Étude de méthodes de transfert d'indicateurs associés à différents découpages du territoire

Application à la ville de Grenoble : estimation des effectifs de population à l'échelle des quartiers à partir de données connues sur IRIS.

Christine Plumejeaud*, Julie Prud'homme*, Paule-Annick Davoine*, Jérôme Gensel*

** Laboratoire d'Informatique de Grenoble,*
681 rue de la passerelle,
38402 Saint Martin d'Hères
{prenom.nom}@imag.fr

RÉSUMÉ. Lors de l'étude d'un phénomène sur un espace géographique, il arrive souvent que les données dont on souhaite analyser les interactions spatiales soient associées à des maillages hétérogènes, c'est-à-dire des partitions spatiales de natures différentes et non-alignées. Il est alors nécessaire de transférer les données sur un support spatial commun permettant une analyse combinée des variables. Cependant, le choix de la méthode de transfert des données est trop souvent dépendant du niveau de connaissances de l'expert, et la mise en œuvre peut s'avérer difficile. Cet article propose une revue d'un ensemble de techniques pour transférer les données, avec les prémices d'une caractérisation des méthodes de transfert, basée sur un ensemble de critères objectifs. Un cas d'étude sur la ville de Grenoble, qui vise au transfert des effectifs de population des IRIS vers les quartiers, vient illustrer la mise en œuvre de plusieurs techniques de transfert et permet de considérer leur efficacité.

ABSTRACT. It is a common fact that the spatial analysis of any complex phenomenon implies the usage of data issued from heterogenous meshes, saying mis-aligned meshed. Then the combination of data requires the transfer of each dataset into a common spatial support. This is known as the Change Of Support Problem (COSP), and despites the great quantity of reports, it is not easy for non-expert in spatial analysis to understand the choices that are offered. This paper proposes a brief overview of the available methods, giving the premices of a caracterization of each method (type of data, usage constraint, result interpretation), grouping them into family. A use case shows how to put into place two of the exposed methods, and reveals the difficulties related to the result interpretation.

MOTS-CLÉS : Maillages – analyse spatiale – variable statistique – transfert – désagrégation – interpolation - agrégation

KEY WORDS: Mesh – spatial analysis – statistical datasets – downscaling – aggregation - interpolation

---

1. Introduction

En France, les données officielles sur les effectifs de la population sont fournies par l'Institut National de la Statistique et des Etudes Economiques (INSEE). Dans un souci de confidentialité statistique des informations, la Commission Nationale de l'Informatique et des Libertés (CNIL) – autorité de contrôle française en matière de protection des données personnelles – impose que les données communiquées soient agrégées. L'INSEE diffuse ainsi les données récoltées selon un découpage du territoire qui lui est propre, appelé Ilots de Regroupement de l'Information Statistique (IRIS), et qui ne correspond pas forcément à une réalité de gestion territoriale. En revanche, les projets d'aménagement urbain, tels que l'installation de crèches, d'écoles, de services de proximité, ou encore la gestion de l'eau repose sur un découpage de la ville correspondant à une réalité urbanistique, politique ou de gestion. Il est donc nécessaire pour les décideurs de disposer de données de population à l'échelle du zonage correspondant au niveau des prises de décision. Toutefois, le transfert des effectifs de population des IRIS vers un zonage de gestion territoriale n'est pas sans poser des problèmes. En effet, les limites des IRIS ne s'alignent pas toujours, voire rarement, aux limites des zonages retenus, tels que les quartiers.

Le problème de transfert de données d'un maillage dans un autre est largement répandu dans la littérature. Des travaux s'attachant à répertorier les méthodes de transfert sont nombreux [GOO 93], [GOT 02], [NOR 02] ; ces travaux orientent généralement leur discours sur l'étude des effets du changement de forme et d'échelle du support sur la variabilité spatiale des phénomènes, problème connu sous le nom de MAUP (*Modifiable Areal Unit Problem*) [OPE 79], [ESP 06]. Mais il n'existe pas à ce jour de système classifiant et caractérisant ces méthodes à l'intention d'un utilisateur non expert en analyse spatiale, même si un certain nombre de travaux s'orientent dans ce sens [GOO 93], [CHA 97], [EUR 99].

Notre hypothèse principale est qu'il doit être possible d'établir une caractérisation de ces méthodes pour systématiser leur mise en œuvre. Cet article a pour objectif de présenter nos premières réflexions quant à l'élaboration d'une méthodologie favorisant la mise œuvre systématique des différentes méthodes de transfert des données sur des maillages hétérogènes.

Après avoir spécifié les hypothèses et le cadrage théorique sur lesquels se fondent nos propositions, il nous semble important de présenter un état de l'art des principales méthodes de transfert d'indicateurs d'un maillage vers un autre. L'estimation des effectifs de population sur les quartiers de la ville de Grenoble nous permet alors d'illustrer l'utilisation de deux des méthodes citées.

2. Le problème du changement de support

La représentation de l'information géographique est nécessairement portée par un objet géométrique qui en constitue le *support*. Il est courant que le recueil des données relatives à un phénomène précis soit effectué sur un support différent du support sur lequel le phénomène s'exprime. Par exemple, la mesure des températures s'opère sur un support ponctuel alors que le phénomène est continu sur

le territoire, sur un support surfacique. D'autre part, avec la multiplication des sources de données, le chercheur en sciences sociales peut avoir besoin de combiner des données qui sont disponibles sur des supports différents. Dans ces cas, il est indispensable d'harmoniser les supports à des fins d'analyse et/ou d'interprétation de la donnée, et donc de transférer les données (ou indicateurs) dans un support commun. Cette harmonisation est identifiée dans la littérature anglo-saxonne sous l'expression *Change Of Support Problem* (COSP) [ARB 89], qui signifie problème de changement de support en français. Le support peut ainsi être de formes différentes : ponctuel, linéaire, polygonal. Le support polygonal a deux aspects : régulier ou irrégulier. Ainsi, un support matriciel (ou grille) est un support polygonal régulier dans lequel on observe la répétition d'un motif géométrique. *A contrario*, un support polygonal irrégulier est composé de cellules de tailles et de formes variées. Ce type de support, polygonal et irrégulier, est aussi généralement nommé *maillage*.

Bien souvent dans le domaine socio-économique, le territoire est découpé selon un maillage, partition totale de celui-ci en mailles irrégulières [GRA 97]. Un maillage peut être constitué à des fins de gestion et d'administration d'un territoire (le maillage des régions sur le territoire français), d'analyse, selon des réalités du terrain (les limites des bassins versants des rivières en France), ou encore pour organiser le recensement de la population (comme les IRIS de l'INSEE). Chacun de ces découpages est associé à un point de vue sur l'analyse du territoire, et dans cette perspective, un ensemble de données spécifiques est collecté sur ces découpages. Le recueil de données comptabilise des effectifs ou bien des flux sur chaque unité du maillage, pour constituer des variables dites quantitatives absolues. Il apparaît que la combinaison de variables issues de maillages différents produit une information supplémentaire utile pour comprendre l'organisation et l'évolution du territoire dans sa globalité. Ainsi, par exemple, le ratio entre la surface d'espaces verts au sol, telle que fournie par l'Agence Européenne de l'Environnement à partir de grilles régulières sur des unités de 100m par 100m, et la densité de population d'une commune constitue une mesure possible de la qualité de vie dans cette commune (m2 d'espaces verts par habitant).

Certains maillages sont alignés, avec des règles de construction entre eux. Ce sont des maillages hiérarchiques ou emboîtés, comme les départements et les régions sur le territoire français où l'agrégation de départements forme les régions. Lorsque deux maillages présentent un emboîtement strict (toute unité de l'un est composée des unités de l'autre), ils forment une hiérarchie sur laquelle le transfert des données du maillage inférieur vers le maillage supérieur est facilité par les règles simples de l'agrégation géométrique [RIG 95]. Il existe également des maillages discontinus qui désignent des partitions partielles du territoire. En effet, des découpages comme les zones ou bassins d'emploi ne couvrent pas la totalité du territoire français, mais révèlent une réalité d'usage de celui-ci. Enfin, on dit que deux maillages sont non-alignés s'ils ne présentent aucune forme d'emboîtement, c'est-à-dire que leurs frontières se chevauchent, comme ce peut être le cas des départements avec les bassins versants, par exemple. Dans ce travail, nous nous

intéressons exclusivement aux opérations permettant de transférer une variable quantitative (absolue ou relative) entre *deux maillages non-alignés*.

En nous basant sur la terminologie adoptée par [MAR 73], nous parlons de changement de support lorsqu'une variable est connue sur un support *source* et qu'elle est transférée, à l'aide de méthodes plus ou moins complexes, sur un support *cible*. Le problème de transfert de données est connu dans la littérature anglaise sous différents noms : « *spatial rescaling* » [NOR 02], ou « *downscaling* », parce que la principale difficulté vient de la désaggrégation d'une variable sur un support d'échelle plus fine, mais non-aligné avec le support source. Les techniques de transfert portent les noms de « *cross-area aggregation* » [FOT 00], ou « *areal interpolation* » [FLO 89], [FIS 95], ou encore « *polygon overlay problem* ».

L'échelle relative des maillages est un caractère important du transfert : les maillages ne présentent pas tous en moyenne la même taille de maille (ou cellule), ni la même variabilité dans la taille de maille. Lorsque la surface des cellules d'un maillage A est inférieure à celle des cellules d'un maillage B, on dira que A est plus fin que B, et si aucune maille de A ne chevauche une maille de B, alors il y a emboîtement. Ce cas est important car il permet par approximation de reconstruire les valeurs de B par agrégation des valeurs de A.

Les méthodes de transfert de données d'un support à un autre sont donc basées sur la création d'un maillage intermédiaire, inclus dans le maillage cible, lorsque celui-ci est plus fin que le maillage source. Les opérations possibles sont les suivantes :

- la *désagrégation* est une transformation qui permet de ventiler la donnée sur des objets (ponctuels ou polygonaux) de taille plus réduite, complètement inclus dans les mailles du support d'origine, grâce à la fabrication d'un support constituant le plus petit commun dénominateur spatial entre le support source et le support cible, par intersection des deux supports.
- *l'agrégation* de données est une transformation obtenue par le cumul des données selon des mailles incluses spatialement dans des mailles de taille supérieure. Elle peut survenir après une désagrégation pour obtenir une estimation de la variable sur un autre maillage ou bien lors du passage d'une échelle locale à une échelle globale, dans le cas de hiérarchie.
- *l'interpolation* permet, à partir d'un support ponctuel (ou parfois polygonal), de reconstruire une surface de densité ayant pour support un maillage régulier, qui peut être aussi fin que voulu.

L'interpolation s'inscrit généralement dans une séquence d'opérations où la donnée est réduite à un semis de points, puis interpolée, puis réagrégée sur le support final. En effet, il est possible et aisé de résumer un polygone en son centre. La donnée portée par le polygone est alors entièrement transférée au centre, qui peut être le centre de l'enveloppe géométrique (centroïde), ou bien un centre géographique localisé dans la maille, choisi par l'utilisateur pour sa pertinence dans l'étude : la plus grande ville, le point d'altitude le plus bas, etc. La surface de

tendance calculée a souvent pour but de donner une représentation continue de mesures discrètes ou ponctuelles. Cependant, cette surface est nécessairement portée par une discrétisation de l'espace, fine, sur laquelle les éléments peuvent être sommés en fonction de leur inclusion dans le maillage cible, ce qui élargit considérablement le panel des techniques de transfert entre maillages non-alignés.

Le graphe de la figure 1 présente schématiquement ces trois opérations. Les transformations dépendent pour beaucoup de la nature des supports source et cible. Mais, ces transformations ont aussi un effet sur l'analyse qui peut être menée sur les données, en raison du changement de taille et de forme du support. En effet, plus les unités spatiales d'analyse sont grandes, moins la variation est importante entre les valeurs prises par les variables d'une unité à l'autre (effet d'échelle – *scale effect*), et la délimitation des entités spatiales influe sur la « lecture » du phénomène (effet de découpage – *zoning effect*) [OPE 79].

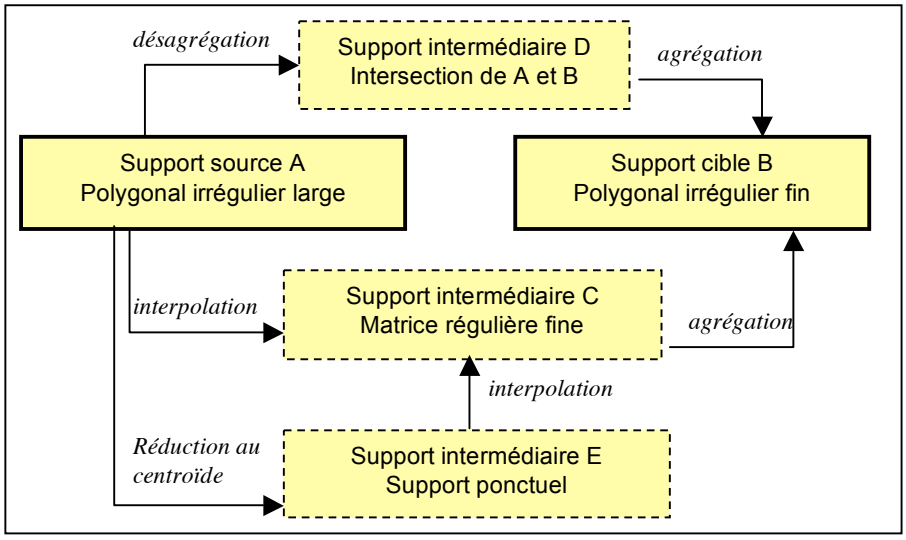

Figure 1. *Graphe des transformations de support pour le transfert des données.*

3. Un aperçu des principales méthodes de « transfert »

Nous présentons ici un aperçu synthétique des différentes familles de techniques de transfert. Cette synthèse est mise en relation avec un ensemble de critères identifiés caractérisant ces méthodes de transfert. Nous avons vu comment l'identification des types de support source et cible amène à choisir un certain type d'opération de transformation. Rappelons toutefois que le type de la variable à transférer joue un rôle : une variable quantitative absolue (un compte d'effectifs sur des unités) ne peut être manipulée de la même manière qu'une variable relative (un ratio), en particulier dans les opérations d'aggrégation. Les taux ne s'additionnent pas pour constituer le taux de la maille supérieure. De même, le transfert s'effectue

dans un contexte, dans lequel on peut choisir de faire intervenir d'autres variables qui ont une influence sur la spatialisation du phénomène : on les appelle variables auxiliaires.

3.1. Opérations d'agrégation

Pour l'aggrégation, nous aurions pu citer les techniques de régionalisation [PUM 05], d'optimisation et de recuit-simulé [OPE 88], ou de clustering. Toutefois ces techniques visent essentiellement à reconstruire un nouveau maillage suivant certaines contraintes posées par l'utilisateur (similarité, homogénéité, continuité, etc.). Or, notre objectif est de transférer une variable vers un maillage cible dont les contours sont définis, et dans ce cadre, la technique *d'agrégation de données matricielles* vers un maillage irrégulier est très opportune. Elle correspond sur la figure 1 à la transformation de C vers B. L'idée est d'agréger des variables connues sur un support matriciel vers un maillage irrégulier. En superposant le maillage cible à la matrice, il est possible de cumuler les effectifs des cellules selon le maillage cible, et de connaître ainsi la valeur de la variable sur le nouveau support, moyennant une marge d'erreur. Cette erreur dépend à la fois de la taille des cellules du maillage intermédiaire, et de la procédure adoptée pour réallouer une cellule à une unité du maillage cible. Par exemple, soit la cellule appartient majoritairement au polygone, dans ce cas on lui affecte l'effectif total porté par la cellule ; soit on réalloue proportionnellement à la surface d'intersection avec le polygone. Pour réduire autant que possible l'erreur, le maillage intermédiaire doit être le plus fin possible. Par ailleurs, une grande part de variabilité dans l'information est perdue puisqu'elle est ré-agrégée dans des unités de plus grande taille et de formes différentes (*scale and zoning effect*). Ce type de méthode est très largement répandu dans le domaine environnemental où les données sur des supports raster (matriciels) abondent, lorsque les données doivent être transférées dans un maillage socio-économique comme la Nomenclature des Unités Territoriales Statistiques (NUTS) [GOM 05].

3.2. Opérations de désagrégation

3.2.1. *Pondération surfacique simple* (*Simple Areal Weighted)*

Cette méthode permet de transférer une variable d'un maillage à un autre sans qu'ils soient emboîtés, par la constrution d'un support intermédiaire *D*, résultant de l'insersection des supports source et cible. Elle suppose l'hypothèse de l'uniformité de densité de la variable dans chaque maille. La ventilation de la variable se fait proportionnellement à la surface d'intersection d'une maille source avec le maillage cible [GOO 80]. Elle est facile à mettre en œuvre et ne nécessite pas un grand échantillon de données. L'algorithme est adopté par la plupart des systèmes d'information géographique (SIG). Par contre, l'hypothèse d'uniformité de densité de la variable est simpliste et ne permet pas de rendre compte des variations locales.

3.2.2. *Pondération surfacique contrôlée par régression* (*Modified Areal Weighted - Regression*)

Cette méthode admet que des relations de corrélation spatiale peuvent exister entre les variables (confère [DRO 06]) : on identifie donc une variable auxiliaire corrélée spatialement avec la variable à transférer, et qui soit connue sur les maillages source et cible. Cette méthode introduit dans la ventilation de la variable sur le support d'intersection *D* une contrainte correspondant à la similarité de répartition entre la variable à transférer et la variable auxiliaire, dite « prédicteur ». Dans une première étape, une étude statistique formalise la ressemblance de dispersion spatiale entre les deux variables sur le support source. La seconde étape résout une équation de régression linéaire entre le prédicteur, dont la dispersion est connue sur le support cible, et la variable à estimer [GOO 93].

L'utilisation d'une variable auxiliaire peut améliorer de façon conséquente les résultats, mais il n'est pas aisé de trouver des données complémentaires qui soient fortement corrélées à la variable étudiée.

3.2.3 *Pondération surfacique contrôlée par zones de contrôle (Modified Areal Weighted – control zones)*

Cette alternative de la pondération surfacique simple se base sur une donnée connue sur un « maillage de contrôle », pour établir des estimations sur le maillage cible. En effet, il est possible de calculer la répartition spatiale de la donnée à transférer à l'aide de la donnée auxilliaire connue sur le maillage de contrôle ou sur des objets géographiques de granularité plus fine que celle du maillage cible. Ce troisième maillage est appelé « zones de contrôle » (*control zones*) par [GOO 89]. On considère que ces zones de contrôle ont une densité constante, [LAN 92].

On considère, en général, que les résultats obtenus par cette méthode sont assez bons. Par exemple, dans le domaine de l'environnement, la tendance actuelle [REI 07] est d'utiliser les données d'image satellitaires (matrices d'occupation du sol) comme zones de contrôle pour le transfert des variables. En effet, ces matrices sont traitées pour reconnaître les types majoritaires d'occupation du sol sur des pixels de 1 ha, ce qui facilite le transfert d'un nombre important de variables : la population est liée aux zones urbaines, la quantité de pesticides aux zones agricoles, etc.

3.3 Opérations d'interpolation étudiées

Les méthodes d'interpolation sont très nombreuses [ARN 00]. [ZAN 05] propose les critères de choix suivants pour les méthodes d'interpolation :

- le type d'algorithme : l'estimation se fait-elle par une fonction mathématique d'ajustement aux données de l'échantillon (méthode déterministe) ou par une méthode stochastique qui utilise une loi probabiliste et qui, ainsi, propose une estimation de son erreur ? ;
- la précision de l'estimation : des algorithmes approximatifs estiment une surface de densité moyenne qui ne conserve pas les valeurs des points de mesure. Les algorithmes exacts conservent les valeurs aux points de mesure ;

- l'extension spatiale de l'estimation qui distingue les méthodes globales des locales ; dans le premier cas, une fonction est ajustée sur la surface d'étude dans sa totalité, alors que dans le second, les estimations sont calculées autour de chaque point de mesure.

Nous en présentons seulement quelques-unes afin d'illustrer la liste de critères présentés ci-dessus, et nous laissons de côté les familles de méthodes suivantes, pour lesquelles nous citons les références :

- les splines [DUB 83],
- les lissages par la méthode des noyaux [BRA 89],
- la triangulation par les polygones de Thiessen (*Triangular Irregular Network*, TIN),
- les arbres hiérarchiques bayésiens, [MUG 00],
- régression par cartes topologiques [BAD 97].

Nous écartons aussi les méthodes de détection d'autocorrélation spatiale par indices globaux (Moran) ou bien locaux (LISA), qui permettent de décrire le jeu de données et d'identifier les zones de ressemblance ou dissimilarité, mais qui ne sont pas à proprement parler des méthodes d'interpolation. Elles peuvent toutefois jouer un rôle dans l'identification de variables auxilliaires et l'identification de schémas spatiaux aidant à l'estimation. Les trois premières méthodes présentées sont déterministes, alors que la dernière est stochastique.

3.3.1 *Surface de tendances (Trend Surface Analysis)*

Les surfaces de tendances sont une interpolation globale polynomiale au sens des moindres carrés. Elle vise à ajuster et à lisser une surface mathématique à partir des valeurs des points échantillonnés. Dans la figure 1, cette méthode correspond au passage de A à E (réduction au centroïde), puis E à C (calcul de surface). La surface créée peut être d'un ordre variable, mais plus l'ordre est élevé, plus l'interprétation des résultats est complexe. Cette méthode est aisée à mettre en œuvre, mais le lissage ainsi effectué peut être excessif et alors masquer des variations locales. Cette méthode n'est pas adaptée si la variable étudiée a une distribution complexe sur le territoire [ZAN 05]. C'est une méthode déterministe, globale, et exacte.

3.3.2 *Moyennes mobiles spatiales - Inverse Distance Weighting (IDW)*

Les filtres spatiaux (*focal functions*) constituent une autre famille de méthodes déterministes, mais locales. L'IDW est un type de filtre spatial, où une grille de carroyage (matrice) aussi fine que possible est appliquée sur l'espace d'étude pour simuler une surface continue. Dans la figure 1, cette opération correspond au passage de A à C, où la valeur des cellules de C prend initialement une valeur proportionnelle à la surface intersectée avec les unités du maillage A. Le calcul de la variable dans chaque cellule de la grille est ensuite pondéré par la valeur des cellules de son voisinage à l'aide d'une fonction décroissante de la distance : ce peut-être une fonction inverse ou bien une exponentielle négative. Le voisinage est défini par l'utilisateur en précisant un rayon de voisinage, ou portée.

La taille du rayon (portée) a un effet sur le niveau de lissage des données : trop grand, on ne voit plus les phénomènes locaux ; trop petit, on les met en exergue. La mise en œuvre nécessite peu d'itérations et repose sur une hypothèse d'auto-corrélation spatiale. La carte résultante est "bruitée" et peut nécessiter une agrégation par classification.

3.3.3 *Interpolation pycnophylactique*

Comme autre type de méthode déterministe, locale, mais cette fois exacte, il existe l'interpolation pycnophylactique, développée par [TOB 79], et qui vise à générer une surface lissée à partir de données connues sur un support polygonal en préservant les effectifs (ou masse) ou le volume sur les régions d'origine.

Elle ressemble aux filtres spatiaux dans sa première étape : une grille de carroyage très fine est superposée à la zone d'étude et les cellules prennent, au départ, la valeur exacte des effectifs de la région à laquelle elles appartiennent (passage de A à C dans la figure 1). Une fenêtre de filtre est déplacée sur la grille et remplace les valeurs de chaque cellule par la moyenne pondérée des cellules voisines (4 ou 8 cellules, au choix de l'utilisateur). Mais spécifiquement, après chaque itération, la méthode pycnophylactique contraint l'ajustement des nouveaux effectifs de manière à garantir la continuité avec les mailles voisines et conserver sur chaque maille de départ la masse (les effectifs) mesurée. Le processus se termine lorsque tout ajustement supplémentaire est plus petit que la tolérance spécifiée par l'utilisateur et que la surface est lissée.

L'interpolation pycnophylactique a l'avantage de conserver les masses, ou effectifs, et d'éviter d'avoir à réduire les mailles irrégulières à leur simple centroïde pour estimer une surface continue. En revanche, c'est un modèle qui gomme les changements brusques (par exemple des poches locales de forte densité de population à l'intérieur d'une maille) [RAS 05], car elle s'appuie sur l'hypothèse d'une progression uniforme de la densité d'effectifs.

3.3.4 Krigeage(s)

Le krigeage est une méthode fréquemment utilisée pour éviter les questions de MAUP et de COSP, alors qu'à l'origine, cette technique, inventée par un ingénieur sud-africain, M. Krige, visait à déterminer la distribution spatiale de minerais à partir d'un ensemble de forages miniers (mesures ponctuelles donc). C'est cependant Matheron [MAT 69] qui a formalisé l'approche en utilisant les corrélations entre les forages pour en estimer la répartition spatiale. Elle modélise la variabilité spatiale d'observations ponctuelles pour créer une surface continue. Le krigeage consiste à déterminer la pondération de chacun des points environnants selon le degré de similarité entre les valeurs. Celle-ci est établie à partir de la covariance entre les points, fonction de la distance entre ceux-ci. Cette modélisation de la similarité spatiale est intégrée dans des opérations d'interpolation spatiale. Le krigeage est ainsi sensible à la distribution spatiale du semis de points, à la portée choisie, et au choix de la fonction d'interpolation.

Il existe plusieurs formes de krigeage qui dépendent de la connaissance que l'utilisateur a des caractéristiques statistiques de la variable étudiée : évolution ou non de la variable avec le temps (stationnarité), connaissance ou non de la moyenne de la variable, pour les plus générales [CRE 93].

Basée sur un algorithme probabiliste, cette méthode fournit les erreurs d'estimations. Par contre, sa mise en œuvre requiert un grand échantillon en entrée (plus de 100 individus) et de nombreux jeux de données ne présentent pas de dépendance spatiale claire. C'est une méthode stochastique, globale et non exacte.

4. Etude de cas : la ville de Grenoble

4.1. Contexte

Concernant la ville de Grenoble, le maillage géographique de référence pour la gestion urbaine ou pour la prise de décision et l'action politique est le quartier. La ville de Grenoble est découpée en une vingtaine de quartiers. Ces derniers correspondent à une réalité urbaine puisqu'ils présentent une certaine homogénéité en termes de bâti, de fonction urbaine, et de profil socio-démographique. En revanche, ce découpage en quartier, créé dans les années 80, ne correspond pas au découpage statistique des IRIS à partir desquels sont rendus disponibles les effectifs de population (voir figure 2).

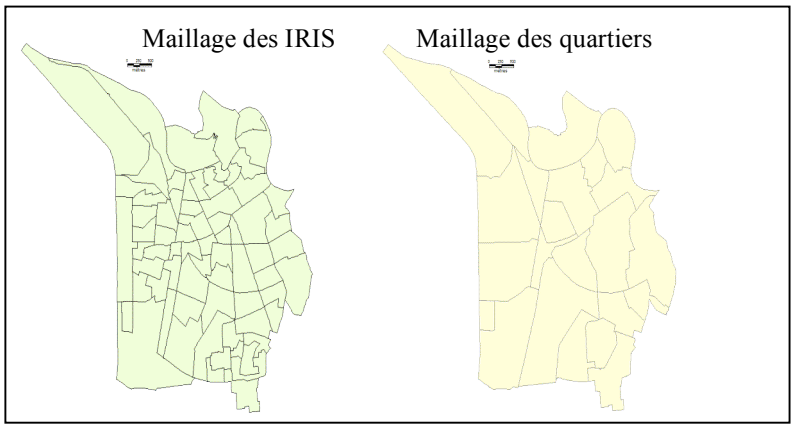

Figure 2. *Différents maillages de la commune de Grenoble.*

Or, la connaissance des effectifs de population par quartier est un problème récurrent lors de la mise en œuvre de projets d'aménagement ou d'action politique. Se pose alors la question de la méthode d'estimation à appliquer. Parmi l'ensemble des méthodes présentées ci-avant, quelles sont les plus adaptées à notre cas de figure ? Sont-elles équivalentes ?

Dans le cas de Grenoble, les supports de données (IRIS et Quartier) sont des maillages de type polygonal irrégulier, de taille variable et non-alignés. Pour

contourner le problème, et transférer l'information portée par les IRIS à l'échelle des quartiers, les services en charge de la gestion urbaine grenobloise utilisent fréquemment la méthode d'agrégation par appartenance géométrique majoritaire des IRIS aux quartiers. Les effectifs d'un IRIS sont ainsi attribués au quartier dans lequel il s'inscrit, ou qu'il recouvre par la majorité de sa surface. Les effectifs des quartiers sont obtenus par le cumul des effectifs des IRIS respectant cette règle. Le quartier « Capuche » (le quartier central) en est un exemple (voir figure 3). Il est composé de plusieurs IRIS partiels et de ce fait, avec cette méthode, l'estimation de la population du quartier inclut la population de bâtiments à fort potentiel de population qui sont situés dans des quartiers voisins. Ceci augmente de manière conséquente les estimations des effectifs de population du quartier.

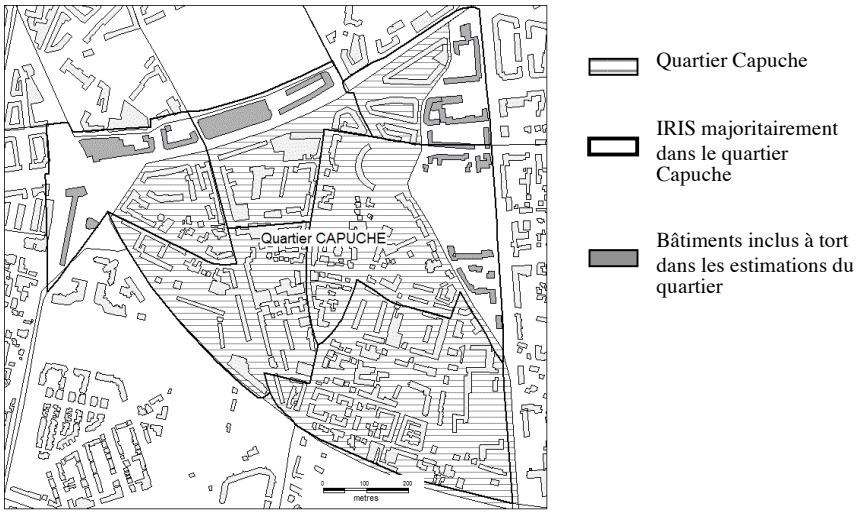

Figure 3. *Surestimation des effectifs de population avec la méthode d'intersection majoritaire pour le quartier de la Capuche à Grenoble.*

Les limites de la méthode de transfert par intersection majoritaire ont été clairement établies. Elle attribue fréquemment des effectifs à des supports cibles qui ne devraient pas en bénéficier. Il semble donc important d'utiliser une autre méthodologie pour obtenir une estimation de la population plus précise à l'échelle des quartiers. La méthode de désagrégation, selon une pondération surfacique contrôlée par zones de contrôle, est retenue.

4.2. Le bâti comme maillage intermédiaire pour l'estimation des effectifs de population sur des maillages hétérogènes

Considérant que les quartiers grenoblois présentent une certaine homogénéité quant à leur morphologie urbaine, nous avons émis l'hypothèse que les données sur les bâtiments pouvaient constituer une variable auxiliaire permettant la pondération des estimations. Les données sur le bâti sont issues de la BDTopo de l'IGN, et les

batiments sont classifiés suivant différentes catégories (habitation, industriel, etc.). Nous ne retenons que la classe dite « habitat ». En fait, l'idée est de calculer un potentiel de population par bâtiment à partir des données disponibles relatives à la hauteur des bâtiments d'habitation et à la géométrie du bâti au sol. Les traitements préalables sont effectués sous MapInfo.

L'évaluation de la population à partir de la hauteur des bâtiments nécessite de :
- connaître la répartition des bâtiments d'habitation en fonction de leur hauteur
- calculer la surface développée des bâtiments qui servira de pondération pour la désagrégation des effectifs de population issus des IRIS.

La caractérisation des bâtiments (identification des bâtiments d'habitation et hauteur moyenne) repose sur la méthode employée par le CERTU de Rouen [CER 05] qui classe les bâtiments en quatre catégories, que nous retrouvons dans la structure urbaine de Grenoble (voir figure 4) :

- Centre-ville : des étages hauts (4 mètres) et des commerces au rez-de-chaussée ;
- Grand collectif : des bâtiments hauts (plus de 15 mètres) ;
- Collectif moyen : des bâtiments entre 9 et 15 mètres ;
- Individuel : des bâtiments de moins de 9 mètres.

Figure 4. *Typologie des bâtiments d'habitation à Grenoble (extrait).*

La figure 5 représente la démarche pour la mise en œuvre de la méthode retenue pour l'estimation de la population grenobloise au niveau des quartiers, en prenant comme support d'origine, les IRIS.

À partir du nombre d'étages, nous pouvons calculer la surface développée des bâtiments $Sdev_{bat}$ (Nombre d'étages par surface au sol). Cette variable sert de coefficient de pondération pour la désagrégation des effectifs de population des IRIS (pop_{IRIS}) sur les bâtiments. Une fois la population de chaque bâtiment estimée (pop_{bat}), on peut sommer la population des batiments inclus dans les quartiers et obtenir une estimation de la population (pop_Q) par quartier.

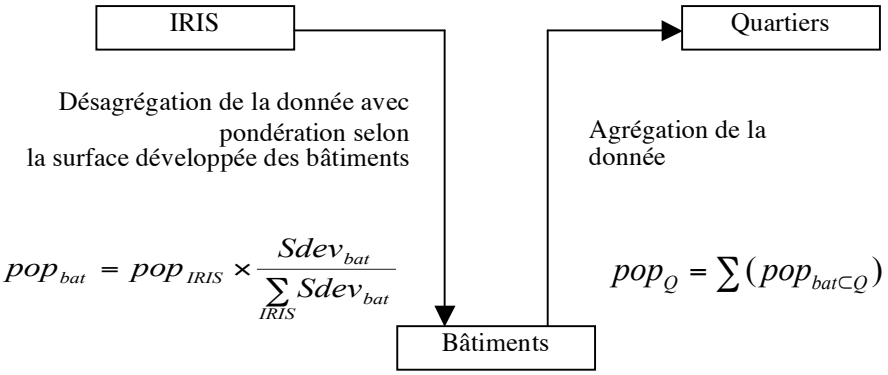

Figure 5. *Démarche d'estimation des effectifs de population de la ville de Grenoble selon la méthode de désagrégation avec pondération par zone de contrôle.*

4.3 *Résultats et analyse*

La carte qui suit (voir figure 6) propose une représentation des résultats d'estimation d'effectifs de population obtenus selon les méthodes de transfert de la donnée décrites précédemment (la désagrégation surfacique majoritaire) - notée *Smaj* - et selon la pondération surfacique avec zone de contrôle (le bâti) - notée *Sbat*

On remarque que les estimations selon les deux méthodes sont proches pour certains quartiers mais pour d'autres, il y a une grande différence (plus de 3500 habitants d'écart pour les plus importants). Le quartier de La Capuche (le quartier central) que nous avons cité en exemple pour les problèmes de non-superposition des limites des IRIS avec les limites des quartiers, fait partie des quartiers concernés par l'écart entre les estimations (près de 3800 habitants de différence entre les deux). Dans le quartier de la Villeneuve 2, un écart important (1106 habitants) existe aussi entre les résultats des deux méthodes, mais cette fois-ci en faveur de la méthode de désagrégation par pondération surfacique utilisant le bâti, ce qui semble très plausible, étant donné que ce quartier possède une forte concentration de logements.

Pour contrôler la qualité des estimations, nous ne disposons des données INSEE de recensement qu'au niveau d'un quartier, la Villeneuve 2. La comparaison avec ces données montre que les résulats par désagrégation surfacique majoritaire sont plus proches de la réalité (qui est 17440 habitants dans le quartier de la Villeneuve 2). Ce résultat est expliqué par une étude approfondie, qui fait alors apparaître que l'utilisation de la forme morphologique des batiments au sol ne donne pas forcément un bon estimateur de la surface développée des bâtiments dans ce quartier d'architecture moderne. En effet, à la Villeneuve 2, l'image satellitaire, à partir de laquelle le bâti a été dessiné, prend en compte les terrasses, balustrades, et les nombreux décrochements que présentent les bâtiments, et augmente donc artificiellement la surface développée des bâtiments. Dans ce cas, la connaissance du nombre de logements par batiment aurait pu être une meilleure variable auxiliaire.

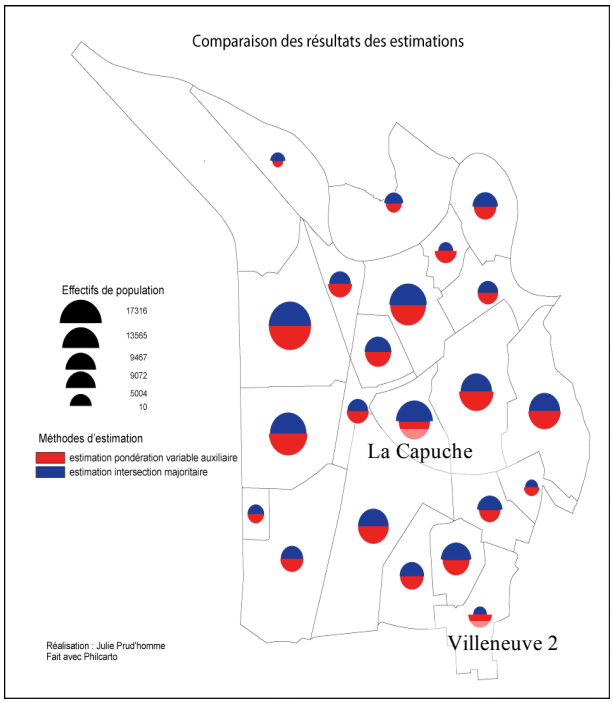

Figure 6. *Cartographie des estimations comparées d'effectifs de population à Grenoble : SMaj (demi cercle supérieur) et SBat (demi cercle inférieur).*

Notre cas d'étude révèle que le transfert basé sur une pondération avec variable auxilliaire sur une zone de contrôle n'est pas forcément plus proche des chiffres d'un recensement exact sur *toutes* les mailles, contrairement à ce qui était attendu théoriquement. Ce phénomène est corroboré par les conclusions d'autres travaux concernant l'emploi de données d'usage de sol issues d'images satellitaires [SCH 06], indiquant que les données satellitaires ne peuvent pas systématiquement se substituer à des données collectées sur le terrain.

5. Conclusion et perspectives

Ce travail a présenté une rapide mise en perspective des différentes méthodes disponibles lorsqu'on souhaite combiner des données issues de supports spatiaux hétérogènes. Ce travail rappelle que le problème est connu dans la littérature sous le nom de changement de support, et il dessine un premier schéma pour le transfert de données. Les résultats d'un transfert dépendent à la fois de la méthode employée, et de la qualité des données utilisées. En effet, la méthode, par son mécanisme de fonctionnement propre, peut induire une transformation de la donnée. La qualité des données utilisées interfère aussi sur l'exactitude et la fiablité des résultats. Un cas d'étude sur la ville de Grenoble vient illustrer les difficultés auxquelles un utilisateur

peut être confronté, et confirme l'aspect très exploratoire (au sens essais/erreurs) que la démarche de transfert implique. Notamment, cette étude montre que la réalisation d'un système automatisant le transfert de variables entre maillages non-alignés doit aussi intégrer la modélisation des erreurs susceptibles de se propager le long de la chaîne de transformation. D'autre part, le choix des variables auxiliaires est délicat car il repose sur une forte part d'expertise : il se pose la question de l'intégration de cette expertise dans le processus d'automatisation.

Il existe un certain nombre de logiciels et de librairies de calcul permettant à un utilisateur averti d'opérer des transferts par diverses méthodes. Mais, ces outils sont rarement connectés directement à une base de données spatiale contenant des données collectées sur des maillages hétérogènes. À terme, notre ambition est de faciliter la démarche du transfert pour un utilisateur non-expert via l'utilisation d'une modélisation adéquate du contexte du transfert (qualité des données, contraintes, types d'opérations). Nous imaginons un système reposant sur une base de données enrichie avec des informations descriptives sur les variables et les supports de données (méta-données), et qui soit connecté à une base de connaissance modélisant les opérations de transfert, avec leurs contraintes respectives.

6. Bibliographie

[ARB 89] ARBIA G. "Statistical Effect of Data Transformations: A Proposed General Framework," in *The Accuracy of Spatial Data Bases*, eds. M. Goodchild and S. Gopal, 1989, London: Taylor and Francis, pp. 249–259.

[ARN 00] ARNAUD, M., EMERY, X., *Estimation et interpolation spatiale : méthodes déterministes et méthodes géostatiques*, Paris, Hermès, 2000, 221 p.

[BAD 97] BADRAN, F., DAIGREMONT, P., and THIRIA S., « Régression par carte topologique »,. In: *Statistiques et méthodes neuronales*, eds. S. Thiria et al., 1997, pp. 207-222.

[BRA 89] BRACKEN, I., & MARTIN, D. The generation of spatial population distributions from census centroid data. *Environment and Planning A*, 1989, 21, 537–543.

[CER 05] CERTU, Méthodes d'estimations de population, Lyon, 2005, 62 p.

[CHA 97] CHARRE J., DUMOLARD, P., LE BERRE, I. *Initiation aux pratiques statistiques en geographie*, 1997 Paris: Masson

[CHA 08] CHAUVET, P., *Aide-mémoire de géostatistique linéaire*, Paris, Mines Paris - Les Presses ParisTech, 2008, 312 p.

[CRE 93] CRESSIE N., *Statistics for spatial data*, John Wiley and Sons, New-York 1991, édition révisée 1993.

[DRO 06] DROESBEKE, J-J., LEJEUNE, M., SAPORTA, G., *Analyse statistique des données spatiales*, Paris, Technip, 2006, 468 p.

[DUB 83] DUBRULE, O. « Two methods with differents objectives : splines and kriging ». *Mathematical geology*. 1983 Vol. 15, pp. 245-255.

[ESP 06] ESPON 3.4.3, The modifiable areas unit problem, Luxembourg, Final report, 2006, 199 p.

[EUR 99] EUROSTAT European Commission Statistical Office – EUROSTAT, GIS Application Development, Final Report, 1999, 74 p.

[FLO 89] FLOWERDEW, R. & GREEN, M. Statistical methods for inference between incompatible zonal systems, in M. Goodchild & S. Gopal, eds, '*The accuracy of spatial data bases*', 1989 Taylor and Francis, London, pp. 239–247.

[FIS 95] FISHER P.F. and LANGFORD M., Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation, in *Environment and Planning A*, vol. 27, 1995, pp. 211-224.

[FOT 00] FOTHERINGHAM, A.S., BRUNSDON, C., and CHARLTON M., *Quantitative Geography*, Sage, London, 2000, pp. 59-60

[GOM 05] GOMEZ O, PARAMO F., The Land and Ecosystem Accounting (LEAC) methodology guidebook, Internal report, July 2005, http://dataservice.eea.europa.eu/download.asp?id=15490

[GOO 80] GOODCHILD, M.F., LAM, N., « Areal interpolation : a variant of the traditionnal spatial problem », *in Geo-processing*, 1980, Vol. 1, pp. 297-312

[GOO 93] GOODCHILD, M.F., ANSELIN, L., DIECHMANN, U., « A general framework for the areal interpolation of socio-economic data », *in Environment and Planning A*, 1993, 25, pp. 383-397

[GOT 02] GOTWAY, C., YOUNG, L, « Combining incompatible spatial data », *in Journal of the American Statistical Association*, 2002, vol. 97, n°458

[GRA 07] GRASLAND, C., « A la recherche d'un cadre théorique et méthodologique pour l'étude des maillages territoriaux » Entretiens Jacques Cartier, « Les découpages du territoire », Lyon, décembre 2007

[LAN 92] LANGFORD, M., UNWIN, D.J., « Generating and mapping population density surface within a GIS », *in The Cartographic Journal*, 1992, 31, pp. 21-26

[MAR 99] MARCEAU, D, « The scale issue in social and natural sciences », *in Canadian Journal of Remote Sens*, 1999, vol. 25, n°4, pp. 347-356

[MAR 73] MARKOFF, J., and SHAPIRO, G. "The Linkage of Data Describing Overlapping Geographical Units," *Historical Methods Newsletter*, 1973, Vol. 7, pp. 34–46.

[MAT 63] MATHERON, G., "Principles of geostatistics", *Economy geology,* 1963, n°58, pp. 1246-1266.

[MUG 00] Mugglin, A.S., Carlin, B.P., Gelfand, A.E., Fully model based approaches for spatially misaligned data, *Journal of the American Statistical Association*, 2000, Vol. 95, pp. 877-887

[NOR, 02] Nordhaus, W.D., « Alternative approaches to spatial rescaling », 2002 (Yale University, New Haven, CT.)

[OPE 79] OPENSHAW, S., and TAYLOR, P., "A Million or so Correlation Coefficients," in Statistical Methods in the Spatial Sciences, 1979, ed. N. Wrigley, London: Pion, pp. 127–144.

[OPE 88] Openshaw, S., « Building an automated modelling system to explore a universe of spatial interaction models », *in Geographical Analysis*, 1988, 20, p.31-46

[PUM 05] PUMAIN D., SAINT-JULIEN, T., *L'Analyse Spatiale, 1. Localisations dans l'espace*, Paris, Armand Colin, collection Cursus, 2005, 167 p.

[RAS 01] RASE W.D., « Volume-preserving interpolation of a smooth surface from polygon-related data », *in Journal of Geographical Systems*, 2001, 3, p.199-213

[REI 07] REIBEL M. & AGRAWAL A., "Areal Interpolation of Population Counts Using Pre-classified Land Cover Data," *Population Research and Policy Review*, 2007, Springer, vol. 26, n°5, pp. 619-633

[RIG 95] RIGAUX, P. et M. Scholl, « Multi-scale partitions: application to spatial and statistical databases », 1995, in *SSD*, pp. 170–183.

[SCH 06] SCHMIT, C., ROUNSEVELL M.D.A., LA JEUNESSE, I., The limitations of spatial land use data in environmental analysis, *Environmental Science & Policy*, Vol. 9, No. 2. (April 2006), pp. 174-188.

[TOB 79] TOBLER, W. A. (1979), 'Smooth pycnopylactic interpolation for geographical regions', *Journal of the American Statistical Association* Vol. 74, pp. 519–530.

[YOO 06] YOO, E.H., KYRIAKIDIS, P.C., « Area-to-point Kringing with inequality-type data », *in Journal of Geographical Systems*, 2006, 8, pp. 357-390

[ZAN 05] ZANINETTI, J.M., *Statistique spatiale, méthodes et applications géomatiques*, Paris, Hermès, 2005, 320 p.