

LE CARROYAGE ET LA GRILLE

MOYENS D'ANALYSE SPATIALE QUANTITATIVE

Philippe Latour ()*

() Gérant de la société Spatialist*

Introduction

Les systèmes d'Information Géographiques (S.I.G.) qui se sont largement démocratisés ces dernières années ne sont encore très largement utilisés que pour créer des illustrations cartographiques de phénomènes variés sans recours à une réelle base méthodologique.

La grande difficulté de l'analyse spatiale consiste à associer des données de toutes sortes à des objets informatiques souvent issus des découpages administratifs et à les représenter sous une forme cartographique la plus rigoureuse possible.

De par leur hétérogénéité en dimensions et formes, ces découpages administratifs constituent une référence géographique très complexe à gérer et tout à fait insatisfaisante si l'on souhaite dépasser le stade de la représentation cartographique visuelle purement qualitative.

Dans le cadre de la gestion de données localisées, le carroyage d'un territoire présente plusieurs avantages immédiats sur les découpages traditionnels :

- Stabilité des unités géographiques de base auxquels sont associées les données, ce qui autorise des analyses historiques fiables
- Homogénéisation des analyses sur des territoires issus de cultures et de systèmes politiques différents (Ref : travaux de l'European Forum for Geostatistics.)
- Simplicité des unités de découpage qui de par leur nature (un carreau) ne peuvent présenter les anomalies couramment constatées avec les communes ou les Iris (morcellements, hétérogénéités...)

Les travaux de recherche spatiale, qui ont pour but la modélisation de phénomènes démographiques, sociologiques, économiques..., se sont toujours heurtés aux problèmes de découpage de l'espace qu'ils ont contourné avec l'usage de systèmes de grille (Loësh, Christaller, Appelbaum, Bertin...).

La grille, une fois résolus les problèmes d'affectation des données aux cellules qui la composent, se comporte comme une matrice mathématique dans laquelle il est possible d'effectuer tous types de calculs avec des opérateurs de type arithmétiques, statistiques ou booléens ; ceci autorise la pratique d'analyses quantitatives à partir de données spatialisées.

1. Les problématiques de l'analyse spatiale

Outre les débats entre cartographes et géographes sur la représentation de phénomènes spatiaux et les travaux approfondis de recherche de R.Brunet et J.Bertin sur les chorèmes et la sémiologie graphique, les analystes sont confrontés à des problèmes pratiques pour traiter, représenter, interpréter des phénomènes qui disposent d'une composante spatiale.

Dans un article de 2 pages, intitulé « La cartographie entre science, art et manipulation » paru dans le Monde Diplomatique de février 2006, Ph.Rekacewicz, cartographe bien connu, ose affirmer : « La carte n'offre... que ce que le cartographe (ou son commanditaire) veut montrer. », et plus loin « Elle ne donne qu'une image tronquée, incomplète, partielle, voire trafiquée de la réalité. »

Le développement des Systèmes d'Information Géographique (S.I.G.) utilisés par des personnes qui ne maîtrisent souvent pas les règles de base de l'analyse spatiale ni les rudiments statistiques indispensables, aboutit à la production d'une profusion de représentations cartographiques dont l'expérience a démontré qu'elles étaient très souvent fausses.

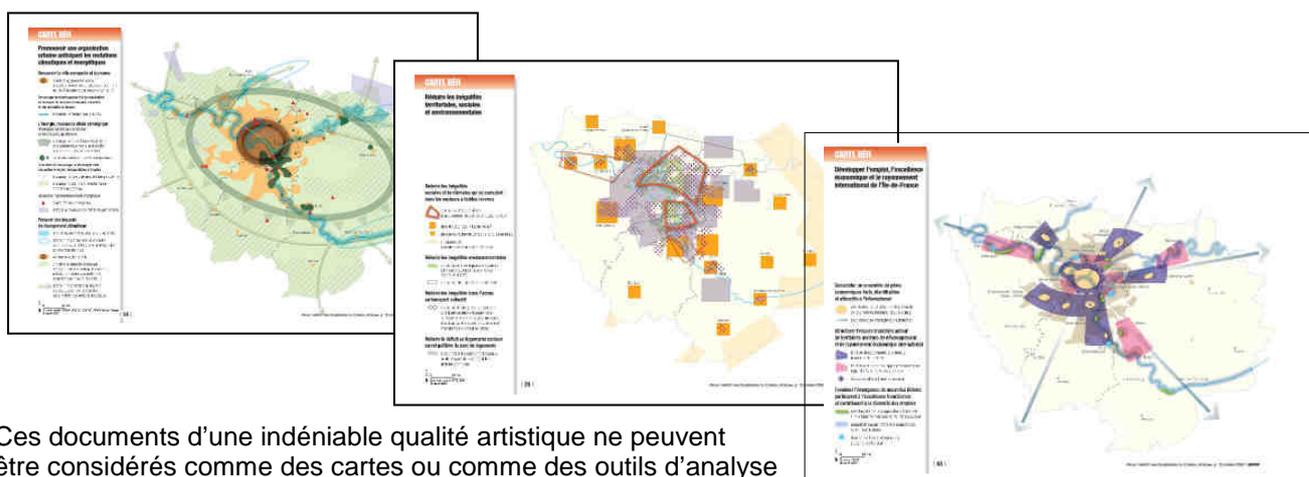
1.1. Les S.I.G.

Les Systèmes d'Information Géographique ont été principalement conçus et sont encore utilisés à plus de 80% pour gérer des implantations d'aménagement de l'espace des collectivités territoriales (planifier et gérer par exemple la voirie, les réseaux téléphoniques, électriques, gaziers...).

Leur utilisation dans le domaine de l'aide à la décision (le géomarketing, entre autres) a largement développé la profession des géomaticiens dont l'expertise se limite malheureusement souvent à savoir manipuler l'un ou l'autre système appartenant à la poignée de S.I.G. couramment commercialisés sur le marché.

Nous donnerons deux exemples de ce type de productions ci-après :

- « Cartes » accompagnant le Schéma Directeur de la région Ile de France, réalisées par l'IAURIF¹ en 2007.

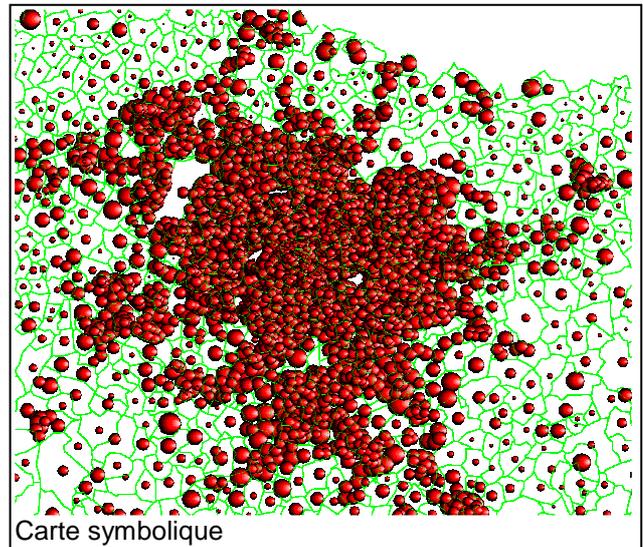
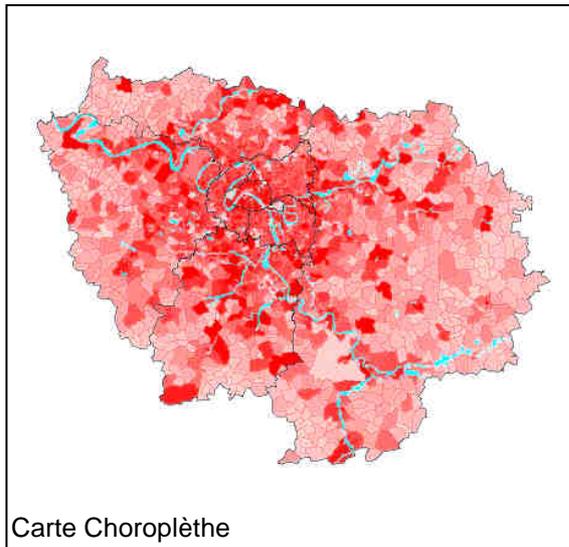


Ces documents d'une indéniable qualité artistique ne peuvent être considérés comme des cartes ou comme des outils d'analyse ou d'aide à la décision. Les élus qui en étaient destinataires ont avoué pour la plupart leur totale incompréhension de la signification de ces beaux dessins.

¹ Institut d'Aménagement et d'Urbanisme : bureau d'études dont la mission essentielle est de réaliser des études destinées aux élus du Conseil Régional

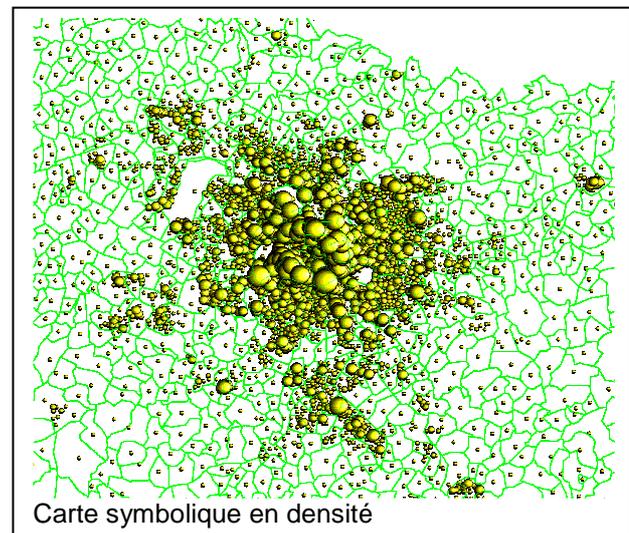
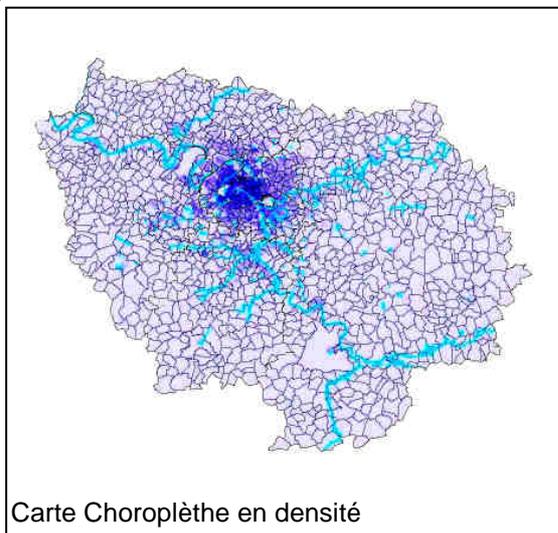
- Représentation de la population des Iris en Ile de France

Si l'on affecte les valeurs brutes de population aux objets géographiques qui constituent la carte de l'Ile de France à l'Iris, nous obtenons les représentations suivantes :



Les territoires (Iris) correspondant à Etrechy (6.206 h.) et Magny en Vexin (5.586 h.) semblent être plus peuplés que Necker 10 avec 5.575 h. L'erreur commise tient au fait que les valeurs ont été chargées sans tenir compte des diversités de taille des territoires concernés (Necker 10 : 0,051 km², Etrechy 5,6km² et Magny 5,44 km²) et que l'on ne peut comparer des territoires dont la superficie varie dans un rapport de 1 à 100.

La seule représentation qui autorise les comparaisons implique un chargement en densité du phénomène étudié² :

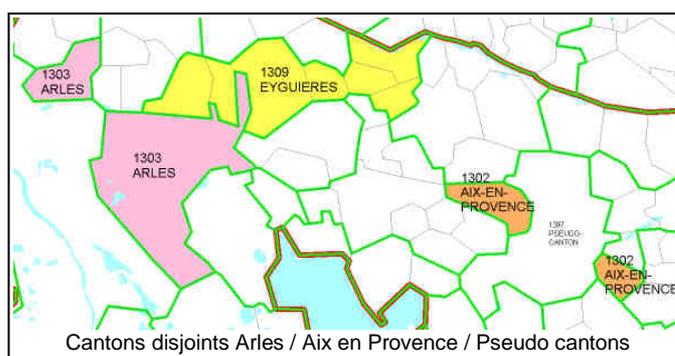
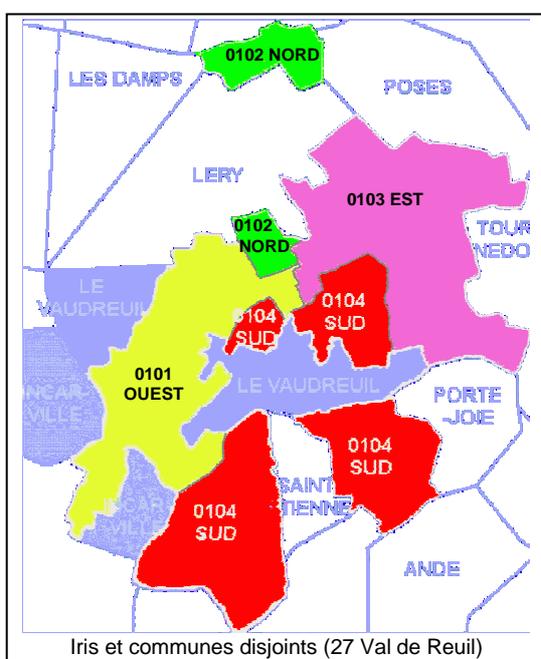
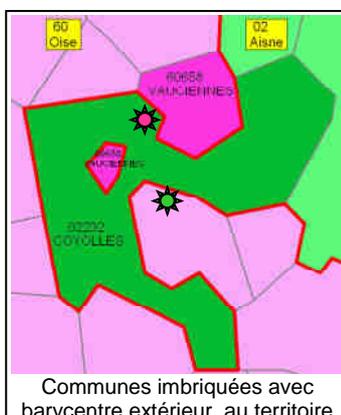
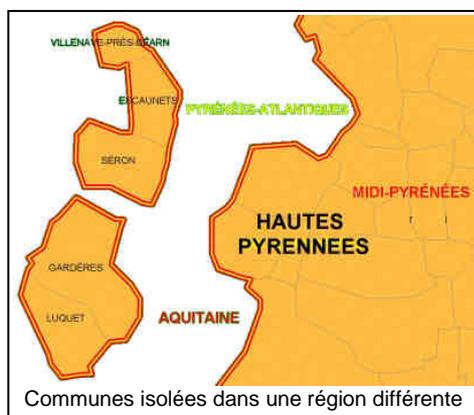


Pratiquement nous avons pu constater que plus de 80% des travaux réalisés en cartographie thématique ne tiennent pas compte des densités et devons par conséquent conclure que les représentations obtenues sont fausses dans la même proportion.

² Toutes les valeurs brutes (populations, chiffre d'affaires, doivent être chargées en densité, seuls les indices, les moyennes ou les % doivent être affectés à des territoires sans tenir compte de leur superficie.

1.2. Les découpages administratifs

Hérités de l'histoire les découpages du territoire présentent des anomalies géométriques dont on trouvera ci après quelques exemples :



Cette forte hétérogénéité des découpages se manifeste non seulement dans leurs formes géométriques, mais aussi dans leurs superficies : Le plus petit Iris en France est Saint Meri 5 (à Paris 4^e) avec 0,33 ha et le plus grand est Saintes Maries de la Mer avec 37.339 ha.

Le rapport entre les deux est de 1 / 110.000, soit le même qu'entre la république de Saint Marin en Italie et la Chine toute entière.

Si on tente le rapprochement sur les chiffres de population, l'Iris Saint Lambert 32 (Paris 15^e) a 1 habitant et Blain (44015) en a 9.041, soit un rapport de 1 à 9.000, équivalent au rapport de population de Jersey et Guernesey avec la Chine.

Il ne viendrait à l'idée d'aucun analyste d'effectuer une comparaison entre la Chine et St Marin ou Guernesey, mais les Irises ou les communes sont l'objet de rapprochements permanents.

A ceci s'ajoute le fait que ces découpages ne sont pas stables dans le temps, chaque année des communes se regroupent ou se scindent, des parcelles sont échangées. On a connu les découpages à l'Ilot, à l'Iris 5000, puis à l'Iris 2000, avec des restructurations intermédiaires en 2007 et 2008. Les comparaisons historiques deviennent donc impossibles dans de nombreux cas.

Pour compliquer le tout, afin de répondre en particulier aux problématiques des collectivités locales³, on envisage la création d'Iris flottants...

³ Qui accédaient autrefois à l'ensemble des données au niveau de l'Ilot.

Au niveau européen on rappellera simplement que les LAU 2 (niveau commune) ont des superficies moyennes qui varient de 5 km² (Malte) à 1.562 km² en Suède et un nombre moyen d'habitants entre 1.195 en Irlande et 33.345 aux Pays Bas. Autre exemple, comment comparer le département Paris (20 communes / 105 km² / 2,16 M. h.) avec la commune de Madrid (8.000 km² et 4,9 M.h.)

1.3. L'association des données aux territoires

Le 13 avril 2010 les débats des « cafés de la statistique » ont porté sur le sujet intitulé « Les enjeux des zonages » avec comme invitée la géographe Denise Pumain.

Ont été relevés tous les problèmes liés à la construction des zonages, à la mobilité des phénomènes qui s'y produisent, à leur multiplicité selon les secteurs (éducation, judiciaire, électoral...), à leurs formes et dimensions et aux biais d'interprétation. Pour un participant la solution idéale est le X,Y, solution purement théorique à deux titres (protection de la vie privée, mais aussi difficulté de représentation des phénomènes à échelle petite ou moyenne).

Il est clair que les zonages sont très intimement liés au phénomène étudié. Les zonages les plus pertinents sont très différents selon que l'on cherche à analyser par exemple :

- L'évolution de la population des ménages
- L'attractivité des universités
- L'implantation de services sociaux
- Les zones de chalandise des hyper marchés ou celles des pharmaciens

La pratique de l'analyse spatiale montre que l'on a toujours intérêt à utiliser le niveau de zonage le plus fin possible pour lequel on dispose de données. Des analyses au niveau d'un département ou d'une région, lorsqu'on introduit une intention de comparaison n'ont aucun sens en raison principalement de l'hétérogénéité que l'on peut rencontrer aussi bien entre les zonages, qu'à l'intérieur même de chaque territoire (Paris et la Lozère, par exemple).

Toutes ces difficultés nous ont conduits à utiliser le carroyage depuis plus d'une dizaine d'années pour effectuer des études de caractère socio-économique.

Notre principale motivation est que le découpage de l'espace en carreaux présente les principaux avantages suivants :

- Libération de la sémantique des découpages administratifs
- Modélisation géométrique de l'espace en unités isotropes qui sont organisées entre elles par un lien mathématique structurel homo-morphique et sont repérées dans l'espace par leurs coordonnées X,Y
- Organisation matricielle des données qui peuvent être traitées avec des opérateurs arithmétiques, statistiques et booléens

2. Le carroyage

2.1. Les théories

Les théories spatiales reposent sur deux notions essentielles : la distance et l'organisation spatiale. Il n'existe pas d'antagonisme entre ces deux approches, mais souvent elles se complètent et nous retrouvons en général les mêmes intervenants dans la recherche.

L'économiste allemand, August Lössch, établit le fondement de l'analyse spatiale en stipulant que l'espace doit être découpé en secteurs homogènes qui puissent être comparables entre eux (1930). Initialement, il utilise des cercles qui représentent la zone de chalandise au centre de laquelle le producteur ou le distributeur vont tenter de s'implanter. La juxtaposition de ces cercles (figurant la concurrence) le conduit, finalement, à opter pour la structure hexagonale dont la géométrie permet, en outre, une couverture exhaustive de l'espace, sans intersections. L'hexagone, quelle que soit sa taille, devient le support géographique de référence des données socio-économiques. A noter que les seules autres structures géométriques qui permettent de découper l'espace en entités homogènes sont le carré et le triangle.

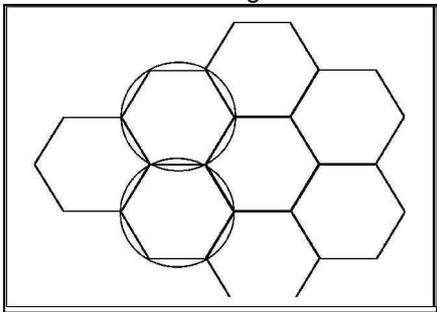


Fig. 1.13 : Analyse de Lösch

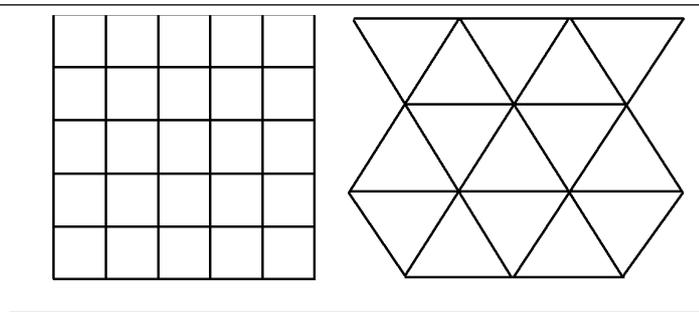


Fig. 1.14 : Découpages possibles de l'espace

Ces travaux constituent la base même des théories de l'analyse spatiale couramment utilisées dans les recherches ultérieures de Léontief, Moses, Isard...

W.Appelbaum, économiste américain, s'en est largement inspiré pour préconiser l'Analyse de Grille en 1966.

Les places centrales

Parallèlement et indépendamment, W.Christaller et G.K.Zipf ont effectué des travaux vers 1930-40 sur l'organisation des réseaux urbains en Allemagne du sud, zone, rappelons-le, peu différenciée sur le plan géographique ; ces travaux les ont conduits à énoncer la théorie des places centrales.

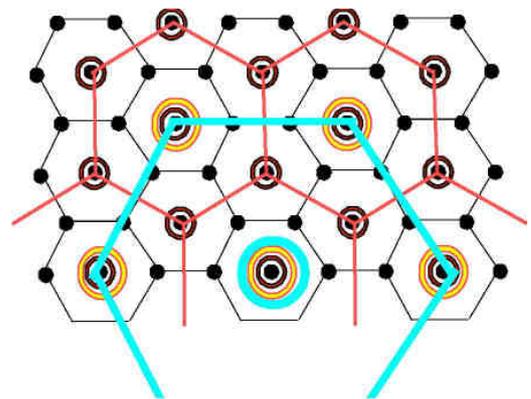
Sans distinguer les apports respectifs de ces scientifiques, nous rappellerons les conclusions de la théorie qui repose sur 2 grands principes : la spécialisation et la hiérarchisation.

Chaque centre urbain possède des spécialités économiques ou administratives et peut se classer vis-à-vis de ses voisins par sa taille de population.

Les observations faites sur le terrain ont conduit à remarquer une organisation géométrique relativement simple des centres (ainsi d'ailleurs que des points de vente) en hexagones emboîtés (voir figure ci-contre).

Chaque petit centre (ou point de vente) se positionne aux sommets d'hexagones (en noir) de façon totalement régulière dans l'espace et est soumis à l'influence directe d'un centre de niveau hiérarchique supérieur. Les centres de niveau 2 s'organisent de la même façon (hexagones en rouge) et ainsi de suite.

Cette disposition s'applique au principe de marché, une autre est destinée à l'organisation des transports où



Loi de Christaller (principe de marché)

les points sont localisés au milieu des cotés des hexagones, ce qui, tout en détériorant légèrement l'accessibilité réduit notablement la taille des infrastructures de desserte. Enfin le principe d'administration enferme les centres dans les hexagones puisque chaque centre est rattaché à un centre de niveau supérieur sans pouvoir mettre les autres en concurrence.

A ces trois principes correspond une évaluation de potentiel du centre de niveau +1 calculée sur base des dépendances illustrée par un coefficient respectif de 3, 4 et 7. (Un centre a un potentiel égal à 3 fois celui du centre de niveau inférieur dans le principe de marché et ainsi de suite...).

Sur la théorie des places centrales s'est greffée une constatation tout à fait simpliste de Pareto qui a donnée naissance à la loi Rang -Taille. Cette loi établit qu'une ville de rang R a une population sensiblement égale à celle de la principale ville d'un pays divisée par la valeur de rang : $Pop_j = Pop_1 / R^k$, ainsi une ville de rang 4 d'un pays où la première ville a 4 M. hab. doit avoir une population d'environ 1 M. hab. Le paramètre k varie entre 0,7 et 1,3 en fonction des pays et des époques considérées.

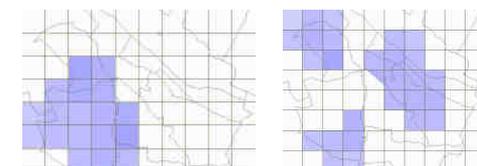
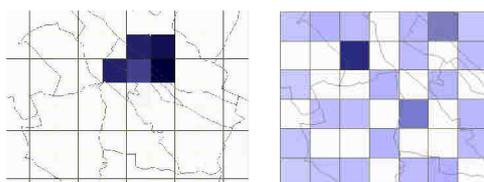
La théorie des places centrales doit être considérée comme un modèle simplificateur de l'organisation du territoire. Sa validité se heurte à 3 aspects principaux de la réalité, d'abord le fait que l'espace géographique n'est pas indifférencié (il existe de nombreuses barrières de toutes natures), ensuite l'apparition de très grands pôles commerciaux modifie le comportement d'achat (regroupements en particulier), enfin l'activité des villes ne se limite pas à la seule satisfaction des populations qui y habitent (par exemple Toulouse possède une activité en aéronautique qui dépasse très largement les besoins des toulousains).

Les phénomènes spatiaux

Dès les années 20 les sociologues de l'école de Chicago ont développé des travaux qui se sont concrétisés avec le modèle qualitatif de Burgess visant à étudier l'évolution dans l'espace de groupes sociaux d'immigrants. Trente ans plus tard les études menées sur la ségrégation (aux Etats-Unis) sont devenues beaucoup plus quantitatives et se sont attachées à la description de répartitions de populations sur bases de divers indices, élaborés également par des géographes, qui permettent de caractériser la distribution d'un phénomène dans l'espace.

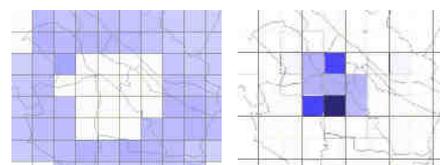
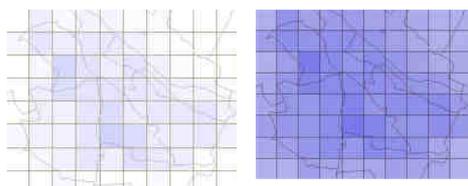
De ces travaux émergent 5 grandes catégories d'indices que nous illustrerons par des exemples opposés (les experts en mathématiques pourront trouver leur formulation en se référant aux auteurs) :

- Agrégation : Duncan 1964 →



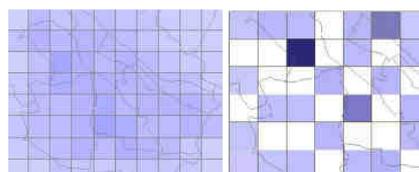
← - Concentration : White 1983

- Centralisation : Duncan & Duncan 1965 →



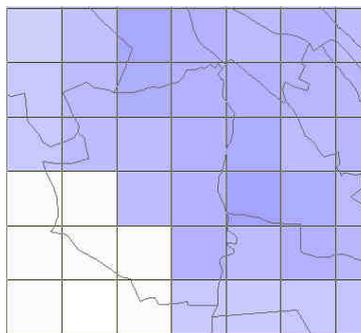
← - Pénétration : Bell 1954

- Distribution : Gini & Thiel 1972 →

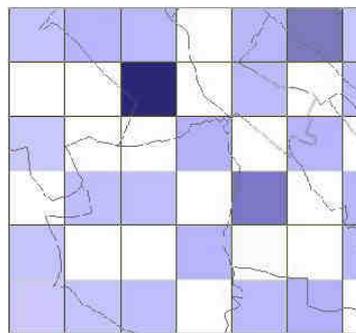


Les études sur la distribution de Gini et Thiel ont amené le développement des notions d'Entropie, fondées sur les principes thermodynamiques du théorème de Carnot, qui mesurent le degré de désordre d'une distribution.

Sachant par ailleurs que des zones territoriales adjacentes ont des configurations relativement comparables et que les caractéristiques d'une zone exerceront nécessairement une influence sur celles des zones contiguës, Geary et Moran ont développé la notion d'autocorrélation spatiale.



$I < 1$ autocorrélation forte
+ entropie relative forte

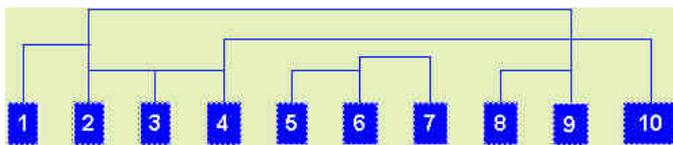


$I = 1$ autocorrélation faible
entropie relative faible

L'indice de Moran fournit un degré d'autocorrélation lié à l'entropie du phénomène étudié.

Les notions d'autocorrélation ont été approfondies par L. Anselin (université d'Illinois) qui a développé un logiciel (Geoda) dont une version bêta peut être téléchargée sur son site et qui possède des fonctions intéressantes dans ce domaine.

Enfin, signalons l'approche intéressante en ACP hiérarchique ascendante avec prise en compte des matrices de contiguïté.

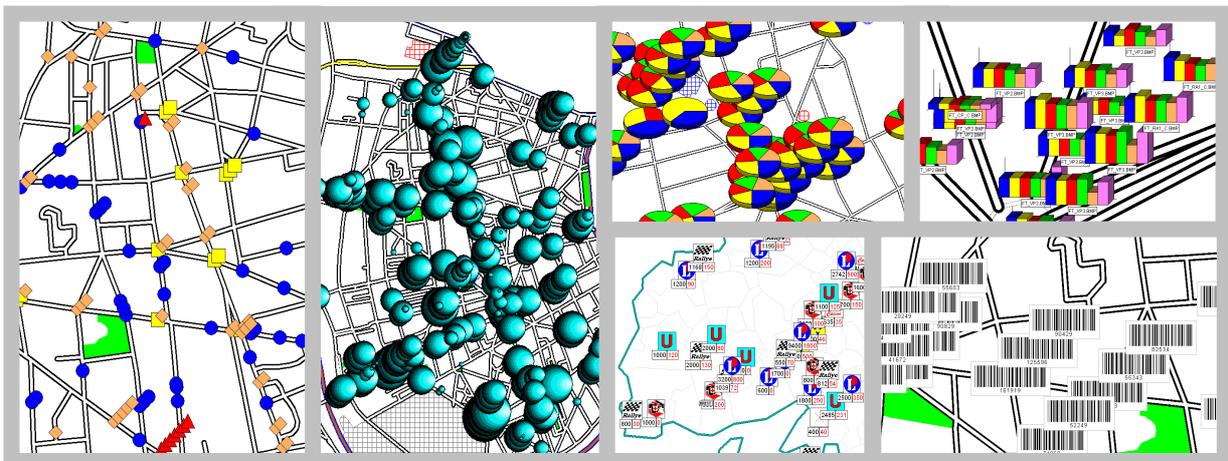


1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	32	33	34	35
36	37	38	39	40	41	42

2.2. Les représentations traditionnelles

Les systèmes d'information géographique sont conçus pour afficher une donnée avec un choix de représentations thématiques diverses.

Parmi les plus répandues, les représentations symboliques sont repérées dans l'espace par leur position X,Y (adresse ou barycentre d'un territoire). Exemples :



Tous ces modes de représentation (points, boules, camemberts, histogrammes, logos, codes barre...) peuvent être qualifiés au mieux de descriptifs, sinon d'anecdotiques, en ce sens où ils ne peuvent décrire qu'un faible volume de données sous une forme fragmentée; ils ne fournissent pas une information synthétique sur la zone étudiée et ne peuvent pas, tels quels, faire l'objet de traitements spatialisés.

Dans la pratique, l'analyse ne doit surtout pas porter sur un examen visuel de représentations de ce type, car le choix des formes, des volumes et des couleurs, la superposition inévitable d'informations fausseront nécessairement le jugement de l'observateur. Il est donc indispensable d'avoir recours à d'autres modes descriptifs dont la lecture ne puisse être mise en question.

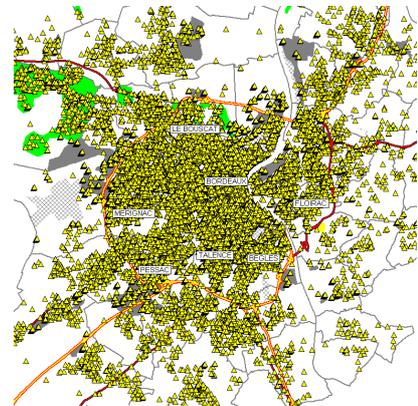
En outre, dans le cadre d'études, il est fréquemment nécessaire de faire appel à des fichiers volumineux, ainsi, par exemple, la population de l'ensemble des communes ou la localisation des entreprises implantées sur une agglomération.

Prenons l'exemple des 45.000 entreprises de Bordeaux; si l'on représente l'information sous forme de points, la lecture devient impossible, aussi bien pour évaluer la quantité d'entreprises et leur concentration (les points se superposent) qu'à fortiori pour représenter des valeurs qui peuvent être attachées à chaque entreprise (ses effectifs, son chiffre d'affaires, ...).

Il est, par conséquent, indispensable d'utiliser une approche différente qui sache fournir l'information requise sous une forme claire, fiable et précise.

La réponse va consister à agglomérer les informations de façon lisible selon un mode et une finesse qui seront liés au volume de données et à l'étendue du territoire étudié.

La méthode d'agglomération forme la base de l'analyse spatiale et elle repose sur une division homogène de l'espace, non pas en hexagones, qui seraient plus complexes à gérer, mais en simples cellules carrées, dont la taille peut se réduire (en théorie) à celle d'un point.



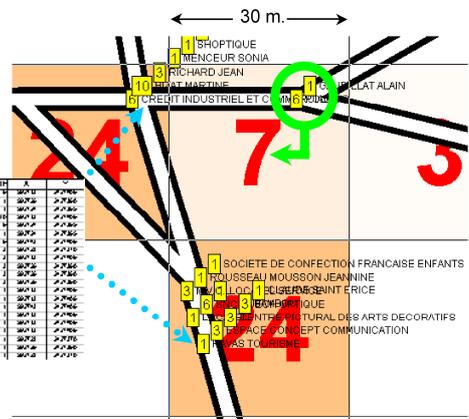
2.3. L'agglomération des données dans la grille

Les données, localisées par leur position X,Y, doivent être affectées aux carreaux (qui seront désignés par le terme «cellules») en fonction des coordonnées X', Y', X''', Y''' de chaque carreau.

Le chargement des données dans ces cellules homogènes, présente deux avantages essentiels :

d'une part, il libère l'utilisateur des contraintes imposées par les découpages administratifs et lui permet de travailler sur des zones plus fines, d'autre part l'information à traiter est mise en densité automatiquement ; ainsi dans l'exemple ci-contre, les effectifs des entreprises sont répartis dans des zones de 30 x 30 m et les valeurs résultantes affectées aux cellules indiquent de façon pertinente la répartition des salariés sur la zone étudiée.

Les données affectées aux cellules peuvent être n'importe quel type d'information disponible dans la base étudiée: le nombre d'entreprises présentes dans le périmètre d'une cellule, la somme des valeurs affectées à une adresse (effectifs, chiffre d'affaires, consommation des ménages...).



Lorsque la source support des données est non plus une adresse, mais une surface, îlot, Iris, commune, canton ou département, (Ex : statistiques de population) la dimension de la surface administrative en question peut être supérieure à celle de la cellule. Dans ce cas, les valeurs devront être réparties en fonction de la surface couverte par la cellule.

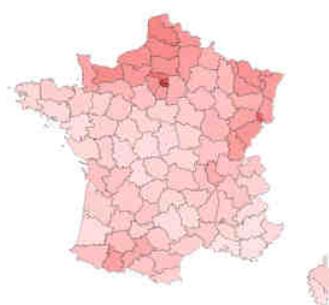
Dans l'hypothèse où les cellules auraient une taille moyenne équivalente ou supérieure aux surfaces administratives considérées, le principe du calcul est absolument identique, car, dans tous les cas, la cellule pourra couper un objet administratif dont il sera nécessaire de distribuer la valeur en proportion des surfaces intersectées.

2.4. Exemple d'application de la grille

Dans le cadre des obligations du service universel applicables à la publiphonie (cabines téléphoniques), il est convenu que l'adjudicataire du service doit s'engager à maintenir et opérer au moins une cabine téléphonique par commune.

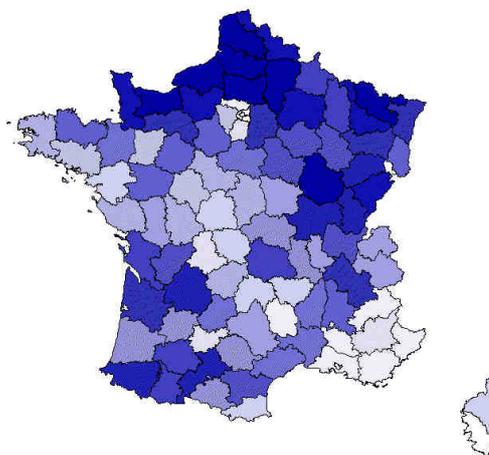
France Télécom souhaitait convaincre l'ARCEP⁴ que cette obligation était trop lourde et a donc produit une carte représentant le nombre minimum de cabines par département.

En fait cette représentation n'est jamais que celle du nombre de communes par département, montrant que certains d'entre eux ont plus de communes que d'autres, sans même prendre en compte les dimensions des communes et des départements.



Nbre min. de cabines en densité

Réalisée en densité, la même carte n'apporte pas d'élément supplémentaire permettant d'évaluer le phénomène.



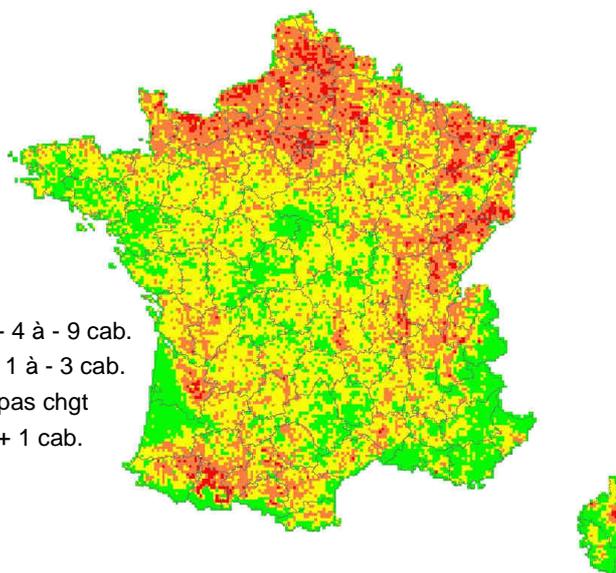
Nbre minimum de cabines par département

Par contre, si l'on considère les motifs qui sous-tendent cette obligation, à savoir principalement d'offrir un moyen de communication qui ne soit pas trop éloigné de n'importe quel point du territoire, afin de préserver un accès facile aux services de secours, on pourra proposer une répartition des cabines tous les 5 km par exemple⁵.

Le résultat obtenu (grille d'un pas de 5 km dont les cellules reçoivent la valeur « 1 ») comparé avec la situation actuelle (par la méthode de corrélation spatiale) nous montre que l'on peut réduire globalement le nombre de cabines sur le territoire métropolitain de 36.590 à 22.789, avec des zones dans lesquelles la réduction pourrait être très importante (jusqu'à 9 cabines) et d'autres dans lesquelles il faudrait en ajouter au moins une.

Une diminution de la distance à 4 km permettrait de réduire le nombre total de 1.000 cabines, en assurant par ailleurs une couverture de l'ensemble du territoire plus optimale que la méthode « 1 cabine par commune »

- - 4 à - 9 cab.
- - 1 à - 3 cab.
- pas chgt
- + 1 cab.



Écarts régression 1 cabine par commune

⁴ Autorité de Régulation des Communications Electroniques et des Postes

⁵ En sorte qu'aucun point du territoire ne soit à plus de 2,5 km d'une cabine

2.5. Traitements statistiques simples dans la grille

L'objectif de ces traitements est non seulement d'obtenir des représentations cartographiques fiables et claires, mais aussi de faire ressortir les valeurs attachées à chaque cellule de la grille.

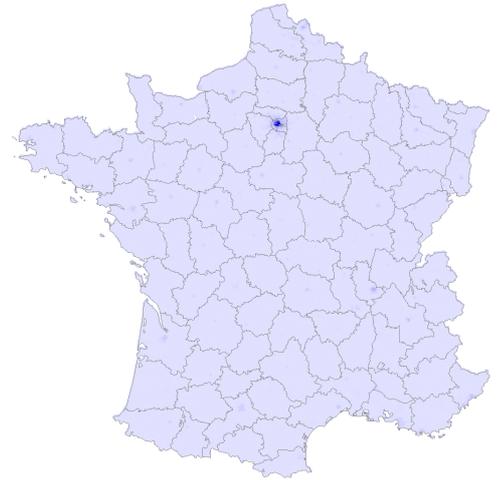
Les traitements réalisés dans la grille produisent des valeurs diverses qui sont ensuite exportables vers des fichiers d'objets ou adresses repérés dans l'espace par leur X,Y

2.5.1. Discrétisation

La répartition des phénomènes socio-économiques dans l'espace fait qu'il est souvent difficile d'identifier les nuances d'une série de données lorsqu'il existe une très forte amplitude dans la répartition des données de la série considérée.

Ainsi, si l'on souhaite représenter l'implantation des ménages en France, la densité varie de 0,01 à plus de 50.000 unités au km².

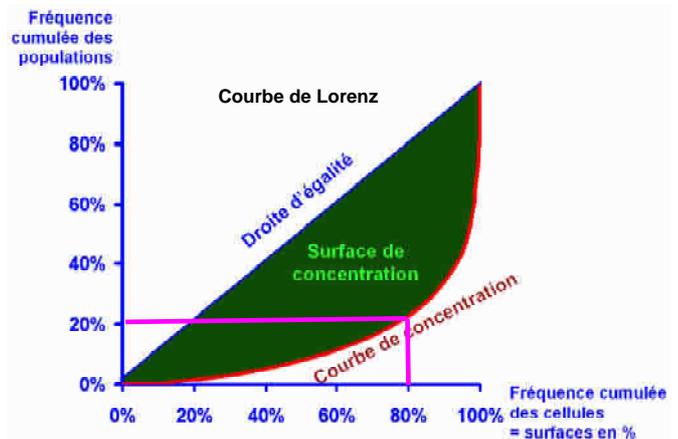
En utilisant un simple affichage en densité, le dégradé de couleurs en mode 32 bits, fait apparaître un point noir sur Paris et un dégradé de couleur quasiment uniforme sur le reste de la France. Un phénomène identique est constaté dans le cadre de l'analyse de la plupart des données économiques à cause du poids considérable de la région parisienne.



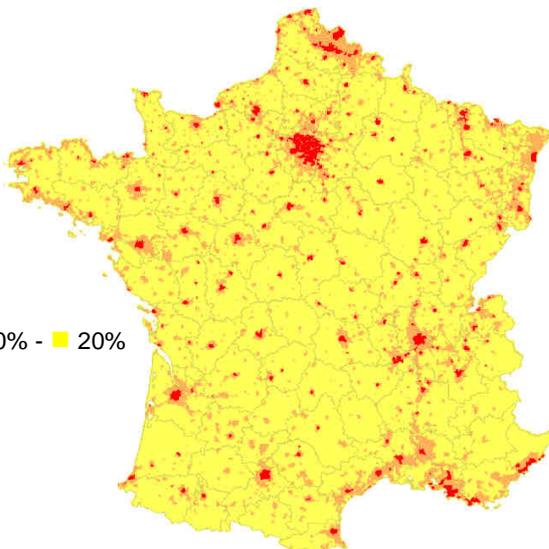
Toutes les méthodes de discrétisation utilisées en statistiques sont applicables dans le cas des grilles de données. Toutefois nous en recommandons une nouvelle qui permet d'obtenir une représentation visuelle particulièrement intéressante des phénomènes et qui repose sur les méthodes de concentration (courbe de Lorenz).

Les densités de ménages sont réparties en 10 classes et triées par ordre décroissant et leur fréquence cumulée est exprimée en %. La fréquence cumulée des cellules correspond à celle des surfaces (toutes les cellules étant identiques).

Nous constatons que la loi de Pareto s'applique bien à la répartition de la densité des ménages (20% des ménages dans les zones à faible densité, correspondent à 80% du territoire).



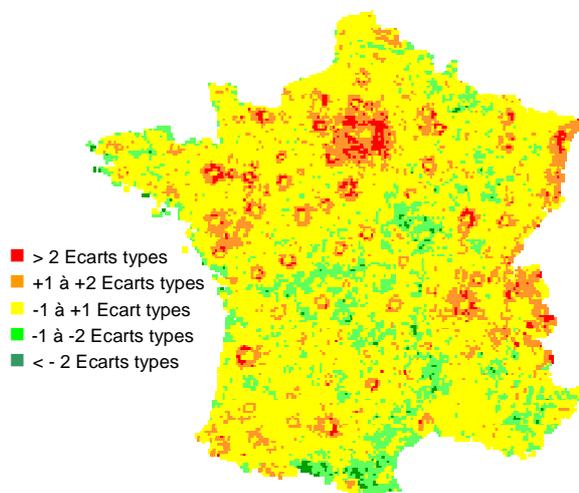
Loi de Pareto : 20% des ménages sur 80% du territoire



La représentation des 5 classes les plus denses en rouge, des 3 classes suivantes en orange et des 2 classes les moins denses en jaune, permet d'identifier visuellement l'implantation de 50% et 80% du phénomène étudié, ici les ménages résidant dans les zones les plus denses.

2.5.2. Corrélation

La comparaison de deux phénomènes est très souvent résolue par une pénétration. En analyse spatiale, cette méthode est trompeuse car elle occulte l'importance des phénomènes constatés.



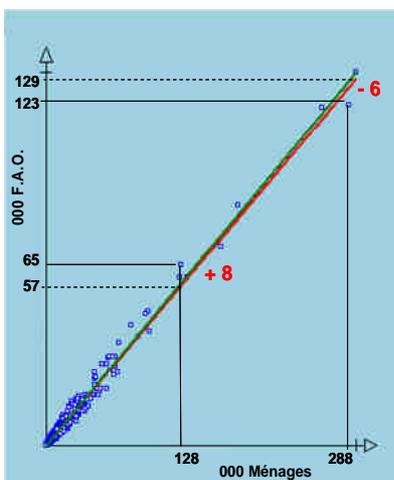
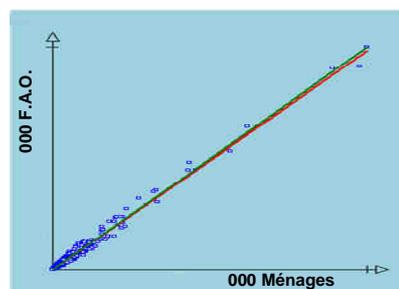
Par exemple nous trouvons ci après une comparaison entre les femmes actives occupées et les ménages (à l'Iris, en densité). La moyenne est de 41% et l'écart type de 9%.

La représentation du pourcentage met en valeur de façon trompeuse des zones dans lesquelles dividende et diviseur sont des nombres très faibles ; ainsi, par exemple, à Douaumont (55) on trouve 2 F.A.O. et 2 ménages, soit un résultat de 100%.

A contrario Vitry (94) est la commune de résidence de 15.000 F.A.O. et 30.000 ménages, donc un quotient de 50% seulement, pourtant Vitry a 2.700 F.A.O. de plus que la moyenne nationale, alors que Douaumont n'en a qu'une.

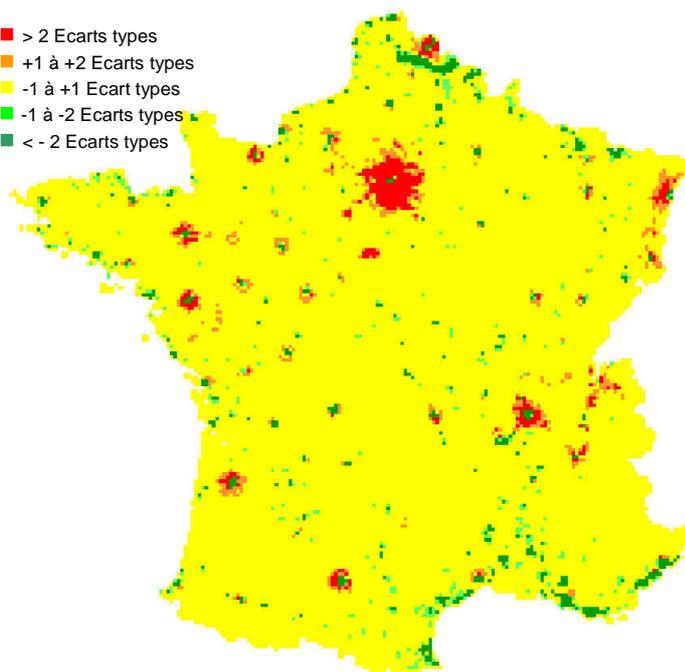
Pour toutes ces opérations de comparaison entre deux variables, nous préconisons de prendre en compte les écarts à la régression (résidus) calculés à partir des 2 matrices.

Sur le graphique ci-contre nous constatons qu'il existe un lien fort entre les 2 variables ($r=0,99$), les droites de régression de x en y et y en x étant quasiment superposées.



L'étude des résidus montre, par contre, que certaines cellules présentent une sur ou une sous représentation du phénomène F.A.O. par rapport à la tendance France entière.

- > 2 Ecart types
- +1 à +2 Ecart types
- -1 à +1 Ecart types
- -1 à -2 Ecart types
- < -2 Ecart types



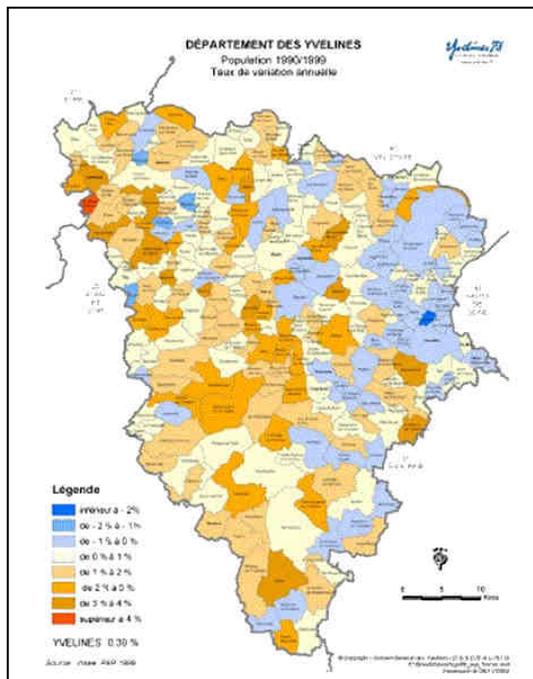
La représentation des résidus en 5 classes sur base des écarts types ($\sigma = 285$) permet de mettre en valeur les territoires dans lesquels le phénomène F.A.O. est significativement sur ou sous représenté, ce qui élimine les toutes petites communes et met en valeur des phénomènes forts comme celui du bassin minier, de la côte méditerranéenne ou des centres des villes les plus importantes (Marseille étant atypique dans le cadre de cet exemple précis).

La relation décrite ici est linéaire, mais on peut envisager aussi bien des relations exponentielles, logarithmiques, ou encore multivariées.

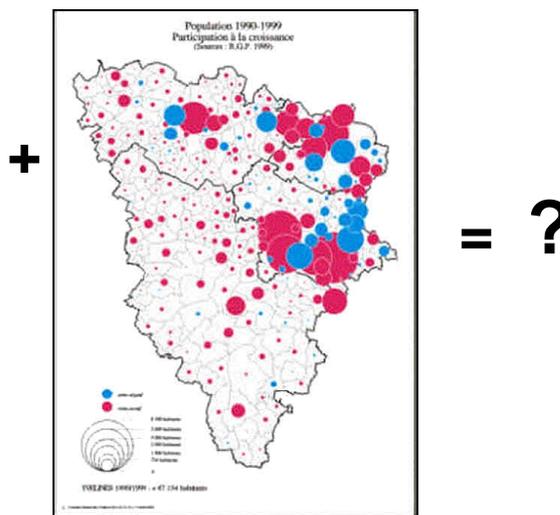
La même démarche d'analyse des résidus statistiques est applicable dans le cadre de l'étude de l'évolution historique d'un phénomène.

Nous prendrons comme exemple une analyse publiée par le conseil général du département des Yvelines.

La carte ci contre compare les populations 1990 / 1999.



Le fait que les communes à forte croissance en % soient concentrées à l'ouest dans lequel les densités sont très faibles a bien évidemment troublé l'analyste qui pour faire comprendre la réalité a juxtaposé une seconde carte intitulée « participation à la croissance » avec des boules pour illustrer les volumes.

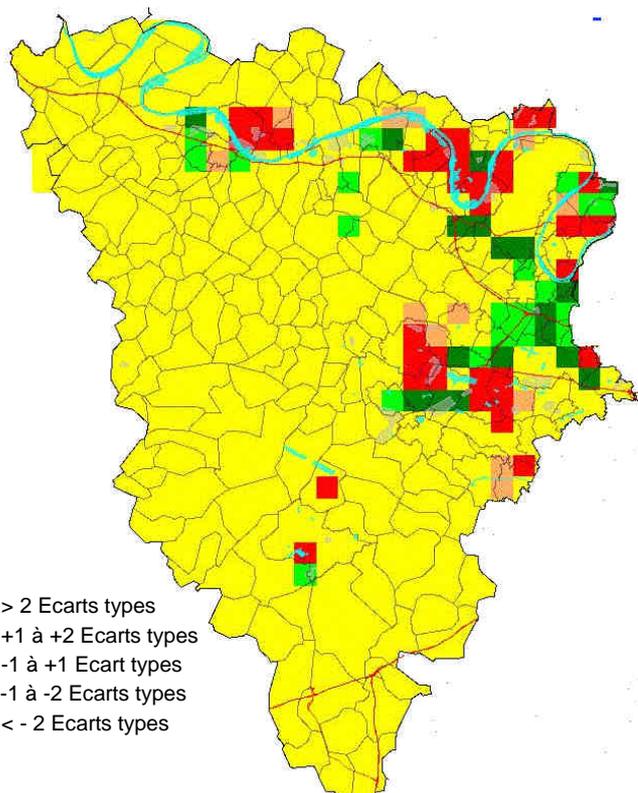


Pour le lecteur la compréhension du phénomène avec ces deux cartes est difficile, surtout s'il souhaite examiner des zones précises comme Rambouillet ou Poissy.

La réponse est apportée par une simple carte des écarts à la régression entre densités 90 et 99, où les cellules ont un pas de 2 km et où l'écart type est de 325.

Cette carte nous montre que l'évolution n'a que très peu concerné le centre et l'ouest du département, alors que les écarts importants sont concentrés dans l'est et dans la zone de Mantes.

Il ne faut pas oublier, bien évidemment, que les écarts représentés sont des différences par rapport à la tendance de l'ensemble de la zone étudiée et que les valeurs obtenues par cellule doivent être rapportées à cette tendance (un résidu négatif peut toutefois correspondre à une augmentation, celle-ci restant plus faible que celle de l'ensemble, bien entendu).



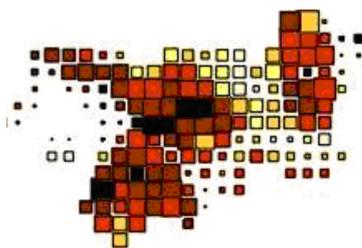
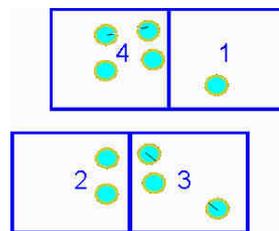
2.6. Réserves sur la grille

Deux critiques principales peuvent être apportées à l'utilisation de la grille

1° Position de la grille

En fonction de la position de la grille les valeurs chargées peuvent être notablement différentes dans les approches à grande échelle.

Par ailleurs toutes les cellules étant identiques il est impossible de tenir compte des proximités entre points pour affecter à la même cellule 2 points proches.⁶



Cellules de tailles différentes

Des tentatives faites avec des cellules de tailles différentes dans une même couche ne constituent en fait qu'un retour à un découpage arbitraire de l'espace proche des territoires administratifs.

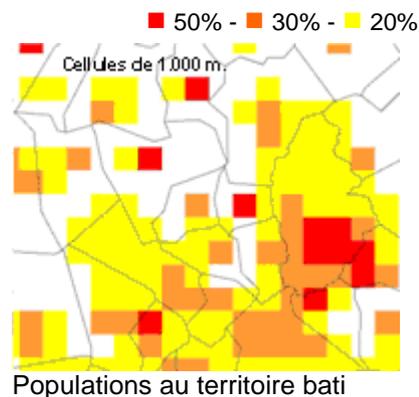
Il est encore envisageable de recourir au lissage pour effectuer des redressements ou de prendre en compte les phénomènes d'autocorrélation spatiale lorsqu'on a affaire à un phénomène dont l'influence déborde le territoire affecté à la cellule.

Toutefois, il nous semble que si nous traitons un phénomène discret (implantation de populations à l'adresse, par exemple), le biais provoqué par la position de la grille est très faible et n'entache pas la qualité générale de l'analyse. Le problème tient surtout à l'homogénéité des données qui sont rapprochées par l'intermédiaire de la grille : ainsi on peut rapprocher les phénomènes Femmes Actives Occupées et Ménages. Par contre on ne pourra pas rapprocher sans précautions particulières les étudiants et les implantations universitaires.

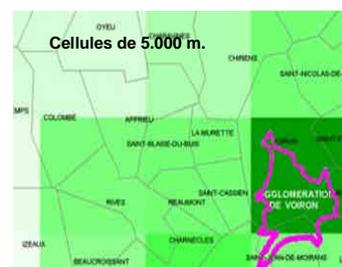
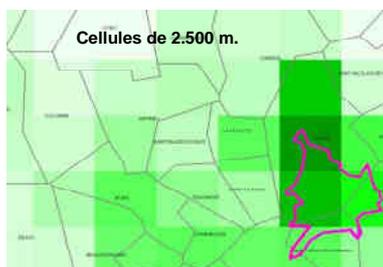
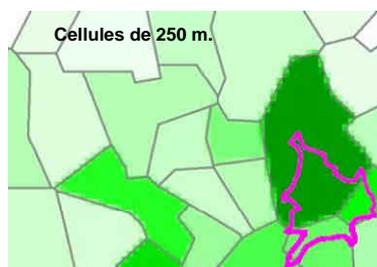
2° Modalités de chargement des données de territoires

L'affectation des valeurs aux cellules peut être influencée par la méthode de repérage des données correspondant à un objet surfacique.

Le principe veut que l'on répartisse les valeurs en proportion des territoires intersectés. Le chargement à partir des barycentres d'objets est à proscrire, mais on peut envisager effectuer par exemple un chargement de variables de population à partir des territoires bâtis. Sous réserve du niveau de précision de la couche cartographique du bâti⁷, il sera possible de concentrer les valeurs sur les zones bâties, mais on fera alors abstraction de l'habitat dispersé. L'approche peut être intéressante dans des zones où le bâti est très concentré et où l'habitat dispersé est faible.



Le pas de la grille a bien entendu un impact non négligeable sur les représentations. Par exemple :



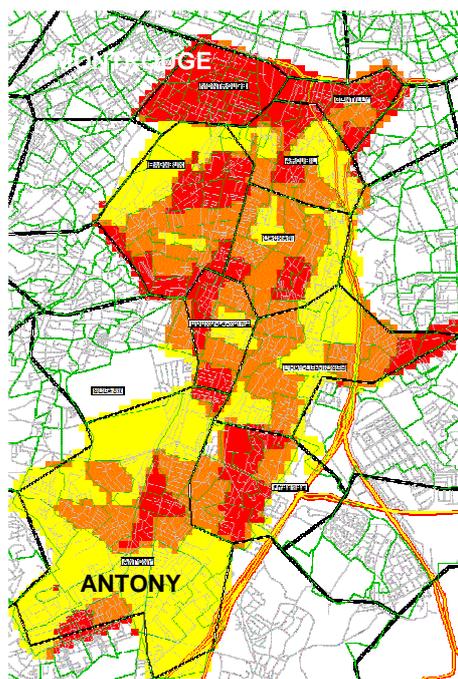
L'agglomération de Voiron (correspondant au bâti) s'étend en fait sur 3 communes jointives

⁶ Le problème est identique lorsqu'on a recours au découpage administratif : une adresse tombe dans un Iris ou dans son voisin et deux points proches peuvent être affectés à des Iris différents

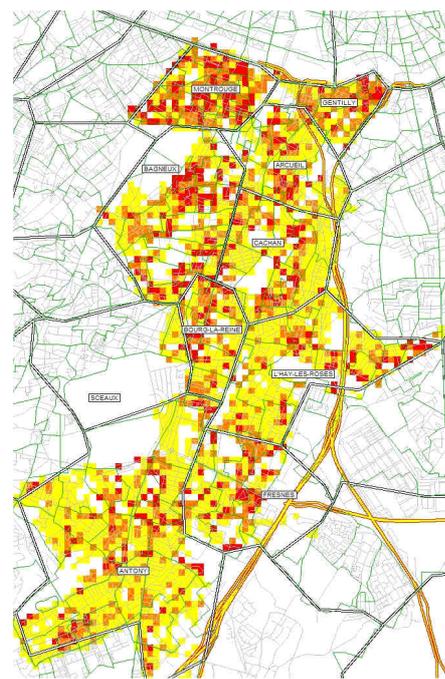
⁷ La précision de Corine Land Cover est en forte progression avec le temps

Le reproche fait à la grille selon lequel l'affectation aux cellules est indifférenciée lorsque les cellules sont entièrement comprises dans l'Iris, par exemple, ne tient pas. En effet lorsque le chargement est fait à l'objet administratif, on ignore s'il existe des concentrations particulières à l'intérieur de l'objet.

A la rigueur, on peut faire appel à des méthodes détournées pour répartir les données, en utilisant par exemple le résultat d'un géocodage à l'adresse de l'annuaire téléphonique, en présupposant que la répartition des populations est homogène avec celle des abonnés.

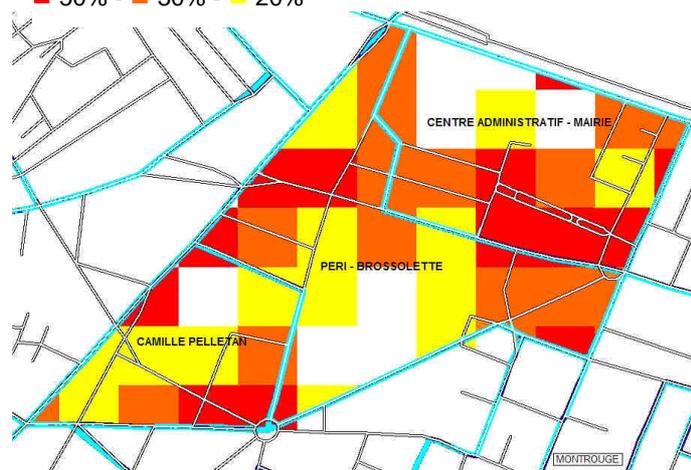


Ménages / Iris / banlieue sud Paris
 ■ 50% - ■ 30% - ■ 20%



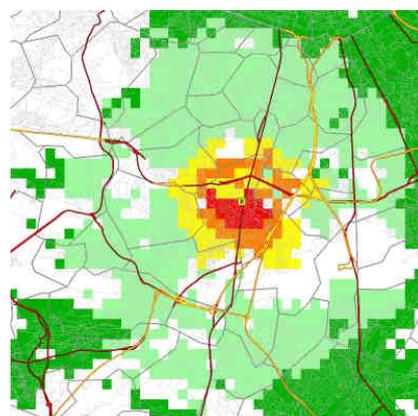
Répartition des ménages / annuaire

Un zoom sur 3 Iris au nord ouest de Montrouge montre les résultats détaillés d'une affectation proportionnelle aux adresses des abonnés et la mise en valeur d'une zone au nord dans laquelle est localisée la mairie



3° Visualisation des représentations

La critique a été faite sur la qualité de lisibilité de l'affichage d'une grille et du repérage des cellules dans l'espace. Il est vrai qu'il est nécessaire de faire appel à un certain niveau d'abstraction pour comprendre la représentation, mais la superposition de P.O.I. ou de voies permet de résoudre la difficulté qui est par ailleurs identique avec les découpages administratifs fins⁸.



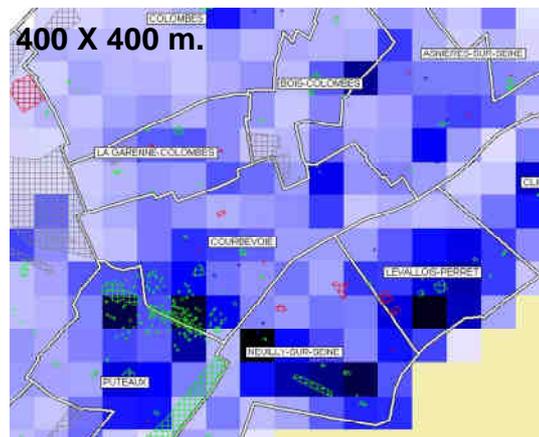
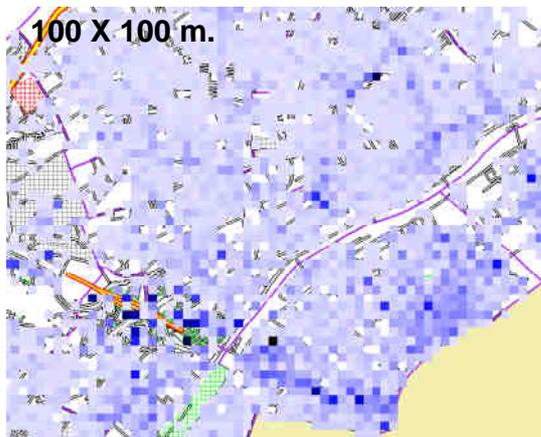
Grille et éléments de repérage

⁸ Le quartier « Ternes » est intuitivement plus facile à localiser que l'Iris « Ternes 11 »

2.7. Pas de la grille

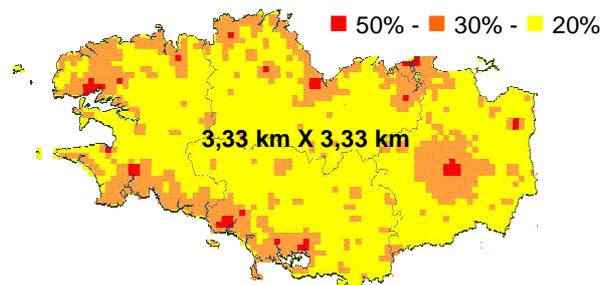
Le pas de la grille ou dimension des cellules doit être adapté à la densité du phénomène et à l'échelle du territoire étudié.

- Cas d'un phénomène dense (populations, implantation d'entreprises...) dans une zone à grande échelle : d'expérience nous avons constaté que la mise en valeur des caractéristiques du phénomène était optimale pour des cellules d'environ 1/10 km². Le pas peut varier pratiquement entre 200 et 400 m. A la limite, on pourra tester 100 m en cas de phénomène très dense.

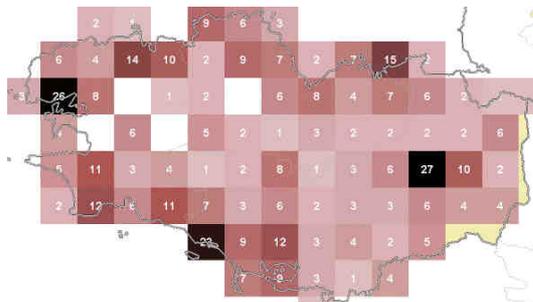


Densité d'entreprises dans le secteur de La Défense (92)

- Pour un phénomène dense et une analyse à petite échelle, le pas pourra se situer entre 2 et 5 km.



Densité des ménages à l'Iris en Bretagne



Densité supermarchés en Bretagne

L'analyse d'un phénomène dispersé (implantation de grandes surfaces, par exemple) dans une zone à petite échelle devra être faite avec un pas de cellule qui pourra atteindre 20 km

3. Exemples d'application de la grille

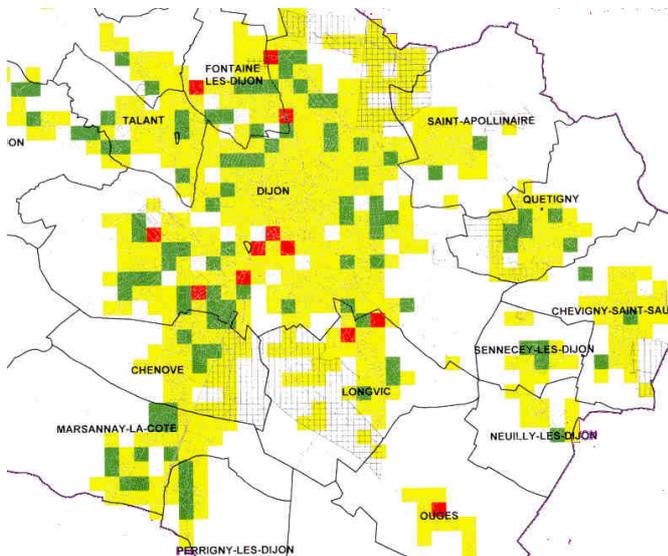
3.1. Clients à risques GDF

GDF souhaitait identifier parmi ses clients ceux qui pourraient présenter une typologie à risques afin de mettre en œuvre une politique préventive pour que les utilisateurs soient mieux sensibilisés à l'entretien des conduites et appareils domestiques connectés en aval de leur compteur.

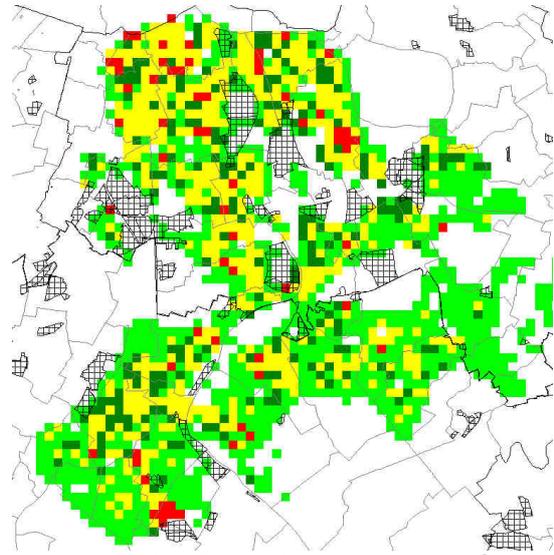
L'étude a porté sur les résultats de 95.000 contrôles effectués en Côte d'or et dans sa banlieue et 260.000 dans la banlieue sud de Paris. (Résultats des DQG / Diagnostics Qualité Gaz).

Les études n'ont identifié aucune relation entre le niveau de risque et les différentes variables du recensement (âge, CSP, statut des occupants, type, ancienneté des logements).

Par contre des zones de concentration de risques ont été nettement matérialisées sur les zones étudiées.



Dijon et sa banlieue – résultats des DQG

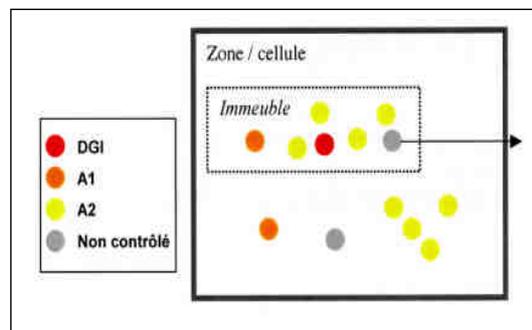


Banlieue sud de Paris – Niveaux de priorité

La taille des cellules est de 1/10 km² en zones urbaines et 10 km² en zones interurbaines

Méthode utilisée : création d'un score pour chaque client, chaque immeuble et chaque cellule, calqué sur les résultats individuels des DQG.

La démarche itérative, sous réserve des tailles d'échantillons contrôlés pour les immeubles et les cellules permet de calculer un score final par client qui indique un risque élevé / moyen / faible ou la nécessité de procéder à des contrôles supplémentaires



Foyer non contrôlé, habitant un immeuble classé à risque situé dans une zone / cellule à faible risque (< 50%)

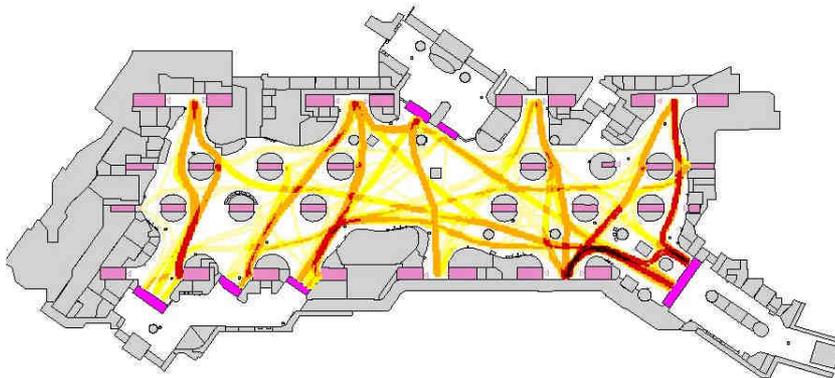
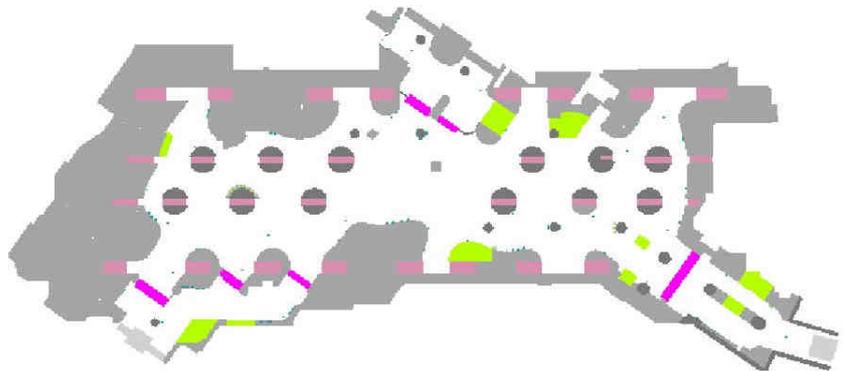
3.2. Implantation de services R.A.P.T.

La RATP recherchait une méthode permettant d'optimiser l'implantation de services dans les stations de métro : boutiques, distributeurs automatiques, photomaton, bornes Internet, points info, poubelles...

Une application (Locserv) a été développée en utilisant les méthodes de Localisation – Allocation développées par J.Baray dans sa thèse doctorale à Rennes I. Le principe consiste à implanter des points d'offre (services) de façon à optimiser une fonction objectif calculée avec un modèle p-médian (rendre minimale la somme des distances entre les points de demande et les points d'offre). Les points de demande sont matérialisés par la localisation des flux de voyageurs.

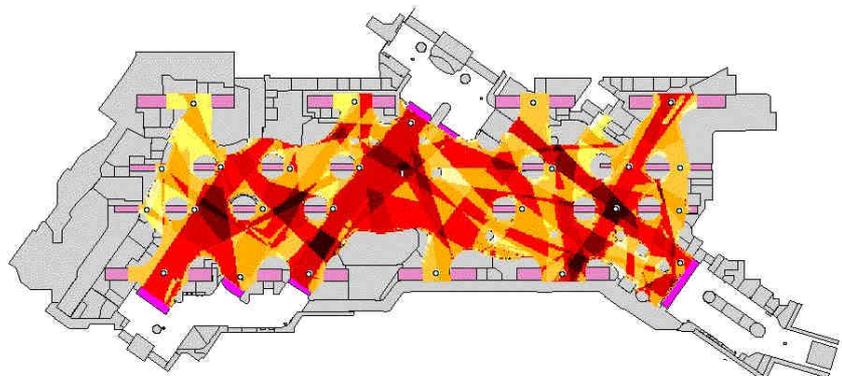
L'exemple porte sur la salle d'échanges RER de Chatelet Les Halles

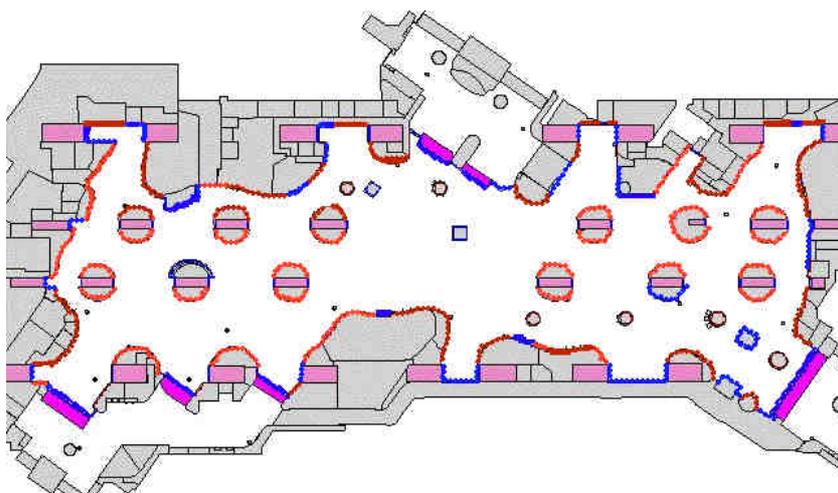
La salle d'échange mesure près de 300 m de long, le fond cartographique a été découpé en 310.000 cellules de 40 cm (espace occupé à un instant t par un voyageur)



Les flux de voyageurs ont été estimés à partir de 2 comptages sur 175.000 et 300.000 personnes interrogées. La connaissance des passages aux points de contrôle permet d'établir des trajectoires rationnelles (sortie la plus proche par le chemin le plus court). Les valeurs attachées aux flux sont réparties dans les cellules.

Une carte des visibilité est établie à partir d'un modèle numérique de terrain et les valeurs sont réparties dans les cellules. Ces valeurs correspondent à la « demande »

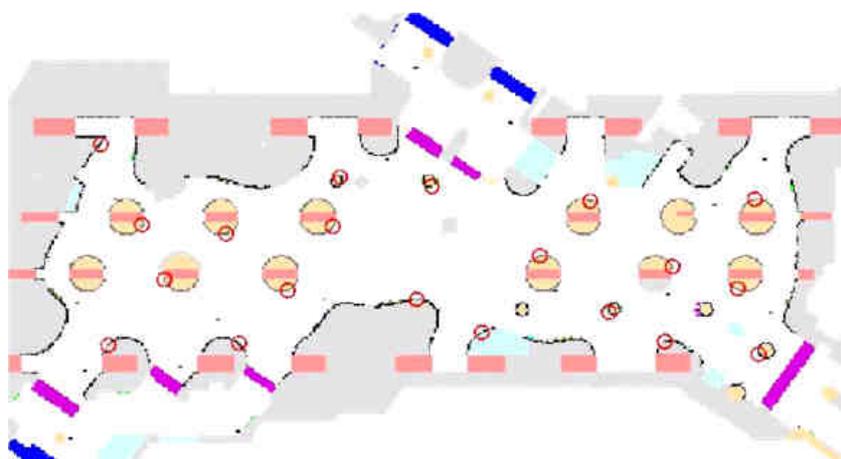
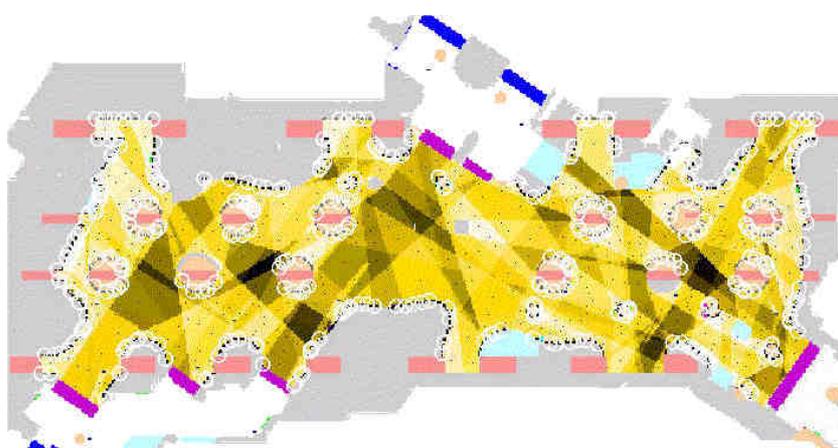




L'exclusion de toutes les zones réservées à la circulation, aux emplacements techniques ou de secours, laisse apparaître les zones d'implantations possibles : « l'offre » (qui peuvent éventuellement être pondérées).

On obtient ainsi une matrice de demande et une matrice d'offres possibles.

Le calcul p-médian peut donc être appliqué sur ces 2 matrices pour obtenir la localisation des points d'offre dont la distance cumulée à tous les points de demande est la plus faible (qui offrent donc la meilleure visibilité).



Exemple de résultat obtenu pour une recherche d'implantation optimale des 20 premiers services à planter : ○

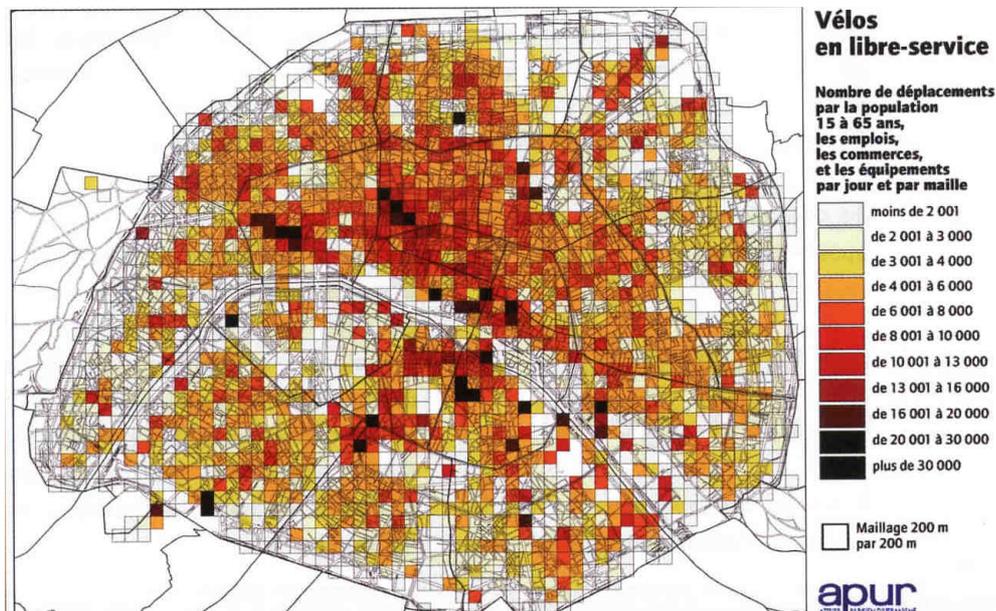
3.3. Etude d'Implantation des stations velib' à Paris par l'APUR

L'APUR⁹ a utilisé la méthode de grille pour effectuer la modélisation de la demande potentielle sur base :

- Densités de population dont l'âge est compris entre 15 et 65 ans
- Localisations de l'emploi
- Localisation des commerces
- Localisation des équipements (Education, sports, tourisme, santé...)

Les données liées aux déplacements sont extraites de l'Enquête Générale Transports du STIF. En fonction de la catégorie EGT des coefficients sont appliqués pour tenir compte de la propension possible à l'usage d'un vélo.

Le cumul des déplacements estimés liés à une adresse est affecté à des cellules de 200 m qui contiennent ces adresses.



Ensemble de la demande potentielle journalière

On trouvera le rapport détaillé de l'étude à l'adresse suivante :

<http://www.apur.org/sites/default/files/documents/216.pdf>

⁹ Méthode proposée par le cabinet Indiggo Altermodal

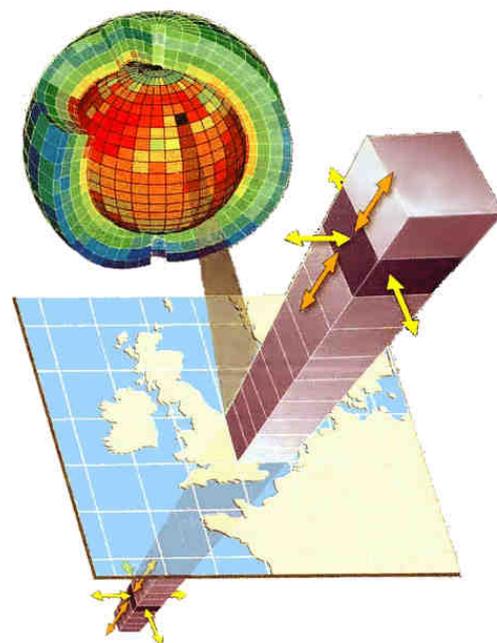
3.4. Application scientifique - Hadley Center

Le Hadley Center, créé en 1990 en Grande Bretagne dans le cadre du Met Office, est un institut de recherche en climatologie dont les objectifs consistent à modéliser :

- Les processus physique, chimiques et biologiques qui impactent le système climatique de la terre
- Les variations entre climats locaux et globaux sur une période s'étendant de - à + 100 ans
- Les variables spécifiques qui induisent des modifications récentes du climat

Même si les phénomènes météorologiques sont de nature continue, on préfère les « discrétiser » selon un maillage qui peut varier en dimension (1 à 5 degrés en longitude et latitude)¹⁰.

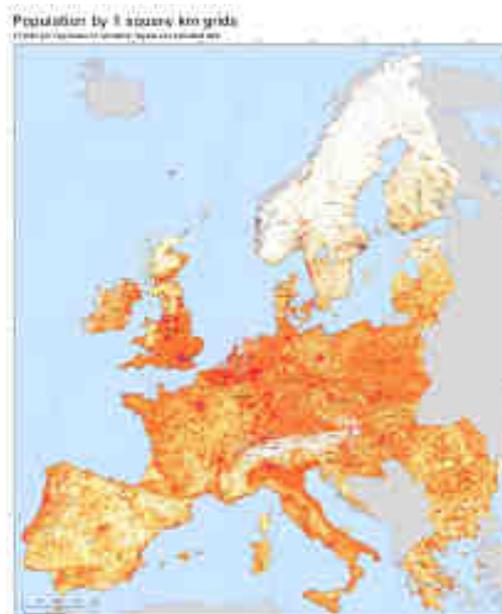
Dans chaque « cube » ainsi formé les modèles de circulation générale des fluides calculent les vents, transferts de chaleur, radiations, l'humidité et l'hydrologie, puis évaluent les interactions avec les cubes voisins.



3.5. Travaux de l'E.F.G.S.

Le Forum Européen pour les Géostatistiques qui rassemble les Instituts statistiques de 32 pays européens promeut le développement des meilleures pratiques de production de données statistiques spatiales dans le cadre des directives Inspire.

Nous renverrons le lecteur à l'exposé de M. J.L.Lipatz, représentant de l'INSEE au Forum en souhaitant, comme futur utilisateur des données du recensement au carreau, d'y avoir accès dès que cela sera possible.



¹⁰ Le modèle le plus récent du Hadley Center (Hadcm3) utilise une grille de 3.75° de longitude / 1.25° de latitude et 19 niveaux en altitude, soit environ 500.000 icosaedres pour l'ensemble du globe.

Bibliographie

- [1] Schéma directeur de la région Ile de France à télécharger sur le site de la région
- [2] Compte rendu de la soirée des Cafés de la statistique (13 avril 2010) « Les enjeux des zonages »
- [3] BÉGUIN H, Méthodes d'analyse géographique quantitative, Litec, 1979.
- [5] BRUNET R, *La carte, mode d'emploi*, Editions Fayard-Reclus, 1987.
- [6] CAUVIN C., RIMBERT S, *La Lecture numérique des cartes thématiques*, Éditions universitaires, 1976
- [7] CICERI M.F MARCHAND B RIMBERT S, *Introduction à l'analyse de l'espace*, Masson, 1977.
- [8] CHARRE J, Statistique et territoire, Espaces, modes d'emploi , RECLUS Montpellier, 1995,
- [9] HAGGETT P, L 'Analyse spatiale en géographie humaine, Armand Colin « U» , 1973 .
- [10] LATOUR P. & LE FLOC'H J. Géomarketing, principes, méthodes et applications, Editions d'organisation 2001
- [11] MERLIN P, Méthodes quantitatives et espace urbain, Masson, 1973.
- [12] RACINE J.B, REYMOND H L'Analyse quantitative en géographie, PUF 1973.
- [13] REKACEWICZ P, « La cartographie, entre science, art et manipulation » Le Monde Diplomatique, février 2006
- [14] RIMBERT S, *Carto-graphies*, Hermès, 1990.
- [15] SANDERS L, L'Analyse des données appliquée à la géographie, RECLUS, 1989.
- [16] VOIRON C, *Analyse spatiale et analyse d'images*, RECLUS « Espaces, modes d'emploi» , 1995.
- [17] SERRES M. Les origines de la géométrie, Flammarion 1993