaluations sont réalisées pour deux cas d'étude : Toulouse et Grenoble-Alpes Metropoles.

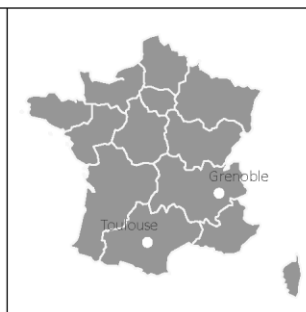
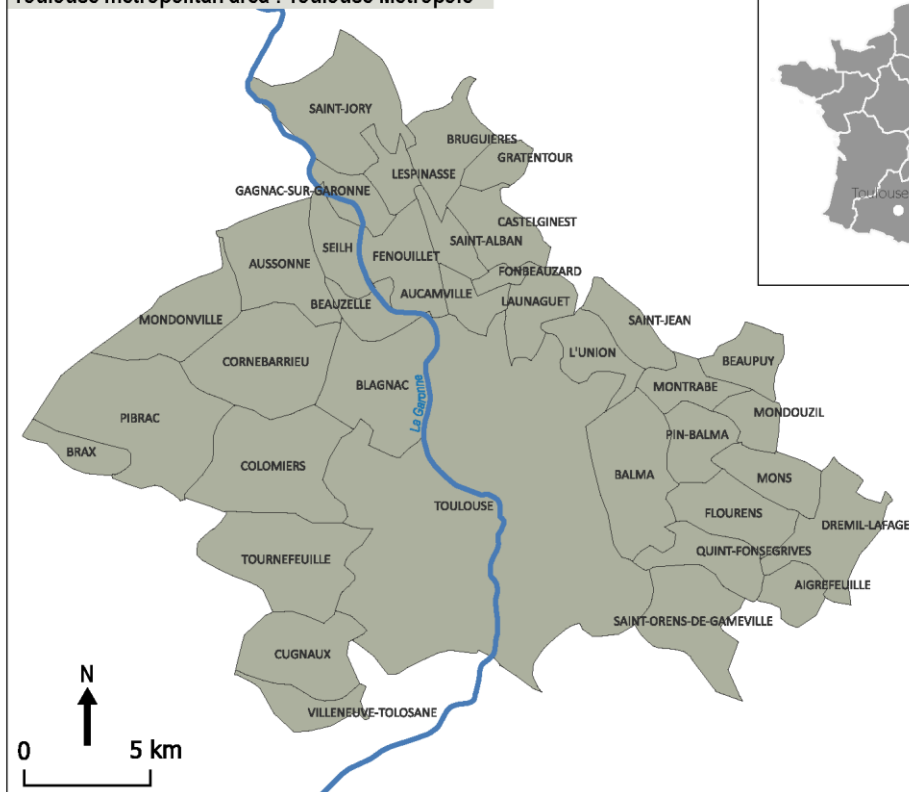
*MAW-CZ est une forme de cartographie dasymétrique à support intermédiaire quantitatif (quantitative proxy-based dasymetric mapping with intermediate support). Parfois appelé « Hybrid dasymetric method » (Pajares et al., 2021),

Deux cas d'étude

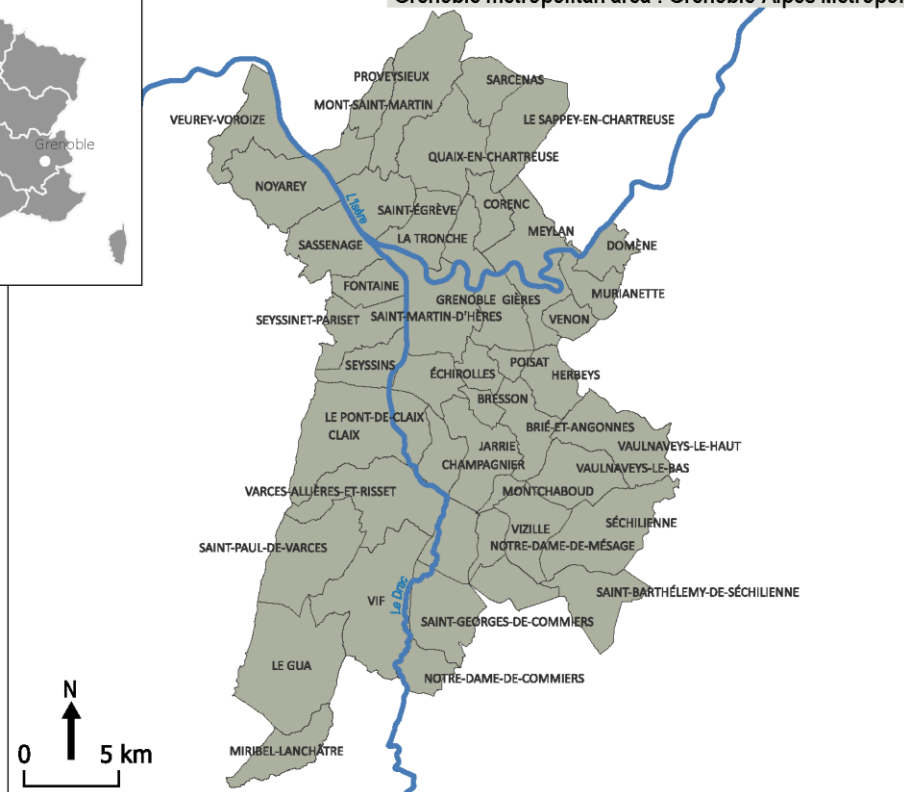
Toulouse Metropole and Grenoble Alpes Metropole

	Toulouse Metropole	Grenoble Alpes Metropole
Number of IRIS (INSEE)	254	201
RSU (via Geoclimate)	24 827	19048
Buildings (via Geoclimate)	355 517	130 946
Inhabitants (FIDELI)	774 971	445 376

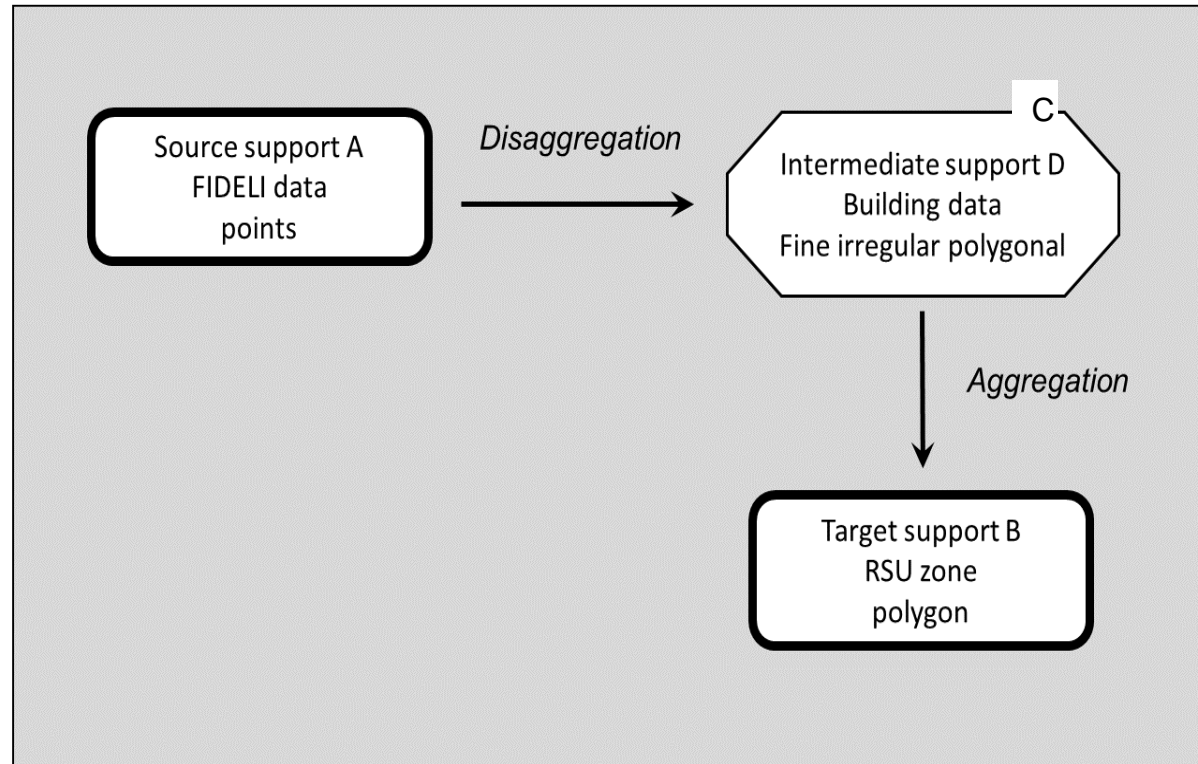
Toulouse metropolitan area : Toulouse Metropole



Grenoble metropolitan area : Grenoble-Alpes Metropole



The Modified Areal Weighting by Control Zones method ou méthode de désagrégation avec pondération par zone de contrôle.

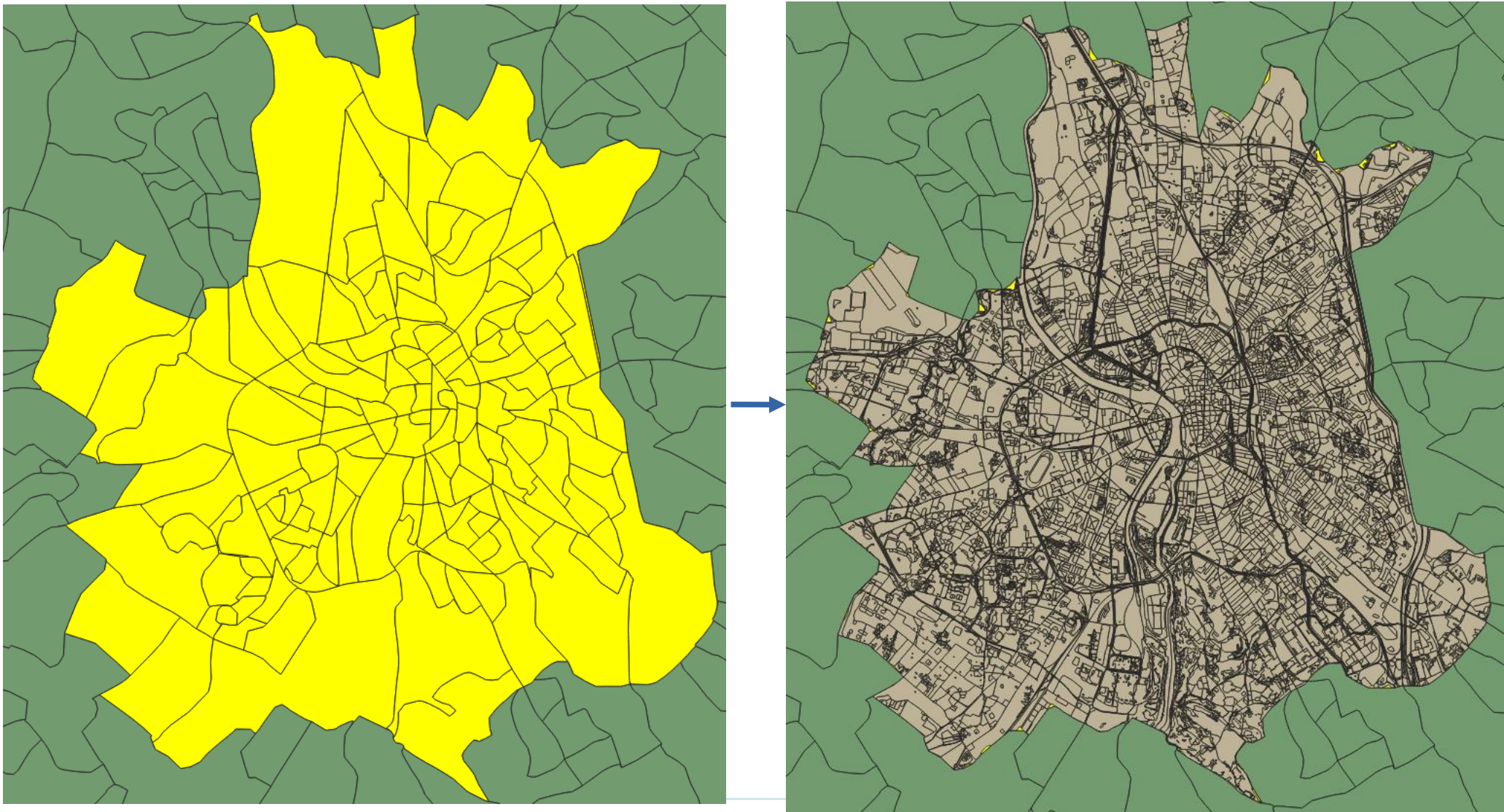


Support transformation data transfer (from Plumejeaud et al., 2009).

IRIS indicates the larger statistical unit (the smallest sub-municipal unit available for public statistical analyses in France), and RSU indicates the smaller topographical reference unit.

IRIS units (left) and RSUs (right) for Toulouse

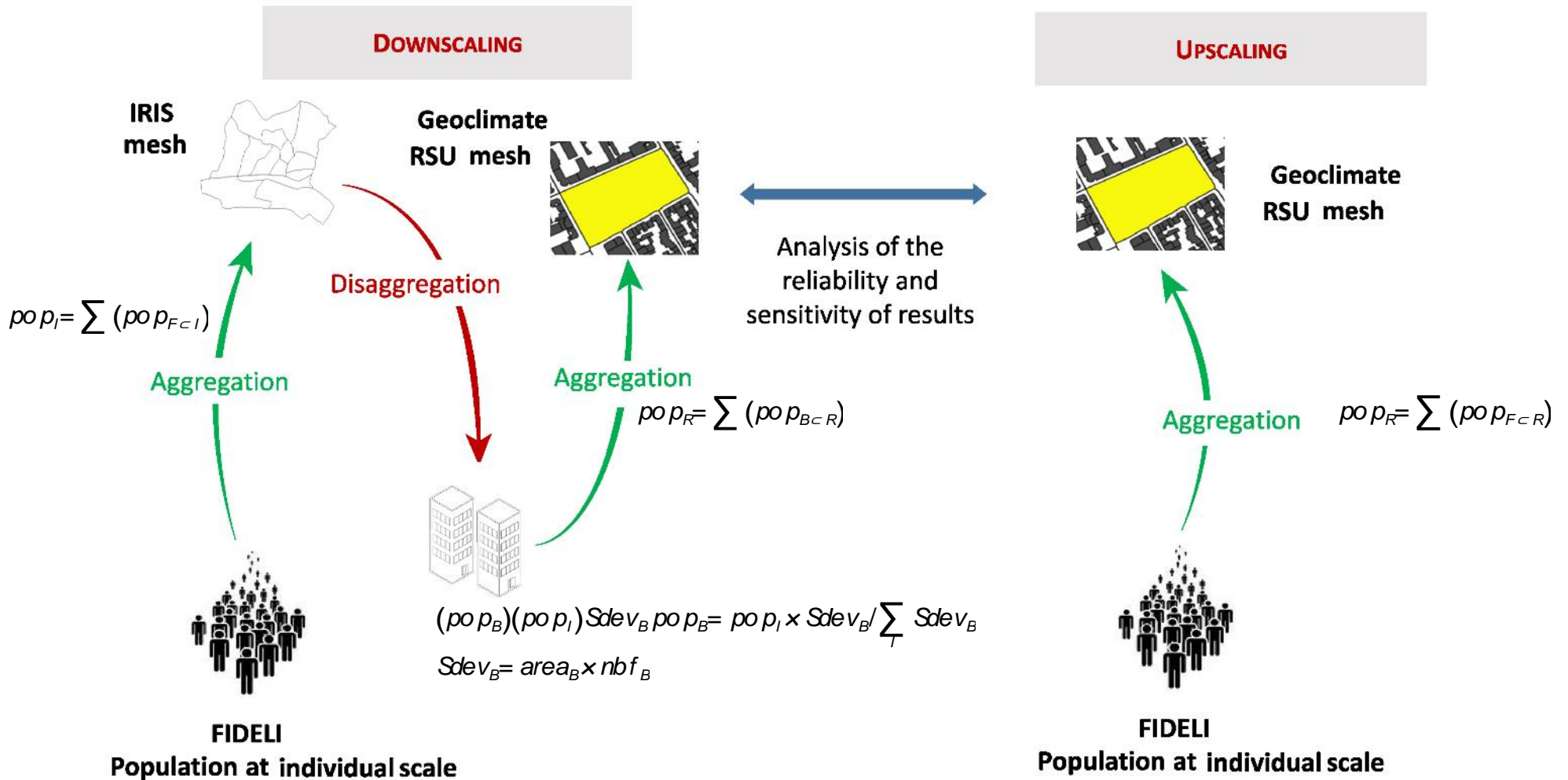
+



Pour les RSU, voir Erwan Bocher, Jérémy Bernard, Elisabeth Le Saux Wiederhold, François Leconte, Gwendall Petit, et al.. Geo-Climat: a Geospatial processing toolbox for environmental and climate studies. Journal of Open Source Software, 2021, 6 (65), pp.3541. (10.21105/joss.03541). (halshs-03359757)

Data processing scheme

+



Results: Descriptive statistics

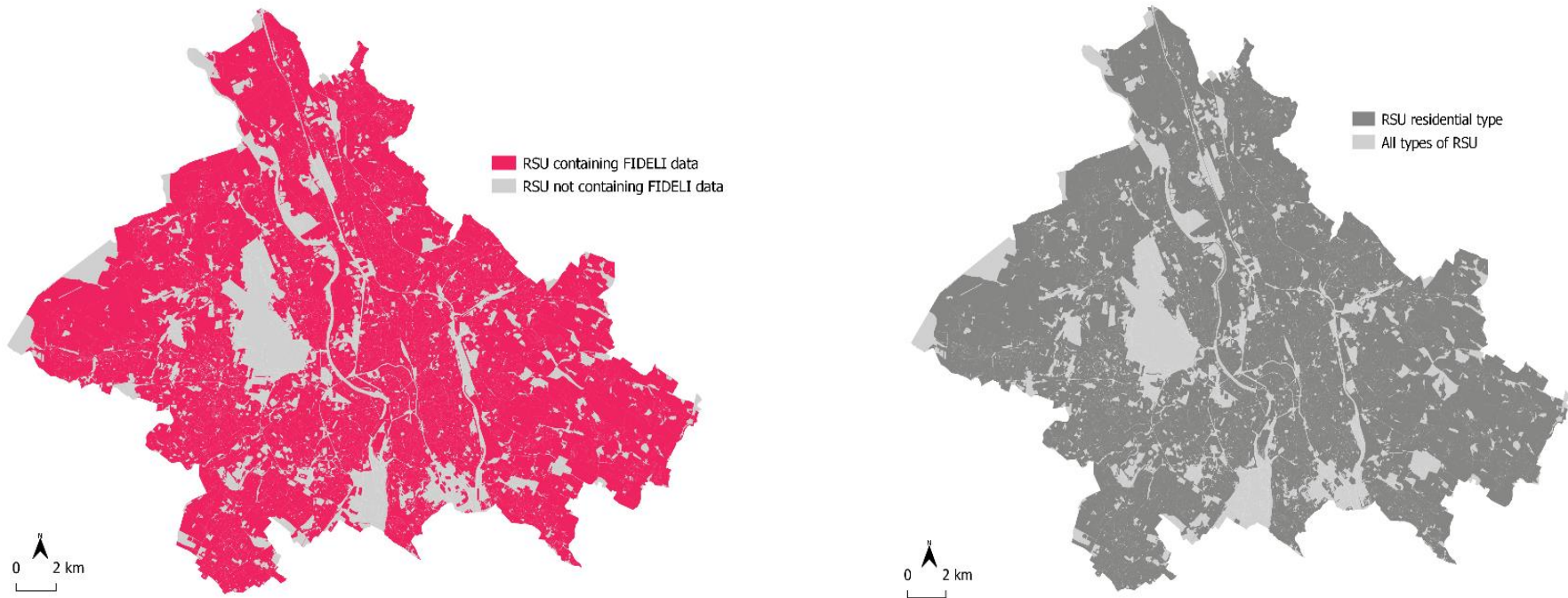
+

	Toulouse		Grenoble-Alpes	
	Upscaling	Downscaling	Upscaling	Downscaling
Total RSUs	24,827		19,048	
Inhabited RSUs	6746 (27.2%)	6302 (25.4%)	4164 (21.9%)	4197 (22%)
Uninhabited RSUs	18,081 (72.8%)	18,525 (74.6%)	14,884 (78.1%)	14,851 (78%)

Differences are explained by people “living” in other type of buildings than residential.

Mapping previous results

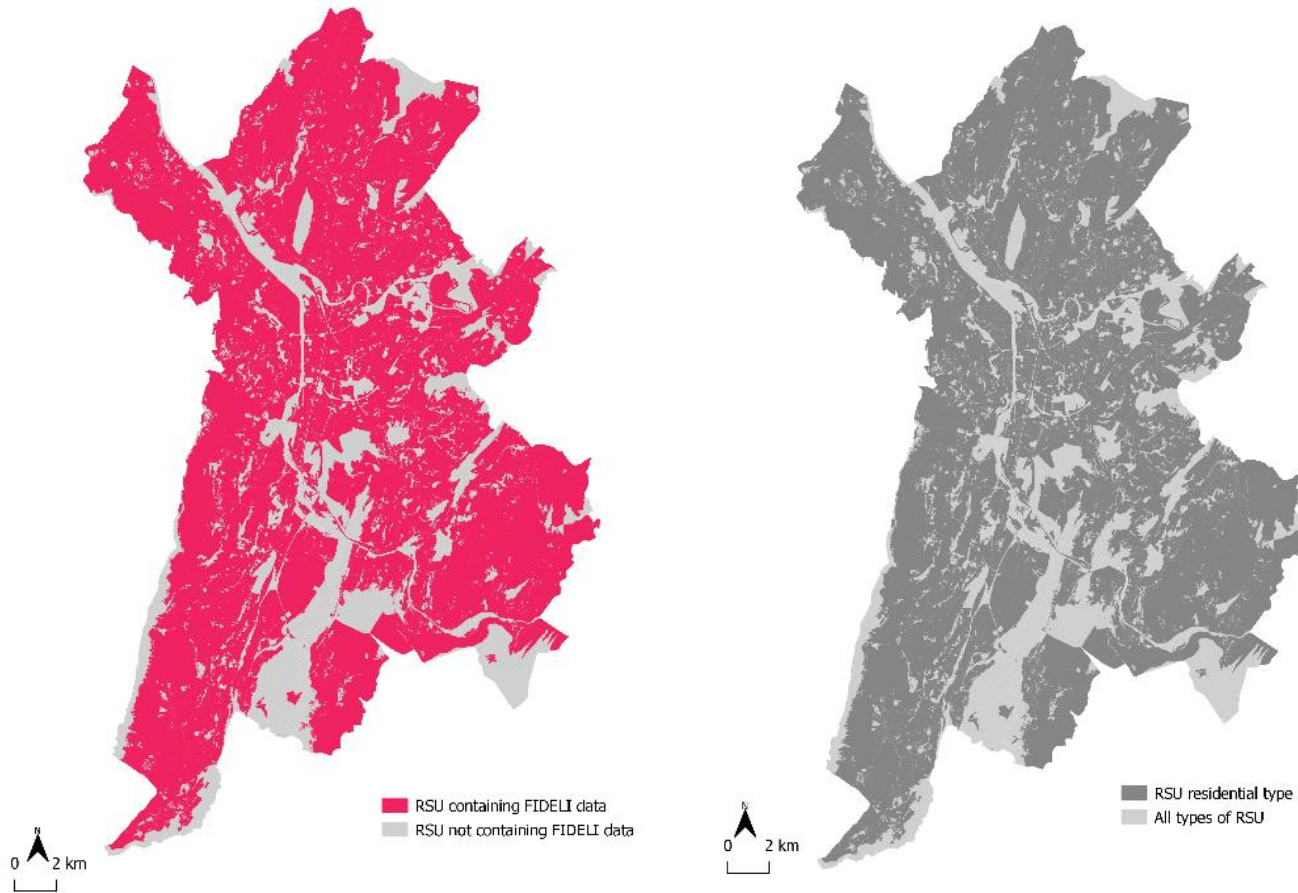
+



Spatial distributions of (a) inhabited and uninhabited RSUs and (b) residential RSUs in the Toulouse Metropole.

Mapping previous results

+



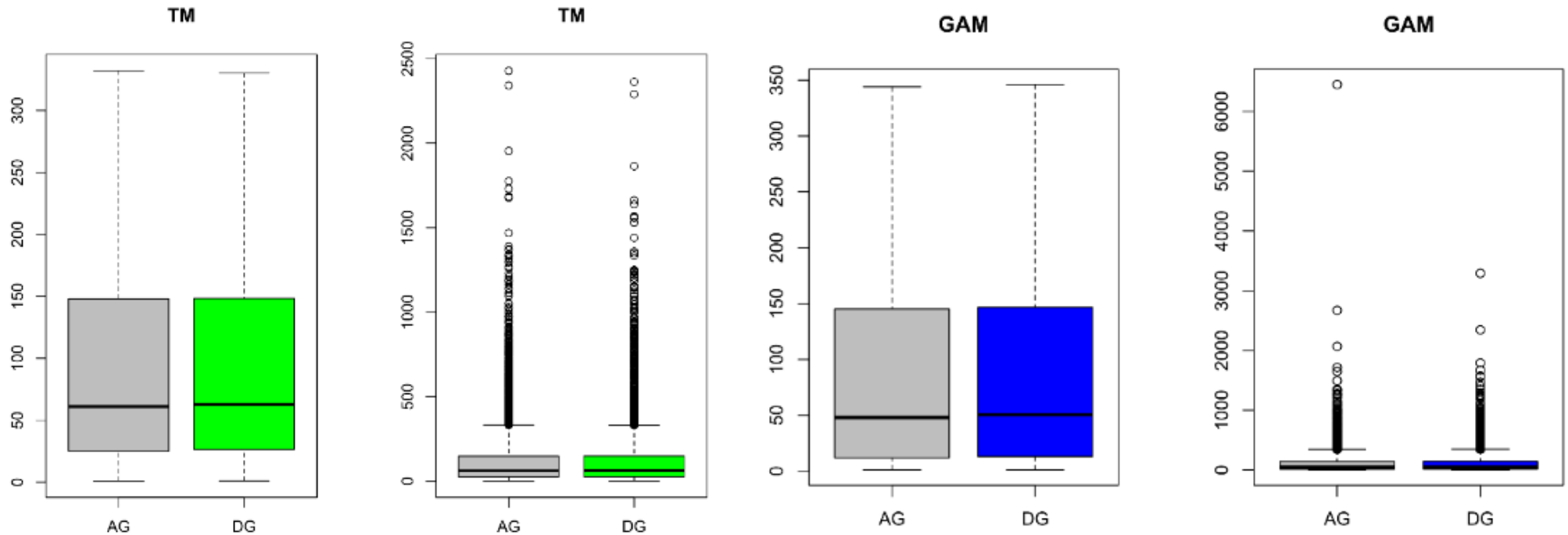
Spatial distributions of (a) inhabited and uninhabited RSUs and (b) residential RSUs in Grenoble-Alpes Metropole.

Results: Descriptive statistics

Statistical Indicators	Toulouse		Grenoble-Alpes	
	Upscaling	Downscaling	Upscaling	Downscaling
Minimum	1	1.1	1	1
1 st quartile	25	26.5	12	13.1
Median	61	62.9	48	50.4
Mean	122	120.8	116	115.2
3 rd quartile	148	148.3	145	146.7
Maximum	2429	2363.8	6451	3291.7
Standard deviation	178.5	170.1	207.5	185.1
Variance	31,862.2	28,958.9	46.2	34,257.8

Results: Descriptive statistics

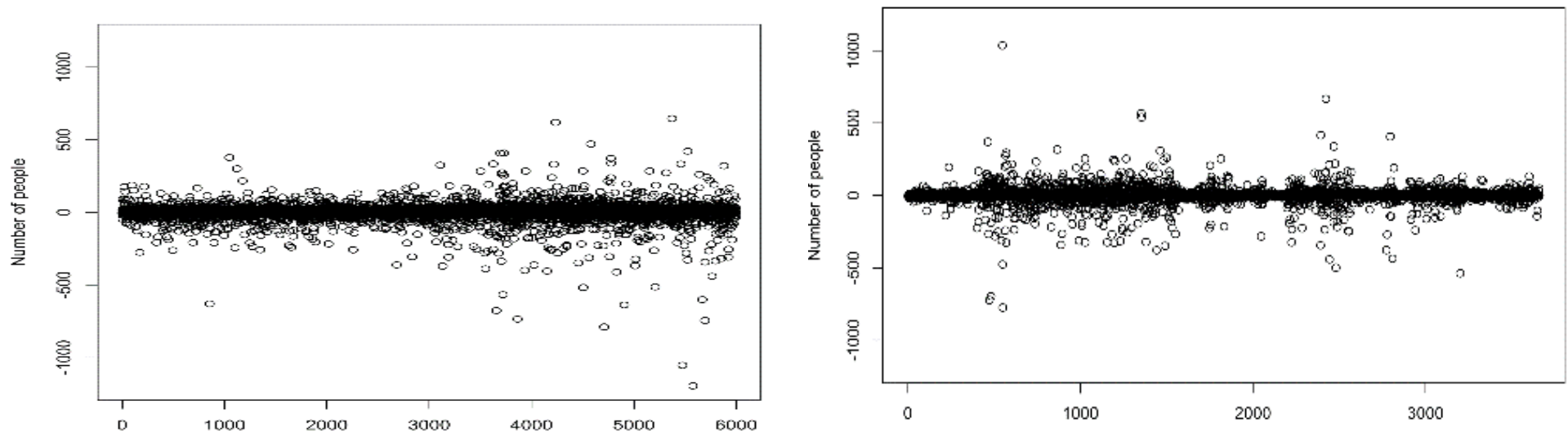
+



Method	Outliers for Toulouse	Outliers for Grenoble-Alpes
Upscaling	8.4% (504/5993)	5.6% (265/4661)
Downscaling	7.9% (474/5993)	5.5% (258/4661)

Résultats : Évaluation des biais

Écarts dans les métropoles de Grenoble-Alpes (à droite) et de Toulouse (à gauche).



Indicators	Toulouse	Grenoble-Alpes
Minimum	0	0
1 st quartile	0.11	0.13
Median	0.23	0.29
Mean	0.58	0.63
3 rd quartile	0.46	0.57
Maximum	332.52	133.47
Standard deviation	5.55	2.68
Variance	30.81	7.20

La figure indique que la plupart des erreurs sont proches de zéro.

Cependant, pour certaines RSU, l'erreur entre les deux méthodes est significative (ex. moyenne de 58% à Toulouse).

Results: Bias evaluation $d_r = D_r - U_r$

Table . Percentage of RSUs under different thresholds. Threshold	Toulouse	Grenoble-Alpes
Less than 5%	12.0%	10.2%
Between 5% and 10%	11.8%	10.2%
Between 10% and 15%	11.4%	9.9%
Between 15% and 30%	25.2%	21.9%
Between 30% and 50%	17.5%	18.7%
Between 50% and 80%	12.0%	13.1%
More than 80%	10.1%	16.0%

23,8 % des RSU ont un taux d'erreur inférieur à 10 % à Toulouse M.

Résultats : Distribution spatiale des erreurs

+

Indice de Moran & autocorrélation spatiale

- L'indice de Moran mesure si les erreurs sont **spatialement regroupées, dispersées** ou **aléatoires**.
- Il évalue l'**autocorrélation spatiale** entre unités voisines. (*mesurer la dépendance spatiale entre les valeurs d'une même variable en différents endroits de l'espace*)

Interprétation :

- ✓ **Moran's $I > 0$** → regroupement spatial (*clustering*) de valeurs similaires
- ✓ **Moran's $I \approx 0$** → distribution aléatoire, absence de structure spatiale
- ✓ **Moran's $I < 0$** → dispersion spatiale de valeurs contrastées (*dispersion*)

Résultats : Distribution spatiale des erreurs

+

Dans nos cas :

- Appliqué aux **erreurs/résidus** pour vérifier l'existence d'un biais spatial ;
 - **Non significatif** → pas de biais géographique structuré ;
 - **Significatif** → présence de structures spatiales locales non captées par le modèle.
-

Résultats : Distribution spatiale des erreurs

+

Les erreurs sont-elles concentrées spatialement, ou réparties aléatoirement ?

Indice de Moran's I calculé sur les erreurs absolues (niveau RSU)

Toulouse Métropole

I = 0,01

$p = 0,36 \rightarrow$ non significatif

✓ Erreurs aléatoires

Aucun biais spatial détecté.
La méthode MAW-CZ ne génère pas de concentration géographique des erreurs à Toulouse.

Grenoble-Alpes Métropole

I = 0,14

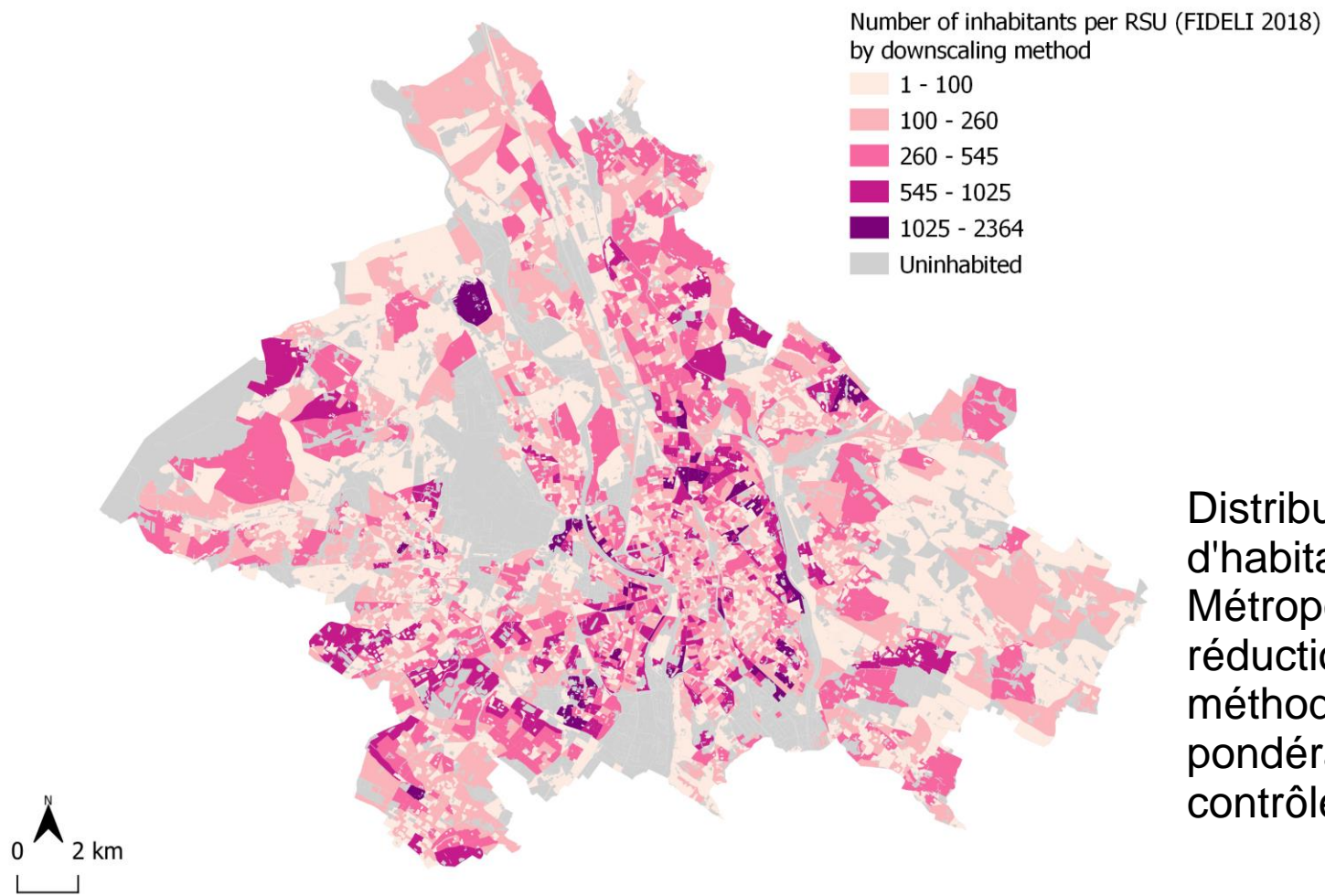
$p < 0,0001 \rightarrow$ significatif

⚠ Léger clustering

Légère concentration spatiale des erreurs dans les zones à morphologie urbaine complexe ou hétérogène (pentes, tissu dense).

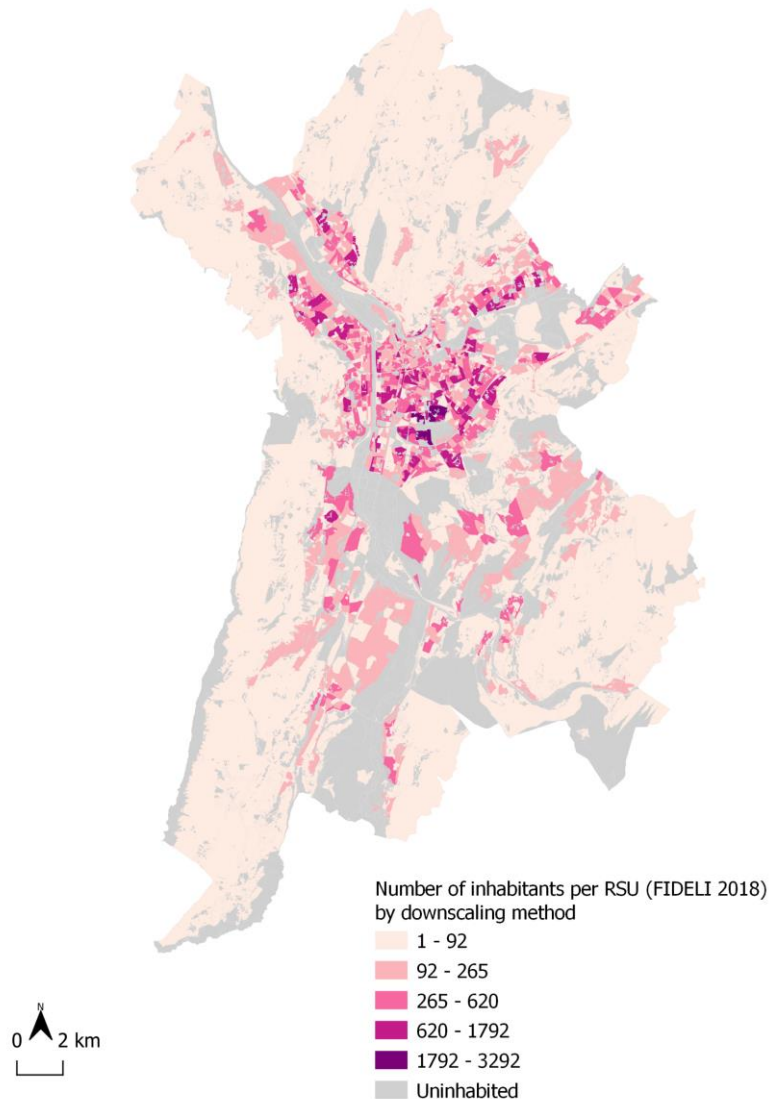
Conclusion : La méthode MAW-CZ ne génère pas de biais spatialement structuré à Toulouse. À Grenoble, le proxy (surface bâtie) capte moins bien la variabilité démographique locale en contexte topographique contraint.

Results: Bias evaluation



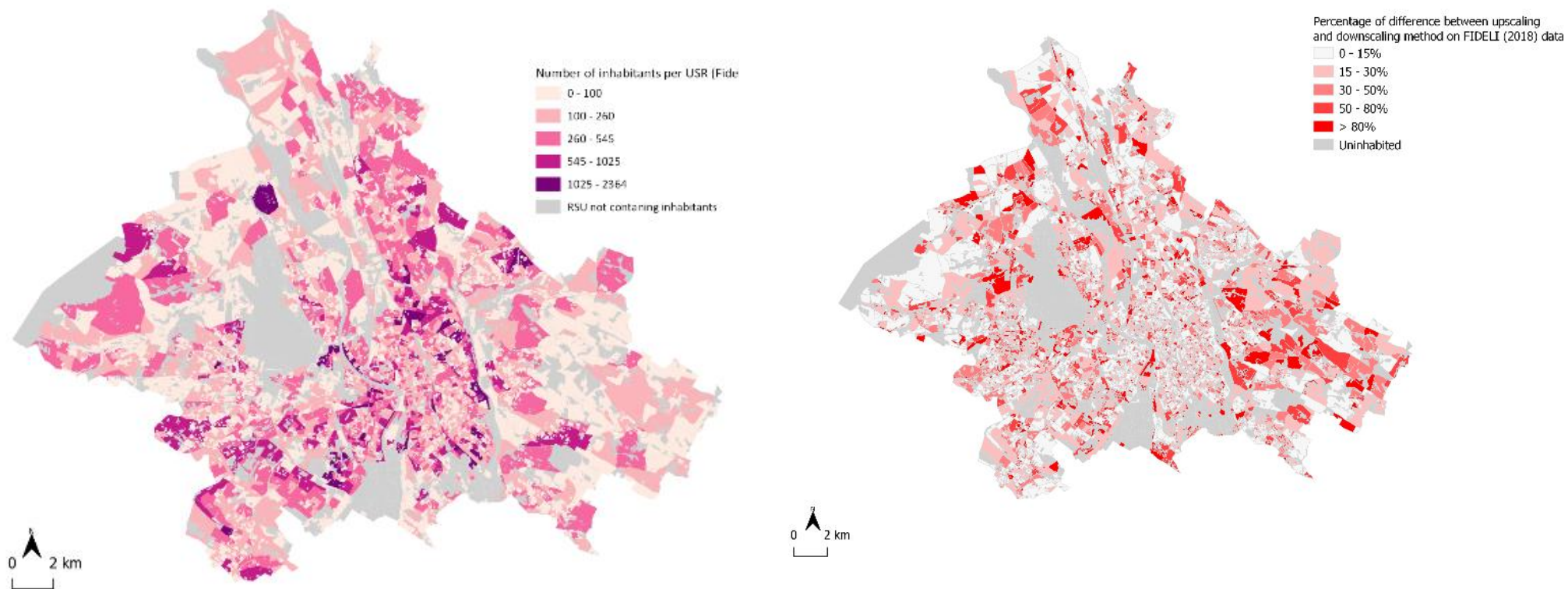
Distribution spatiale du nombre d'habitants de Toulouse Métropole par RSU lors de la réduction d'échelle avec la méthode de désagrégation avec pondération par zone de contrôle.

Results: Bias evaluation



Distribution spatiale du nombre d'habitants de Toulouse Métropole par RSU lors de la réduction d'échelle avec la méthode de désagrégation avec pondération par zone de contrôle.

Results: Bias evaluation



Distribution spatiale du nombre d'habitants de Toulouse Métropole par RSU lors de la réduction d'échelle par la méthode de désagrégation avec pondération par zone de contrôle (gauche) et carte du taux d'erreur par rapport au jeu de données de référence obtenu en agrégeant les données FIDELI (à droite)

ex. pour les RSU de 1 à 97 habitants,
94% des RSU ont été classées dans la
bonne catégorie dans les métropoles de
Toulouse et de Grenoble-Alpes.

Results: Confusion matrix

	Toulouse - Downscaling					
Upscaling	1-97	98-255	256-520	521-956	957-2012	
1-97	94%	6.2%	0.2%	<0.1%	0%	
98-255	13%	78%	8.7%	0.1%	0%	
256-520	2.3%	22%	72%	3.3%	0%	
521-956	1.2%	4.2%	27%	61%	6.6%	
957-2012	0%	6.7%	2.2%	27%	62%	
2012-6351	0%	0%	0%	0%	50%	

	Grenoble - Downscaling					
Upscaling	1-97	98-255	256-520	521-956	957-2012	2012-6351
1-97	94%	5.5%	<0.1%	0%	0%	0%
98-255	14%	75%	11%	0.4%	0%	0%
256-520	3.7%	21%	68%	6.9%	0.3%	0%
521-956	1.2%	4.8%	23%	65%	6%	0%
957-2012	0%	0%	3.3%	37%	60%	0%
2012-6351	0%	0%	0%	0%	0%	100%

Conclusions — Mise en perspective

✓ Convergence avec la littérature sur les biais dasymétriques

La sous-estimation des RSU très peuplés est un biais structurel documenté universellement dans la littérature dasymétrique (Eicher & Brewer, 2001 ; Cockx & Canters, 2015 ; Mennis & Hultgren, 2006). Notre étude confirme ce résultat à l'échelle fine IRIS → RSU, avec une précision qui se dégrade à 60 % au-delà de 500 hab. — cohérent avec le constat que "les paramètres globaux imposent un effet de moyenne masquant l'hétérogénéité intrinsèque" (Cockx & Canters, 2015).

★ Contribution originale : première évaluation statistique spatialement explicite

Contrairement aux études existantes (Mennis, 2009 ; Pajares et al., 2021), nous disposons des données individuelles FIDELI comme référence de vérité terrain. Cela permet pour la première fois de quantifier non seulement l'amplitude des erreurs mais aussi leur structure spatiale via Moran's I et LISA — une évaluation absente de la littérature antérieure sur la MAW-CZ.

◆ Hétérogénéité spatiale des erreurs : résultat contexte-dépendant

L'absence de biais spatial à Toulouse ($I = 0,01$) et le léger clustering à Grenoble ($I = 0,14$) montrent que la robustesse de la méthode dépend de la morphologie urbaine. Ce résultat nuance les conclusions de Schug et al. (2021) et Pajares et al. (2021), qui montrent que le volume bâti améliore les estimations mais reste insuffisant en tissu dense ou topographiquement contraint.

→ Portée et transférabilité

La MAW-CZ offre une alternative économique et open source, basée sur GeoClimate/OpenStreetMap, transférable à d'autres contextes nationaux disposant d'unités statistiques comparables à l'IRIS. Elle est particulièrement adaptée aux études de vulnérabilité urbaine à fine échelle lorsque les données 3D ou de machine learning ne sont pas disponibles.

Conclusions

Les résultats indiquent que cette méthode permet de transférer efficacement la population avec une précision relativement élevée pour les RSU de petite et moyenne taille.

La comparaison entre les résultats à fine échelle et à une maille plus large a révélé que la méthode est performante pour les RSU comptant peu d'habitants qu'elle classe correctement avec un pourcentage élevé.

La distribution spatiale des RSU biaisées s'est avérée homogène, ce qui suggère que les erreurs ne sont pas concentrées dans des régions spécifiques mais plutôt réparties uniformément sur l'ensemble des territoires.

Cette étude contribue aux efforts en cours pour intégrer les données sociales et environnementales à des résolutions spatiales plus fines, fournissant des informations précieuses pour les urbanistes, les décideurs politiques et les chercheurs dans leurs efforts pour relever les défis urbains et environnementaux complexes