

Novembre 2013

“FICHES MÉTHODOLOGIQUES
Méthodes statistiques d’allocation spatiale :
interpolation de données surfaciques”

Céline VIGNES, Sarah RIMBOURG, Anne RUIZ-GAZEN et Christine
THOMAS-AGNAN

GREMAQ - TSE - UT1C



FICHES MÉTHODOLOGIQUES

Méthodes statistiques d'allocation spatiale : interpolation de données surfaciques

Céline VIGNES, Sarah RIMBOURG, Anne RUIZ-GAZEN et Christine THOMAS-AGNAN
GREMAQ - TSE - UT1C

en collaboration avec la DREAL Midi-Pyrénées (commanditaire) :
Sorkorn MARIGOT et Stephan MASSE

19 novembre 2013

Table des matières

I	Introduction et présentation des données	1
1	Introduction	3
1.1	Notations	4
1.2	Agrégation ou désagrégation	4
1.3	Information auxiliaire	4
1.4	Nature extensive ou intensive d'une variable	6
1.5	Propriété pycnophylactique	7
1.6	Uniformité sur les sources et uniformité sur les cibles (variables intensives uniquement)	8
1.7	Homogénéité dans une zone source (variables extensives uniquement)	8
1.8	Choix de la méthode	9
1.9	Synthèse des méthodes dasymétriques	9
2	Présentation des données	11
2.1	Variable d'intérêt Y	11
2.2	Information auxiliaire	12
2.2.1	Couche des Routes (BD TOPO)	12
2.2.2	Couche Corine Land Cover	12
2.3	Cas de désagrégation : variable d'intérêt connue au niveau de la commune	14
2.4	Cas d'agrégation : variable d'intérêt connue au niveau de la parcelle	15
2.4.1	Liaison entre l'aire des parcelles et le nombre de logements	16
II	Méthodes dasymétriques	19
I	Allocation proportionnelle à l'aire - <i>Areal weighting interpolation</i> (DAW)	21
I.1	Formules d'allocation	21
I.1.1	Cas d'une variable extensive	22
I.1.2	Cas d'une variable intensive	24
I.2	Implications de la formule d'allocation	25
I.2.1	Cas de l'agrégation (variable extensive)	26
I.2.2	Cas de la désagrégation (variable extensive)	27
I.3	Exemples de calcul	28
I.3.1	Comment ventiler les logements d'une même parcelle dans plusieurs carreaux ? Détails du calcul pour la parcelle 31424000BN0062	28
I.3.2	Détails du calcul pour le carreau AJ10 (exemple 1)	29
II	Allocation proportionnelle à une variable auxiliaire - <i>Ordinary dasymetric weighting (with auxiliary information)</i> (DAX)	31
II.1	Formules d'allocation	32
II.1.1	Cas d'une variable extensive	32
II.1.2	Cas d'une variable intensive	34
II.2	Exemple de calcul : détails du calcul pour le carreau AB7	35
III	Méthode dasymétrique avec zones de contrôle - <i>Dasymetric weighting with control zones</i> (DAC)	37
III.1	Formule d'allocation (cas d'une variable extensive)	39
III.1.1	Etape préliminaire : Estimation de la densité de Y sur les zones de contrôle C_k	39
III.1.2	Etape 1 : Prédiction de Y sur les zones d'intersection A_{ij}	40
III.1.3	Etape 2 : Prédiction de Y sur les zones cibles T_j	41
III.1.4	Ecriture synthétique	41
III.2	Exemples de calcul	43

III.2.1	Densité de logements pour les zones artificialisées (CLC_1)	43
III.2.2	Détails du calcul pour le carreau AA11	43
IV	Méthode dasymétrique avec zones de contrôle et variable auxiliaire - <i>Two steps dasymetric weighting (with auxiliary information and control zones) (DA2)</i>	47
IV.1	Formule d'allocation (cas d'une variable extensive)	49
IV.1.1	Etape 1 : Prédiction de Y sur les zones d'intersection A_{ijk}	49
IV.1.2	Etape 2 : Prédiction de Y sur les zones cibles T_j	50
IV.2	Exemple de calcul : Détails du calcul pour le carreau S9	52
IV.2.1	Etape 1 : Prédiction de Y sur les zones d'intersection A_{ij} ($commune \cap S9$)	53
IV.2.2	Etape 2 : Prédiction de Y sur la zone cible T_j (le carreau S9)	54
IV.3	Lien entre les méthodes DA2 et DAC	54
III	Autre méthode	55
V	Allocation via le centroïde de la source - <i>Point-in-Polygon (PiP)</i>	57
V.1	Formule d'allocation (cas d'une variable extensive)	58
V.2	Exemple de calcul : Détails du calcul pour le carreau AJ10	59
	Bibliographie	60
	Annexes	63
A	Dictionnaire des données	67
A.1	Couche des communes	67
A.2	Couche des parcelles	68
A.3	Couche des routes	69
A.4	Couche Corine Land Cover	70
A.5	Tableau des logements	71
B	Réalisation du carroyage sous MapInfo	73
B.1	Zone d'étude	73
B.2	Ouverture du module et délimitation de la grille	74
B.3	Choix des paramètres du carroyage	75
B.4	Définition de la fenêtre optimale	76
B.5	Suppression des carreaux n'intersectant pas la zone d'étude	77
C	Calcul du nombre de logements au niveau des parcelles et au niveau des communes	81
C.1	Table <i>logement</i> : création de la variable <i>nlogements</i>	81
C.2	Table des parcelles	82
C.2.1	Calcul du nombre de logements par parcelle	82
C.2.2	Suppression des 2 437 parcelles sans coordonnées	83
C.3	Table des communes	84
D	Mise en œuvre de l'allocation proportionnelle à l'aire (DAW)	85
D.1	Préparation de la table : Calcul de l'aire des parcelles	85
D.2	Etape 1 : Prédiction de Y sur les zones d'intersection A_{ij}	86
D.3	Etape 2 : Prédiction de Y sur les zones cibles T_j	86
E	Mise en œuvre de l'allocation proportionnelle à une variable auxiliaire (DAX)	89
E.1	Préparation de la table	90
E.1.1	Sélection des bonnes routes	90
E.1.2	Suppression des routes extérieures aux communes	90
E.1.3	Assembler les routes	91
E.1.4	Calcul de la longueur de route	91
E.1.5	Ajout du nombre de logements par commune	97
E.2	Etape 1 : Prédiction de Y sur les zones d'intersection A_{ij}	97
E.3	Etape 2 : Prédiction de Y sur les zones cibles T_j	98

F	Estimation de la densité de logement sur les zones de contrôle	101
F.1	Préparation de la table	101
F.1.1	Création des zones Corine Land Cover au niveau 1	101
F.1.2	Création des intersections commune-zone Corine Land Cover	102
F.2	Calcul de la densité de logement par zone CLC	104
F.2.1	Calcul du nombre de logements au niveau des intersections communes-CLC	104
F.2.2	Calcul de la densité de logement par zone CLC	105
G	Mise en œuvre de l'allocation proportionnelle à une zone de contrôle (DAC)	107
G.1	Préparation de la table	107
G.1.1	Table des intersections communes-CLC	107
G.1.2	Table des intersections communes-carreaux-CLC	108
G.2	Etape 1 : Prédiction de Y sur les zones d'intersection A_{ij}	110
G.2.1	Etape 1a : Prédiction grossière de Y sur les zones d'intersection A_{ij}	110
G.2.2	Etape 1b : Prédiction grossière de Y sur les zones sources S_i	111
G.2.3	Etape 1c : Prédiction de Y sur les zones d'intersection A_{ij}	112
G.3	Etape 2 : Prédiction de Y sur les zones cibles T_j	113
H	Mise en œuvre de la méthode dasymétrique avec zone de contrôle et variable auxiliaire (DA2)	115
H.1	Préparation de la table	115
H.1.1	Table des intersections commune-CLC, <i>communes_inter_CLC.tab</i>	115
H.1.2	Calcul de la longueur de route	116
H.2	Etape 1 : Prédiction de Y sur les zones d'intersection A_{ijk}	120
H.2.1	Etape 1a : Prédiction de Y sur les intersections sources-zones de contrôle $A_{i,k}$ par DAX	120
H.2.2	Etape 1b : Prédiction de Y sur les zones d'intersection A_{ijk} par DAW	121
H.3	Etape 2 : Prédiction de Y sur les zones cibles T_j	122
I	Preuve du lien entre les méthodes DA2 et DAC	125
J	Mise en œuvre de la méthode Point-in-Polygon (PiP)	127
J.1	Préparation de la table	127
J.1.1	Calcul des coordonnées des centroïdes des parcelles	127
J.1.2	Création d'une table avec pour objets les centroïdes des parcelles	128
J.2	Prédiction de Y sur les zones cibles T_j	129

Première partie

Introduction et présentation des données

Chapitre 1

Introduction

L'analyse de données socio-économiques nécessite souvent de combiner des bases de données provenant de différentes sources administratives, données collectées sur des unités spatiales différentes, selon des découpages administratifs variés comme par exemple sur le territoire français les îlots, les Iris¹, les parcelles ou les communes. Il est donc souvent nécessaire de transformer les données provenant d'unités spatiales d'origine ("sources") en données associées à des unités spatiales "cibles". Le carroyage est un mode de partition spécifique de l'espace avec pour unité de base un carreau. Il s'agit alors de ventiler toutes les données initiales sur cette nouvelle partition cible unique définie par les carreaux. Le carroyage est actuellement en plein essor en France, notamment à l'Insee et dans d'autres instituts producteurs de données comme la Dreal, ainsi qu'en Europe avec la directive "INSPIRE" (INfrastructure for SPatial InfoRmation).

Ce rapport traite de l'interpolation de données surfaciques ("areal interpolation" en anglais), nom générique donné aux méthodes statistiques d'allocation spatiale. Il existe quatre principaux types de méthodes :

- la méthode des poids proportionnels ou allocation proportionnelle à l'aire ("areal weighting interpolation" en anglais) et les méthodes dérivées connues sous le nom de "dasymétriques",
- l'allocation via le centroïde de la source ("point-in-polygon" en anglais),
- l'interpolation basée sur des méthodes de régression,
- les techniques de lissage.

Ces méthodes se distinguent par le type de la variable d'intérêt (continue ou discrète, extensive ou intensive), la vérification de la propriété pycnophylactique, la présence d'information auxiliaire ou encore le recours à des hypothèses simplificatrices (cf. explications p.3 et suivantes).

Ce rapport fait suite à un travail de bibliographie extensive de l'ensemble des méthodes statistiques d'allocation spatiale mené par le GREMAQ (nous renvoyons le lecteur à cet article scientifique [2] pour une revue bibliographie plus détaillée des méthodes traitées dans ce rapport ainsi que pour la présentation des méthodes de régression qui ne sont pas traitées ici). A visée plus didactique que l'article pré-cité, ce rapport propose une revue de la littérature sur les méthodes dasymétriques et l'allocation via le centroïde de la source. Il se veut pédagogique, de ce fait les formules sont systématiquement expliquées dans le texte de sorte que le lecteur n'est pas obligé de les maîtriser pour comprendre les méthodes présentées.

Les notions générales sont développées dans cette première partie, qui présente également le jeu de données servant à illustrer chaque méthode. Les différentes méthodes sont ensuite présentées en détail sous forme de fiches : les méthodes dasymétriques dans une seconde partie et l'allocation via le centroïde de la source (point-in-polygon) dans une troisième partie. Les mises en œuvre des différentes méthodes dans le logiciel MapInfo (version 8.0) sont présentées en annexes sous forme de pas-à-pas très détaillés fournissant des extraits du code MapBasic et de nombreuses copies d'écran des manipulations "par les menus". Ces annexes sont disponibles sur demande².

1. Ilot regroupé pour l'information statistique

2. Contacter Christine THOMAS-AGNAN christine.thomas@tse-fr.eu

1.1 Notations

La variable d'intérêt Y est connue au niveau des zones sources $\{S_i\}_{i=1,\dots,I}$. On note y_{S_i} la valeur de Y sur la zone source S_i , c'est la valeur réelle, connue.

On cherche à prédire \hat{Y}_{T_j} la valeur de Y au niveau des zones cibles $\{T_j\}_{j=1,\dots,J}$. Les zones sources et les zones cibles ne sont pas nécessairement imbriquées et, en général, leurs limites ne coïncident pas.

La zone d'intersection entre la zone source S_i et la zone cible T_j est notée A_{ij} : $A_{ij} = S_i \cap T_j$. Il y a I_j intersections A_{ij} entre la zone cible T_j (j fixé) et les zones sources S_i ($i=1, \dots, I_j$), autrement dit T_j peut être considérée comme l'union des I_j intersections A_{ij} (j fixé ; $i=1, \dots, I_j$) : $T_j = \bigcup_{i=1}^{I_j} A_{ij}$.

La figure 1.1 (p. 5) montre un exemple de ces deux partitions d'une même région d'intérêt, d'une part en 3 zones sources et, d'autre part, en 6 zones cibles de forme quelconque. Un zoom sur les zones d'intersection A_{ij} est représenté pour la zone cible T_5 qui intersecte les trois zones sources.

La figure 1.2 (p. 5) illustre le cas particulier d'un carroyage, avec les 3 mêmes zones sources et 12 carreaux (zones cibles). Les zones d'intersection A_{ij} sont représentées pour le carreau T_5 qui intersecte les trois zones sources.

On note $|S_i|$ l'aire de la zone source S_i .

1.2 Agrégation ou désagrégation

Selon la taille respective des zones sources et des zones cibles, on se situera dans un cas d'agrégation ou de désagrégation de l'information.

Si les zones sources sont de taille inférieure aux zones cibles, et donc plus nombreuses, on se situe dans un cas d'agrégation : l'information provenant de plusieurs zones sources sera agrégée pour constituer l'information de la zones cible qui les contient.

A l'inverse, si les zones sources sont de taille supérieure aux zones cibles, et donc moins nombreuses (comme c'est le cas dans les figures 1.1 et 1.2), on se situe dans un cas de désagrégation : l'information provenant d'une zone source est à ventiler dans plusieurs zones cibles.

Il est important de noter que beaucoup de cas pratiques comprennent à la fois de l'agrégation et de la désagrégation selon la région considérée.

1.3 Information auxiliaire

En complément de la variable d'intérêt Y , on peut disposer d'une information auxiliaire sur laquelle se baser pour améliorer le calcul de Y sur les zones cibles.

L'information auxiliaire peut se présenter sous la forme d'une variable X quantitative. Pour que l'utilisation de cette variable auxiliaire X améliore le calcul de Y , elle doit être suffisamment corrélée³ à Y . Elle pourra être connue à des niveaux différents (sources, cibles, intersections) selon la méthode considérée.

L'information auxiliaire peut également être de nature catégorielle (comme par exemple un découpage rural/urbain ou selon la nature du sol). Dans ce cas, un découpage géographique selon les différents niveaux de cette variable catégorielle génère un troisième jeu de zones, dites zones de contrôle.

Les conditions sur les zones de contrôle différeront selon la méthode considérée (cf. méthodes DAC p. 37 et DA2 p. 47).

3. Attention, on ne fait pas référence à la corrélation usuelle en raison de l'autocorrélation spatiale présente dans les données.

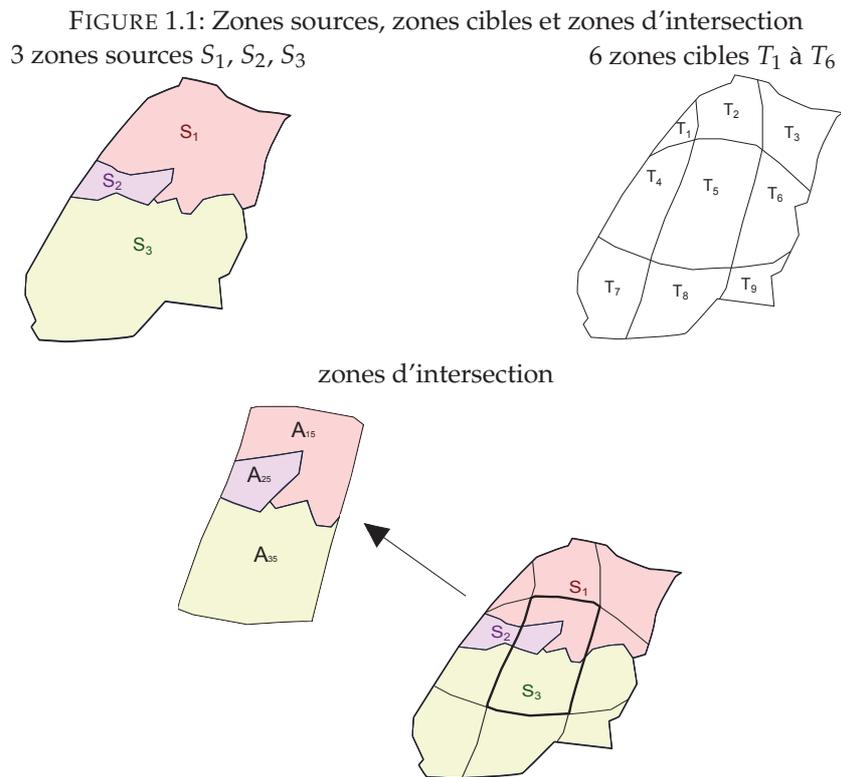
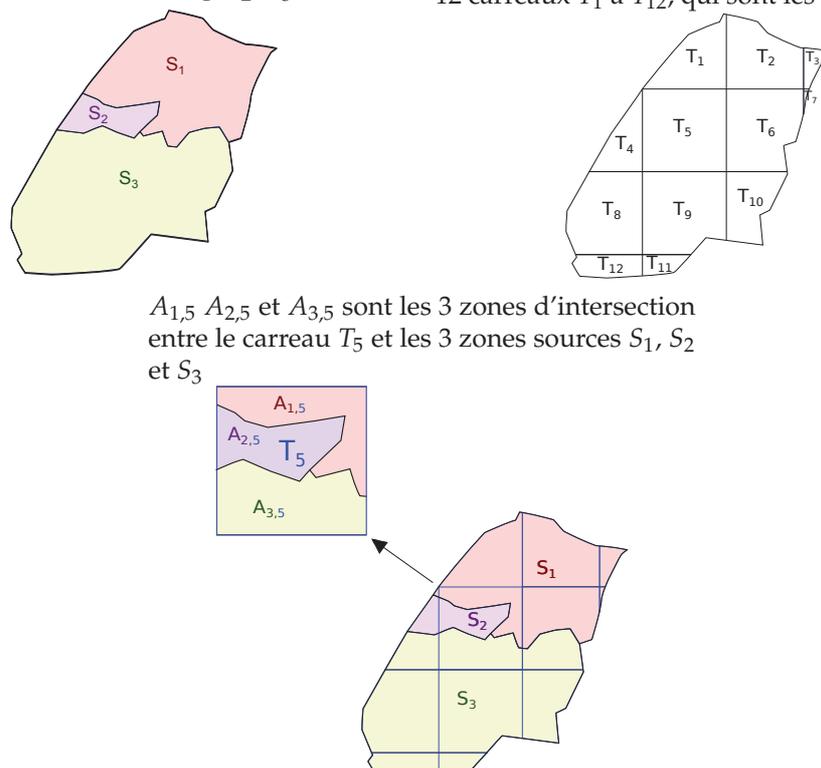


FIGURE 1.2: Zones sources, zones cibles et zones d'intersection dans le cas d'un carroyage

3 zones sources S_1, S_2, S_3 12 carreaux T_1 à T_{12} , qui sont les zones cibles



1.4 Nature extensive ou intensive d'une variable

La nature d'une variable est une des premières questions à se poser au moment de choisir une méthode, les formules étant différentes. Une variable peut être extensive ou intensive.

Afin de déterminer la nature d'une variable Y , on doit se demander comment calculer la valeur de Y sur l'ensemble d'une zone géographique. Prenons comme exemple la zone géographique de la figure 1.1 (p. 5), que nous appellerons S , qui est l'union de trois zones S_1, S_2 et S_3 : $S = S_1 \cup S_2 \cup S_3$.

Dans le cas d'une variable qui est un effectif (par exemple la population, le nombre de bâtiments, ...), la valeur de Y sur l'ensemble de la zone géographique S est la somme de la valeur de Y sur chaque sous-ensemble de S : $Y_S = Y_{S_1} + Y_{S_2} + Y_{S_3}$. On dit alors que la variable Y est **extensive**.

Dans le cas d'une variable qui est un ratio (par exemple la densité de population, le prix moyen du logement, ...), la valeur de Y sur l'ensemble de la zone géographique ne se calcule pas comme la somme de la valeur de Y sur chaque sous-ensemble de S mais comme une somme pondérée : $Y_S = w_{S_1} Y_{S_1} + w_{S_2} Y_{S_2} + w_{S_3} Y_{S_3}$ avec $w_{S_1}, w_{S_2}, w_{S_3}$ les poids associés respectivement à S_1, S_2 et S_3 . On dit que la variable Y est **intensive** si sa valeur sur l'ensemble d'une zone géographique S est la somme pondérée de la valeur de Y sur chaque sous-ensemble de S et qu'au moins un des poids n'est pas égal à 1 (pour ne pas se ramener au cas d'une variable extensive).

Si la variable Y représente la densité de population, le poids w_{S_i} sera la proportion de l'aire totale appartenant à la source S_i : $w_{S_i} = |S_i|/|S|$. Si Y est le prix moyen du logement, le poids w_{S_i} sera la proportion de logements appartenant à S_i : $w_{S_i} = n_i/n$. Le détail du calcul des poids est fourni dans l'encadré en infra (page suivante).

D'une manière plus formelle, la nature d'une variable se définit selon la forme que prend la relation entre la valeur de la variable sur un domaine donné Ω et sa valeur sur une partition⁴ $\{\Omega_k\}_k$ de Ω , on appelle cette relation "formule d'agrégation".

Une variable Y est **extensive** si

$$Y_\Omega = \sum_k Y_{\Omega_k}$$

c'est-à-dire si la valeur de Y sur l'ensemble du domaine (Y_Ω) est égale à la somme des valeurs de Y sur les sous-ensembles du domaine (les Y_{Ω_k}).

Une variable Y est **intensive** si

$$Y_\Omega = \sum_k w_{\Omega_k} Y_{\Omega_k} \quad \text{avec au moins un } w_{\Omega_k} \text{ différent de 1}$$

c'est-à-dire si la valeur de Y sur l'ensemble du domaine (Y_Ω) est égale à la somme pondérée des valeurs de Y sur les sous-ensembles du domaine (les Y_{Ω_k}). Les poids w_{Ω_k} doivent être indépendants de Y , la plupart du temps ils sont positifs et leur somme est égale à 1.

Une variable ne peut pas être simultanément extensive et intensive mais elle peut être ni l'une ni l'autre. Par exemple, une variable qui est une fonction non linéaire (comme un carré ou un logarithme) d'une variable extensive n'est ni extensive ni intensive. Considérons que le prix maximum des maisons dans un quartier est la variable d'intérêt. Pour chercher à connaître le prix maximum des maisons dans deux quartiers différents, on ne prendra ni la somme ni la somme pondérée des prix maximums des maisons dans ces deux quartiers mais la valeur maximale de ces deux prix maximums. Notons que les variables n'étant ni extensives ni intensives sont plutôt rares en sciences sociales mais elles peuvent se rencontrer en physique (domaine d'où proviennent ces notions). Notons également que le ratio de deux variables extensives est une variable intensive.

4. Une famille $\{\Omega_k\}_k$ de sous-ensembles de Ω forme une partition de Ω si : (i) Ω est l'union des Ω_k ; (ii) les Ω_k sont disjoints, i.e. ils n'ont pas de partie commune; (iii) Ω_k n'est pas vide.

Calcul des poids pour des variables intensives

1. Densité de population

On considère Y la densité de population, définie comme $Y_\Omega = Z_\Omega / |\Omega|$ avec Z_Ω le nombre d'habitants dans la zone Ω et $|\Omega|$ l'aire de la zone Ω .

Comme Z est extensive, on a $Z_\Omega = \sum_k Z_{\Omega_k}$.

En exprimant Z en fonction de Y , on obtient : $|\Omega|Y_\Omega = \sum_k |\Omega_k|Y_{\Omega_k}$.

En divisant par $|\Omega|$, on obtient : $Y_\Omega = \sum_k \frac{|\Omega_k|}{|\Omega|} Y_{\Omega_k}$.

Y est bien intensive avec $w_{\Omega_k} = \frac{|\Omega_k|}{|\Omega|}$.

2. Prix moyen du logement

On considère à présent que la variable Y représente le prix moyen du logement, définie comme $Y_{\Omega_k} = Z_{\Omega_k} / n_k$ avec Z_{Ω_k} la somme des prix des logements de la zone Ω_k ($Z_{\Omega_k} = \sum_{i=1}^{n_k} Z_i$) et n_k le nombre de logements de la zone Ω_k .

Comme Z est extensive, on a $Z_\Omega = \sum_k Z_{\Omega_k}$.

En exprimant Z en fonction de Y , on obtient : $nY_\Omega = \sum_k n_k Y_{\Omega_k}$.

En divisant par n , on obtient : $Y_\Omega = \sum_k \frac{n_k}{n} Y_{\Omega_k}$.

Y est bien intensive avec $w_{\Omega_k} = \frac{n_k}{n}$.

Il est important de remarquer qu'une variable extensive est toujours associée à une variable intensive, et réciproquement. Transformons une variable Y extensive en une variable \tilde{Y} intensive sous la forme $\tilde{Y}_A = Y_A / |A|$, comme pour une densité de population. Comme Y est extensive, on a $Y_\Omega = \sum_k Y_{\Omega_k}$.

$$\tilde{Y}_\Omega = \frac{Y_\Omega}{|\Omega|} = \frac{\sum_k Y_{\Omega_k}}{|\Omega|} = \sum_k \frac{Y_{\Omega_k}}{|\Omega|} = \sum_k \frac{|\Omega_k|}{|\Omega|} \times \frac{Y_{\Omega_k}}{|\Omega_k|} = \sum_k \frac{|\Omega_k|}{|\Omega|} \times \tilde{Y}_{\Omega_k} = \sum_k w_{\Omega_k} \times \tilde{Y}_{\Omega_k} \quad \text{avec} \quad w_{\Omega_k} = \frac{|\Omega_k|}{|\Omega|}$$

\tilde{Y} est donc bien intensive.

À l'inverse, on peut transformer une variable Y intensive dont les poids sont proportionnels à l'aire en une variable \tilde{Y} extensive sous la forme $\tilde{Y}_A = |A| \times Y_A$. Comme Y est intensive, on a $Y_\Omega = \sum_k w_{\Omega_k} \times Y_{\Omega_k}$ avec $w_{\Omega_k} = \frac{|\Omega_k|}{|\Omega|}$.

$$\tilde{Y}_\Omega = |\Omega| \times Y_\Omega = |\Omega| \times \sum_k \frac{|\Omega_k|}{|\Omega|} Y_{\Omega_k} = \sum_k |\Omega_k| Y_{\Omega_k} = \sum_k \tilde{Y}_{\Omega_k}$$

\tilde{Y} est donc bien extensive.

1.5 Propriété pycnophylactique

La propriété pycnophylactique est une propriété de préservation des masses, l'idée sous-entendue est que l'information sur les zones sources dont on dispose "en entrée" du calcul puisse être retrouvée intacte "en sortie". Plus formellement, la propriété pycnophylactique stipule que la valeur prédite de Y sur la zone source S_i (\hat{Y}_{S_i}) est égale à la valeur réelle de Y sur S_i (y_{S_i}) :

$$\hat{Y}_{S_i} = y_{S_i}$$

La zone source S_i peut être considérée comme l'union des zones d'intersection A_{ij} (i fixé ; $j=1, \dots, J_i$) : $S_i = \bigcup_{j=1}^{J_i} A_{ij}$.

La plupart des méthodes (sauf l'allocation via le centroïde de la source, ou PiP) permettent de prédire une valeur sur les intersections A_{ij} et donc de reconstituer une valeur prédite de Y sur les sources (soit \hat{Y}_{S_i}) :

$$\hat{Y}_{S_i} = \sum_{j : S_i \cap T_j \neq \emptyset} \hat{Y}_{A_{ij}}$$

Pour ces méthodes, la formule d'agrégation peut s'écrire de deux façons différentes selon la nature de Y .

Ainsi, si Y est extensive, la formule d'agrégation s'écrit : $y_{S_i} = \sum_{j : S_i \cap T_j \neq \emptyset} Y_{A_{ij}}$ et la propriété pycnophylactique est vérifiée si la valeur prédite de Y est telle que :

$$\sum_{j : S_i \cap T_j \neq \emptyset} \hat{Y}_{A_{ij}} = y_{S_i}$$

Si Y est intensive, la formule d'agrégation s'écrit : $y_{S_i} = \sum_{j : S_i \cap T_j \neq \emptyset} w_j Y_{A_{ij}}$ et la propriété pycnophylactique est vérifiée si la valeur prédite de Y est telle que :

$$\sum_{j : S_i \cap T_j \neq \emptyset} w_j \hat{Y}_{A_{ij}} = y_{S_i}$$

Pour les méthodes qui satisfont à la propriété pycnophylactique, on peut vérifier que le total des valeurs de Y sur les zones cibles ($\sum_j \hat{Y}_{T_j}$) est bien égal, aux erreurs d'arrondis près, au total des valeurs de Y sur les zones sources ($\sum_i y_{S_i}$).

1.6 Uniformité sur les sources et uniformité sur les cibles (variables intensives uniquement)

On dit qu'une variable Y intensive est :

- **uniforme sur les sources** si sa valeur sur les zones d'intersection A_{ij} (notée $Y_{A_{ij}}$) est égale à sa valeur sur les zones sources S_i (notée y_{S_i}) :

$$Y_{A_{ij}} = y_{S_i} \quad \text{pour tout } j \text{ tel que la cible } T_j \text{ intersecte la source } S_i$$

- **uniforme sur les cibles** si sa valeur sur les zones d'intersection A_{ij} (notée $Y_{A_{ij}}$) est égale à sa valeur sur les zones cibles T_j (notée y_{T_j}) :

$$Y_{A_{ij}} = y_{T_j} \quad \text{pour tout } i \text{ tel que la source } S_i \text{ intersecte la cible } T_j$$

Si la variable est uniforme respectivement sur les sources ou sur les cibles, il n'y a pas de différence entre les valeurs qu'elle prend sur les différentes zones d'intersection, sa valeur sera la même sur toutes les zones d'intersection et égale à sa valeur sur la source ou, respectivement, sur la cible. Par exemple, considérons une zone source S_i (i fixé) intersectée par deux carreaux T_j et $T_{j'}$ et une variable Y intensive. Y est uniforme sur les sources si $Y_{A_{ij}} = Y_{A_{ij'}} = y_{S_i}$.

1.7 Homogénéité dans une zone source (variables extensives uniquement)

On dit qu'une variable Y extensive est **homogène** sur les sources si sa valeur sur les zones d'intersection A_{ij} (notée $Y_{A_{ij}}$) rapportée à leur aire (notée $|A_{ij}|$) est égale à sa valeur sur les zones sources S_i (notée y_{S_i}) rapportée à leur aire (notée $|S_i|$) :

$$\frac{Y_{A_{ij}}}{|A_{ij}|} = \frac{y_{S_i}}{|S_i|} = \text{constante pour tout } j \text{ tel que la cible } T_j \text{ intersecte la source } S_i,$$

ce qui revient à dire que la valeur de Y sur une zone est proportionnelle à l'aire de cette zone.

Si la variable est répartie de manière homogène dans les zones sources, il n'y a pas de différence entre les ratios sur les différentes zones d'intersection, leur valeur sera la même sur toutes les zones d'intersection et égale à la valeur du ratio sur la source. Par exemple, considérons une zone source S_i (i fixé) intersectée par deux carreaux T_j et $T_{j'}$ et une variable Y extensive. Y est homogène sur les sources si $\frac{Y_{A_{ij}}}{|A_{ij}|} = \frac{Y_{A_{ij'}}}{|A_{ij'}|} = \frac{y_{S_i}}{|S_i|} = \text{constante}$.

On remarque que si une variable extensive est homogène sur les sources, la variable intensive correspondante est uniforme sur les sources, et réciproquement.

En pratique, on ne peut pas vérifier l'homogénéité car la valeur de Y sur les zones d'intersection ($Y_{A_{ij}}$) est inconnue. On s'intéresse donc à la corrélation entre la variable Y et l'aire des zones sources.

L'homogénéité implique la corrélation positive mais la réciproque n'est pas vraie. En vérifiant la corrélation, on n'est pas assuré d'avoir l'homogénéité, par contre si la corrélation est faible, cela implique qu'il n'y a pas d'homogénéité. Par exemple, si la variable d'intérêt sur une zone A s'exprime comme une fonction linéaire de l'aire de A ($Y_A = \alpha + \beta|A|$ avec α et β des réels non nuls et $\beta > 0$), il y a corrélation linéaire positive entre Y et son aire mais pas homogénéité (à cause du terme α qui est non-nul). Concrètement, le coût d'acheminement de l'eau dans une zone pourrait s'exprimer sous cette forme comme la somme d'un coût fixe (α) et d'un coût variable proportionnel à l'aire de la zone ($\beta|A|$) ; par exemple la taille des conduits qui dépend de l'aire de la zone).

1.8 Choix de la méthode

Le tableau 1.1 (p. 10) présente une synthèse des différentes méthodes détaillées dans les fiches de la partie suivante. Il est important de préciser que le type de la variable d'intérêt n'est pas déterminant pour le choix de la méthode car, comme cela a été expliqué au paragraphe 1.4 (p. 6) à toute variable extensive correspond une variable intensive et réciproquement [2].

1.9 Synthèse des méthodes dasymétriques

On dispose d'une variable d'intérêt Y extensive⁵ connue au niveau des zones sources $\{S_i\}_{i=1,\dots,I}$, dont on cherche à prédire la valeur au niveau des zones cibles $\{T_j\}_{j=1,\dots,J}$.

Les méthodes dasymétriques reposent sur un principe d'allocation proportionnelle des données aux zones d'intersection A_{ij} et se décomposent en deux étapes.

La première étape consiste à prédire la valeur de Y sur les zones d'intersection A_{ij} :

$$\hat{Y}_{A_{ij}} = w_{A_{ij}} \times y_{S_i}$$

avec $w_{A_{ij}}$ la proportion de la variable Y sur la zone source S_i qui est allouée à la zone d'intersection A_{ij} .

Cette pondération diffère selon le type de méthode choisie :

- $\frac{|A_{ij}|}{|S_i|}$ pour l'allocation proportionnelle à l'aire (DAW),
- $\frac{x_{A_{ij}}}{x_{S_i}}$ pour l'allocation proportionnelle à une variable auxiliaire X (DAX),
- $\frac{\sum_k |A_{ijk}| \hat{D}_{C_k}}{\sum_l \sum_m |A_{ilm}| \hat{D}_{C_m}}$ pour la méthode dasymétrique avec zones de contrôle C_k (DAC),
- $\sum_k \frac{|A_{ijk}|}{|A_{i.k}|} \times \frac{x_{A_{i.k}}}{x_{S_i}}$ pour la méthode dasymétrique avec zones de contrôle C_k et variable auxiliaire X (DA2).

La deuxième étape consiste à prédire la valeur de Y dans la zone cible T_j en sommant la valeur de Y sur les zones d'intersection A_{ij} pour toutes les sources S_i qui intersectent la cible T_j (car $T_j = \bigcup_{i=1}^I A_{ij}$) :

$$\hat{Y}_{T_j} = \sum_{i: S_i \cap T_j \neq \emptyset} \hat{Y}_{A_{ij}}$$

Remerciements

Ce travail a été partiellement financé par l'Agence Nationale de la Recherche à travers le projet ModULand (ANR-11-BSH1-005) et par la DREAL Midi-Pyrénées.

5. Si la variable d'intérêt est intensive, on utilisera la variable extensive qui lui correspond (cf. paragraphe 1.4 p. 6)

Méthode	Information auxiliaire		Propriété pycnophylactique	Conditions d'utilisation	Principe d'allocation aux zones d'intersection sources-cibles	Page
	Nature	Niveau spatial				
Allocation via le centroïde de la source (PiP)	aucune information auxiliaire		non satisfaite	concentration de la localisation du phénomène d'intérêt près des centroïdes des sources	allocation des données d'une source à une cible si la cible contient le centroïde de la source	57
Allocation proportionnelle à l'aire (DAW)	aucune information auxiliaire		satisfaite	variable d'intérêt homogène (cas extensif) ou uniforme (cas intensif) sur les sources	proportionnellement à leur aire	21
Allocation proportionnelle à une variable auxiliaire (DAX)	variable auxiliaire	connue au niveau des zones d'intersection sources-cibles	satisfaite	bonne corrélation positive entre variable auxiliaire et variable d'intérêt	proportionnellement à la variable auxiliaire	31
Méthode dasymétrique avec zones de contrôle (DAC)	zones de contrôle		satisfaite	densité de la variable d'intérêt connue ou estimée sur les zones de contrôle. Variable d'intérêt homogène (cas extensif) ou uniforme (cas intensif) sur les zones de contrôle	proportionnellement à une estimation "grossière" de la variable d'intérêt utilisant la densité sur les zones de contrôle	37
Méthode dasymétrique avec zones de contrôle et variable auxiliaire (DA2)	variable auxiliaire	connue au niveau des intersections sources-zones de contrôle	satisfaite	bonne corrélation positive entre variable auxiliaire et variable d'intérêt. Variable d'intérêt homogène (cas extensif) ou uniforme (cas intensif) sur les intersections sources-zones de contrôle	allocation aux intersections sources-zones de contrôle proportionnellement à une variable auxiliaire, puis allocation aux intersections sources-cibles-zones de contrôle proportionnellement à leur aire.	47

TABLE 1.1: Tableau récapitulatif

Chapitre 2

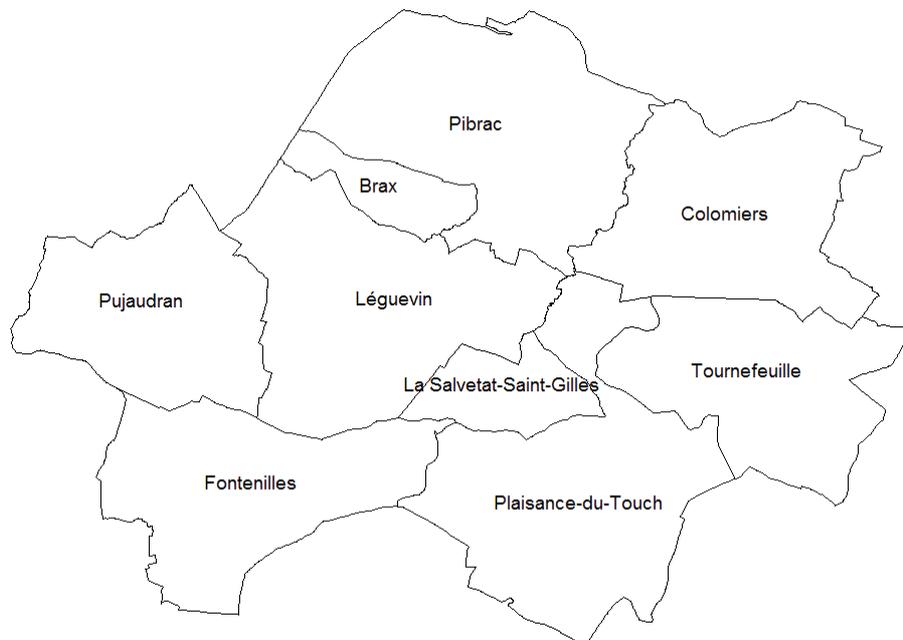
Présentation des données

Chaque méthode sera appliquée sur un même jeu de données de référence provenant de la DREAL Midi-Pyrénées. L'objectif est d'effectuer un carroyage de 500 m x 500 m (projection Lambert 93) et de calculer le nombre de logements au carreau (ou la densité de logements, selon la méthode considérée).

Ces prédictions carroyées du nombre de logements peuvent permettre par exemple d'estimer la population située dans des zones à risques (sismique, inondation) ou encore d'étudier les dynamiques d'urbanisation ou de consommation d'espace.

Pour des facilités d'illustration et de traitement, le territoire d'étude est restreint à 9 communes de l'ouest toulousain (cf. figure 2.1). Mais les méthodologies sont applicables à des territoires plus vastes.

FIGURE 2.1: Couche des 9 communes de l'ouest toulousain



2.1 Variable d'intérêt Y

Nous disposons donc de deux variables d'intérêt Y :

1. le **nombre de logements**, $Nb_logements$, qui est une variable *extensive*,
2. la **densité de logements**, $Densité_lgts$, qui est une variable *intensive* (cf. 1.4 p. 6).

Ces données sont issues des fichiers MAJIC (cf. annexe A p. 67) et sont connues au niveau parcellaire. Cependant, selon les méthodes à illustrer, nous serons amenés à prétendre que la variable d'intérêt est connue seulement au niveau de la commune.

La variable $Nb_logements$ a été construite en sommant les variables $nlocmaison$ (nombre de locaux de type maison individuelle) et $nlocappt$ (nombre de locaux de type appartement) pour n'inclure que les locaux étant réellement des

logements. Elle diffère donc de la variable *nlocal* qui est le nombre de locaux quel que soit leur type et ajoute aux logements les locaux de type commercial ou industriel (*nloccom*) et dépendance (*nlocdep*).

La variable *Densité_lgts* est construite à partir de la variable *Nb_logements* comme le ratio entre le nombre de logements d'une zone donnée et son aire :

$$Densité_lgts_{parc\grave{e}lle} = \frac{Nb_logements_{parc\grave{e}lle}}{Aire_{parc\grave{e}lle}}$$

2.2 Information auxiliaire

Différentes informations auxiliaires sont issues de deux couches MapInfo :

- la couche des routes (issue de la BD TOPO version 2 - décembre 2009), qui sert à construire la variable auxiliaire de longueur des routes (cf. annexe E p. 89),
- la couche Corine Land Cover 2006, qui sert à construire des zones de contrôles (cf. annexe G p. 107).

2.2.1 Couche des Routes (BD TOPO)

La BD TOPO est une base de données produite par l'Institut Géographique National (IGN), c'est la composante topographique du Référentiel géographique à Grande Echelle (RGE). Elle fournit une description vectorielle 3D (structurée en objets) des éléments du territoire et de ses infrastructures ainsi qu'une information d'adressage.

Les objets de la BD TOPO sont structurés en thèmes (par exemple : le réseau routier, le réseau hydrographique, les bâtiments, la structure administrative, etc.). Les informations dont nous disposons sont issues de la classe ROUTE (cf. figure 2.2) qui permet notamment de distinguer les différents types de route de la 2x2 voies au chemin ou sentier.

FIGURE 2.2: Superposition de la couche des communes et de la couche des routes



2.2.2 Couche Corine Land Cover

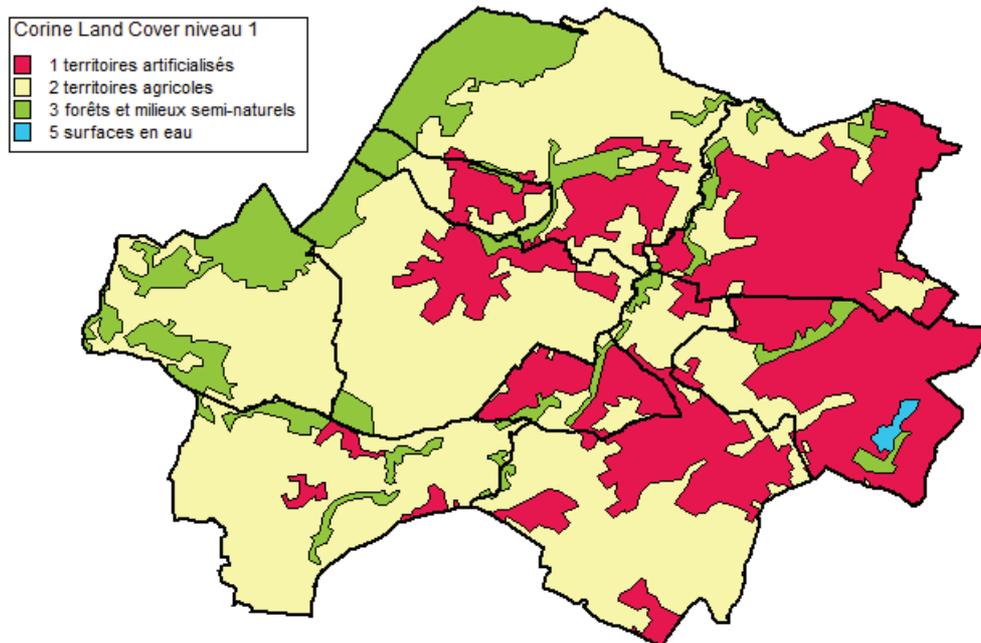
La base CORINE Land Cover est une base de données européenne d'occupation biophysique des sols, produite dans le cadre du programme européen CORINE de coordination de l'information sur l'environnement. Le producteur pour la France est le Service de l'observation et des statistiques du ministère chargé de l'environnement. Les données sont issues de l'interprétation visuelle d'images satellitaires, avec des données complémentaires d'appui (en particulier les BD ORTHO et CARTO de l'IGN). L'occupation biophysique du sol prévaut à son utilisation, ainsi la nature

des objets (forêts, cultures, surfaces en eau, roches affleurantes...) est privilégiée par rapport à leur fonction socio-économique (agriculture, habitat...).

La nomenclature CLC est hiérarchisée en 3 niveaux :

- niveau 1 (cf. figure III.1), qui sera utilisé dans nos exemples : niveau le plus grossier, comprenant 5 catégories (4 présentes dans nos données : 1 = territoires artificialisés, 2 = territoires agricoles, 3 = forêts et milieux semi-naturels, 4 = zones humides, 5 = surface en eau).
- niveau 2 : niveau intermédiaire comprenant 15 catégories (10 présentes dans nos données).
- niveau 3 : niveau le plus fin, il comprend 44 catégories (16 présentes dans nos données).

FIGURE 2.3: Superposition de la couche des communes et de la couche Corine Land Cover niveau 1

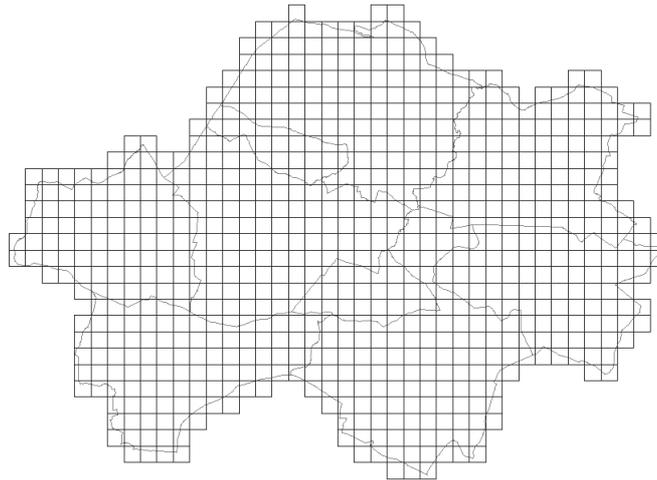


Notons que la superficie minimale des unités cartographiées est de 25 hectares.

Afin de reprendre les notations de l'introduction avec cet exemple, nous allons distinguer deux cas : celui de la désagrégation où la variable d'intérêt est connue au niveau de la commune, et celui de l'agrégation où elle est connue au niveau de la parcelle (cf. 1.2 p. 4).

2.3 Cas de désagrégation : variable d'intérêt connue au niveau de la commune

FIGURE 2.4: Superposition de la couche des communes et de la grille



On considère ici que la variable d'intérêt Y , le nombre ou la densité de logements, est connue au niveau communal. Les 9 communes de l'ouest toulousain constituent donc les 9 zones sources S_i ($i = 1, \dots, 9$). Comme on cherche à connaître le nombre (ou la densité) de logements au carreau, les zones cibles T_j ($j = 1, \dots, 1\ 160$) sont les 747 carreaux¹. Ces 747 carreaux intersectent au moins une commune (cf. figure 2.4), le tableau 2.1 donne le nombre de carreaux selon le nombre de communes qu'ils intersectent. On peut constater que la très grande majorité des carreaux intersecte une seule commune (i.e. ces carreaux sont entièrement inclus dans une commune). 131 carreaux sont à la frontière de deux communes et treize carreaux intersecte trois ou quatre communes. Au total, les données comptent 905 intersections commune-carreau² (notées $commune \cap carreau$).

TABLE 2.1: Nombre de carreaux selon le nombre de communes qu'ils intersectent

Nombre de communes intersectées	Nombre de carreaux	% de carreaux
aucune commune	413	35,6
une seule commune	603	52,0
deux communes	131	11,3
trois communes	12	1,0
quatre communes	1	0,1
Total	1 160	100,0

En faisant correspondre les notations précédentes aux données, on a donc :

- $y_{S_i} = Nb_logements_{commune}$ (ou $Densité_lgts_{commune}$), variable connue pour i (commune) = $1, \dots, 9$
- $Y_{T_j} = Nb_logements_{carreau}$ (ou $Densité_lgts_{carreau}$), à calculer pour j (carreau) = $1, \dots, 747$
- $A_{ij} = commune \cap carreau$ (avec 905 intersections non vides).

Le tableau 2.2 fournit les caractéristiques des neuf communes de la zone d'étude. Aux extrêmes, on distingue deux communes, Colomiers et Tournefeuille, très urbaines et densément peuplées et deux communes très rurales ayant une faible densité de population, Pujaudran et Fontenilles.

1. Table *grille* : 40 colonnes (numérotées de A à AN) et 29 lignes (numérotées de 1 à 29), la grille est donc composée de $40 \times 29 = 1\ 160$ carreaux, que ceux-ci intersectent ou non les communes

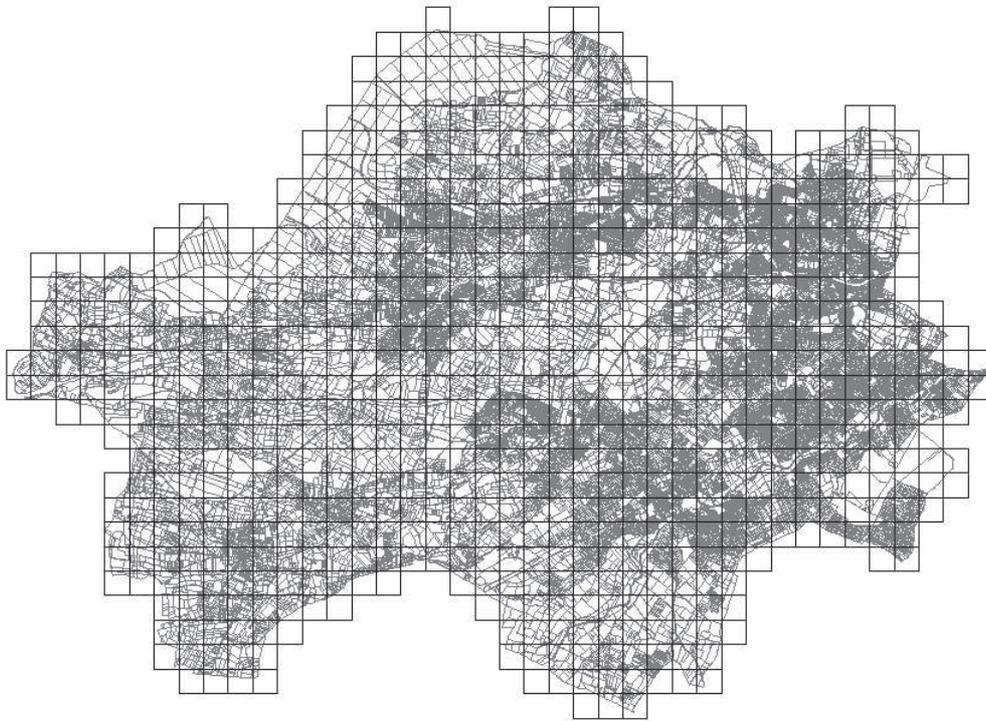
2. Table *communes_inter_carreaux*

TABLE 2.2: Caractéristiques des communes

Nom	Population	Aire (en m^2)	Nb_logements	Densité_lgts ($/km^2$)
Brax	2 373	4 466 614	891	199
Colomiers	32 110	20 995 365	14 273	680
Fontenilles	3 440	20 210 100	1 332	66
La Salvetat-Saint-Gilles	6 311	5 753 103	2 331	405
Léguevin	7 483	24 562 559	3 143	128
Pibrac	7 712	25 997 608	2 860	110
Plaisance-du-Touch	15 265	26 724 679	6 431	241
Pujaudran	1 217	17 659 003	498	28
Tournefeuille	25 444	18 186 755	10 668	587
Total	101 355	164 555 786	42 427	258

2.4 Cas d'agrégation : variable d'intérêt connue au niveau de la parcelle

FIGURE 2.5: Superposition de la couche des parcelles et de la grille



Ici, la variable d'intérêt Y , le nombre (ou la densité) de logements, est connue au niveau de la parcelle, les zones sources S_i sont donc les 50 709 parcelles (cf. figure 2.5).

Les parcelles ont une aire comprise entre 0 et $590\,716\,m^2$, avec une moyenne de $2\,915\,m^2$ et un écart-type de $9\,860\,m^2$. 25 % des parcelles ont une aire inférieure à $400\,m^2$, 50 % une aire inférieure à $762\,m^2$ et 75 % une aire inférieure à $1\,876\,m^2$. Les parcelles de petite taille³ ou de taille moyenne sont donc largement majoritaires sur la zone d'étude.

Comme on cherche à connaître le nombre de logements au carreau, les zones cibles T_j sont les 741 carreaux⁴. La taille des zones sources est inférieure à celle des carreaux, on se trouve donc bien dans un cas d'agrégation.

Il y a 63 269 intersections parcelle-carreau (notées $parcelle \cap carreau$) : 40 112 intersections correspondant exactement à la parcelle (parcelles totalement incluses dans un carreau, $parcelle \cap carreau = parcelle$ dans ce cas-là) et 23 157 intersections correspondant à des découpages de parcelles. Ces 23 157 intersections parcelle-carreau correspondent à 10 595 parcelles intersectées différentes. Les 50 709 parcelles se répartissent donc en 40 112 parcelles totalement incluses dans un carreau et 10 595 parcelles intersectées.

3. Notons que parmi les 4 073 parcelles de moins de $90\,m^2$, 230 parcelles contiennent entre 1 et 110 logements (192 parcelles avec un logement, 25 parcelles avec 2 à 10 logements et 13 parcelles avec plus de 10 logements, pour un total de 730 logements) alors qu'il est peu probable que des parcelles de si petite taille soient, dans la réalité, occupées par des logements.

4. Dans le cas de désagrégation, les communes intersectent 747 carreaux mais seulement 741 carreaux sont intersectés par les parcelles, l'explication concernant cette différence est donnée dans l'annexe B p. 73.

Les logements des parcelles entièrement incluses dans un carreau (33 328 sur 42 359) seront affectés au carreau sans erreur, le problème d'allocation provenant uniquement des découpages de parcelles sur plusieurs carreaux. Les 10 595 parcelles intersectées comptent au total 9 031 logements, une parcelle comporte entre 0 et 158 logements, avec une moyenne de 0,9 logement (écart-type=5,1). La plupart des parcelles intersectées ne comporte aucun logement (6 203 parcelles, soit 58,5%) ou un seul (4 048 parcelles, soit 32,8%).

En faisant correspondre les notations précédentes aux données, on a donc :

- $S_i = \text{parcelle}$, d'aire $|S_i|$. Il y a $I = 50\,709$ parcelles.
- $y_{S_i} = \text{Nb_logements}_{\text{parcelle}}$ connue pour i (parcelle) = $1, \dots, 50\,709$
- $Y_{T_j} = \text{Nb_logements}_{\text{carreau}}$ à calculer pour j (carreau) = $1, \dots, 741$
- $A_{ij} = \text{parcelle} \cap \text{carreau}$, d'aire $|A_{ij}|$. Il y a 63 269 intersections non vides.

Un carreau contient entre 1 et 455 parcelles, avec une moyenne de 85,4 parcelles et un écart-type de 88,3.

Parmi les 10 595 parcelles intersectées, une parcelle intersecte entre 2 et 8 carreaux (cf. tableau 2.3), avec une moyenne de 2,2 carreaux (écart-type=0,5). Près de 9 parcelles intersectées sur dix (9 261 parcelles, soit 87,41%) "chevauchent" deux carreaux, près d'une sur sept intersecte 3 ou 4 carreaux (1 390 parcelles, soit 13,12%) et seulement 14 parcelles intersectent 5 carreaux ou plus (soit 0,13%). Trois parcelles intersectent 7 ou 8 carreaux, ce sont des parcelles de très grande aire⁵.

TABLE 2.3: Nombre de parcelles selon le nombre de carreaux qu'elles intersectent (hors parcelles totalement incluses dans un carreau)

Nombre de carreaux intersectés	Nombre de parcelles	% de parcelles
2	9 261	87,41
3	725	6,84
4	595	5,62
5	9	0,08
6	2	0,02
7	1	0,01
8	2	0,02
Total	10 595	100,00

2.4.1 Liaison entre l'aire des parcelles et le nombre de logements

La liaison entre le nombre de logements et l'aire des parcelles ne semble pas très bonne, en particulier dans les zones fortement urbanisées où les parcelles qui contiennent des immeubles, et donc un nombre élevé de logements, sont assez petites (cf. figures 2.6 et 2.7 représentant le carreau AI11 à Colomiers qui contient plusieurs parcelles d'environ 450 m² avec plusieurs dizaines de logements qui sont des immeubles).

5. La parcelle intersectant 7 carreaux (identifiant = 31557000AZ0005) est située à Tournefeuille, elle comporte un seul logement, son aire est de 41,88 ha. Les deux parcelles intersectant 8 carreaux (identifiants = 323340000A0695 et 323340000A0696) sont situées à Pujaudran, elles ne comportent aucun logement, leurs aires respectives sont de 40,79 et 42,34 ha.

FIGURE 2.6: Exemple du carreau AI11 contenant des petites parcelles avec des immeubles



FIGURE 2.7: Photographie aérienne et parcelles cadastrales du centre du carreau AI11 (source : Géoportail)



Deuxième partie

Méthodes dasymétriques

Fiche I

Allocation proportionnelle à l'aire - *Areal weighting interpolation* (DAW)

Conditions d'utilisation

variable cible Y :

- extensive, avec une répartition homogène dans les zones sources S_i
- ou intensive, et uniforme sur les sources S_i

pas d'information auxiliaire

Notons que ces conditions d'uniformité ou d'homogénéité sur les zones sources sont drastiques et sont rarement remplies en pratique.

Principe [1, 5]

Allouer les données aux zones d'intersection A_{ij} proportionnellement à leur aire.

Propriété pycnophylactique satisfaite (cf. p. 7).

Ecriture synthétique

Variable extensive : $\hat{Y}_{T_j} = \sum_i : s_i \cap T_j \neq \emptyset \frac{|A_{ij}|}{|S_i|} \times y_{S_i}$. Variable intensive : $\hat{Y}_{T_j} = \sum_i : s_i \cap T_j \neq \emptyset \frac{|A_{ij}|}{|T_j|} \times y_{S_i}$.

Cette fiche s'articule en deux parties. La première partie explique de manière détaillée la formule d'allocation et sa mise en œuvre dans le cas d'une variable extensive (exemple : nombre de logements, connu au niveau des parcelles) puis dans celui d'une variable intensive (exemple : densité de logements, connue au niveau des parcelles). La seconde partie s'intéresse aux implications de la formule d'allocation tout d'abord dans le cas de l'agrégation (source = parcelle, cible = carreau) puis dans le cas de la désagrégation (source = commune, cible = carreau). A des fins de simplicité, la seconde partie n'est déclinée que dans le cas d'une variable extensive : le nombre de logements.

Notons que l'allocation proportionnelle est la seule méthode d'interpolation de données surfaciques proposée dans MapInfo.

I.1 Formules d'allocation

Exemples

La méthode d'allocation proportionnelle à l'aire sera illustrée par deux exemples, selon la nature de la variable d'intérêt Y :

- exemple 1 : Y ="nombre de logements" (noté $Nb_logements_{carreau}$), cette variable est *extensive*,

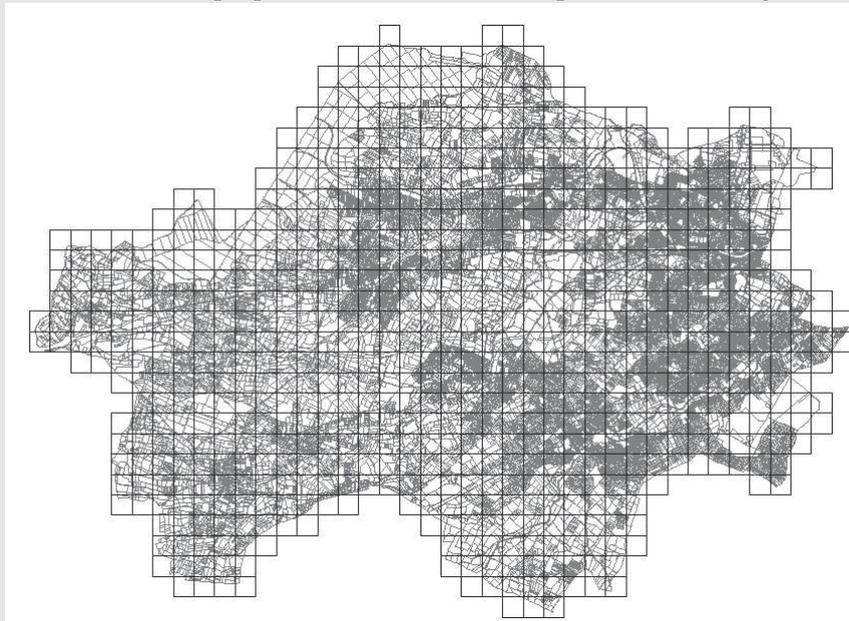
– exemple 2 : Y ="densité de logements" (notée $Densité_lgts_{carreau}$), cette variable est *intensive*.

Dans ces deux exemples, nous chercherons à prédire la variable Y au niveau du carreau à partir de l'information connue au niveau de la parcelle (respectivement notée $Nb_logements_{parcelle}$ et $Densité_lgts_{parcelle}$). Les zones sources sont donc les 50 709 parcelles (cf. figure I.1) et les zones cibles les 741 carreaux. La taille des zones sources étant inférieure à celle des carreaux, on se trouve dans un cas d'agrégation (pour plus d'information sur les données cf. 2.4 p. 15).

Application des notations aux données étudiées

- $S_i =$ parcelle, d'aire $|S_i|$. Il y a $I = 50\,709$ parcelles.
- $y_{S_i} = Nb_logements_{parcelle}$ ou $Densité_lgts_{parcelle}$, connue pour i (parcelle) = $1, \dots, 50\,709$
- $Y_{T_j} = Nb_logements_{carreau}$ ou $Densité_lgts_{carreau}$, à calculer pour j (carreau) = $1, \dots, 741$
- $A_{ij} =$ parcelle \cap carreau, d'aire $|A_{ij}|$. Il y a 63 269 intersections non vides (dont 40 112 qui correspondent exactement à la parcelle, i.e. parcelles entièrement incluses dans un carreau).

FIGURE I.1: Superposition de la couche des parcelles et de la grille



I.1.1 Cas d'une variable extensive

1ère étape : Prédiction de Y sur les zones d'intersection A_{ij}

Dans le cas d'une variable extensive, l'hypothèse supplémentaire à émettre est l'homogénéité de la variable Y au sein des zones sources (i.e. elle se répartit de manière homogène dans l'ensemble d'une même zone source). Cette hypothèse permet de supposer que les données sources peuvent être allouées aux zones d'intersection proportionnellement à leur aire.

Ainsi, la proportion de la valeur de Y sur la zone source S_i (i.e. la proportion de y_{S_i}) qui est allouée à la zone d'intersection A_{ij} est déterminée proportionnellement à $|A_{ij}|$, l'aire de la zone d'intersection. Le coefficient de pondération appliqué à y_{S_i} est donc défini comme le ratio de l'aire de la zone d'intersection A_{ij} ($|A_{ij}|$) et de celle de la zone source S_i ($|S_i|$).

La valeur prédite de Y pour la zone d'intersection A_{ij} , notée $\hat{Y}_{A_{ij}}$, s'écrit donc :

$$\hat{Y}_{A_{ij}} = \frac{|A_{ij}|}{|S_i|} \times y_{S_i}$$

Exemple 1

La première étape consiste à calculer le nombre de logements au niveau des intersections parcelles-carreaux ($\hat{Y}_{A_{ij}}$).

Pour cela, on calcule tout d'abord l'aire des intersections parcelles-carreaux ($|A_{ij}|$). Puis on calcule le nombre de logements au niveau des intersections parcelles-carreaux ($\hat{Y}_{A_{ij}}$) à partir de cette aire, de celle des parcelles ($|S_i|$) et du nombre de logements de la parcelle (y_{S_i}).

Ainsi avec les variables de l'exemple, la formule précédente s'écrit :

$$Nb_logements_{parcelle \cap carreau} = \frac{Aire_{parcelle \cap carreau}}{Aire_{parcelle}} \times Nb_logements_{parcelle}$$

2e étape : Prédiction de Y sur les zones cibles T_j

Rappelons qu'une zone cible T_j (j fixé) est l'union de toutes les zones d'intersection A_{ij} de cette zone cible avec les sources S_i : $T_j = \bigcup_i A_{ij}$.

En conséquence de la propriété d'extensivité, la valeur de Y dans la zone cible T_j (\hat{Y}_{T_j}) s'obtient en sommant la valeur de Y sur les zones d'intersection A_{ij} ($\hat{Y}_{A_{ij}}$) pour toutes les sources S_i qui intersectent la cible T_j , c'est-à-dire en sommant sur i les $\hat{Y}_{A_{ij}}$:

$$\hat{Y}_{T_j} = \sum_{i : S_i \cap T_j \neq \emptyset} \hat{Y}_{A_{ij}}$$

Exemple 1

Cette deuxième étape consiste à calculer le nombre de logements au niveau des carreaux (\hat{Y}_{T_j}), à partir du nombre de logements au niveau des intersections parcelles-carreaux ($\hat{Y}_{A_{ij}}$) pour les parcelles S_i qui intersectent chaque carreau T_j .

Ainsi avec les variables de l'exemple, la formule précédente s'écrit :

$$Nb_logements_{carreau} = \sum_{parcelles} Nb_logements_{parcelle \cap carreau}$$

Ecriture synthétique

$$\hat{Y}_{T_j} = \sum_{i : S_i \cap T_j \neq \emptyset} \frac{|A_{ij}|}{|S_i|} \times y_{S_i}$$

Exemple 1

Avec les variables de l'exemple, la formule synthétique d'allocation s'écrit :

$$Nb_logements_{carreau} = \sum_{parcelles} \frac{Aire_{parcelle \cap carreau}}{Aire_{parcelle}} \times Nb_logements_{parcelle}$$

Les figures I.2 et I.3 donnent le résultat du carroyage avec la méthode DAW.

FIGURE I.2: Extrait du carroyage DAW (agrégation) : Nombre de logements calculé au carreau

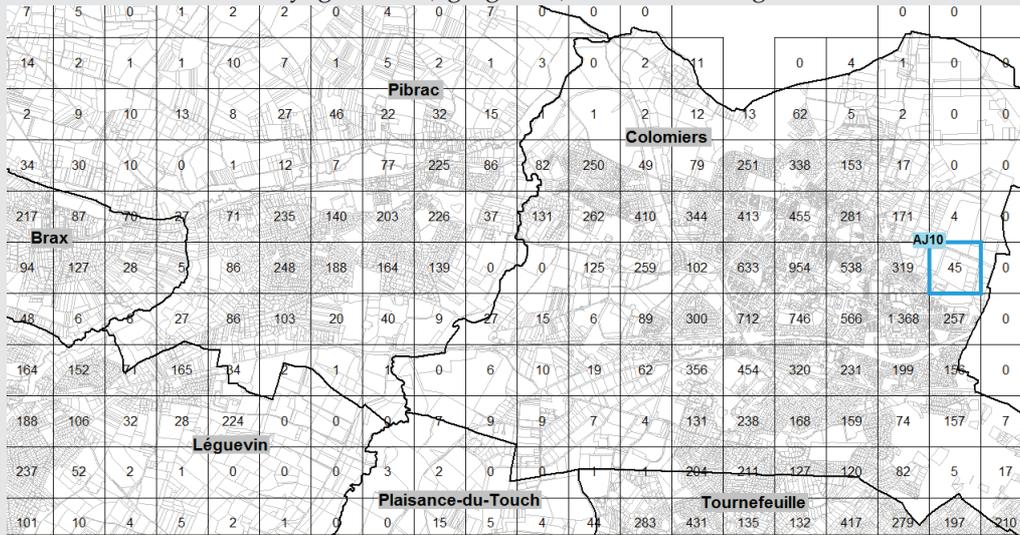
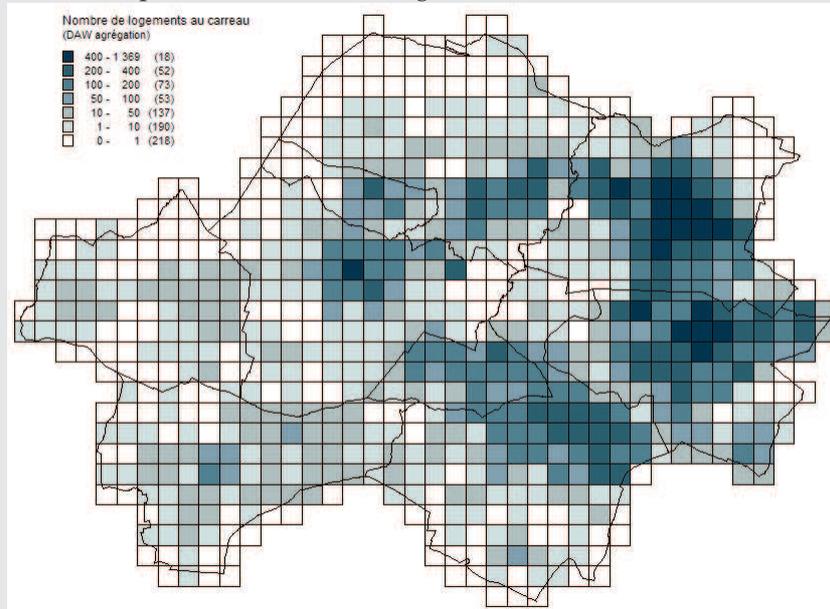


FIGURE I.3: Carte choroplèthe du nombre de logements au carreau, méthode DAW (agrégation)



I.1.2 Cas d'une variable intensive

1ère étape : Prédiction de Y sur les zones d'intersection A_{ij}

Dans le cas d'une variable intensive, l'hypothèse supplémentaire à émettre est que cette variable est uniforme sur les sources. C'est-à-dire que sa valeur sur la zone d'intersection A_{ij} est égale à sa valeur sur la zone source S_i : $y_{A_{ij}} = y_{S_i}$.

Cette hypothèse permet de calculer la valeur prédite de Y pour la zone d'intersection A_{ij} , notée $\hat{Y}_{A_{ij}}$:

$$\hat{Y}_{A_{ij}} = y_{S_i}$$

Exemple 2 Dans cet exemple, la variable d'intérêt Y doit être intensive, nous avons choisi la densité de logement qui est le ratio entre le nombre de logements d'une zone donnée et son aire :

$$y_{S_i} = \text{Densité_lgts}_{parc\grave{e}lle} = \frac{Nb_logements_{parc\grave{e}lle}}{Aire_{parc\grave{e}lle}}$$

La première étape consiste à calculer la densité de logements au niveau des intersections parcelles-carreaux ($\hat{Y}_{A_{ij}}$). Comme nous venons de le voir, elle est égale à la densité de logements au niveau des parcelles (y_{S_i}).

Ainsi avec les variables de l'exemple, la formule précédente s'écrit :

$$Densité_lgts_{parcelle \cap carreau} = Densité_lgts_{parcelle}$$

2ème étape : Prédiction de Y sur les zones cibles T_j

Rappelons qu'une zone cible T_j (j fixé) est l'union de toutes les zones d'intersection A_{ij} de cette zone cible avec les sources S_i : $T_j = \cup_i A_{ij}$.

En conséquence de la propriété d'intensivité, la valeur de Y dans la zone cible T_j (\hat{Y}_{T_j}) s'obtient par une somme pondérée de la valeur de Y sur les zones d'intersection A_{ij} ($\hat{Y}_{A_{ij}}$) pour toutes les sources S_i qui intersectent la cible T_j :

$$\hat{Y}_{T_j} = \sum_{i : S_i \cap T_j \neq \emptyset} \frac{|A_{ij}|}{|T_j|} \times \hat{Y}_{A_{ij}}$$

Le coefficient de pondération est déterminé suivant le principe de l'allocation proportionnelle à l'aire comme le ratio de l'aire de la zone d'intersection A_{ij} et de celle de la zone cible T_j .

Exemple 2

Cette deuxième étape consiste à calculer la densité de logements au niveau des carreaux (\hat{Y}_{T_j}), à partir de la densité de logements au niveau des intersections parcelles-carreaux ($\hat{Y}_{A_{ij}}$) pour les parcelles S_i qui intersectent chaque carreau T_j .

Ainsi avec les variables de l'exemple, la formule précédente s'écrit :

$$Densité_lgts_{carreau} = \sum_{parcelles} \frac{Aire_{parcelle \cap carreau}}{Aire_{carreau}} \times Densité_lgts_{parcelle \cap carreau}$$

Ecriture synthétique

$$\hat{Y}_{T_j} = \sum_{i : S_i \cap T_j \neq \emptyset} \frac{|A_{ij}|}{|T_j|} \times y_{S_i}$$

Exemple 2

Avec les variables de l'exemple, la formule synthétique d'allocation s'écrit :

$$Densité_lgts_{carreau} = \sum_{parcelles} \frac{Aire_{parcelle \cap carreau}}{Aire_{carreau}} \times Densité_lgts_{parcelle}$$

I.2 Implications de la formule d'allocation

Dans cette partie, seront détaillées les implications de la formule d'allocation tout d'abord dans le cas de l'agrégation puis dans le cas de la désagrégation. Pour ne pas alourdir la lecture, seul le cas d'une variable extensive sera exposé mais toute cette partie pourrait être déclinée pour une variable intensive. Dans les exemples choisis, la variable d'intérêt (Y) est le nombre de logements. Dans le cas de l'agrégation, on considèrera l'information connue au niveau des parcelles (= sources) et dans le cas de la désagrégation, on considèrera l'information connue au niveau des communes (= sources). Les zones cibles seront dans les deux cas les carreaux.

I.2.1 Cas de l'agrégation (variable extensive)

Dans le cas de l'agrégation, les zones sources sont de taille inférieure aux zones cibles. L'ensemble des zones sources peut donc se décomposer en deux ensembles : celui des zones sources totalement incluses dans une zone cible T_j (noté I_0^j) et celui des zones sources intersectant des zones cibles (noté I_1 , $I = I_0^j \cup I_1$).

1ère étape : Prédiction de Y sur les zones d'intersection A_{ij}

Pour les zones sources appartenant à I_0^j , c'est-à-dire complètement incluses dans une zone cible ($S_i \subset T_j$), la zone d'intersection est égale à la zone source : $A_{ij} = S_i$. Dans ces cas-là, la valeur prédite de Y pour la zone d'intersection A_{ij} est égale à sa valeur sur la zone source S_i : $\hat{Y}_{A_{ij}} = y_{S_i}$.

Pour les zones sources appartenant à I_1 , c'est-à-dire intersectant plusieurs zones cibles, l'affectation aux zones d'intersection se fait proportionnellement à leur aire, comme dans le cas général (cf. I.1.1).

La valeur prédite de Y pour la zone d'intersection A_{ij} s'écrit donc :

$$\hat{Y}_{A_{ij}} = \begin{cases} y_{S_i} & \text{si } i \in I_0^j \text{ (i.e. } S_i \subset T_j) \\ \frac{|A_{ij}|}{|S_i|} \times y_{S_i} & \text{si } i \in I_1 \end{cases}$$

2e étape : Prédiction de Y sur les zones cibles T_j

La valeur de Y dans la zone cible T_j (\hat{Y}_{T_j}) s'obtient en sommant la valeur prédite de Y sur les zones d'intersection pour toutes les sources S_i qui intersectent la cible T_j , qu'elles soient totalement incluses dans T_j (i.e. $i \in I_0^j$) ou qu'elles intersectent plusieurs zones cibles dont T_j (i.e. $i \in I_1$) :

$$\hat{Y}_{T_j} = \sum_{i \in I_0^j} y_{S_i} + \sum_{i \in I_1} \hat{Y}_{A_{ij}}$$

Ecriture synthétique

$$\hat{Y}_{T_j} = \sum_{i \in I_0^j} y_{S_i} + \sum_{i \in I_1} \frac{|A_{ij}|}{|T_j|} \times y_{S_i}$$

La seule source d'erreur dans la prédiction \hat{Y}_{T_j} provient du second terme, c'est-à-dire des zones sources qui intersectent plusieurs zones cibles. En effet, pour les zones sources "entières" on ne fait qu'affecter la valeur de Y de la zone source dans la zone cible.

Exemple 1 Avec les variables de l'exemple, la formule synthétique d'allocation s'écrit :

$$Nb_logements_{carreau} = \sum_{\substack{\text{parcelles} \\ \text{entières}}} Nb_logements_{parcelle} + \sum_{\substack{\text{parcelles} \\ \text{intersectées}}} \frac{Aire_{parcelle \cap carreau}}{Aire_{parcelle}} \times Nb_logements_{parcelle}$$

L'ensemble I_0^j est dans cet exemple composé des 40 112 parcelles totalement incluses dans un carreau (parcelles "entières") et l'ensemble I_1 est composé des 23 157 intersections parcelle-carreau qui sont des découpages de parcelles entre plusieurs carreaux (ces intersections correspondent à 10 595 parcelles différentes, dont 6 203 qui ne comportent aucun logement). Pour les parcelles "entières", il n'y a aucune source d'erreur car on affecte simplement au carreau le nombre de logements de la parcelle. Ces parcelles "entières" totalisent 33 328 logements. Pour les parcelles "intersectées" (qui totalisent 9 031 logements), l'hypothèse de répartition uniforme des logements sur la parcelle est nécessaire pour une allocation proportionnelle à l'aire.

I.2.2 Cas de la désagrégation (variable extensive)

Dans le cas de la désagrégation, les zones sources sont de taille supérieure aux zones cibles. L'ensemble des zones cibles peut donc se décomposer en deux ensembles : celui des zones cibles totalement incluses dans une zone source (noté J_0^i) et celui des zones cibles intersectant plusieurs zones sources (noté J_1 , $J = J_0^i \cup J_1$).

Exemple 3 Les exemples 1 et 2 correspondant à des cas d'agrégation, on introduit un nouvel exemple en considérant que la variable d'intérêt Y est connue au niveau de la commune. Les zones sources sont donc les 9 communes et les zones cibles les 747 carreaux. La taille des zones sources étant supérieure à celle des carreaux, on se trouve bien dans un cas de désagrégation (pour plus d'information sur les données cf. 2.3 p. 14).

Application des notations aux données étudiées

- $S_i = \text{commune}$, d'aire $|S_i|$. Il y a $I = 9$ communes.
- $y_{S_i} = \text{Nb_logements}_{\text{commune}}$, connue pour i (commune) = 1, ..., 9
- $Y_{T_j} = \text{Nb_logements}_{\text{carreau}}$, à calculer pour j (carreau) = 1, ..., 747
- $A_{ij} = \text{commune} \cap \text{carreau}$, d'aire $|A_{ij}|$. Il y a 63 269 intersections non vides.

1ère étape : Prédiction de Y sur les zones d'intersection A_{ij}

L'affectation des données aux zones d'intersection se fait proportionnellement à leur aire, comme dans le cas général (cf. I.1.1). Cependant, les zones cibles appartenant à J_0^i , c'est-à-dire complètement incluses dans une zone source S_i ($T_j \subset S_i$), forment un cas particulier où la zone d'intersection est égale à la zone cible : $A_{ij} = T_j$.

La valeur prédite de Y pour la zone d'intersection A_{ij} s'écrit donc :

$$\hat{Y}_{A_{ij}} = \begin{cases} \frac{|T_j|}{|S_i|} \times y_{S_i} & \text{si } j \in J_0^i \text{ (i.e. } T_j \subset S_i) \\ \frac{|A_{ij}|}{|S_i|} \times y_{S_i} & \text{si } j \in J_1 \end{cases}$$

2e étape : Prédiction de Y sur les zones cibles T_j

Pour les zones cibles incluses dans une zone source (i.e. $j \in J_0^i$), la valeur de Y dans la zone cible T_j a été déjà été calculée à l'étape précédente.

Pour les zones cibles intersectant plusieurs zones sources (i.e. $j \in J_1$), la valeur de Y dans la zone cible T_j s'obtient en sommant la valeur prédite de Y sur les zones d'intersection pour toutes les sources S_i qui intersectent la cible T_j :

$$\hat{Y}_{T_j} = \sum_i \hat{Y}_{A_{ij}} \quad \text{si } j \in J_1$$

Ecriture synthétique

$$\hat{Y}_{T_j} = \begin{cases} \frac{|T_j|}{|S_i|} \times y_{S_i} & \text{si } j \in J_0^i \text{ (i.e. } T_j \subset S_i) \\ \sum_i \frac{|A_{ij}|}{|S_i|} \times y_{S_i} & \text{si } j \in J_1 \end{cases}$$

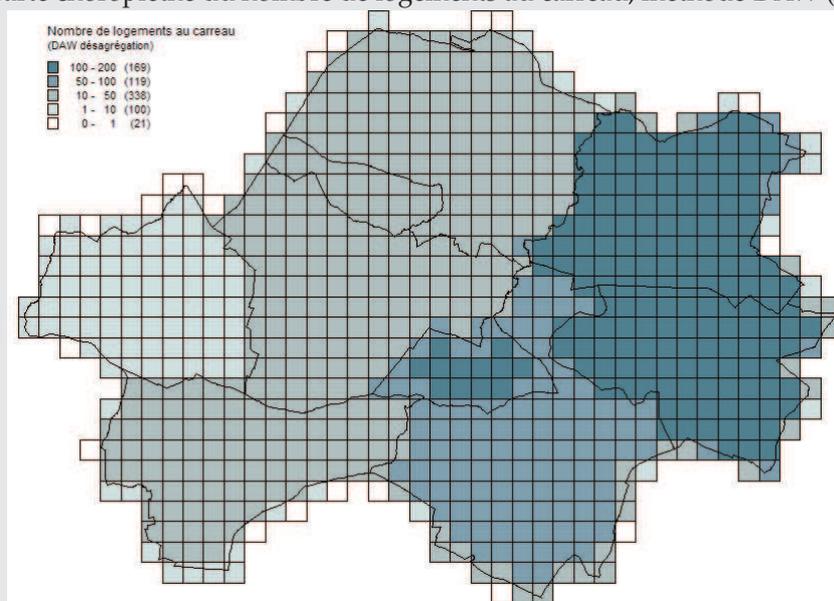
Exemple 3

Avec les variables de l'exemple, la formule synthétique d'allocation s'écrit :

$$\text{Nb_logements}_{\text{carreau}} = \begin{cases} \frac{\text{Aire}_{\text{carreau}}}{\text{Aire}_{\text{commune}}} \times \text{Nb_logements}_{\text{commune}} & \text{si le carreau est inclus dans une commune} \\ \sum_{\text{communes}} \frac{\text{Aire}_{\text{carreau} \cap \text{commune}}}{\text{Aire}_{\text{commune}}} \times \text{Nb_logements}_{\text{commune}} & \text{si carreau intersecte plusieurs communes} \end{cases}$$

Notons que tous les carreaux totalement inclus dans une même commune ont une valeur constante (car leur aire est constante). La figure I.4 donne le résultat du carroyage avec la méthode DAW dans le cas de la désagrégation.

FIGURE I.4: Carte choroplèthe du nombre de logements au carreau, méthode DAW (désagrégation)



I.3 Exemples de calcul

I.3.1 Comment ventiler les logements d'une même parcelle dans plusieurs carreaux ? Détails du calcul pour la parcelle 31424000BN0062

La parcelle 31424000BN0062¹ se situe à Plaisance-du-Touch. Elle intersecte 4 carreaux : AC19, AD19, AC20 et AD20 (cf. figure I.5). Son aire est de 9 219,33 m², elle contient 63 logements. Le tableau I.1 fournit les données nécessaires au calcul du nombre de logements pour les zones d'intersection entre la parcelle BN0062 et les quatre carreaux.

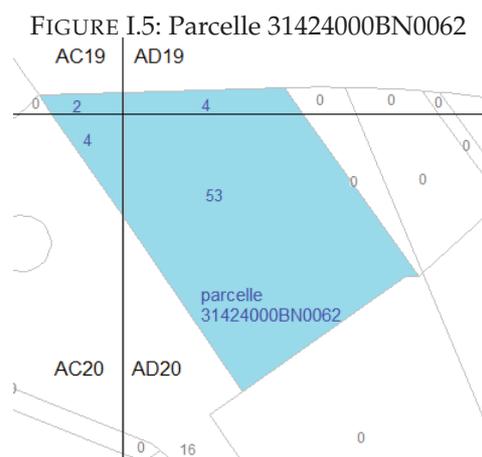


TABLE I.1: Exemple DAW : données nécessaires au calcul pour la parcelle BN0062

Zone cible T_j	$ A_{BN0062 \cap j} $
carreau	$Aire_{BN0062 \cap \text{carreau}}$
AC19	240,769
AD19	615,298
AC20	564,045
AD20	7799,56

1. Pour améliorer la lisibilité de l'exemple, l'identifiant de cette parcelle - 31424000BN0062 - est raccourci en BN0062

Prédiction de Y sur les zones d'intersection A_{ij} ($BN0062 \cap \text{carreau}$)

$$Nb_logements_{BN0062 \cap AC19} = \frac{Aire_{BN0062 \cap AC19}}{Aire_{BN0062}} \times Nb_logements_{BN0062} = \frac{240,769}{9\,219,33} \times 63 = 1,645$$

$$Nb_logements_{BN0062 \cap AD19} = \frac{Aire_{BN0062 \cap AD19}}{Aire_{BN0062}} \times Nb_logements_{BN0062} = \frac{615,298}{9\,219,33} \times 63 = 4,205$$

$$Nb_logements_{BN0062 \cap AC20} = \frac{Aire_{BN0062 \cap AC20}}{Aire_{BN0062}} \times Nb_logements_{BN0062} = \frac{564,045}{9\,219,33} \times 63 = 3,854$$

$$Nb_logements_{BN0062 \cap AD20} = \frac{Aire_{BN0062 \cap AD20}}{Aire_{BN0062}} \times Nb_logements_{BN0062} = \frac{7\,799,56}{9\,219,33} \times 63 = 53,298$$

I.3.2 Détails du calcul pour le carreau AJ10 (exemple 1)

On se place dans le cadre de l'exemple 1, où la variable d'intérêt est le nombre de logements, les zones sources sont les parcelles et les zones cibles, les carreaux.

Le carreau AJ10 se situe à Colomiers. Il intersecte 6 parcelles² dont le nombre de logements est supérieur à 0 (cf. figure I.6) : deux qui sont entièrement contenues dans ce carreau (BI0033 et BI035, en clair sur la figure), deux parcelles qui intersectent également le carreau AI10 (BI0218 et BI0129) et deux parcelles qui intersectent également le carreau AJ11 (BH0002 et BH0003). Le tableau I.2 fournit les données nécessaires au calcul du nombre de logements pour le carreau AJ10.

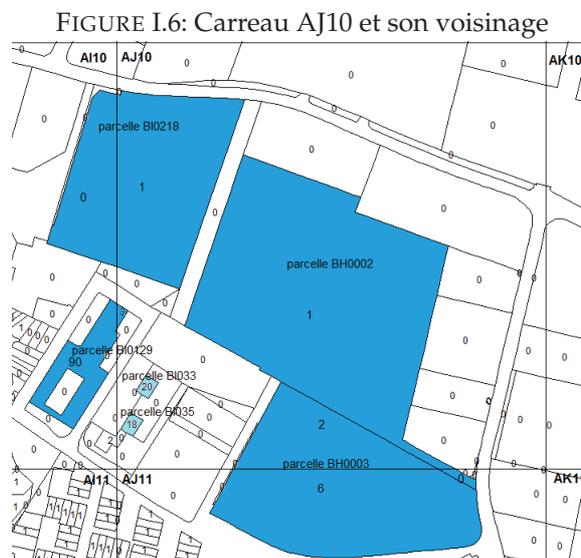


TABLE I.2: Exemple DAW : données nécessaires au calcul pour le carreau AJ10

Zone source S_i	$ S_i $	y_{S_i}	$ A_i \cap AJ10 $
parcelle	Aire _{parcelle}	Nb_logements _{parcelle}	Aire _{parcelle} \cap AJ10
BH0002	62 854,80	1	62 468,48
BH0003	39 666,04	8	12 231,15
BI0033	373,668	20	373,668
BI0035	379,69	18	379,69
BI0129	8 132,73	108	208,139
BI0218	34 890,44	1	24 374,66

2. Pour améliorer la lisibilité de l'exemple, la partie de l'identifiant des parcelles commune aux parcelles de l'exemple a été tronquée. Par exemple, l'identifiant de la parcelle BI0033 est en réalité 31149000BI033

1ère étape : Prédiction de Y sur les zones d'intersection A_{ij} (parcelle \cap AJ10)

$$Nb_logements_{BH0002 \cap AJ10} = \frac{Aire_{BH0002 \cap AJ10}}{Aire_{BH0002}} \times Nb_logements_{BH0002} = \frac{62\,468,48}{62\,854,80} \times 1 = 0,994$$

$$Nb_logements_{BH0003 \cap AJ10} = \frac{Aire_{BH0003 \cap AJ10}}{Aire_{BH0003}} \times Nb_logements_{BH0003} = \frac{12\,231,15}{39\,666,04} \times 8 = 2,467$$

$$Nb_logements_{BI0033 \cap AJ10} = Nb_logements_{BI0033} = 20 \quad \text{car cette parcelle est totalement incluse dans le carreau AJ10}$$

$$Nb_logements_{BI0035 \cap AJ10} = Nb_logements_{BI0035} = 18 \quad \text{car cette parcelle est totalement incluse dans le carreau AJ10}$$

$$Nb_logements_{BI0129 \cap AJ10} = \frac{Aire_{BI0129 \cap AJ10}}{Aire_{BI0129}} \times Nb_logements_{BI0129} = \frac{208,139}{8\,132,73} \times 108 = 2,764$$

$$Nb_logements_{BI0218 \cap AJ10} = \frac{Aire_{BI0218 \cap AJ10}}{Aire_{BI0218}} \times Nb_logements_{BI0218} = \frac{24\,374,66}{34\,890,44} \times 1 = 0,699$$

2ème étape : Prédiction de Y sur la zone cible T_j (le carreau AJ10)

$$\begin{aligned} Nb_logements_{AJ10} &= Nb_logements_{BH0002 \cap AJ10} + Nb_logements_{BH0003 \cap AJ10} + Nb_logements_{BI0033 \cap AJ10} \\ &\quad + Nb_logements_{BI0035 \cap AJ10} + Nb_logements_{BI0129 \cap AJ10} + Nb_logements_{BI0218 \cap AJ10} \\ &= 0,994 + 2,467 + 20 + 18 + 2,764 + 0,699 = 44,923 \end{aligned}$$

Fiche II

Allocation proportionnelle à une variable auxiliaire - *Ordinary dasymetric weighting (with auxiliary information) (DAX)*

Conditions d'utilisation

variable cible Y : extensive ou intensive

variable auxiliaire X :

- de même nature que Y (extensive si Y est extensive ou intensive si Y est intensive),
- connue au niveau des zones d'intersection A_{ij}
- corrélée à Y (une bonne corrélation positive est nécessaire pour que cette méthode donne de bons résultats)

Remarque : on ne fait pas référence ici à la corrélation usuelle en raison de l'autocorrélation spatiale présente dans les données.

Principe [6, 8]

Allouer les données aux zones d'intersection A_{ij} proportionnellement à la variable auxiliaire X .

Propriété pycnophylactique satisfaite (cf. p. 7).

Ecriture synthétique

$$\hat{Y}_{T_j} = \sum_{i: S_i \cap T_j \neq \emptyset} \frac{x_{A_{ij}}}{x_{S_i}} \times y_{S_i} \quad \text{avec} \quad x_{S_i} = \sum_j x_{A_{ij}} \quad \text{si } Y \text{ et } X \text{ extensives}$$

$$\hat{Y}_{T_j} = \sum_{i: S_i \cap T_j \neq \emptyset} \frac{x_{A_{ij}}}{x_{S_i}} \times y_{S_i} \quad \text{avec} \quad x_{S_i} = \sum_j \omega_{A_{ij}} x_{A_{ij}} \quad \text{si } Y \text{ et } X \text{ intensives}$$

La méthode d'allocation proportionnelle à l'aire (DAW, cf. fiche I p.21) suppose une répartition homogène de la variable d'intérêt (dans le cas extensif) dans les zones sources. Cette hypothèse est rarement vérifiée en pratique. Les différentes méthodes dasymétriques cherchent à utiliser une ou plusieurs informations auxiliaires permettant de rendre compte de la variabilité de la variable d'intérêt au sein des zones sources. Dans le cas de l'allocation proportionnelle à une variable auxiliaire (DAX) présentée ici, l'information auxiliaire est une variable, de même nature que la variable d'intérêt, connue au niveau des intersections sources-cibles. Pour que cette méthode donne de bons résultats, la variable auxiliaire et la variable d'intérêt doivent être fortement corrélées.

Exemple

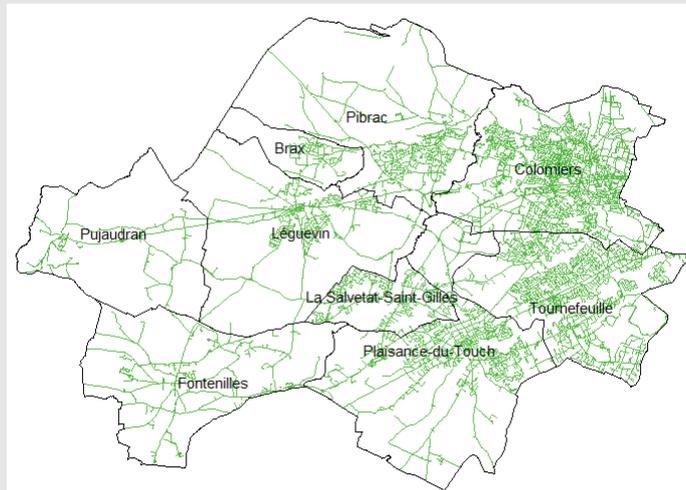
Pour illustrer cette méthode, nous chercherons à connaître le nombre de logements au carreau (noté $Nb_logements_{carreau}$) à partir de l'information connue au niveau communal (noté $Nb_logements_{commune}$). La variable Y , le nombre de logements, est extensive.

Dans cet exemple, les zones sources sont donc les 9 communes et les zones cibles les 747 carreaux. La taille des zones sources est supérieure à celle des carreaux, on se trouve donc dans un cas de désagrégation.

Nous disposons d'une information auxiliaire : la couche routes de la BD Topo. La variable auxiliaire X est la longueur de routes, elle est obtenue au niveau des zones d'intersection commune-carreau ($Longueur_routes_{commune \cap carreau}$) après quelques manipulations (cf. annexe E p.89). La figure II.1 présente la superposition de la couche des communes et de la couche des routes, après sélection des routes à 1 ou 2 chaussées et suppression des tronçons de route extérieurs aux communes.

Cette variable auxiliaire est corrélée de façon satisfaisante au nombre de logements : les zones contenant beaucoup de routes ont tendance à contenir beaucoup de logements, et inversement.

FIGURE II.1: Superposition de la couche des communes (sources) et de la couche des routes (BD Topo, info. auxiliaire = longueur des routes)



Application des notations aux données étudiées

- $y_{S_i} = Nb_logements_{commune}$ connue pour i (commune) = 1, ..., 9
- $Y_{T_j} = Nb_logements_{carreau}$ à calculer pour j (carreau) = 1, ..., 747
- $A_{ij} = commune \cap carreau$, il y a 905 intersections non vides.
- $x_{A_{ij}} = Longueur_routes_{commune \cap carreau}$ connues sur les 905 intersections $commune \cap carreau$.

II.1 Formules d'allocation

II.1.1 Cas d'une variable extensive

1ère étape : Prédiction de Y sur les zones d'intersection A_{ij}

Suivant le principe d'allocation proportionnelle à la variable auxiliaire X et en supposant que X est corrélée à Y , la proportion de la valeur de Y sur la zone source S_i (i.e. la proportion de y_{S_i}) qui est allouée à la zone d'intersection A_{ij} est déterminée proportionnellement à X . Ainsi le coefficient de pondération appliqué à y_{S_i} est défini comme le ratio de la valeur de la variable auxiliaire X sur l'intersection A_{ij} ($x_{A_{ij}}$) et de sa valeur sur la source S_i (x_{S_i}).

La valeur prédite de Y pour la zone d'intersection A_{ij} , notée $\hat{Y}_{A_{ij}}$, s'écrit donc :

$$\hat{Y}_{A_{ij}} = \frac{x_{A_{ij}}}{x_{S_i}} \times y_{S_i} \quad \text{avec} \quad x_{S_i} = \sum_j x_{A_{ij}}$$

Exemple

Cette première étape consiste à calculer le nombre de logements au niveau des intersections communes-carreaux ($\hat{Y}_{A_{ij}}$).

Pour cela, on calcule tout d'abord la longueur des routes au niveau des intersections communes-carreaux ($x_{A_{ij}}$). On en déduit la longueur des routes au niveau communal (x_{S_i}). Enfin on calcule le nombre de logements au niveau des intersections communes-carreaux ($\hat{Y}_{A_{ij}}$) à partir de ces éléments et du nombre de logements de la commune (y_{S_i}).

Ainsi avec les variables de l'exemple, la formule précédente s'écrit :

$$Nb_logements_{commune \cap carreau} = \frac{Longueur_routes_{commune \cap carreau}}{Longueur_routes_{commune}} \times Nb_logements_{commune}$$

$$\text{avec} \quad Longueur_routes_{commune} = \sum_{carreaux} Longueur_routes_{carreau \cap commune}$$

2e étape : Prédiction de Y sur les zones cibles T_j

Rappelons qu'une zone cible T_j (j fixé) est l'union de toutes les zones d'intersection A_{ij} de cette zone cible avec les sources S_i : $T_j = \bigcup_i A_{ij}$.

En conséquence de la propriété d'extensivité, la valeur de Y dans la zone cible T_j (\hat{Y}_{T_j}) s'obtient en sommant la valeur de Y sur les zones d'intersection A_{ij} ($\hat{Y}_{A_{ij}}$) pour toutes les sources S_i intersectant la cible T_j , c'est-à-dire en sommant sur i les $\hat{Y}_{A_{ij}}$:

$$\hat{Y}_{T_j} = \sum_{i : S_i \cap T_j \neq \emptyset} \hat{Y}_{A_{ij}}$$

Exemple

Cette deuxième étape consiste à calculer le nombre de logements au niveau des carreaux (\hat{Y}_{T_j}), à partir du nombre de logements au niveau des intersections communes-carreaux ($\hat{Y}_{A_{ij}}$) pour les communes S_i qui intersectent chaque carreau T_j .

Ainsi avec les variables de l'exemple, la formule précédente s'écrit :

$$Nb_logements_{carreau} = \sum_{communes} Nb_logements_{communes \cap carreau}$$

Ecriture synthétique

$$\hat{Y}_{T_j} = \sum_{i : S_i \cap T_j \neq \emptyset} \frac{x_{A_{ij}}}{x_{S_i}} \times y_{S_i} \quad \text{avec} \quad x_{S_i} = \sum_j x_{A_{ij}}$$

Exemple

Avec les variables de l'exemple, la formule synthétique d'allocation s'écrit :

$$Nb_logements_{carreau} = \sum_{communes} \frac{Longueur_routes_{commune \cap carreau}}{Longueur_routes_{commune}} \times Nb_logements_{commune}$$

Les figures II.2 et II.3 donnent le résultat du carroyage avec la méthode DAX.

FIGURE II.2: Extrait du carroyage DAX : Nombre de logements calculé au carreau (superposition avec la couche des routes)

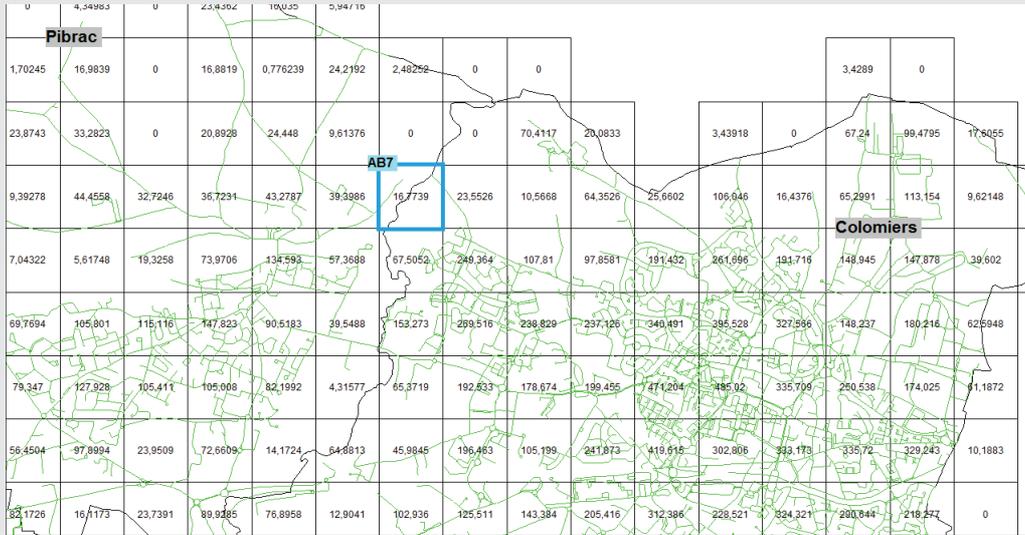
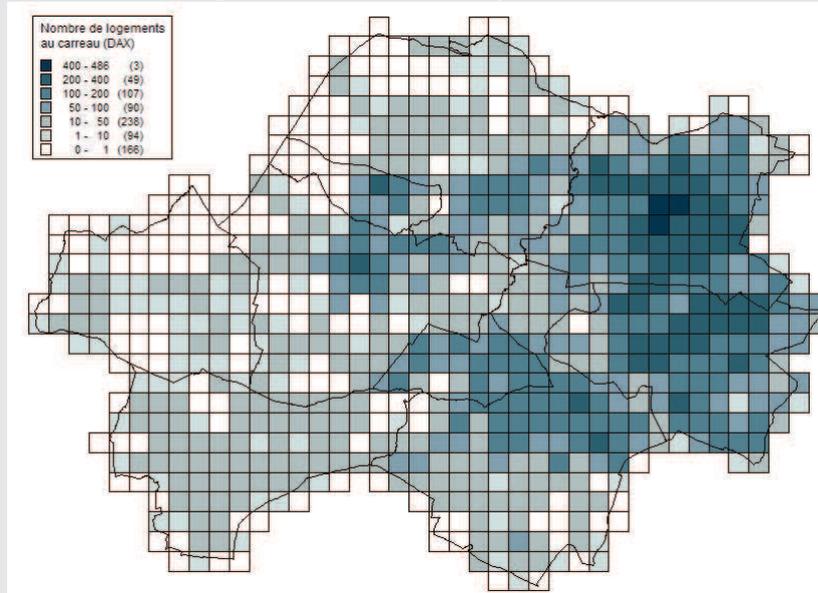


FIGURE II.3: Carte choroplèthe du nombre de logements au carreau, méthode DAX



II.1.2 Cas d'une variable intensive

Si la variable d'intérêt Y est intensive (par exemple la densité de logements) et si on dispose d'une information auxiliaire sous la forme d'une variable intensive X , alors la formule de prédiction dans les zones cibles T_j reste la même :

$$\hat{Y}_{T_j} = \sum_{i: S_i \cap T_j \neq \emptyset} \frac{x_{A_{ij}}}{x_{S_i}} \times y_{S_i}$$

En revanche, la formule d'agrégation pour X diffère (car c'est une variable intensive et non plus extensive) :

$$x_{S_i} = \sum_j \omega_{A_{ij}} x_{A_{ij}}$$

II.2 Exemple de calcul : détails du calcul pour le carreau AB7

Le carreau AB7 intersecte les communes de Pibrac et de Colomiers (cf. figure II.4). Le tableau II.1 fournit les données nécessaires au calcul du nombre de logements pour le carreau AB7.

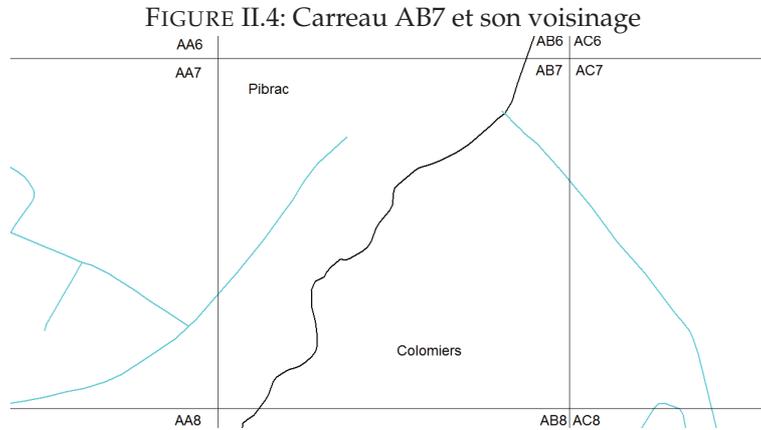


TABLE II.1: Exemple DAX : données nécessaires au calcul pour le carreau AB7

Zone source S_i	x_{S_i}	y_{S_i}	$x_{A_i \cap AB7}$
commune	Longueur_routes _{commune}	Nb_logements _{commune}	Longueur_routes _{commune} \cap AB7
Pibrac	90 589,36	2 860	296,664
Colomiers	259 147,56	14 273	134,502

1ère étape : Prédiction de Y sur les zones d'intersection A_{ij} ($commune \cap AB7$)

$$Nb_logements_{Pibrac \cap AB7} = \frac{Longueur_routes_{Pibrac \cap AB7}}{Longueur_routes_{Pibrac}} \times Nb_logements_{Pibrac} = \frac{296,664}{90\,589,36} \times 2\,860 = 9,366$$

$$Nb_logements_{Colomiers \cap AB7} = \frac{Longueur_routes_{Colomiers \cap AB7}}{Longueur_routes_{Colomiers}} \times Nb_logements_{Colomiers} = \frac{134,502}{259\,147,56} \times 14\,273 = 7,408$$

2ème étape : Prédiction de Y sur la zone cible T_j (le carreau AB7)

$$Nb_logements_{AB7} = Nb_logements_{Pibrac \cap AB7} + Nb_logements_{Colomiers \cap AB7} = 9,366 + 7,408 = 16,774$$

Fiche III

Méthode dasymétrique avec zones de contrôle - *Dasymetric weighting with control zones*) (DAC)

Conditions d'utilisation

variable cible Y : extensive ou intensive

information auxiliaire : zones de contrôle C_k

- la densité de Y sur C_k , notée D_{C_k} avec $D_{C_k} = y_{C_k}/|C_k|$, est connue (fixée par un expert ou estimée statistiquement, éventuellement à partir d'un autre jeu de données),
- Y est homogène (si de nature extensive) ou uniforme (si de nature intensive) sur les zones de contrôle C_k .

On note A_{ijk} l'intersection entre la zone de contrôle C_k et la zone d'intersection entre la source S_i et la cible T_j (notée A_{ij}) : $A_{ijk} = A_{ij} \cap C_k = S_i \cap T_j \cap C_k$.

Principe [4]

Allouer les données aux zones d'intersection A_{ij} proportionnellement à une estimation "grossière" de la variable d'intérêt Y utilisant la densité de Y sur des zones de contrôle.

Propriété pycnophylactique satisfaite (cf. p. 7).

Ecriture synthétique

$$\hat{Y}_{T_j} = \sum_{i : S_i \cap T_j \neq \emptyset} \frac{\sum_k |A_{ijk}| \hat{D}_{C_k}}{\sum_l \sum_m |A_{ilm}| \hat{D}_{C_m}} \times y_{S_i} \quad \text{si } Y \text{ extensive}$$

La méthode d'allocation proportionnelle à l'aire (DAW, cf. fiche I p.21) suppose une répartition homogène de la variable d'intérêt (dans le cas extensif) dans les zones sources. Cette hypothèse est rarement vérifiée en pratique. Les différentes méthodes dasymétriques cherchent à utiliser une ou plusieurs informations auxiliaires permettant de rendre compte de la variabilité de la variable d'intérêt au sein des zones sources. Dans le cas de la méthode dasymétrique avec zones de contrôle (DAC) présentée ici, l'information auxiliaire prend la forme de zones de contrôle sur lesquelles la variable d'intérêt est homogène (dans le cas extensif). Cette hypothèse est moins contraignante que celle d'une homogénéité sur les zones sources. On suppose de plus que la densité de la variable d'intérêt sur les zones de contrôle est connue.

Exemple

Pour illustrer cette méthode, nous chercherons à connaître le nombre de logements au carreau (noté $Nb_logements_{carreau}$) à partir de l'information connue au niveau communal (noté $Nb_logements_{commune}$). La variable d'intérêt Y , le nombre de logements, est extensive.

Dans cet exemple, les zones sources sont donc les 9 communes et les zones cibles les 747 carreaux. La taille des zones sources est supérieure à celle des carreaux, on se trouve donc dans un cas de désagrégation.

Nous disposons d'une information auxiliaire par zones de contrôle sur la nature du sol : la couche Corine Land Cover 2006 (notée CLC). Ces données sont utilisées ici au niveau 1 (cf. figure III.1), soit le niveau le plus grossier, comprenant 5 catégories dont 4 présentes dans nos données :

- 1 = territoires artificialisés,
- 2 = territoires agricoles,
- 3 = forêts et milieux semi-naturels,
- 5 = surface en eau.

FIGURE III.1: Superposition de la couche des communes (sources) et de la couche Corine Land Cover niveau 1 (zones de contrôle)

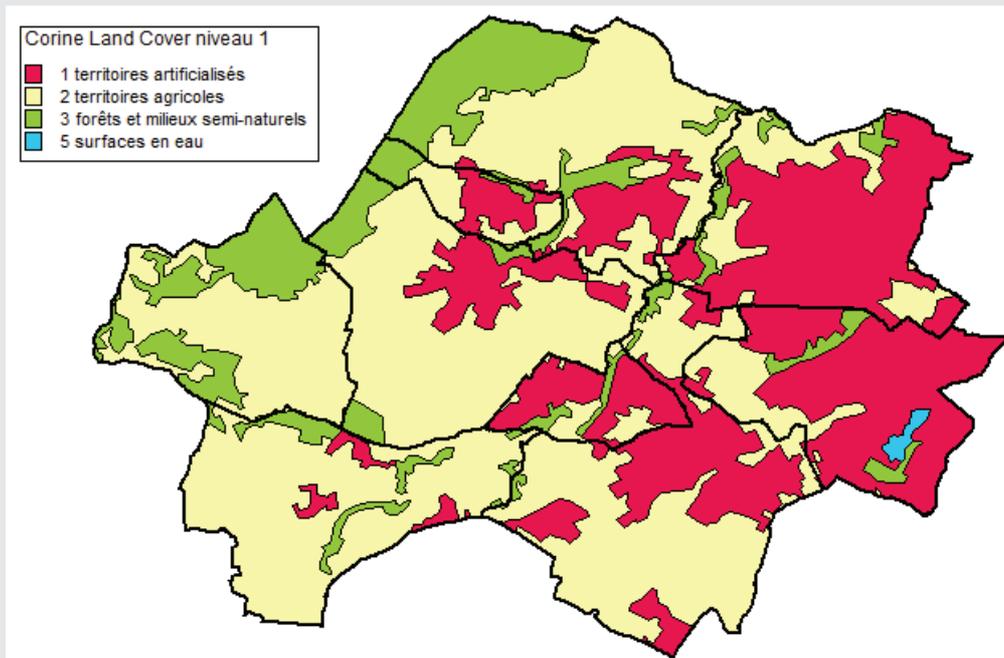


TABLE III.1: Aire des CLC par commune (en ha) et % de l'aire totale de la commune (Source : site du ministère du développement durable, contour communal de la BD Carto IGN 2006 en système de projection RGF93)

Commune	1	%	2	%	3	%	5	%	Total	%
Brax	164	37	178	40	104	23	0	0	446	100
Colomiers	1 530	73	449	21	128	6	0	0	2 107	100
Fontenilles	112	6	1 717	85	186	9	0	0	2 014	100
Léguevin	394	16	1 833	74	244	10	0	0	2 472	100
Pibrac	377	14	1 488	57	744	29	0	0	2 609	100
Plaisance-du-Touch	919	34	1 704	64	46	2	0	0	2 668	100
La Salvetat St Gilles	385	66	136	23	61	10	0	0	583	100
Tournefeuille	1 346	74	321	18	109	6	43	2	1 818	100
Pujaudran	0	0	1 130	64	638	36	0	0	1 768	100
Total	5 226	32	8 957	54	2 259	14	43	0	16 485	100

Codage CLC : 1 = territoires artificialisés, 2 = territoires agricoles, 3 = forêts et milieux semi-naturels, 5 = surface en eau

Application des notations aux données étudiées

- $C_k = CLC_k$ (zone de contrôle Corine Land Cover) pour $k = 1, 2, 3, 5$.
- $D_{C_k} = Densité_Igt_{CLC_k}$ connue pour $k = 1, 2, 3, 5$.
- $y_{S_i} = Nb_logements_{commune}$ connue pour i (commune) = $1, \dots, 9$.
- $Y_{T_j} = Nb_logements_{carreau}$ à calculer pour j (carreau) = $1, \dots, 747$.
- $A_{ijk} = commune \cap carreau \cap CLC_k$, il y a 1 467 intersections non vides.
- $A_{ij} = commune \cap carreau$, il y a 905 intersections non vides.

III.1 Formule d'allocation (cas d'une variable extensive)

III.1.1 Etape préliminaire : Estimation de la densité de Y sur les zones de contrôle C_k

La densité de Y sur la zone de contrôle C_k , notée D_{C_k} , est ici définie comme le rapport entre la valeur de Y sur la zone de contrôle C_k et l'aire de C_k ($D_{C_k} = y_{C_k} / |C_k|$).

On suppose que la densité de Y sur la zone de contrôle C_k est connue. Elle est soit fixée par un expert soit déterminée statistiquement par :

$$\hat{D}_{C_k} = \frac{\sum_{i: S_i \in C_k} y_{S_i}}{\sum_{i: S_i \in C_k} |S_i|}$$

avec $S_i \in C_k$ signifiant : " S_i inclus dans C_k " ou bien "le centroïde de S_i appartient à C_k " ou bien "un pourcentage fixé de S_i est couvert par C_k " ou encore " C_k est la zone majoritaire sur S_i ".

Exemple : Calcul de la densité de logement à partir de l'information disponible au niveau parcellaire

Comme ici nous disposons du nombre de logements au niveau parcellaire, la densité de logement sur les zones de contrôle sera estimée statistiquement à partir de cette information disponible au niveau des parcelles.

Pour déterminer la densité de logement sur une zone de contrôle Corine Land Cover donnée, on considère toutes les communes qui l'intersectent puis on calcule le rapport entre la somme du nombre de logements sur l'intersection entre ces communes et la zone CLC considérée et la somme de l'aire de ces intersections :

$$Densité_Igt_{CLC_k} = \frac{\sum_{communes} Nb_logements_{commune \cap CLC_k}}{\sum_{communes} Aire_{commune \cap CLC_k}}$$

pour toutes les communes intersectant la zone de contrôle Corine Land Cover considérée (CLC_k).

Le tableau III.2 reprend ces éléments pour les quatre zones Corine Land Cover et fournit la densité de logements estimée. Notons que l'information concernant l'absence de logement sur les surfaces en eau (lac de la base de la Ramée à Tournefeuille) assure ici d'éviter une prédiction non nulle sur les zones constituées d'eau.

TABLE III.2: Calcul des densités pour chaque zone Corine Land Cover

CLC_k	$\sum_{comm} Aire_{commune \cap CLC_k}$ (en km^2)	$\sum_{comm} Nb_logements_{commune \cap CLC_k}$	$Densité_Igt_{CLC_k}$ (en logements/ km^2)
1 "territoires artificialisés"	52,30	38 311	733
2 "territoires agricoles"	89,57	3 381	38
3 "forêts et milieux semi-naturels"	22,68	734	32
5 "surfaces en eau"	0,43	0	0

III.1.2 Etape 1 : Prédiction de Y sur les zones d'intersection A_{ij}

Etape 1a : Prédiction grossière de Y sur les zones d'intersection A_{ij}

On peut prédire de façon grossière la valeur de Y sur une zone donnée en multipliant l'aire de cette zone par la densité de Y sur la zone de contrôle intersectant cette zone. Ainsi, la prédiction grossière de Y sur A_{ijk} (intersection entre la source S_i , la cible T_j et la zone de contrôle C_k) est obtenue en multipliant l'aire de A_{ijk} par la densité de Y sur C_k (\hat{D}_{C_k}). Si on considère que la zone d'intersection A_{ij} (intersection entre la source S_i et la cible T_j) se décompose sur plusieurs zones de contrôle C_k ($A_{ij} = \bigcup_k A_{ijk}$), la prédiction grossière de Y sur A_{ij} (notée $\tilde{Y}_{A_{ij}}$) est obtenue en sommant les prédictions précédentes (prédiction grossière de Y sur A_{ijk}) pour toutes les zones de contrôle C_k intersectant la source S_i et la cible T_j :

$$\tilde{Y}_{A_{ij}} = \sum_k |A_{ijk}| \hat{D}_{C_k} \quad \text{avec} \quad A_{ij} = \bigcup_k A_{ijk}$$

Exemple

Cette toute première étape consiste à calculer, de façon *grossière*, le nombre de logements au niveau d'une intersection communes-carreaux ($\tilde{Y}_{A_{ij}}$) comme le produit de la densité de logement sur les zones Corine Land Cover recouvrant l'intersection communes-carreaux par l'aire de cette intersection, puis en sommant ces produits sur les zones Corine Land Cover :

$$\tilde{N}b_logements_{commune \cap carreau} = \sum_{CLC_k} Aire_{commune \cap carreau \cap CLC_k} \times Densité_lgt_{CLC_k}$$

pour toutes les zones Corine Land Cover intersectant la commune et le carreau considérés.

Etape 1b : Prédiction grossière de Y sur les zones sources S_i

En considérant que la zone source S_i est intersectée par les zones cibles T_l (soit $S_i = \bigcup_l A_{il}$), on peut prédire de façon grossière la valeur de Y sur S_i (notée \tilde{Y}_{S_i}), comme la somme des prédictions grossières de Y sur les intersections A_{il} :

$$\tilde{Y}_{S_i} = \sum_l \tilde{Y}_{A_{il}} = \sum_l \left(\sum_k |A_{ilk}| \hat{D}_{C_k} \right) \quad \text{avec} \quad S_i = \bigcup_l A_{il}$$

Exemple

Cette étape consiste à calculer, de façon *grossière*, le nombre de logements au niveau de la commune (noté \tilde{Y}_{S_i}). On calcule comme précédemment le nombre de logements au niveau des intersections communes-carreaux comme des sommes de produits d'une aire par une densité, puis on somme sur les carreaux intersectant la commune considérée :

$$\begin{aligned} \tilde{N}b_logements_{commune} &= \sum_{carreaux} \tilde{N}b_logements_{commune \cap carreau} \\ &= \sum_{carreaux} \left(\sum_{CLC_k} Aire_{commune \cap carreau \cap CLC_k} \times Densité_lgt_{CLC_k} \right) \end{aligned}$$

pour toutes les zones Corine Land Cover et tous les carreaux intersectant la commune considérée.

Etape 1c : Prédiction de Y sur les zones d'intersection A_{ij} à partir des prédictions grossières de Y sur A_{ij} et S_i

La proportion de la valeur de Y sur la zone source S_i (i.e. la proportion de y_{S_i}) qui est allouée à la zone d'intersection A_{ij} est déterminée proportionnellement aux estimations grossières de Y . Ainsi le coefficient de pondération appliqué à y_{S_i} est défini comme le ratio de la valeur de l'estimation grossière de Y sur l'intersection A_{ij} et de sa valeur sur la source S_i .

La valeur prédite de Y pour la zone d'intersection A_{ij} , notée $\hat{Y}_{A_{ij}}$, s'écrit donc :

$$\hat{Y}_{A_{ij}} = \frac{\tilde{Y}_{A_{ij}}}{\tilde{Y}_{S_i}} \times y_{S_i} = \frac{\sum_k |A_{ijk}| \hat{D}_{C_k}}{\sum_l \sum_m |A_{ilm}| \hat{D}_{C_m}} \times y_{S_i}$$

Exemple

Cette étape consiste à calculer de façon définitive le nombre de logements au niveau des intersections communes-carreaux ($\hat{Y}_{A_{ij}}$) à partir des prédictions grossières effectuées précédemment. Avec les variables de l'exemple, la formule précédente s'écrit :

$$\begin{aligned} Nb_logements_{commune \cap carreau} &= \frac{\tilde{Nb_logements}_{commune \cap carreau}}{\tilde{Nb_logements}_{commune}} \times Nb_logements_{commune} \\ &= \frac{\sum_{CLC_k} Aire_{commune \cap carreau \cap CLC_k} \times Densité_Igt_{CLC_k}}{\sum_{carreaux} \sum_{CLC_m} Aire_{commune \cap carreau \cap CLC_m} \times Densité_Igt_{CLC_m}} \times Nb_logements_{commune} \end{aligned}$$

III.1.3 Etape 2 : Prédiction de Y sur les zones cibles T_j

Rappelons qu'une zone cible T_j (j fixé) est l'union de toutes les zones d'intersection A_{ij} de cette zone cible avec les sources S_i : $T_j = \bigcup_i A_{ij}$.

En conséquence de la propriété d'extensivité, la valeur de Y dans la zone cible T_j (\hat{Y}_{T_j}) s'obtient en sommant la valeur de Y sur les zones d'intersection A_{ij} ($\hat{Y}_{A_{ij}}$) pour toutes les sources S_i intersectant la cible T_j , c'est-à-dire en sommant sur i les $\hat{Y}_{A_{ij}}$:

$$\hat{Y}_{T_j} = \sum_{i : S_i \cap T_j \neq \emptyset} \hat{Y}_{A_{ij}}$$

Exemple

Cette dernière étape consiste à calculer le nombre de logements au niveau des carreaux (\hat{Y}_{T_j}), à partir du nombre de logements au niveau des intersections communes-carreaux ($\hat{Y}_{A_{ij}}$) pour les communes S_i qui intersectent chaque carreau T_j .

Ainsi avec les variables de l'exemple, la formule précédente s'écrit :

$$Nb_logements_{carreau} = \sum_{\substack{\text{communes} \\ \text{intersectant carreau}}} Nb_logements_{communes \cap carreau}$$

III.1.4 Ecriture synthétique

$$\hat{Y}_{T_j} = \sum_{i : S_i \cap T_j \neq \emptyset} \frac{\sum_k |A_{ijk}| \hat{D}_{C_k}}{\sum_l \sum_m |A_{ilm}| \hat{D}_{C_m}} \times y_{S_i}$$

Exemple

Avec les variables de l'exemple, la formule synthétique d'allocation s'écrit :

$$Nb_logements_{carreau} = \sum_{communes} \frac{\sum_{CLC_k} Aire_{communes \cap carreau \cap CLC_k} \times Densité_Igt_{CLC_k}}{\sum_{carreaux} \sum_{CLC_m} Aire_{commune \cap carreau \cap CLC_m} \times Densité_Igt_{CLC_m}} \times Nb_logements_{commune}$$

Les figures III.2, III.3 et III.4 donnent le résultat du carroyage avec la méthode DAC. Sur la dernière figure, on remarque une bonne correspondance entre les prédictions et les zones Corine Land Cover, avec un nombre de logements prédits élevé dans les zones artificialisées et plus faible dans les autres zones.

FIGURE III.2: Extrait du carroyage DAC : Nombre de logements calculé au carreau (superposition avec la couche Corine Land Cover)

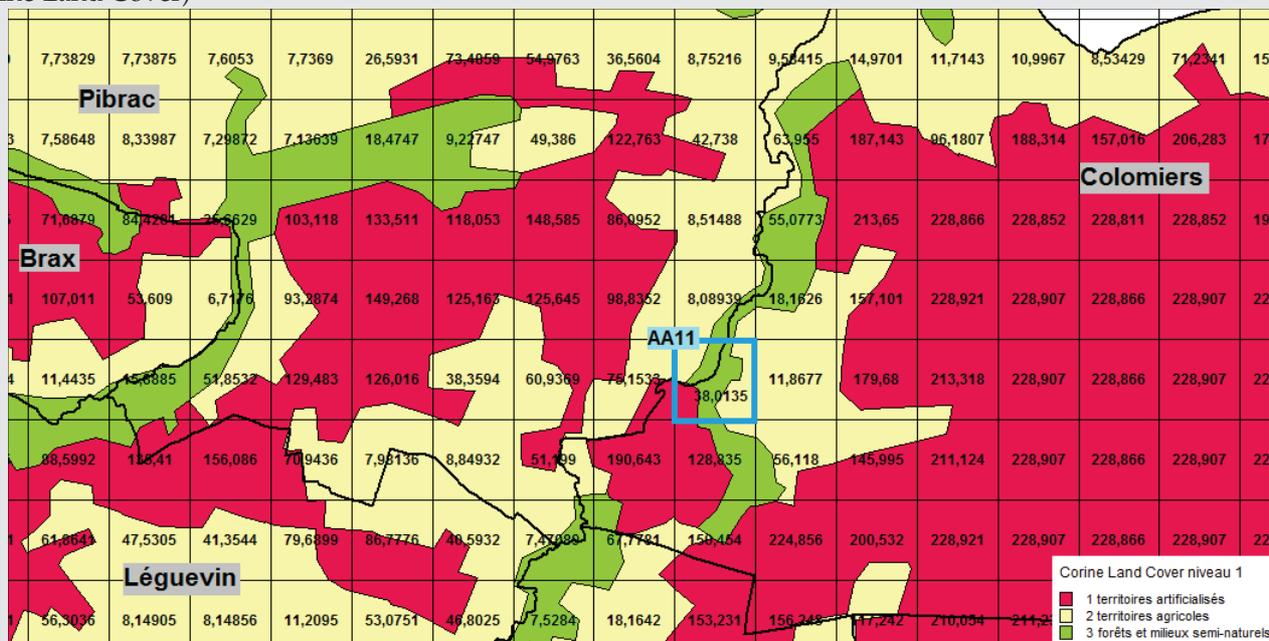


FIGURE III.3: Carte choroplèthe du nombre de logements au carreau, méthode DAC

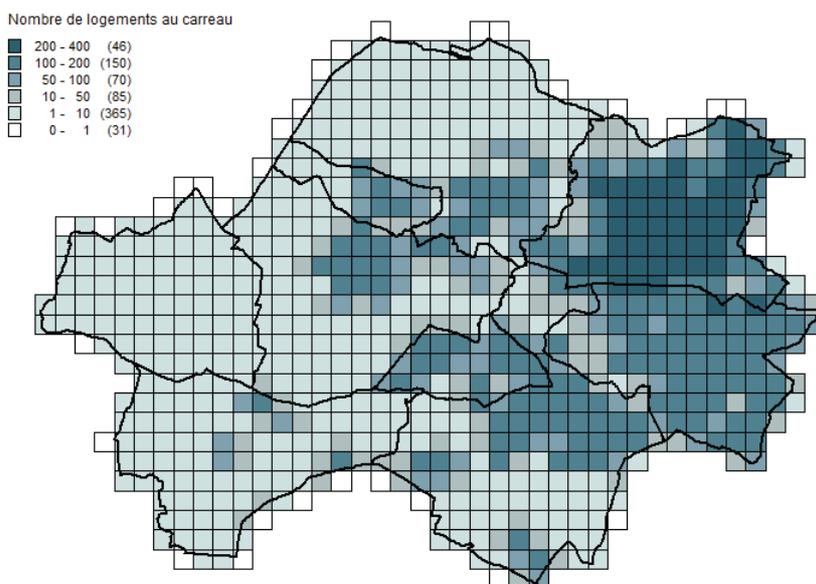
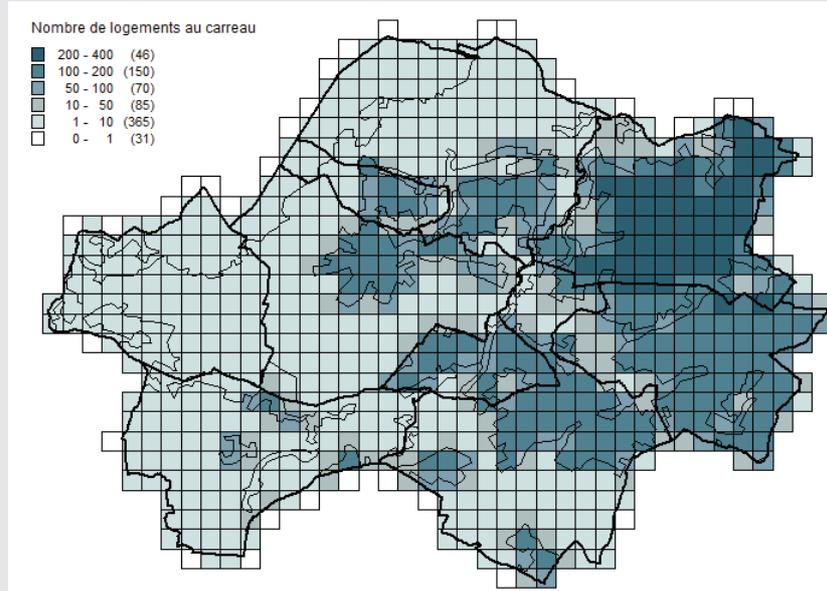


FIGURE III.4: Carte choroplèthe du nombre de logements au carreau, méthode DAC, avec superposition de la couche Corine Land Cover (matérialisée par les contours uniquement)



III.2 Exemples de calcul

III.2.1 Densité de logements pour les zones artificialisées (CLC_1)

TABLE III.3: Exemple DAC : données nécessaires au calcul de la densité de logements pour la zone de contrôle CLC_1

$S_i \cap CLC_k$ $commune \cap CLC_1$	$ S_i \cap CLC_k $ $Aire_{commune \cap CLC_1}$	$y_{S_i \cap CLC_k}$ $Nb_logements_{commune \cap CLC_1}$
$Brax \cap CLC_1$	1,66	804
$Colomiers \cap CLC_1$	15,30	13 850
$Fontenilles \cap CLC_1$	1,10	483
$La\ Salvetat-Saint-Gilles \cap CLC_1$	3,86	2 161
$Léguévin \cap CLC_1$	3,95	2 720
$Pibrac \cap CLC_1$	3,69	2 339
$Plaisance-du-Touch \cap CLC_1$	9,20	5 871
$Tournefeuille \cap CLC_1$	13,54	10 083

$$\begin{aligned}
 Densité_lgt_{CLC_1} &= \frac{Nb_logements_{Brax \cap CLC_1} + Nb_logements_{Colomiers \cap CLC_1} + \dots + Nb_logements_{Tournefeuille \cap CLC_1}}{Aire_{Brax \cap CLC_1} + Aire_{Colomiers \cap CLC_1} + \dots + Aire_{Tournefeuille \cap CLC_1}} \\
 &= \frac{804 + 13\,850 + 483 + 2\,161 + 2\,720 + 2\,339 + 5\,871 + 10\,083}{1,66 + 15,30 + 1,10 + 3,86 + 3,95 + 3,69 + 9,20 + 13,54} = \frac{38\,311}{52,30} = 733
 \end{aligned}$$

III.2.2 Détails du calcul pour le carreau AA11

Le carreau AA11 intersecte les communes de Pibrac et de Colomiers, trois zones Corine Land Cover sont présentes dans chaque intersection commune-carreau (cf. figure III.5). Le tableau III.4 fournit les données nécessaires au calcul du nombre de logements pour le carreau AA11 par la méthode DAC.

FIGURE III.5: Carreau AA11

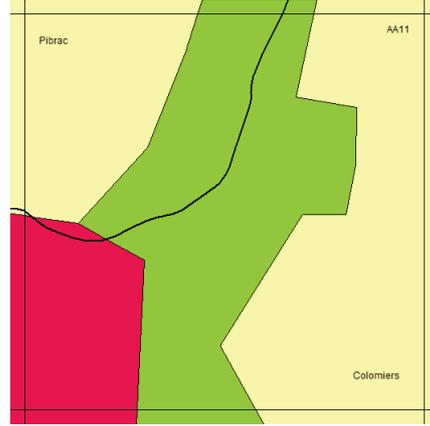


TABLE III.4: Exemple DAC : données nécessaires au calcul pour le carreau AA11

Zone	D_{C_k} Densité_lgt $_{CLC_k}$	y_{S_i} Nb_logements $_{commune}$	$ A_{i \cap AA11 \cap k} $ Aire $_{commune \cap AA11 \cap CLC_k}$
Pibrac		2 860	
$Pibrac \cap CLC_1$	733		0,001137
$Pibrac \cap CLC_2$	38		0,041558
$Pibrac \cap CLC_3$	32		0,029577
Colomiers		14 273	
$Colomiers \cap CLC_1$	733		0,031476
$Colomiers \cap CLC_2$	38		0,082126
$Colomiers \cap CLC_3$	32		0,064084

1ère étape : Prédiction de Y sur les zones d'intersection A_{ij} ($commune \cap AA11$)

Etape 1a : Prédiction grossière de Y sur les zones d'intersection A_{ij} ($commune \cap AA11$) On commence par calculer les prédictions grossières de Y sur les zones d'intersection A_{ijk} ($commune \cap AA11 \cap CLC_k$) :

$$\begin{aligned} \tilde{N}b_logements_{Pibrac \cap AA11 \cap CLC_1} &= Aire_{Pibrac \cap AA11 \cap CLC_1} \times Densité_lgt_{CLC_1} = 0,001137 \times 733 = 0,8334 \\ \tilde{N}b_logements_{Pibrac \cap AA11 \cap CLC_2} &= Aire_{Pibrac \cap AA11 \cap CLC_2} \times Densité_lgt_{CLC_2} = 0,041558 \times 38 = 1,5792 \\ \tilde{N}b_logements_{Pibrac \cap AA11 \cap CLC_3} &= Aire_{Pibrac \cap AA11 \cap CLC_3} \times Densité_lgt_{CLC_3} = 0,029577 \times 32 = 0,9465 \\ \tilde{N}b_logements_{Colomiers \cap AA11 \cap CLC_1} &= Aire_{Colomiers \cap AA11 \cap CLC_1} \times Densité_lgt_{CLC_1} = 0,031476 \times 733 = 23,0719 \\ \tilde{N}b_logements_{Colomiers \cap AA11 \cap CLC_2} &= Aire_{Colomiers \cap AA11 \cap CLC_2} \times Densité_lgt_{CLC_2} = 0,082126 \times 38 = 3,1208 \\ \tilde{N}b_logements_{Colomiers \cap AA11 \cap CLC_3} &= Aire_{Colomiers \cap AA11 \cap CLC_3} \times Densité_lgt_{CLC_3} = 0,064084 \times 32 = 2,0507 \end{aligned}$$

Puis on somme ces prédictions sur les CLC pour obtenir les prédictions grossières de Y sur les A_{ij} ($commune \cap AA11$) :

$$\begin{aligned} \tilde{N}b_logements_{Pibrac \cap AA11} &= \tilde{N}b_logements_{Pibrac \cap AA11 \cap CLC_1} + \tilde{N}b_logements_{Pibrac \cap AA11 \cap CLC_2} \\ &\quad + \tilde{N}b_logements_{Pibrac \cap AA11 \cap CLC_3} = 0,8334 + 1,5792 + 0,9465 = 3,3591 \\ \tilde{N}b_logements_{Colomiers \cap AA11} &= \tilde{N}b_logements_{Colomiers \cap AA11 \cap CLC_1} + \tilde{N}b_logements_{Colomiers \cap AA11 \cap CLC_2} \\ &\quad + \tilde{N}b_logements_{Colomiers \cap AA11 \cap CLC_3} = 23,0719 + 3,1208 + 2,0507 = 28,2434 \end{aligned}$$

Etape 1b : Prédiction grossière de Y sur les zones sources S_i ($commune$) De la même façon¹, en sommant les prédictions grossières de Y sur les zones d'intersection A_{ijk} sur tous les carreaux et toutes les zones CLC intersectés par chaque commune, on obtient les prédictions grossières pour les communes :

$$\begin{aligned} \tilde{N}b_logements_{Pibrac} &= 3\,511,358 \\ \tilde{N}b_logements_{Colomiers} &= 11\,426,939 \end{aligned}$$

1. Les données nécessaires à ce calcul ne sont pas fournies ici.

Etape 1c : Prédiction de Y sur les zones d'intersection A_{ij} ($commune \cap AA11$) Le nombre de logements au niveau des intersections commune-carreau est calculé à partir des prédictions grossières effectuées précédemment :

$$Nb_logements_{Pibrac \cap AA11} = \frac{\tilde{Nb_logements}_{Pibrac \cap AA11}}{\tilde{Nb_logements}_{Pibrac}} \times Nb_logements_{Pibrac} = \frac{3,3591}{3\,511,358} \times 2\,860 = 2,736$$

$$Nb_logements_{Colomiers \cap AA11} = \frac{\tilde{Nb_logements}_{Colomiers \cap AA11}}{\tilde{Nb_logements}_{Colomiers}} \times Nb_logements_{Colomiers} = \frac{28,2434}{11\,426,939} \times 14\,273 = 35,278$$

2ème étape : Prédiction de Y sur la zone cible T_j (le carreau AA11)

$$\begin{aligned} Nb_logements_{AA11} &= Nb_logements_{Pibrac \cap AA11} + Nb_logements_{Colomiers \cap AA11} \\ &= 2,736 + 35,278 = 38,014 \end{aligned}$$

Fiche IV

Méthode dasymétrique avec zones de contrôle et variable auxiliaire - *Two steps dasymetric weighting (with auxiliary information and control zones) (DA2)*

Conditions d'utilisation

variable cible Y : extensive ou intensive

information auxiliaire 1 : zones de contrôle C_k

- on s'intéresse aux intersections entre les sources S_i et les zones de contrôle C_k , elles sont notées A_{ik} ($A_{i,k} = S_i \cap C_k$),
- Y est homogène (si de nature extensive) ou uniforme (si de nature intensive) sur les intersections sources-zones de contrôle $A_{i,k}$.

information auxiliaire 2 : variable auxiliaire X

- de même nature que Y ,
- connue au niveau des intersections sources-zones de contrôle $A_{i,k}$,
- corrélée à Y (une bonne corrélation positive est nécessaire pour que cette méthode donne de bons résultats).

Remarque : on ne fait pas référence ici à la corrélation usuelle en raison de l'autocorrélation spatiale présente dans les données.

Principe [3]

Allouer les données aux zones d'intersection entre les sources et les zones de contrôle ($A_{i,k}$) proportionnellement à une variable auxiliaire puis aux intersections entre les sources, les cibles et les zones de contrôle (A_{ijk}) par affectation proportionnelle à l'aire.

Propriété pycnophylactique satisfaite (cf. p. 7).

Ecriture synthétique

$$\hat{Y}_{T_j} = \sum_i \sum_k \frac{|A_{ijk}|}{|A_{i,k}|} \times \frac{x_{A_{i,k}}}{x_{S_i}} \times y_{S_i} \quad \text{si } Y \text{ et } X \text{ extensives}$$

La méthode d'allocation proportionnelle à l'aire (DAW, cf. fiche I p.21) suppose une répartition homogène de la variable d'intérêt (dans le cas extensif) dans les zones sources. Cette hypothèse est rarement vérifiée en pratique. Les différentes méthodes dasymétriques cherchent à utiliser une ou plusieurs informations auxiliaires permettant de rendre compte de la variabilité de la variable d'intérêt au sein des zones sources. Dans le cas de la méthode dasymétrique avec zones de contrôle et variable auxiliaire (DA2) présentée ici, on dispose de deux informations auxiliaires : des zones de contrôle et une variable auxiliaire.

Par rapport à l'allocation proportionnelle à une variable auxiliaire (DAX, cf. fiche II p.31), cette méthode présente pour avantage de connaître l'information auxiliaire au niveau des intersections sources-zones de contrôle (voire des zones de contrôle si celles-ci sont imbriquées dans les sources) et non plus au niveau des intersections sources-cibles. (Notons que cet argument n'est pas pertinent dans nos exemples puisque l'information auxiliaire utilisée, la longueur de routes, peut être calculée à n'importe quel niveau géographique).

Par rapport à la méthode dasymétrique avec zones de contrôle (DAC, cf. fiche III p.37), cette méthode présente pour avantage de ne pas utiliser la densité de la variable d'intérêt sur les zones de contrôle. On suppose seulement que la variable d'intérêt est homogène (dans le cas extensif) sur les intersections sources-zones de contrôle.

On note $A_{i,k}$ l'intersection entre la source S_i et la zone de contrôle C_k : $A_{i,k} = S_i \cap C_k$. Si les zones de contrôle sont imbriquées dans les sources, l'intersection $A_{i,k}$ est confondue avec la zone de contrôle C_k .

On note A_{ijk} l'intersection entre la zone de contrôle C_k et la zone d'intersection entre la source S_i et la cible T_j (notée A_{ij}) : $A_{ijk} = A_{ij} \cap C_k = S_i \cap T_j \cap C_k$.

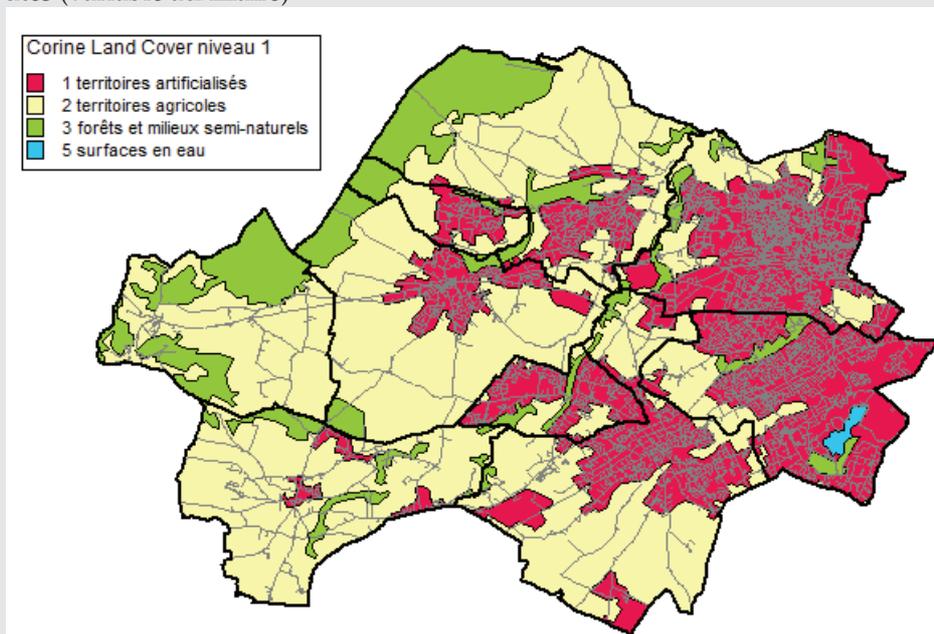
Exemple

Pour illustrer cette méthode, nous chercherons à connaître le nombre de logements au carreau (noté $Nb_logements_{carreau}$) à partir de l'information connue au niveau communal (noté $Nb_logements_{commune}$). La variable d'intérêt Y , le nombre de logements, est extensive.

Dans cet exemple, les zones sources sont donc les 9 communes et les zones cibles les 747 carreaux. La taille des zones sources est supérieure à celle des carreaux, on se trouve donc dans un cas de désagrégation.

Nous disposons de deux sortes d'information auxiliaire (cf. figure IV.1), une première par zones de contrôle sur la nature du sol : la couche Corine Land Cover 2006 (notée CLC, cf. Fiche DAC p.37), ainsi qu'une deuxième sous forme de variable auxiliaire X : la longueur de routes obtenue à partir de la couche routes de la BD Topo (cf. Fiche DAX p.31).

FIGURE IV.1: Superposition de la couche des communes (sources), de la couche des Corine Land Cover (zones de contrôle) et des routes (variable auxiliaire)



Application des notations aux données étudiées

- $C_k = CLC_k$ (zone de contrôle Corine Land Cover) pour $k = 1, 2, 3, 5$.
- $y_{S_i} = Nb_logements_{commune}$ connue pour i (commune) = $1, \dots, 9$.
- $Y_{T_j} = Nb_logements_{carreau}$ à calculer pour j (carreau) = $1, \dots, 747$.
- $A_{i,k} = commune \cap CLC_k$, il y a 27 intersections non vides.
- $A_{ijk} = commune \cap carreau \cap CLC_k$, il y a 1 467 intersections non vides.
- $A_{ij} = commune \cap carreau$, il y a 905 intersections non vides.
- $x_{A_{i,k}} = Longueur_routes_{commune} \cap CLC_k$ connue sur les intersections $commune \cap CLC_k$.

IV.1 Formule d'allocation (cas d'une variable extensive)**IV.1.1 Etape 1 : Prédiction de Y sur les zones d'intersection A_{ijk}** **Etape 1a : Prédiction de Y sur les intersections sources-zones de contrôle $A_{i,k}$ par DAX**

On considère les intersections sources-zones de contrôle comme un nouvel ensemble de zones cibles et on prédit Y sur ces intersections en utilisant la méthode d'allocation proportionnelle à une variable auxiliaire (DAX).

Ainsi, suivant le principe d'allocation proportionnelle à la variable auxiliaire X et en supposant que X est corrélée à Y , la proportion de la valeur de Y sur la zone source S_i (i.e. la proportion de y_{S_i}) qui est allouée à l'intersection $A_{i,k}$ est déterminée proportionnellement à X . Plus précisément, le coefficient de pondération appliqué à y_{S_i} est défini comme le ratio de la valeur de la variable auxiliaire X sur l'intersection $A_{i,k}$ ($x_{A_{i,k}}$) et de sa valeur sur la source S_i (x_{S_i}).

La valeur prédite de Y pour l'intersection source-zone de contrôle $A_{i,k}$, notée $\hat{Y}_{A_{i,k}}$, s'écrit donc :

$$\hat{Y}_{A_{i,k}} = \frac{x_{A_{i,k}}}{x_{S_i}} \times y_{S_i} \quad \text{avec} \quad x_{S_i} = \sum_k x_{A_{i,k}}$$

Exemple

Cette première étape consiste à calculer le nombre de logements au niveau des intersections communes-Corine Land Cover ($\hat{Y}_{A_{i,k}}$). Il s'agit de répartir les logements d'une commune entre les différentes zones Corine Land Cover qui la composent (proportionnellement à la longueur de route).

Pour cela, on calcule tout d'abord la longueur des routes au niveau des intersections communes-Corine Land Cover ($x_{A_{i,k}}$). On en déduit la longueur des routes au niveau communal (x_{S_i}). Enfin on calcule le nombre de logements au niveau des intersections communes-Corine Land Cover ($\hat{Y}_{A_{i,k}}$) à partir de ces éléments et du nombre de logements de la commune (y_{S_i}).

Ainsi avec les variables de l'exemple, la formule précédente s'écrit :

$$Nb_logements_{commune \cap CLC_k} = \frac{Longueur_routes_{commune \cap CLC_k}}{Longueur_routes_{commune}} \times Nb_logements_{commune}$$

$$\text{avec} \quad Longueur_routes_{commune} = \sum_{CLC_k} Longueur_routes_{commune \cap CLC_k}$$

Etape 1b : Prédiction de Y sur les zones d'intersection A_{ijk} par DAW

On considère les intersections sources-zones de contrôle comme un nouvel ensemble de zones sources et on prédit Y sur les intersections entre les cibles et ces premières intersections en utilisant la méthode d'allocation proportionnelle à l'aire (DAW).

Ainsi, la proportion de la valeur de Y sur l'intersection source-zone de contrôle A_{ik} (i.e. la proportion de $\hat{Y}_{A_{i,k}}$) qui est allouée à l'intersection A_{ijk} est déterminée proportionnellement à son aire ($|A_{ijk}|$)¹. Le coefficient de pondération

1. Notons que l'hypothèse d'homogénéité de la variable Y au sein des intersections sources-zones de contrôle permet de supposer que les données au niveau source-zone de contrôle ($\hat{Y}_{A_{i,k}}$) peuvent être allouées aux intersections sources-carreaux-zone de contrôle proportionnellement à leur aire.

appliqué à $\hat{Y}_{A_{i,k}}$ est donc défini comme le ratio de l'aire de l'intersection A_{ijk} ($|A_{ijk}|$) et de celle de l'intersection $A_{i,k}$ ($|A_{i,k}|$).

La valeur prédite de Y pour la zone d'intersection A_{ijk} , notée $\hat{Y}_{A_{ijk}}$, s'écrit donc :

$$\hat{Y}_{A_{ijk}} = \frac{|A_{ijk}|}{|A_{i,k}|} \times \hat{Y}_{A_{i,k}}$$

Exemple

Cette étape consiste à calculer le nombre de logements au niveau des intersections communes-carreaux-Corine Land Cover ($\hat{Y}_{A_{ijk}}$). Après avoir réparti les logements d'une commune entre les différentes zones Corine Land Cover qui la composent, il s'agit maintenant de redistribuer les logements à l'intérieur de chaque zone Corine Land Cover, c'est-à-dire dans les différents carreaux qui la composent.

On calcule tout d'abord l'aire des intersections communes-carreaux-Corine Land Cover ($|A_{ijk}|$). Puis on calcule le nombre de logements au niveau des intersections communes-carreaux-Corine Land Cover ($\hat{Y}_{A_{ijk}}$) à partir de cette aire, de l'aire des intersections communes-Corine Land Cover ($|A_{i,k}|$) et du nombre de logements précédemment prédit au niveau des intersections communes-Corine Land Cover ($\hat{Y}_{A_{i,k}}$).

Ainsi avec les variables de l'exemple, la formule précédente s'écrit :

$$Nb_logements_{commune \cap carreau \cap CLC_k} = \frac{Aire_{commune \cap carreau \cap CLC_k}}{Aire_{commune \cap CLC_k}} \times Nb_logements_{commune \cap CLC_k}$$

IV.1.2 Etape 2 : Prédiction de Y sur les zones cibles T_j

Rappelons qu'une zone cible T_j (j fixé) est l'union de toutes les zones d'intersection A_{ijk} de cette zone cible avec les sources S_i et les zones de contrôle C_k : $T_j = \bigcup_i \bigcup_k A_{ijk}$.

En conséquence de la propriété d'extensivité, la valeur de Y dans la zone cible T_j (\hat{Y}_{T_j}) s'obtient en sommant la valeur de Y sur les zones d'intersection A_{ijk} ($\hat{Y}_{A_{ijk}}$) pour toutes les sources S_i et les zones de contrôle C_k intersectant la cible T_j , c'est-à-dire en sommant sur i et sur k les $\hat{Y}_{A_{ijk}}$:

$$\hat{Y}_{T_j} = \sum_i \sum_k \hat{Y}_{A_{ijk}}$$

Exemple

Cette dernière étape consiste à calculer le nombre de logements au niveau des carreaux (\hat{Y}_{T_j}), à partir du nombre de logements au niveau des intersections communes-carreaux-CLC ($\hat{Y}_{A_{ijk}}$) pour les communes S_i et les zones Corine Land Cover C_k qui intersectent chaque carreau T_j .

Ainsi avec les variables de l'exemple, la formule précédente s'écrit :

$$Nb_logements_{carreau} = \sum_{communes} \sum_{CLC_k} Nb_logements_{communes \cap carreau \cap CLC_k}$$

Ecriture synthétique

$$\hat{Y}_{T_j} = \sum_i \sum_k \frac{|A_{ijk}|}{|A_{i,k}|} \times \frac{x_{A_{i,k}}}{x_{S_i}} \times y_{S_i}$$

Exemple

Avec les variables de l'exemple, la formule synthétique d'allocation s'écrit :

$$Nb_logements_{carreau} = \sum_{communes} \sum_{CLC} \frac{Aire_{communes \cap carreau \cap CLC_k}}{Aire_{communes \cap CLC_k}} \times \frac{Longueur_routes_{communes \cap CLC_k}}{Longueur_routes_{communes}} \times Nb_logements_{communes}$$

Les figures IV.2, IV.3 et ?? donnent le résultat du carroyage avec la méthode DA2. Sur la dernière figure, on remarque une bonne correspondance entre les prédictions et les zones Corine Land Cover, avec un nombre de logements prédits élevé dans les zones artificialisées et plus faible dans les autres zones.

FIGURE IV.2: Extrait du carroyage DA2 : Nombre de logements calculé au carreau (superposition avec la couche Corine Land Cover)

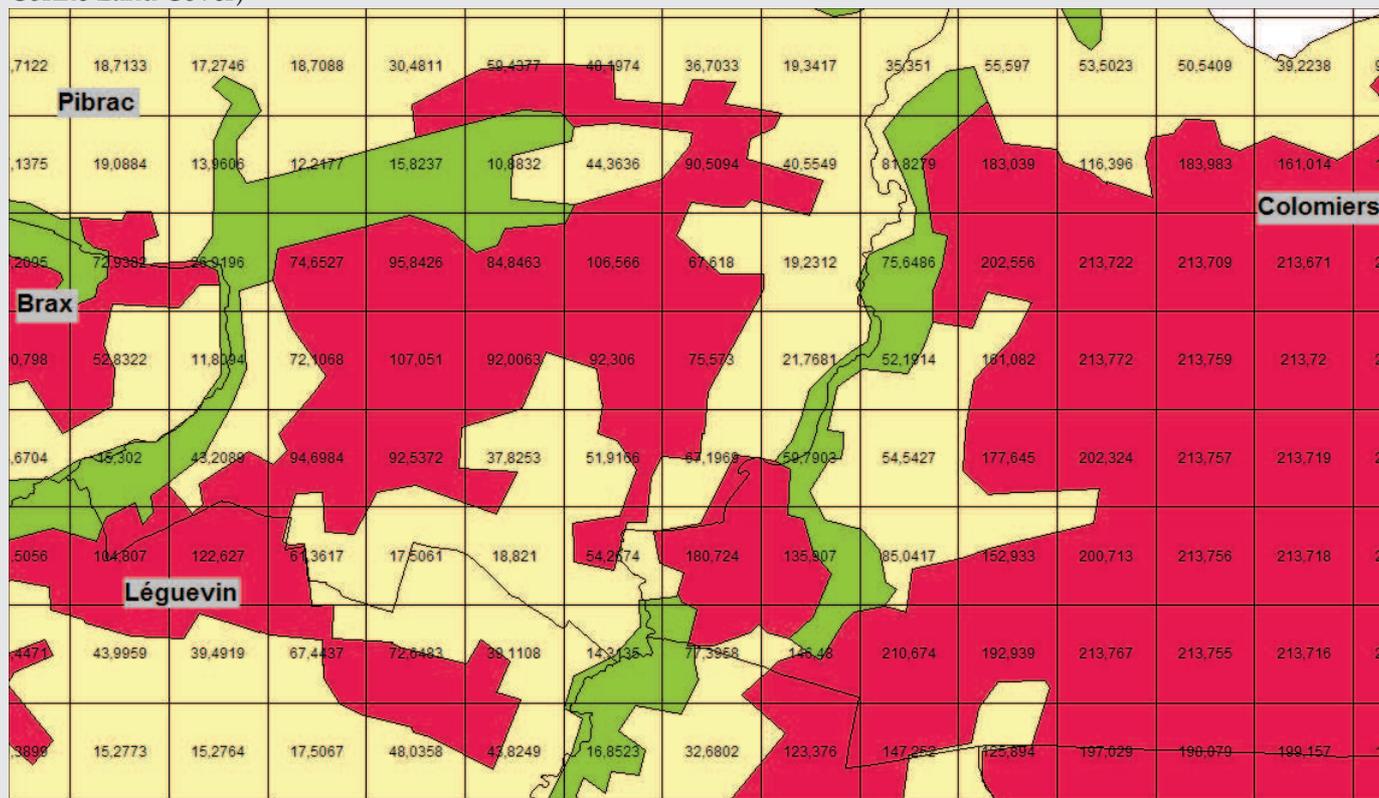


FIGURE IV.3: Carte choroplèthe du nombre de logements au carreau, méthode DA2

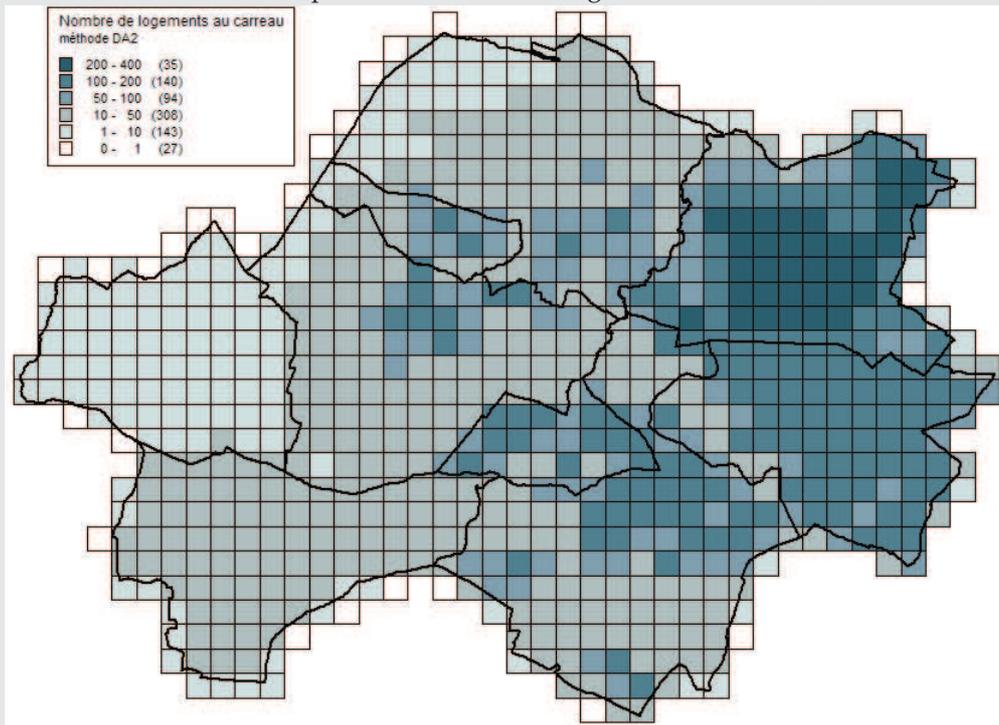
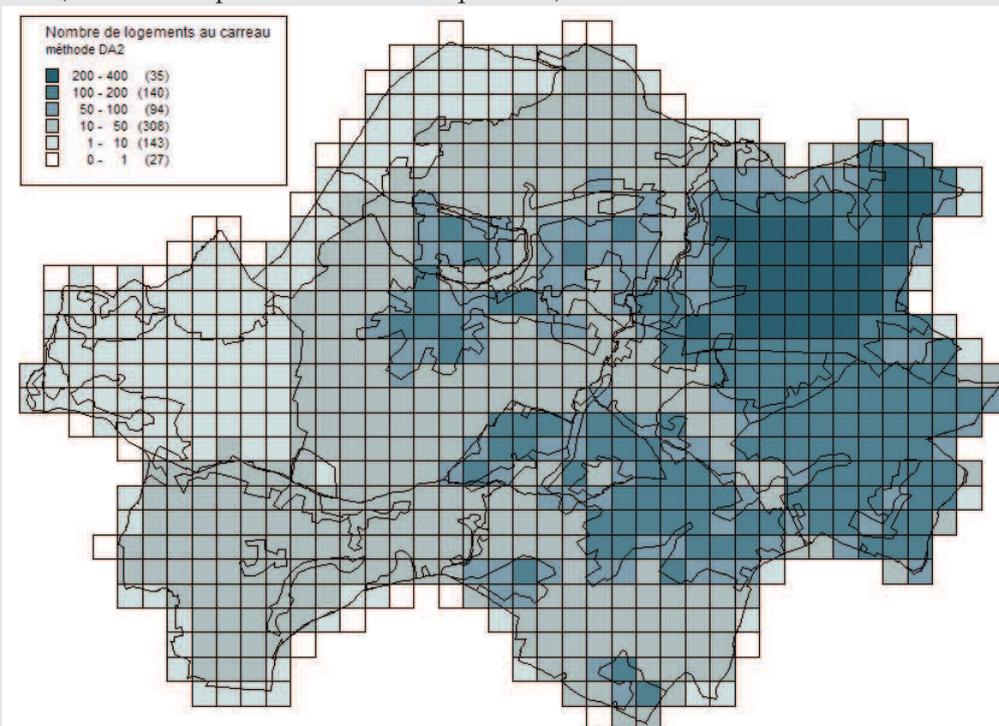


FIGURE IV.4: Carte choroplèthe du nombre de logements au carreau, méthode DA2, avec superposition de la couche Corine Land Cover (matérialisée par les contours uniquement)



IV.2 Exemple de calcul : Détails du calcul pour le carreau S9

Le carreau S9 intersecte les communes de Pibrac et de Brax, deux zones Corine Land Cover sont présentes dans chaque intersection commune-carreau (cf. figure IV.5). Le tableau IV.1 fournit les données nécessaires au calcul du

nombre de logements pour le carreau S9 par la méthode DA2.

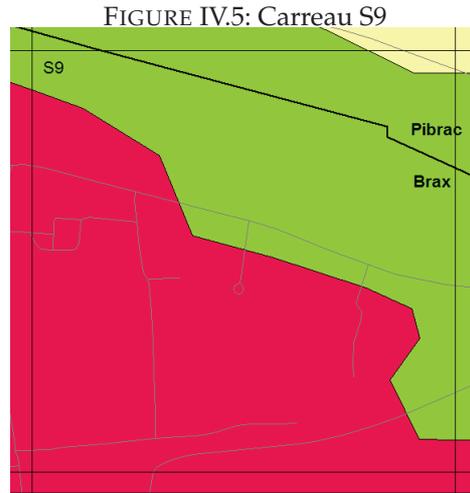


TABLE IV.1: Exemple DA2 : données nécessaires au calcul pour le carreau S9

zone	Nb_logements _{commune}	Aire	Longueur_Route
Pibrac	2 860		90 203,65
Pibrac ∩ CLC ₂		14 891 767,08	35 378,98
Pibrac ∩ CLC ₃		7 479 165,03	5 227,47
Pibrac ∩ CLC ₂ ∩ S9		2 027,25	
Pibrac ∩ CLC ₃ ∩ S9		22 829,33	
Brax	891		18 626,65
Brax ∩ CLC ₁		1 659 738,17	16 102,63
Brax ∩ CLC ₃		1 086 937,07	1 010,89
Brax ∩ CLC ₁ ∩ S9		139 514,98	
Brax ∩ CLC ₃ ∩ S9		85 585,95	

IV.2.1 Etape 1 : Prédiction de Y sur les zones d'intersection A_{ij} ($commune \cap S9$)

Etape 1a : Prédiction de Y sur les intersections sources-zones de contrôle $A_{i,k}$ ($commune \cap CLC_k$) par DAX

On commence par calculer le nombre de logements au niveau des intersections commune-CLC à partir des longueurs de route et du nombre de logements de la commune :

$$Nb_logements_{Pibrac \cap CLC_2} = \frac{Longueur_routes_{Pibrac \cap CLC_2}}{Longueur_routes_{Pibrac}} \times Nb_logements_{Pibrac} = \frac{35\,378,98}{90\,203,65} \times 2\,860 = 1\,121,7275$$

$$Nb_logements_{Pibrac \cap CLC_3} = \frac{Longueur_routes_{Pibrac \cap CLC_3}}{Longueur_routes_{Pibrac}} \times Nb_logements_{Pibrac} = \frac{5\,227,47}{90\,203,65} \times 2\,860 = 165,7423$$

$$Nb_logements_{Brax \cap CLC_1} = \frac{Longueur_routes_{Brax \cap CLC_1}}{Longueur_routes_{Brax}} \times Nb_logements_{Brax} = \frac{16\,102,6}{18\,626,65} \times 891 = 770,2629$$

$$Nb_logements_{Brax \cap CLC_3} = \frac{Longueur_routes_{Brax \cap CLC_3}}{Longueur_routes_{Brax}} \times Nb_logements_{Brax} = \frac{1\,010,89}{18\,626,65} \times 891 = 48,3556$$

Etape 1b : Prédiction de Y sur les zones d'intersection A_{ijk} ($commune \cap S9 \cap CLC_k$) par DAW

On calcule ensuite le nombre de logements au niveau des intersections commune-CLC-carreau à partir du nombre de logements au niveau des intersections commune-CLC calculé précédemment et des aires de ces deux types d'in-

tersections :

$$Nb_logements_{Pibrac \cap S9 \cap CLC_2} = \frac{Aire_{Pibrac \cap S9 \cap CLC_2}}{Aire_{Pibrac \cap CLC_2}} \times Nb_logements_{Pibrac \cap CLC_2} = \frac{2\,027,25}{14\,891\,767,08} \times 1\,121,7275 = 0,1527$$

$$Nb_logements_{Pibrac \cap S9 \cap CLC_3} = \frac{Aire_{Pibrac \cap S9 \cap CLC_3}}{Aire_{Pibrac \cap CLC_3}} \times Nb_logements_{Pibrac \cap CLC_3} = \frac{22\,829,33}{7\,479\,165,03} \times 165,7423 = 0,5059$$

$$Nb_logements_{Brax \cap S9 \cap CLC_1} = \frac{Aire_{Brax \cap S9 \cap CLC_1}}{Aire_{Brax \cap CLC_1}} \times Nb_logements_{Brax \cap CLC_1} = \frac{139\,514,98}{1\,659\,738,17} \times 770,2629 = 64,7471$$

$$Nb_logements_{Brax \cap S9 \cap CLC_3} = \frac{Aire_{Brax \cap S9 \cap CLC_3}}{Aire_{Brax \cap CLC_3}} \times Nb_logements_{Brax \cap CLC_3} = \frac{85\,585,95}{1\,086\,937,07} \times 48,3556 = 3,8075$$

IV.2.2 Etape 2 : Prédiction de Y sur la zone cible T_j (le carreau S9)

$$\begin{aligned} Nb_logements_{S9} &= Nb_logements_{Pibrac \cap S9 \cap CLC_2} + Nb_logements_{Pibrac \cap S9 \cap CLC_3} + Nb_logements_{Brax \cap S9 \cap CLC_1} \\ &\quad + Nb_logements_{Brax \cap S9 \cap CLC_3} \\ &= 0,1527 + 0,5059 + 64,7471 + 3,8075 = 69,2132 \end{aligned}$$

IV.3 Lien entre les méthodes DA2 et DAC

Il est à noter que la méthode dasymétrique avec zones de contrôle et variable auxiliaire (DA2) est un cas particulier de la méthode dasymétrique avec zones de contrôle (DAC, cf. p. 37) où la densité de la variable auxiliaire $X (= x_{C_k} / |C_k|)$ est utilisée à la place de la densité de Y et où on suppose que les zones de contrôle sont imbriquées dans les zones sources ($A_{i,k} = S_i \cap C_k = C_k$). La preuve est donnée dans l'annexe I p.125.

Troisième partie
Autre méthode

Fiche V

Allocation via le centroïde de la source - *Point-in-Polygon (PiP)*

Conditions d'utilisation

variable cible Y :

- extensive ou intensive
- cette méthode donne de bons résultats si le phénomène d'intérêt est concentré près des centroïdes des zones sources.

pas d'information auxiliaire

Principe [7]

Allouer les données d'une source à une cible si la cible contient le centroïde de la source.

Propriété pycnophylactique non satisfaite (cf. p. 7) car cette méthode ne définit pas de valeur de Y sur les intersections A_{ij} mais prédit directement Y sur les cibles T_j .

Ecriture synthétique

$$\hat{Y}_{T_j} = \sum_{z_i \in T_j} y_{S_i} \quad \text{avec } z_i \text{ le centroïde de } S_i \quad \text{si } Y \text{ extensive}$$

Dans cette méthode, on fait comme si les données surfaciques étaient concentrées en un seul point, le centroïde de la zone.

Exemple

Pour illustrer cette méthode, nous chercherons à connaître le nombre de logements au carreau (noté $Nb_logements_{carreau}$) à partir de l'information connue au niveau des parcelles (noté $Nb_logements_{parcelle}$). La variable d'intérêt Y , le nombre de logements, est extensive.

Dans cet exemple, les zones sources sont donc les 50 709 parcelles et les zones cibles les 741 carreaux. La taille des zones sources est inférieure à celle des carreaux, on se trouve donc dans un cas d'agrégation.

Application des notations aux données étudiées

- $S_i = \text{parcelle}$, il y a $I = 50\,709$ parcelles
- $z_i = \text{centroïde}_{parcelle}$
- $y_{S_i} = Nb_logements_{parcelle}$, connue pour i (parcelle) = $1, \dots, 50\,709$
- $Y_{T_j} = Nb_logements_{carreau}$, à calculer pour j (carreau) = $1, \dots, 741$

V.1 Formule d'allocation (cas d'une variable extensive)

En conséquence de la propriété d'extensivité, la valeur de Y dans la zone cible T_j (\hat{Y}_{T_j}) s'obtient en sommant la valeur de Y sur les sources S_i (y_{S_i}) pour toutes les sources dont le centroïde z_i est inclus dans la cible T_j :

$$\hat{Y}_{T_j} = \sum_{z_i \in T_j} y_{S_i}$$

Exemple

On calcule le nombre de logements au niveau des carreaux (\hat{Y}_{T_j}) en sommant le nombre de logements au niveau des parcelles (y_{S_i}) pour les parcelles S_i dont le centroïde est inclus dans le carreau T_j .

Avec les variables de l'exemple, la formule d'allocation s'écrit :

$$Nb_logements_{carreau} = \sum_{\substack{\text{parcelles dont} \\ \text{centroïde inclus dans carreau}}} Nb_logements_{parcelle}$$

Les figures V.1 et V.2 donnent le résultat du carroyage avec la méthode PiP.

FIGURE V.1: Extrait du carroyage PiP : Nombre de logements calculé au carreau

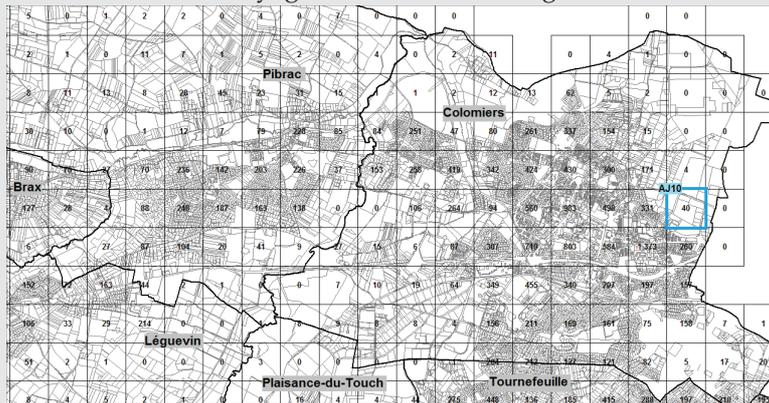
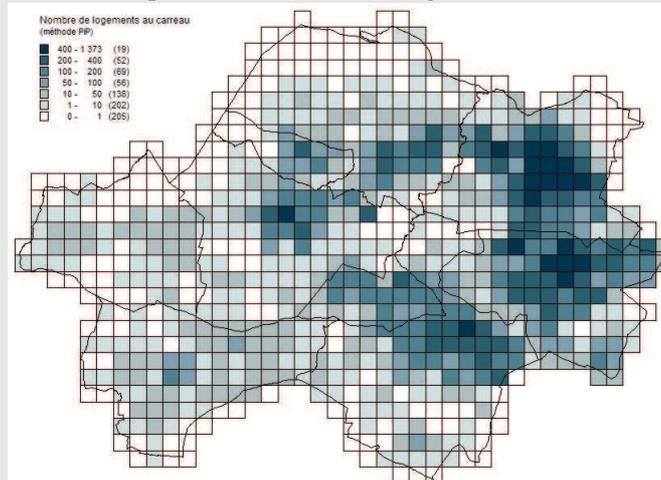


FIGURE V.2: Carte choroplèthe du nombre de logements au carreau, méthode PiP



V.2 Exemple de calcul : Détails du calcul pour le carreau AJ10

Le carreau AJ10 se situe à Colomiers, il a été utilisé comme exemple pour la méthode DAW (cf. I.3.2 p.29). Il intersecte six parcelles avec un nombre de logements supérieur à 0, parmi lesquelles trois parcelles dont le centroïde est inclus dans le carreau (BI0033, BI0035, BI0218 et BH0002, cf. figure V.3). Le tableau V.1 fournit le nombre de logements de chaque parcelle intersectant le carreau AJ10.

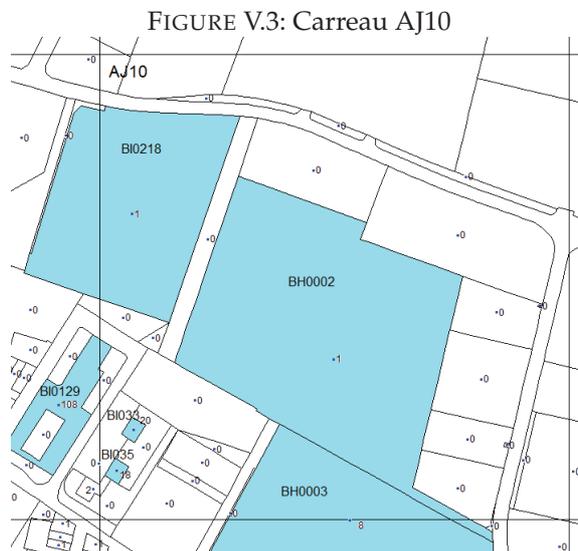


TABLE V.1: Exemple PiP : données nécessaires au calcul pour le carreau AJ10

Zone source S_i parcelle	y_{S_i} $Nb_logements_{parcelle}$
BH0002	1
BH0003	8
BI0033	20
BI0035	18
BI0129	108
BI0218	1

$$\begin{aligned}
 Nb_logements_{AJ10} &= Nb_logements_{BI0033} + Nb_logements_{BI0035} + Nb_logements_{BI0218} + Nb_logements_{BH0002} \\
 &= 20 + 18 + 1 + 1 = 40
 \end{aligned}$$

Nous observons que le nombre de logements prédit pour le carreau AJ10 dans le cas d'agrégation avec la méthode PiP (40) est très proche de celui prédit avec la méthode DAW (44,9).

Bibliographie

Bibliographie

- [1] L.M. Bloom, P.J. Pedler, and G.E. Wragg. Implementation of enhanced areal interpolation using mapinfo. *Computers and Geosciences*, 22(5) :459–466, 1996.
- [2] Van Huyen Do, Christine Thomas-Agnan, and Anne Vanhems. Spatial reallocation of areal data - a review. *Submitted to RERU*, 2013.
- [3] Ian N. Gregory. The accuracy of areal interpolation techniques : standardizing 19th and 20th century census data to allow long-term comparisons. *Computers, Environment and Urban Systems*, 26(4) :293–314, 2002.
- [4] Jeremy Mennis and Torrin Hultgren. Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science*, 33(3) :179–194, 2006.
- [5] Michael Reibel and Aditya Agrawal. Areal interpolation of population counts using pre-classified land cover data. *Population Research and Policy Review*, 26 :619–633, 2007.
- [6] Michael Reibel and Michael E Bufalino. Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A*, 37(1) :127–139, 2005.
- [7] Yukio Sadahiro. Accuracy of count data estimated by the point-in-polygon method. *Geographical Analysis*, 32(1) :64–89, 2000.
- [8] Paul R. Voss, David D. Long, and Roger B. Hammer. When census geography doesn't work : using ancillary information to improve the spatial interpolation of demographic data. CDE Working Paper No. 99-26, 1999.

Annexes

Annexe A

Dictionnaire des données

Cette annexe décrit l'ensemble des données brutes, en donnant le dictionnaire des variables sous forme de tableau et une figure représentant la couche MapInfo (sauf pour la table des logements qui nous a été fournie au format texte).

A.1 Couche des communes

La couche des communes est un fichier MapInfo nommé *9-com-Ouest-Toulouse.tab*. Il comporte 9 observations (communes) et 11 variables.

TABLE A.1: Dictionnaire des variables - couche des communes

Nom variable	Description	Type
id	Identifiant de la commune	qualitative
prec_plani	Précision planimétrique	qualitative (codée numériquement)
nom	Nom de la commune	qualitative
code_insee	Code Insee de la commune (code postal)	qualitative (codée numériquement)
statut	Statut administratif de la commune	qualitative
canton	Canton de la commune	qualitative
arrondisst	Arrondissement de la commune	qualitative
depart	Département de la commune	qualitative (codée numériquement)
region	Région de la commune	qualitative
population	Population de la commune	quantitative
multican	Précise si la commune est située sur plusieurs cantons	qualitative

FIGURE A.1: Couche des communes



A.2 Couche des parcelles

La couche des parcelles est un fichier MapInfo nommé *parcelles_9_com.tab*. Il comporte 53 146 observations (parcelles) et 10 variables.

Cette couche est un extrait de la BD PARCELLAIRE de l'IGN qui fournit l'information cadastrale numérique, géoréférencée et continue sur l'ensemble du territoire français. Elle est réalisée à partir de l'assemblage du plan cadastral dématérialisé.

TABLE A.2: Dictionnaire des variables - couche des parcelles

Nom variable	Description	Type
numero	Numéro de parcelle dans le plan	qualitative (codée numériquement)
feuille	Numéro de feuille	qualitative (codée numériquement)
section	Section cadastrale	qualitative
code_dep	Code du département	qualitative (codée numériquement)
nom_com	Nom de la commune	qualitative
code_com	Code de la commune	qualitative (codée numériquement)
com_abs	Code commune absorbée	qualitative (codée numériquement)
code_arr	Code de l'arrondissement	qualitative (codée numériquement)
extraction	Code de l'extraction	qualitative
recette	Date de recette de la couche	qualitative

FIGURE A.2: Couche des parcelles



A.3 Couche des routes

La couche des routes est un fichier MapInfo nommé *BD-TOPO-routes-9com.tab*. Il comporte 22 367 observations (tronçons de routes) et 34 variables.

TABLE A.3: Dictionnaire des variables - couche des routes

Nom variable	Description	Type
Id	Identifiant du tronçon	qualitative
Prec_plani	Précision planimétrique	qualitative (codée numériquement)
Prec_alti	Précision altimétrique	qualitative (codée numériquement)
Nature	Nature de la voie	qualitative
Numéro	Numéro de la voie	qualitative
Nom_rue_G	Nom du côté gauche de la rue	qualitative
Nom_rue_D	Nom du côté droit de la rue	qualitative
Importance	Importance du tronçon de route pour le trafic routier	qualitative (codée numériquement)
Cl_admin	Classement administratif	qualitative
Gestion	Gestionnaire administratif de la voie	qualitative
Mise_service	Date de mise en service	qualitative
It_vert	Appartenance à un itinéraire vert	qualitative
It_europe	Numéro de l'itinéraire européen	qualitative
Fictif	Indique si la géométrie du tronçon est significative	qualitative
Franchissement	Nature du franchissement	qualitative
Largeur	Largeur de la chaussée	quantitative
Nom_iti	Nom de l'itinéraire	qualitative
Nb_voies	Nombre de voies	quantitative
Pos_sol	Position par rapport au sol	qualitative (codée numériquement)
Sens	Sens de circulation de la voie	qualitative
Inseecom_G	Numéro Insee de la commune à gauche de la voie	qualitative (codée numériquement)
Inseecom_D	Numéro Insee de la commune à droite	qualitative (codée numériquement)
Codevoie_G	Identifiant du côté gauche de la voie	qualitative
Codevoie_D	Identifiant du côté droit de la voie	qualitative
Typ_adres	Type d'adressage de la voie	qualitative
Bornedeb_G	Borne gauche de début de voie	quantitative
Bornedeb_D	Borne droite de début de voie	quantitative
Bornefin_G	Borne gauche de fin de voie	quantitative
Bornefin_D	Borne droite de fin de voie	quantitative
Etat	Indique si le tronçon est en construction	qualitative
Z_ini	Altitude au sommet initial du tronçon	quantitative
Z_fin	Altitude au sommet final	quantitative
Extraction_IGN	Code de l'extraction	qualitative
Recette	Date de recette de la couche	qualitative

La variable la plus importante pour nos exemples est la variable qualitative *Nature*, qui prend pour valeurs dans notre zone d'étude : Bac auto / Bretelle / Chemin Escalier / Piste cyclable / Quasi-autoroute / Route à 1 chaussée / Route à 2 chaussées / Sentier.

FIGURE A.3: Couche des routes



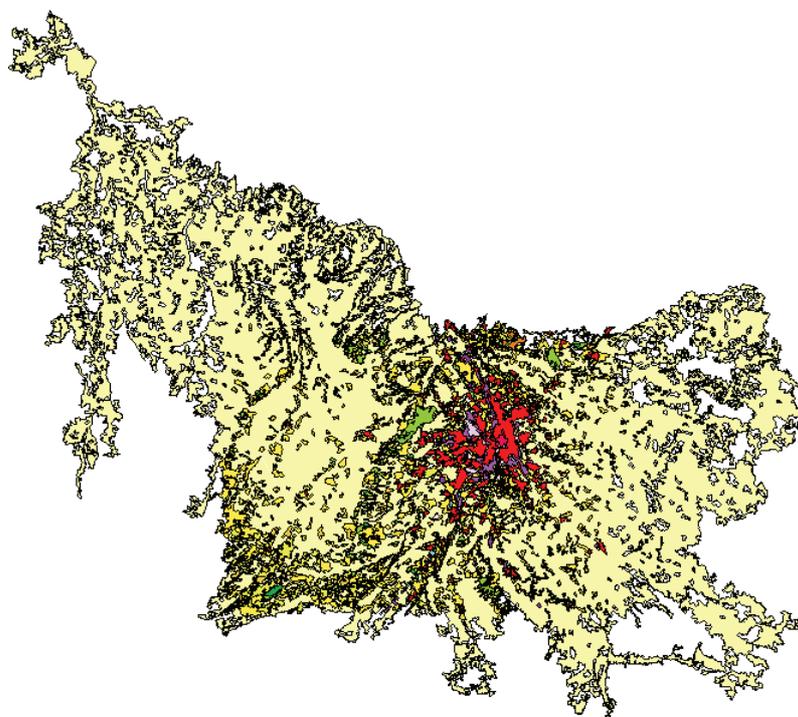
A.4 Couche Corine Land Cover

La couche Corine Land Cover est un fichier MapInfo nommé *CLC06_RGF_autourcommunes.tab*. Il comporte 1 637 observations (zones CLC) et 3 variables.

TABLE A.4: Dictionnaire des variables - couche Corine Land Cover

Nom variable	Description	Type
id	Identifiant de la zone CLC	qualitative
code_06	Code CLC au niveau 3	qualitative (codée numériquement)
area_ha	Aire de la zone CLC (en hectares)	quantitative

FIGURE A.4: Couche Corine Land Cover



A.5 Tableau des logements

Le tableau des logements est un fichier ASCII (texte) nommé *parcelles_9_communes_cle7f1b65.txt*. Il comporte 55 898 observations (parcelles) et 17 variables. Il fournit pour chaque parcelle cadastrale (clé primaire) des informations sur le nombre de locaux (logements ou autres) et l'occupation du sol. Il est extrait de la table des propriétés non bâties des fichiers MAJIC II (données 2009). Le nombre de logements d'une commune est calculable en sommant les logements des parcelles qui la composent.

La source MAJIC

Les fichiers MAJIC (Mise A Jour des Informations Cadastres), ou fichiers fonciers, sont produits par le CETE-Nord-Picardie à partir des fichiers fiscaux fournis par la DGFIP. Ce sont des fichiers anonymisés et redressés à des fins statistiques, de recherche scientifique et historique. Ils contiennent des informations sur les propriétés bâties (immeubles, locaux, maisons, appartements, etc.) et non bâties (terrains, parcelles, etc.).

TABLE A.5: Dictionnaire des variables - table des logements

Nom variable	Description	Type
idpar	Identifiant de la parcelle	qualitative
idcom	Identifiant de la commune	qualitative (codée numériquement)
idcomtxt	Nom de la commune	qualitative
dcntpa	Contenance de la parcelle (en centiares)	quantitative
dcntsol	Surface de type sol (en m^2)	quantitative
dcnttab	Surface de type terrain à bâtir (en m^2)	quantitative
dcntagri	Surface de type agricole (en m^2)	quantitative
dcntbois	Surface de type bois (en m^2)	quantitative
dcntnat	Surface de type naturel (en m^2)	quantitative
dcnteau	Surface de type eau (en m^2)	quantitative
nlocal	Nombre de locaux sur la parcelle	quantitative
nlocmaison	Nombre de locaux de type maison	quantitative
nlocappt	Nombre de locaux de type appartement	quantitative
nloccom	Nombre de locaux de type commercial ou industriel	quantitative
nlocdep	Nombre de locaux de type dépendances	quantitative
y_sref	Coordonnée Y du localisant	quantitative
x_sref	Coordonnée X du localisant	quantitative

