

Le spatial est il toujours spécial? le point de vue des méthodes

Hélène Mathian

UMR 5600 EVS – plateforme ISIG



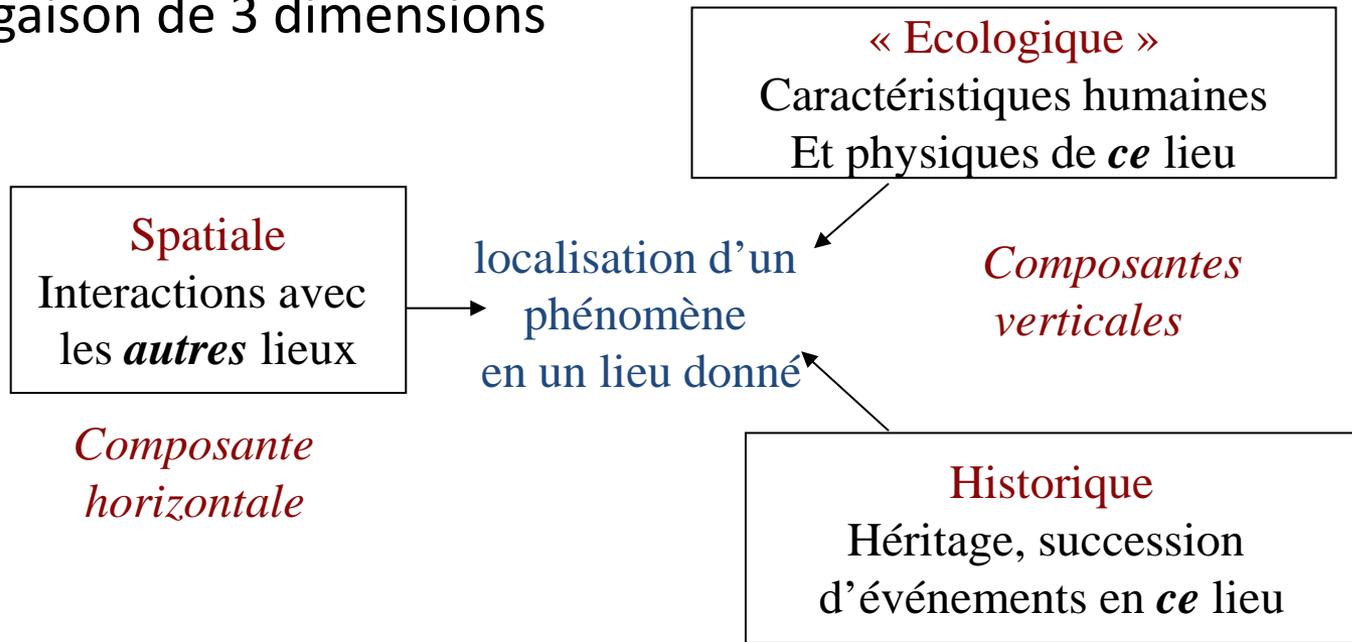
***Collecter et produire des données pour la recherche en SHS
Fréjus, 15-18 novembre 2016***

1. des données aux méthodes : catégorisation des méthodes dédiées aux données spatiales
2. exemple de déroulement : vers l'explication statistique
3. synthèse des (R)évolutions (?) en géographie (quantitative) et retour sur les implications métier

Itinéraire: des Statistiques à la GéoDataScience ?

Voyage au pays de l'analyse spatiale

- Expliquer la localisation d'un phénomène en un lieu.
Conjugaison de 3 dimensions



- Ensemble de techniques et de modèles qui appliquent des structures formelles, généralement quantitatives, à des systèmes dans lesquels la principale variable évolue de façon significative à travers l'espace (Longley, Batty, 1996)

D'après Durand-Dastès (Géopoint 1990)
et Pumain, Saint-Julien (L'analyse spatiale, 1997,
cursus, Colin)

Nouvelles données – nouvelles méthodes ?

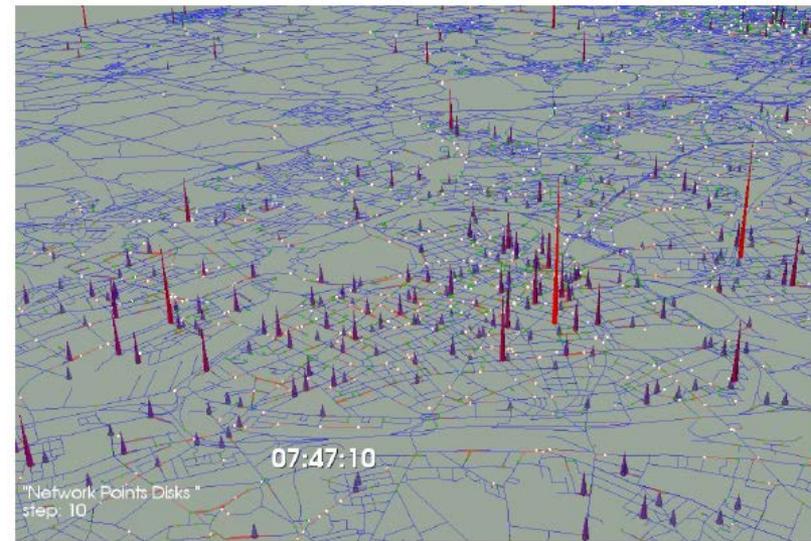
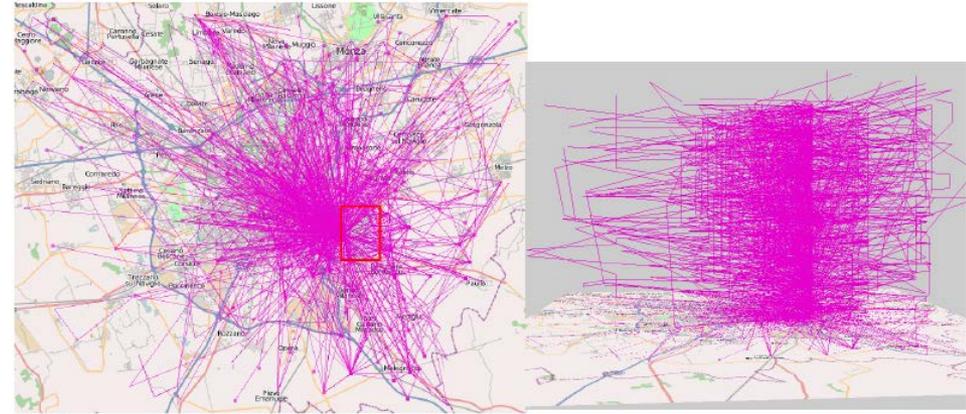


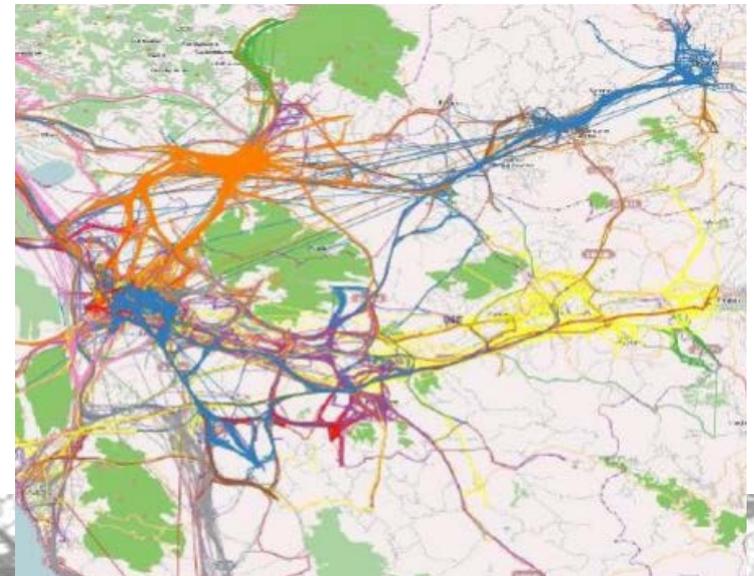
Figure 13 : visualiser la « fourmilière urbaine » MIRO, Banos & al

Examples of movement data: calls of WIND customers in Milano

- 10 days; 550,000 customers; about 5,000,000 calls

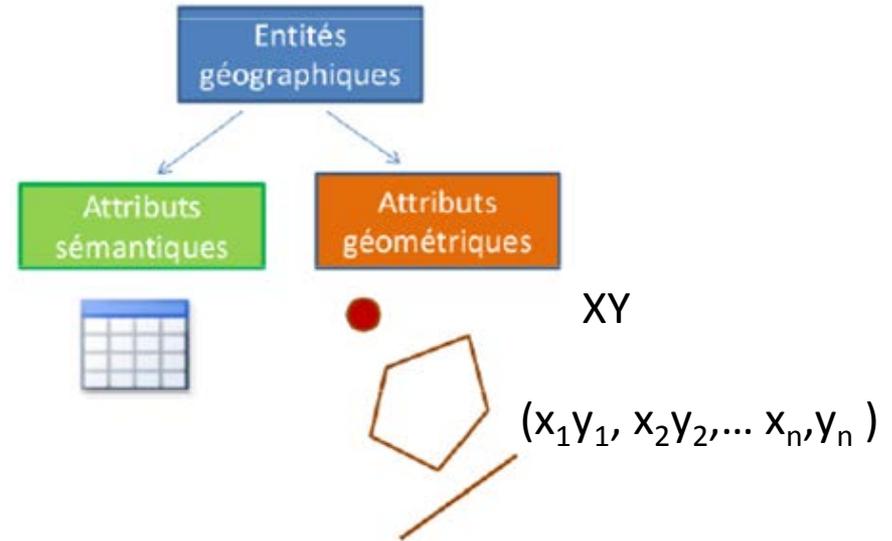


- Nouvelles données => **micro-données** (points ou maillages)
- Mais changement de niveau d'observation => changement de d'analyse ?

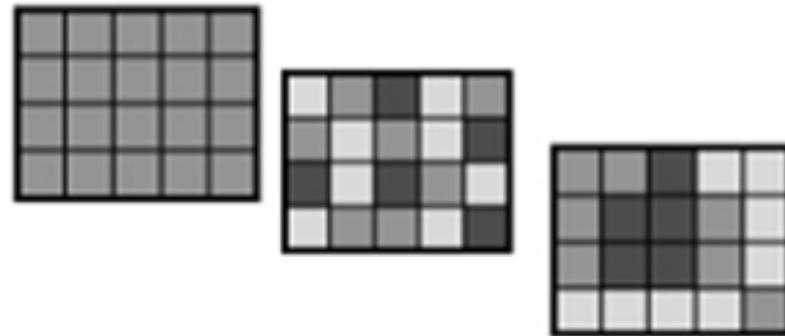
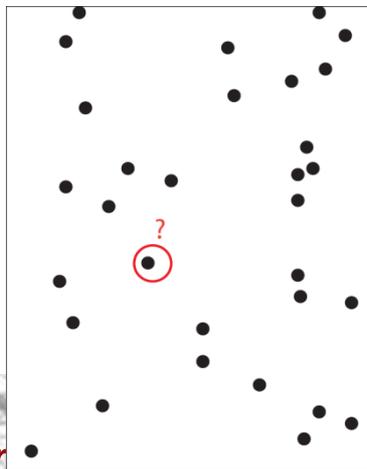


Des données aux méthodes

- Données géographiques:



- Catégories de méthodes éprouvées



Les points « événements » ou « cas »

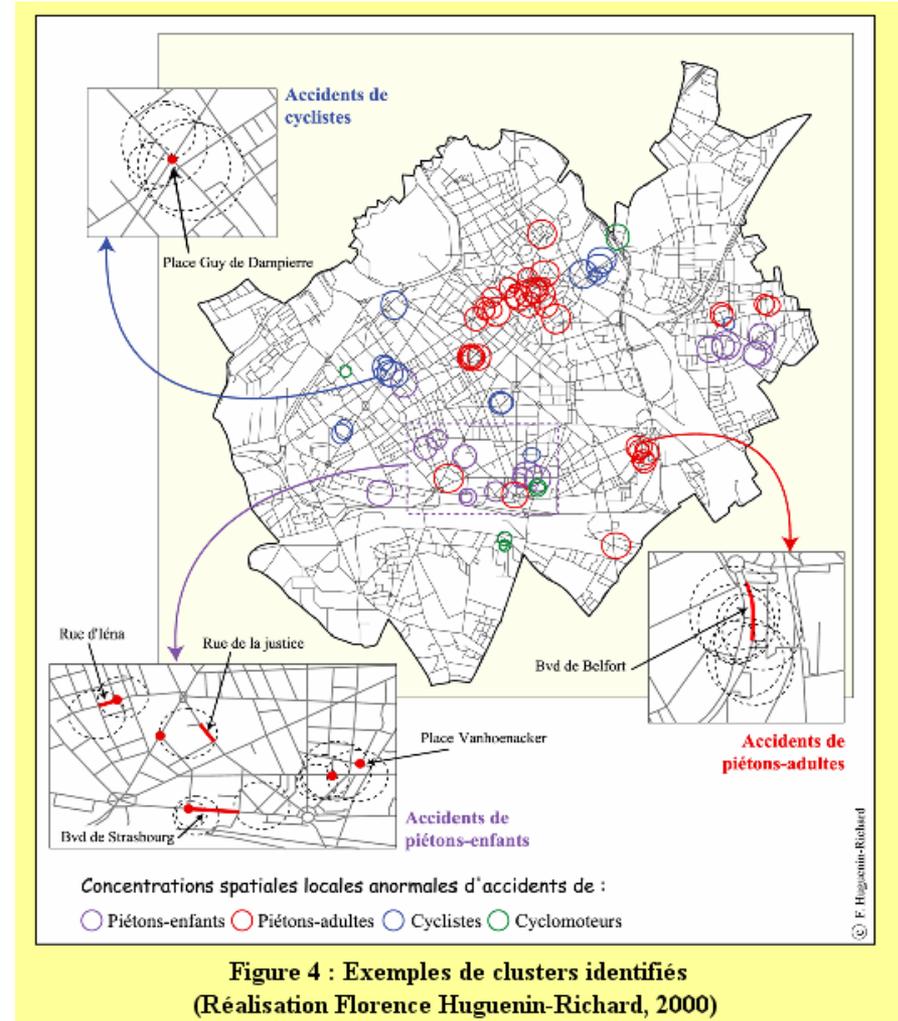
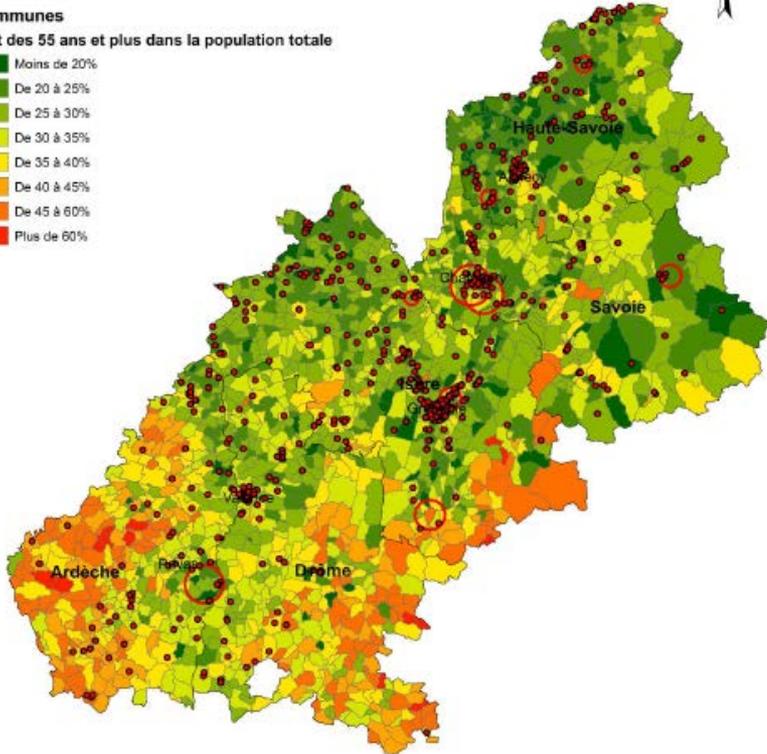
Population de 55 ans et plus et clusters détectés de SLA

- Cas de SLA entre 2003 et 2011
- Cluster détecté par SaTScan
- Préfecture
- Limites départementales

Communes

Part des 55 ans et plus dans la population totale

- Moins de 20%
- De 20 à 25%
- De 25 à 30%
- De 30 à 35%
- De 35 à 40%
- De 40 à 45%
- De 45 à 60%
- Plus de 60%



Les cas d'une maladie
~ population résidente, emplois...

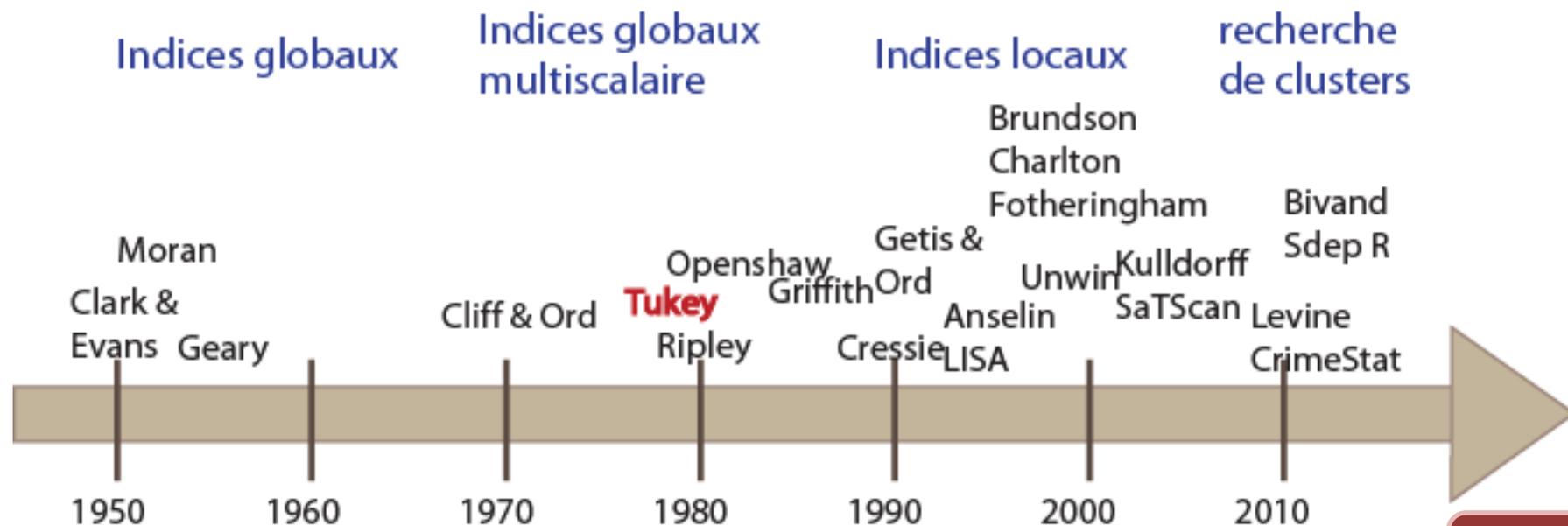
Les « événements » accidents
~ réseau viaire
~ différenciation par type

1 -des données
aux méthodes

Les points « événements » ou « cas »

- Méthodes qui proviennent d'un dialogue ancien entre statistique et géographie

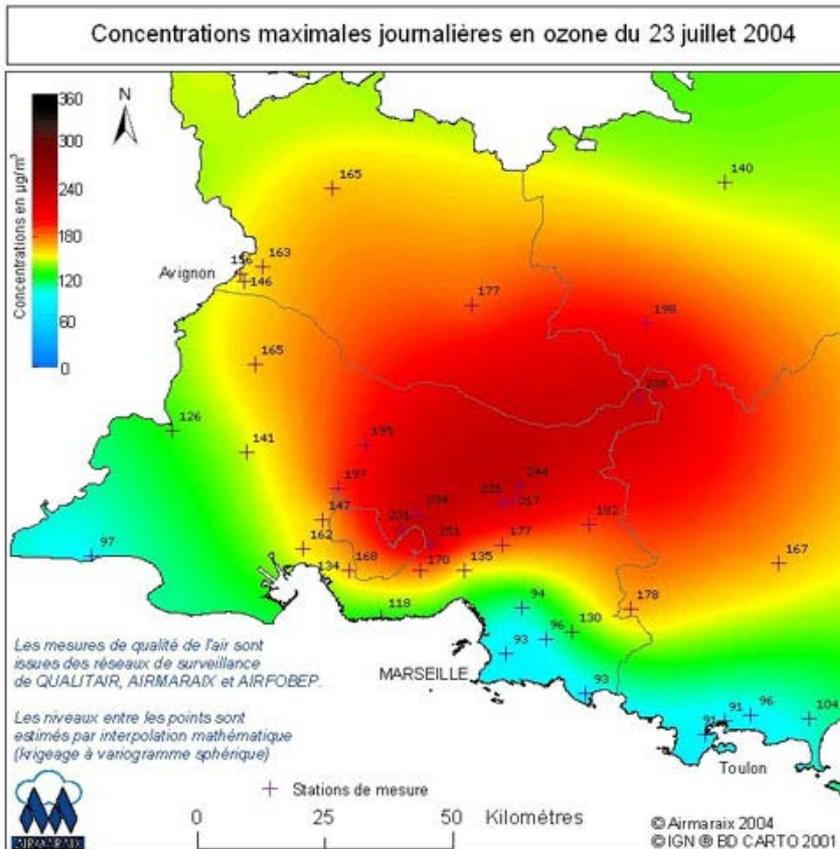
=> *statistiques spatiales*



Les points « échantillonnés »

Echantillons d'un champ continu : phénomène environnementaux (qualité de l'air, pollution, température..)

=> **Géostatistique** (méthodes d'interpolation spatiale, Krigeage par exemple Matheron, 1962)



6, Fréjus

1 -des données aux méthodes

Les points d'un réseau, de trajectoires

- Analyse de la structure => indices issus de *la théorie des graphes*

- Représentation des relations

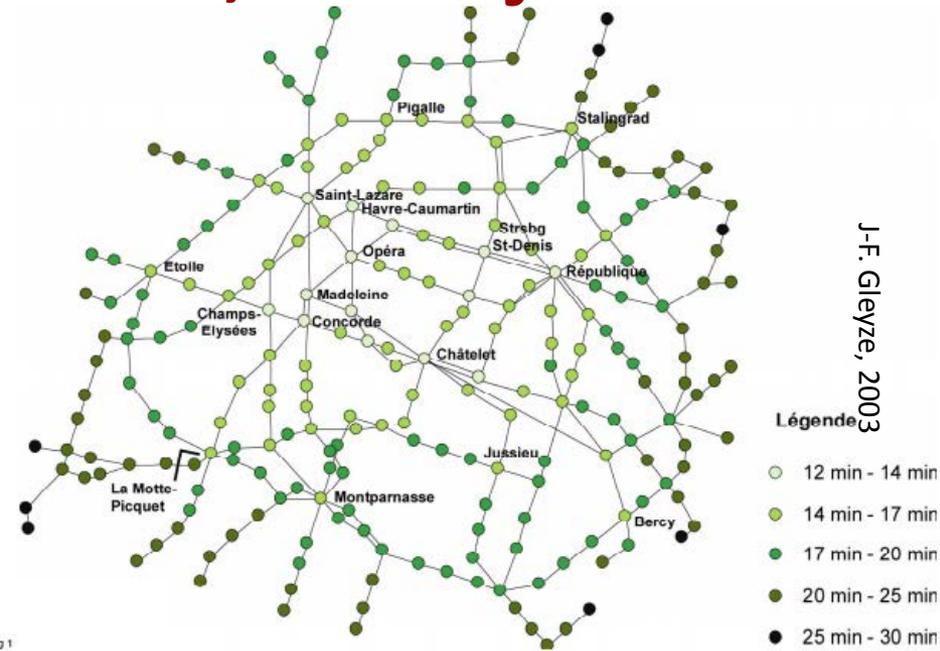
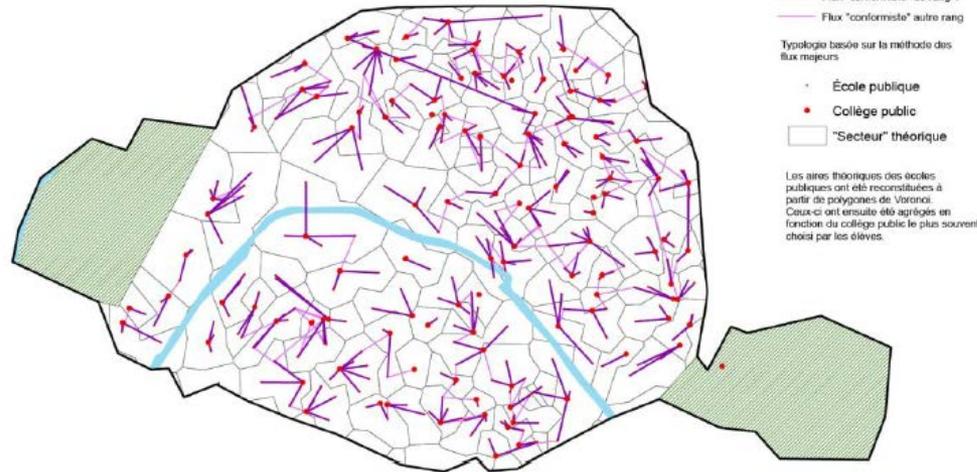
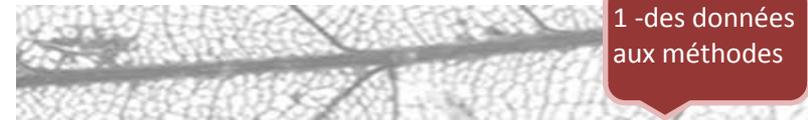


figure B.9.4 : indice d'éloignement moyen sur le réseau de métro parisien



0 2 km

Source : Éducation nationale, 2012



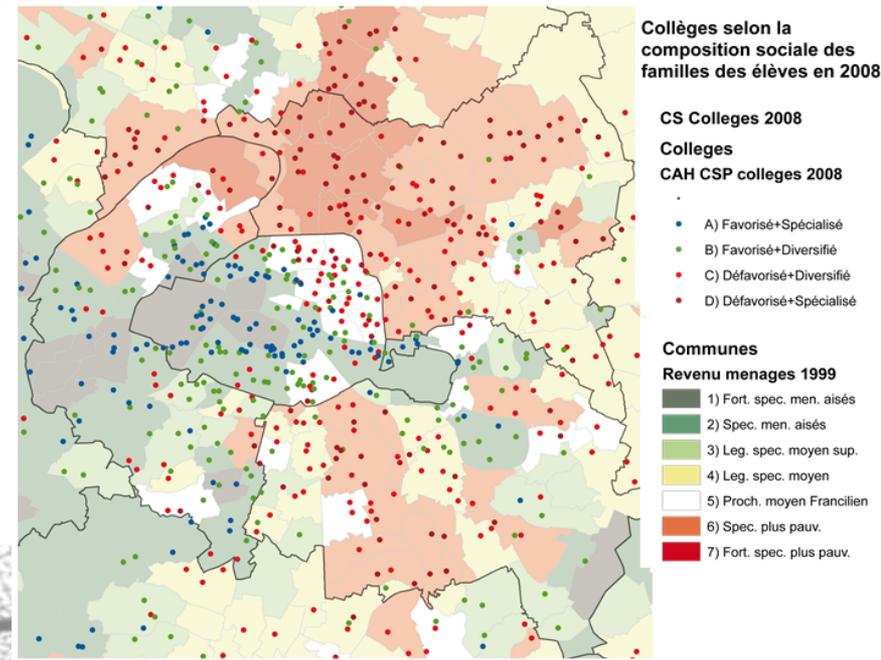
1 -des données aux méthodes

Les points « entités complexes »

Exemple des collèges à Paris=> processus s'inscrivant dans une inter-relation complexe entre acteurs de différents niveaux, logiques top-down et bottom-up:

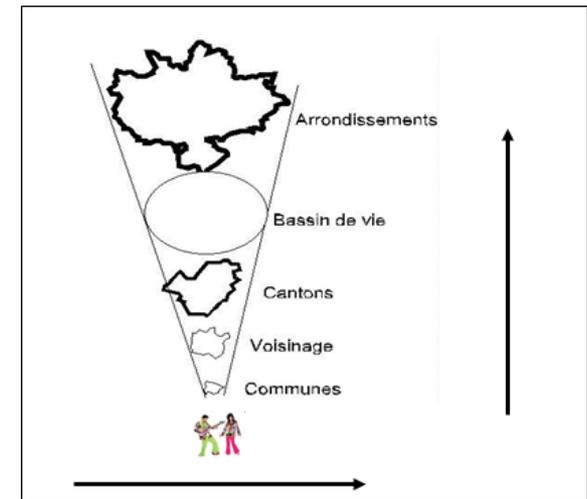
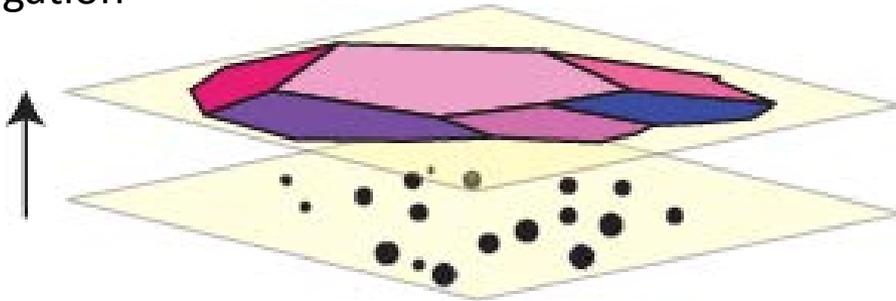
➤ analyse de la distribution spatiale avec intégration d'éléments de contexte

➤ Mixte de ***statistiques spatiales*** et de méthodes ***statistiques appliquées à des données géographiques***



Des points aux zones

Agrégation



- Quel niveau géographique?
- Problème du MAUP (Modifiable Areal Problem)
- Problème de l'autocorrélation spatiale (violation des conditions stat

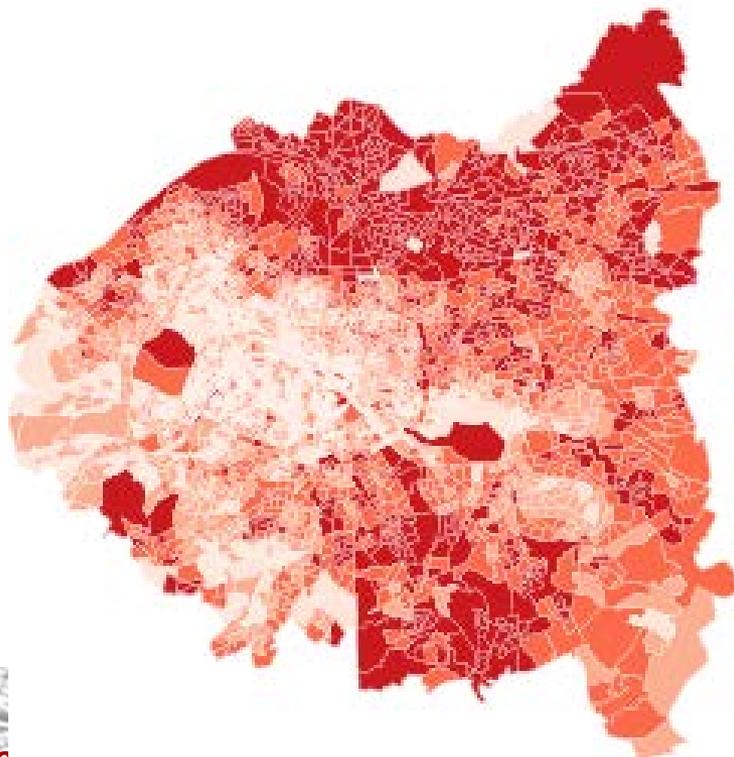
		THEMATIQUE			
		1	2	3	4
ESPACE	Région1	7	4		
	i				

1 -des données aux méthodes

Données géographiques agrégées et statistiques spatiales

- Méthodes de statistiques spatiales: indice locaux d'autocorrélation spatiale (régions significatives)

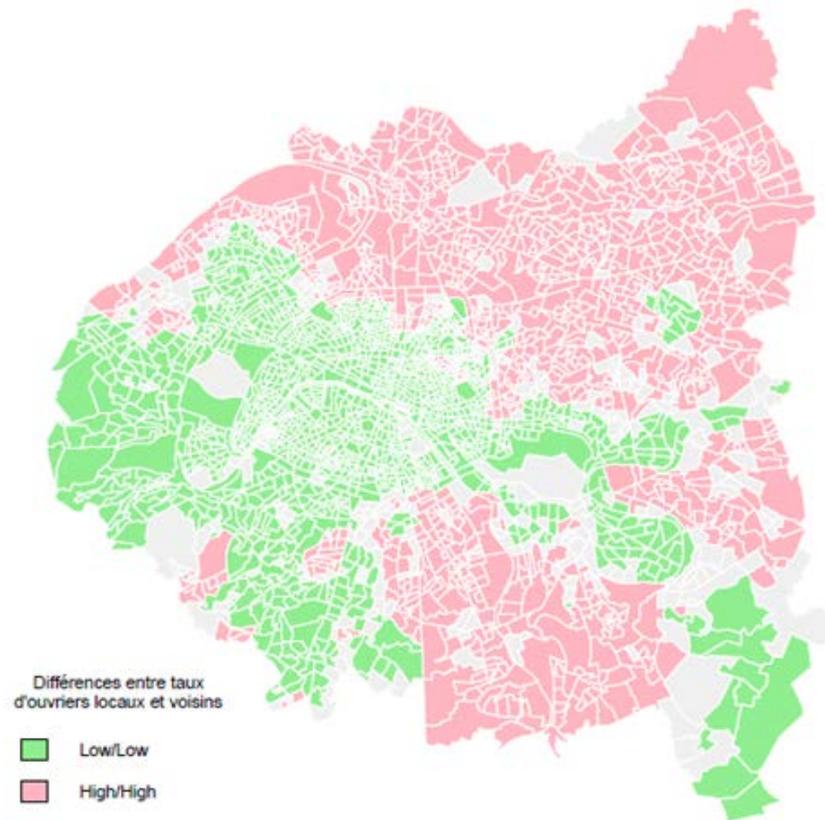
% d'ouvriers en 2007 par IRIS



% d'ouvriers



Taux d'ouvriers : Hotspots & «Lowspots»

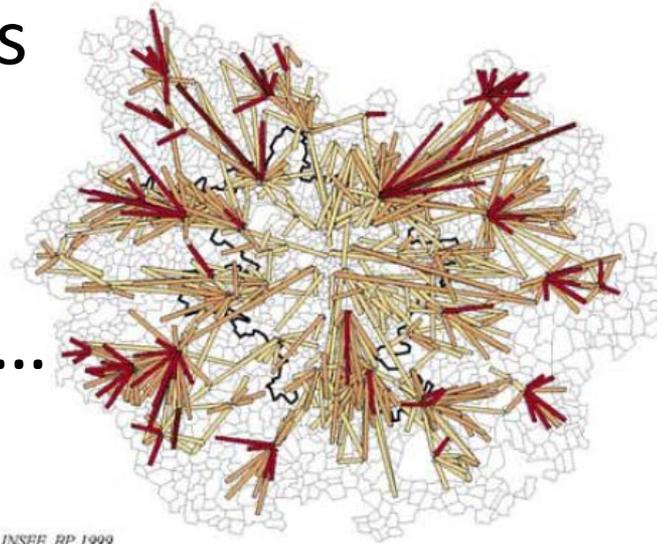


Différences entre taux d'ouvriers locaux et voisins

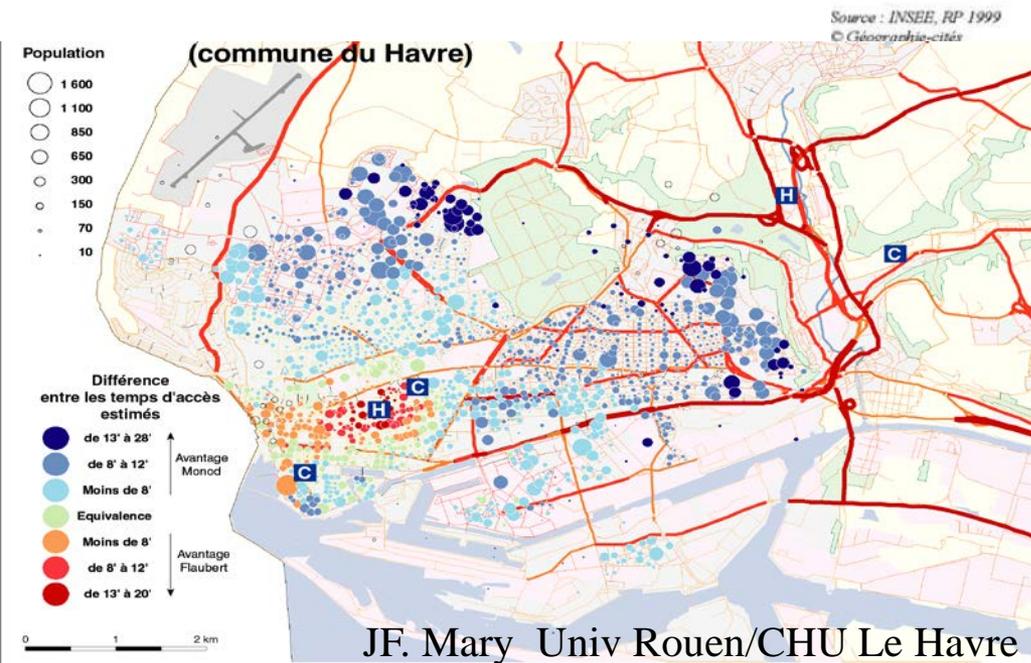
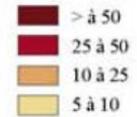


Les Liens entre les lieux : Origine-Destination

- Analyser les structures en intégrant les fonctionnements => accessibilités, flux,...



Part du flux parmi les ouvriers actifs ayant un emploi et résidant dans la commune de départ (en %)

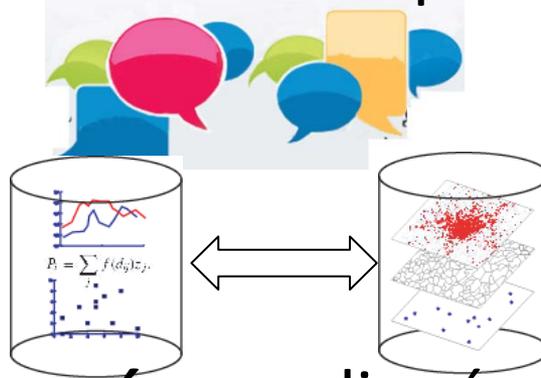


Méthodes spécifiques aux données géographiques

➤ les *statistiques spatiales*

➤ *l'analyse géostatistique*

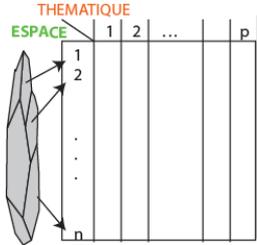
Nourrissent les méthodes exploratoires des masses de données (Fouille de données spatiales et GeoDataViz)



➤ ***l'analyse de données*** appliquée aux données géographiques (exploratoire et confirmatoire)

Prise en compte de la dimension spatiale dans l'analyse des données

Espace support: \longleftrightarrow Espace actif



region (nuts3)



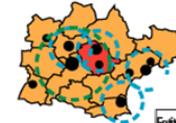
region x

neighbouring regions



neighbouring regions

regional urban system

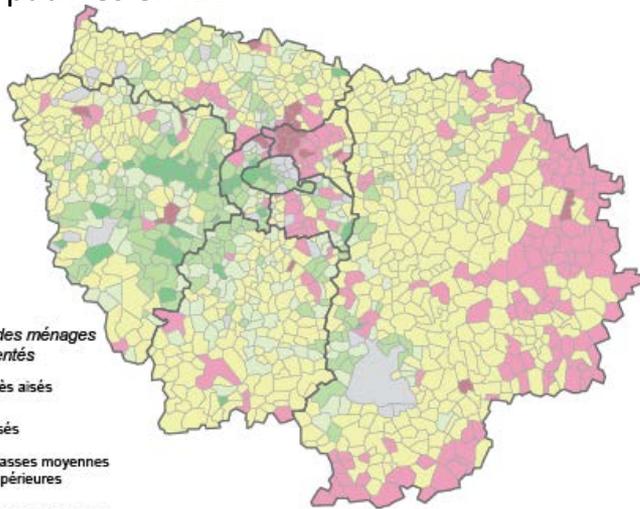


population size

area under urban attractivity

	entité		vois.		aire influence villes		
Entités spatiales élémentaires	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	...
A							
B							
C							
D							
...							

Riches et pauvres en IDF

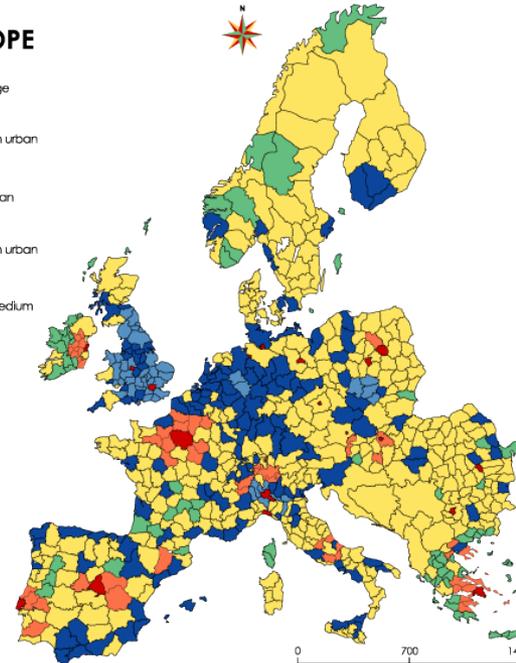


Catégorie des ménages sur-représentés

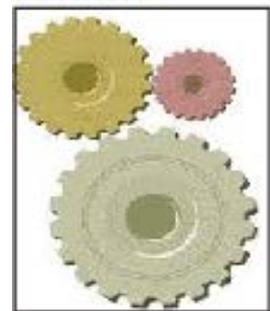
- 1 Très aisés
- 2 aisés
- 3 Classes moyennes supérieures
- 4 Classes moyennes
- 5 Pas de sur-représentation significative
- 6 Pauvres
- 7 Les plus pauvres

SETTLEMENT STYLES IN EUROPE

- regions dominated by a large metropolis
- polycentric regions with high urban and rural densities
- rural areas under metropolitan influence
- polycentric regions with high urban densities
- rural areas with small and medium sized towns
- remote rural areas



ESPACE



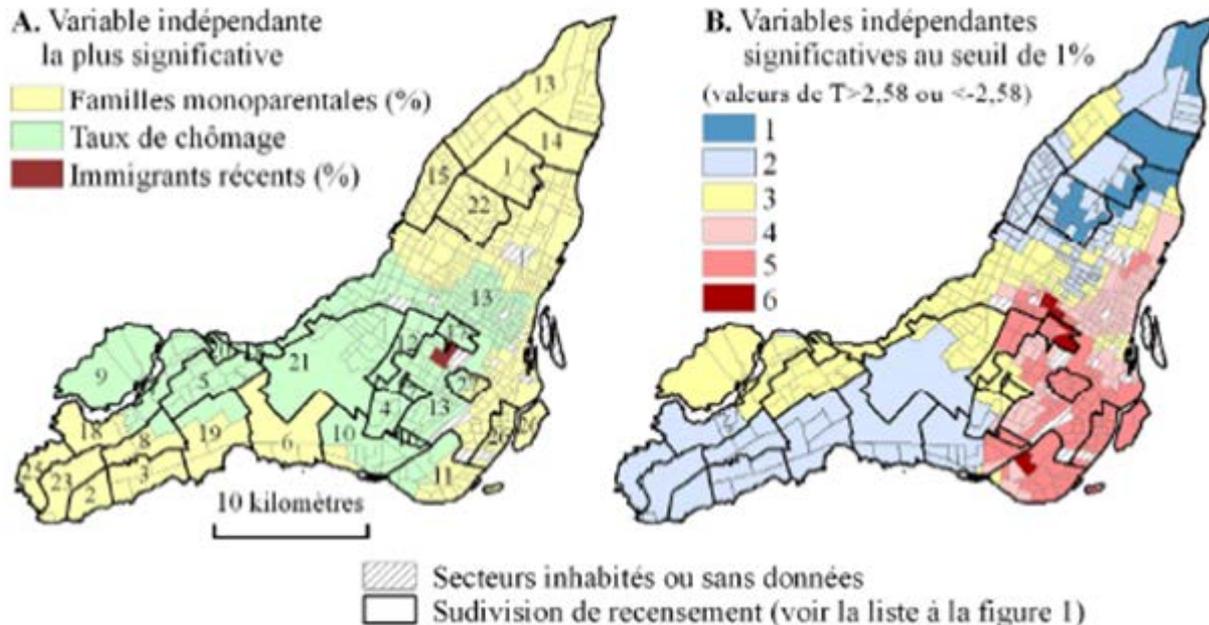
Exemple de méthodes « adaptées »

- Classification avec contrainte de contiguïté
- Régression géographique pondérée (GWR)

Expliquer la différence de revenus moyens dans les secteurs de l'île de Montréal

Résultats : le modèle géographiquement pondéré (GWR)

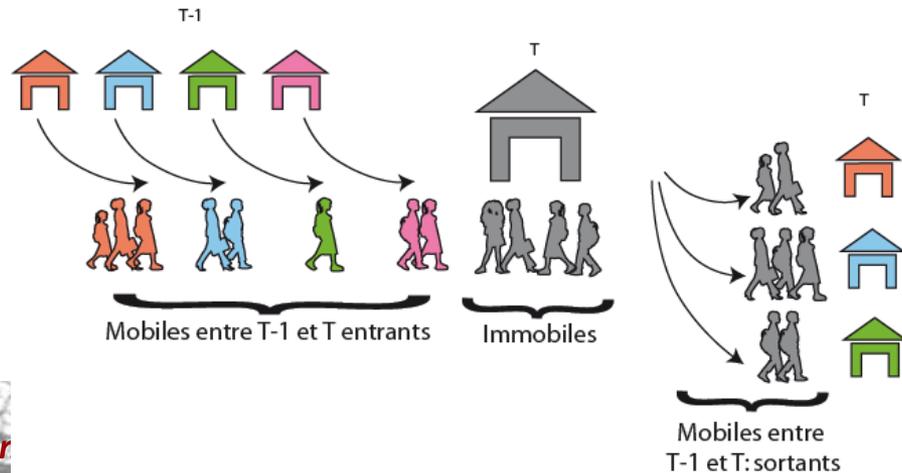
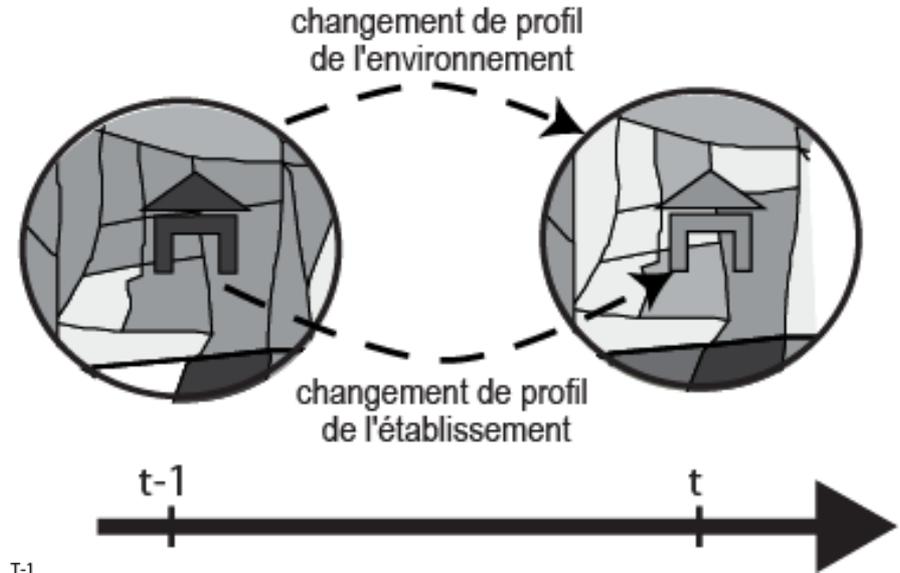
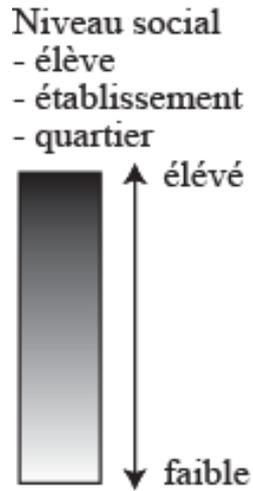
Le modèle GWR : les variables indépendantes significatives du modèle



Source : Statistique Canada, recensement de 2001

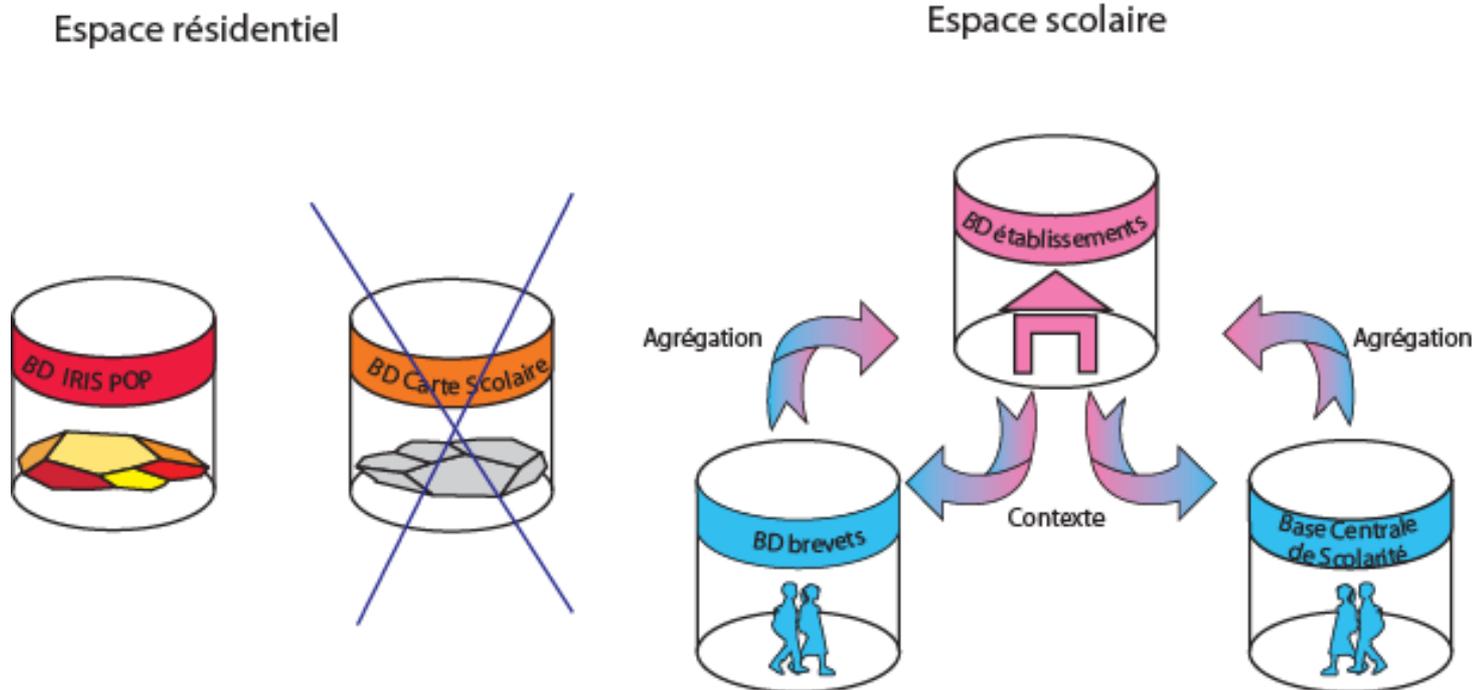
Apparicio & al. , INRS

Exemple d'un déroulement: Espace social et espace scolaire



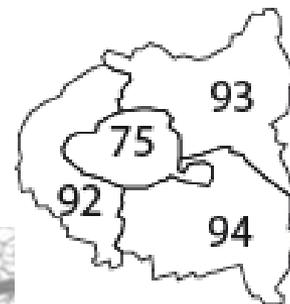
Exemple d'un déroulement: Espace social et espace scolaire

- 3 bases de données en relation (© DEPP)

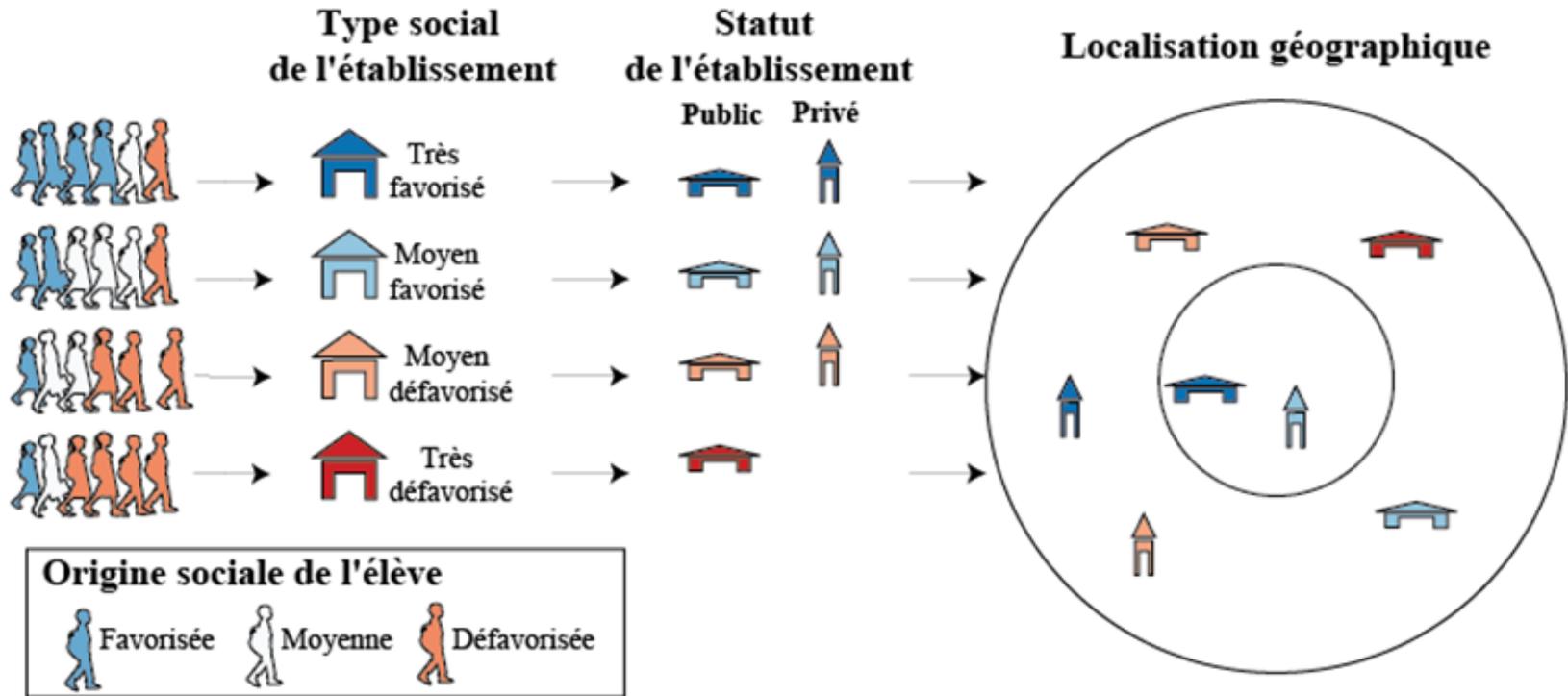


- multi-niveaux : élèves et établissements
- multi-dates : 2004 à 2011

- Focus sur Paris+petite couronne en 2010



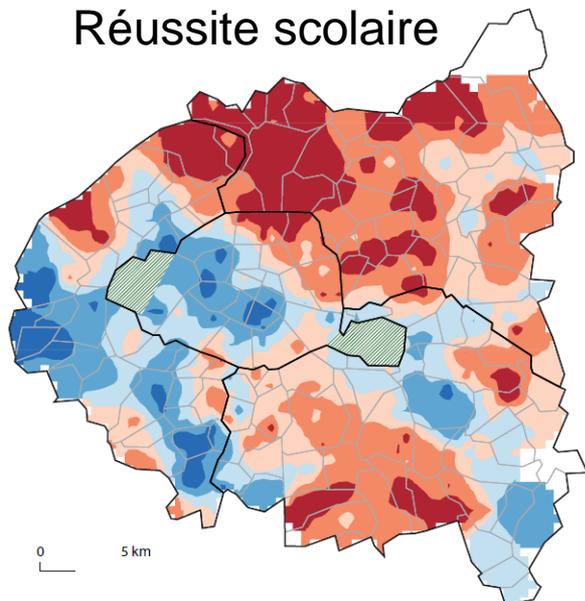
Espace des collégiens: intégration de différents niveaux



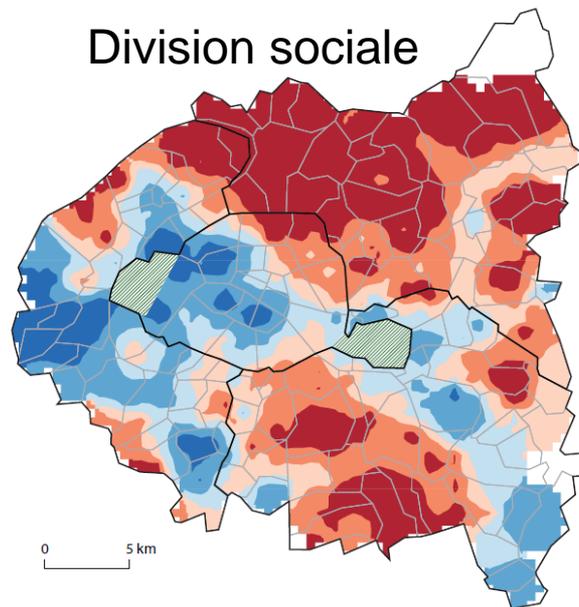
Espace des collèges: réussite scolaire et division sociale

Collèges publics

Réussite scolaire

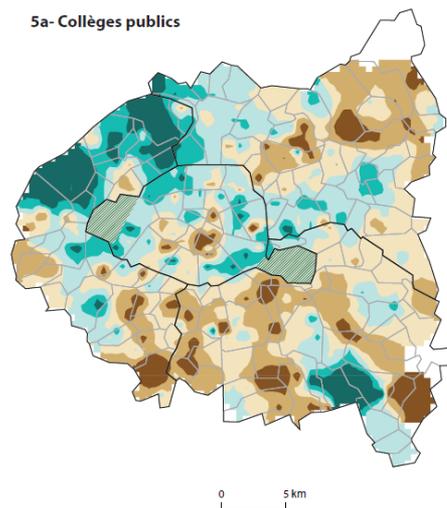


Division sociale



Source : Education nationale, 2011
Cartographie : Henry Clesielski, UMR Géographie-Cités, 2012

5a- Collèges publics

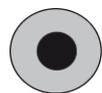


Source : Education nationale, 2011
Cartographie : Henry Clesielski, UMR Géographie-Cités, 2012

Modélisation de la réussite à l'échelon individuel

Modèle logistique (Logit)

Élèves du public



Variables	Modalité	coefficient	significativité
Sexe	Garçon	-0,15	****
	Fille	0,15	
Origine sociale	Favorisé	0,67	****
	Moyenne	-0,10	
	Défavorisée	-0,57	
Age	Avance	1,43	****
	Normal	0,21	
	Retard	-1,64	
Type social du collègue	Favorisé	0,67	****
	Moyen favorisé	0,24	
	Moyen défavorisé	-0,26	
	Défavorisé	-0,64	
Couronne d'implantation du collège	Paris	0,01	NS
	Banlieue proche	-0,01	

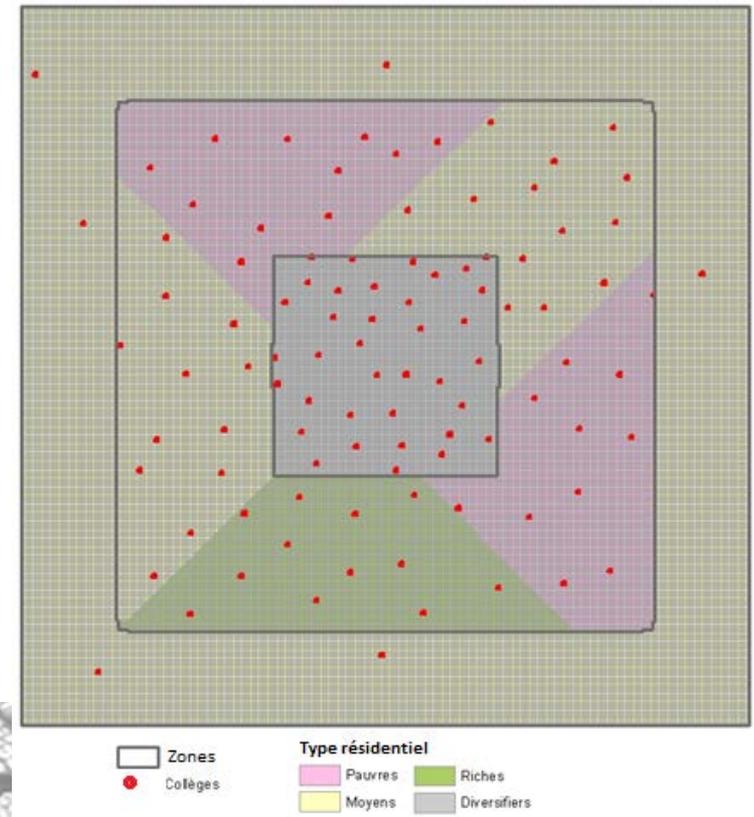
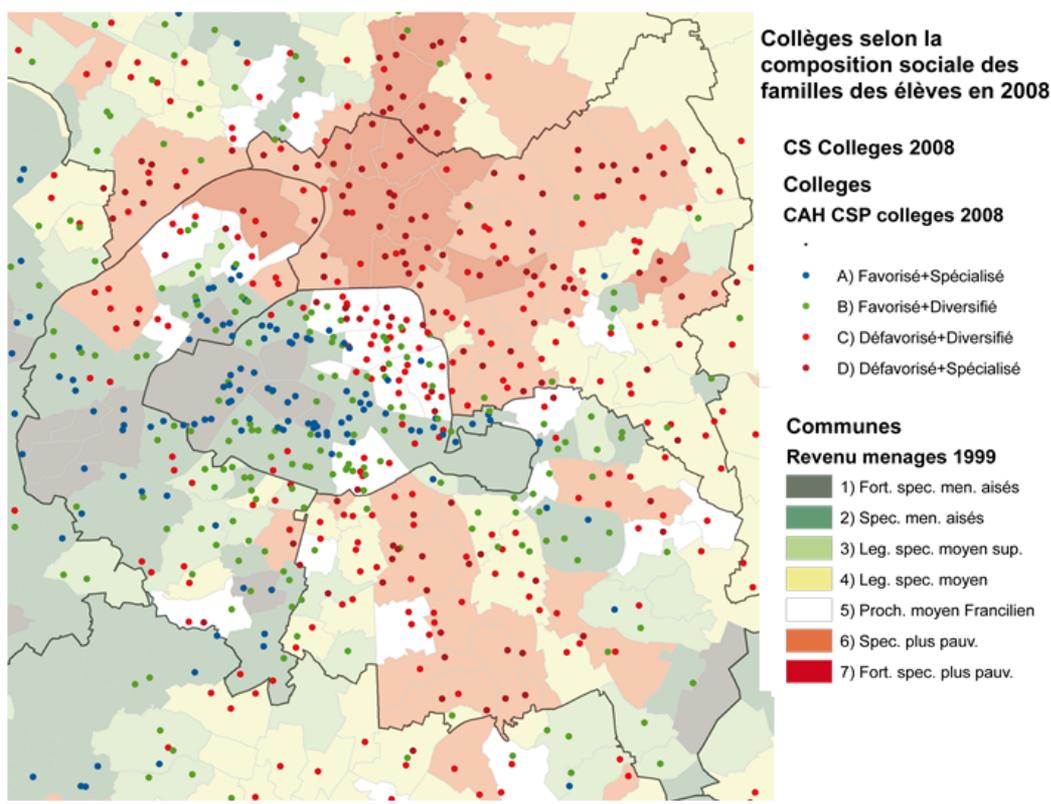
2 types d'approches exploratoires

- Statistique

Analyser des données empiriques: « faire parler les données », qu'est-il arrivé dans la « réalité »?

- Par simulation

Simuler une « expérience de pensée »: explorer les mécanismes du changement : « que se passerait il si? »



Les (R) évolutions méthodologiques en géographie

Repères *Imbrications des évolutions méthodologiques et technologiques*

1970 Analyse quantitative / Analyse spatiale

1980 Systèmes d'information géographique

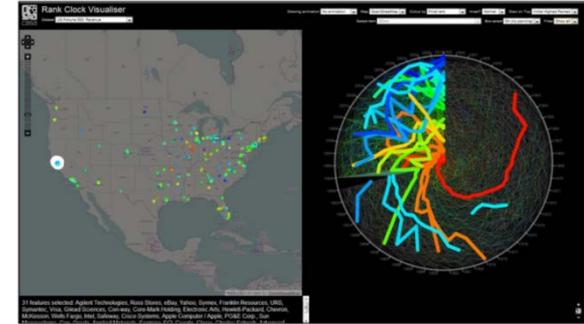
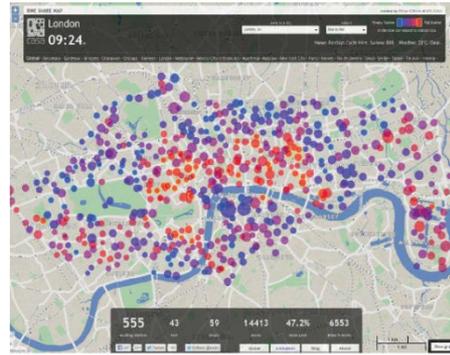
1990 Simulation informatique

2000 Environnement Web

2010 masses de données, humanités numériques ..

=> Renouveau des méthodes, ré-emploi dans des cadres différents

le renouveau de l'exploration carto& graphique



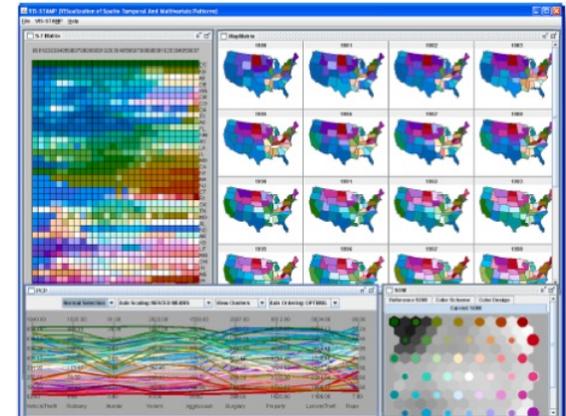
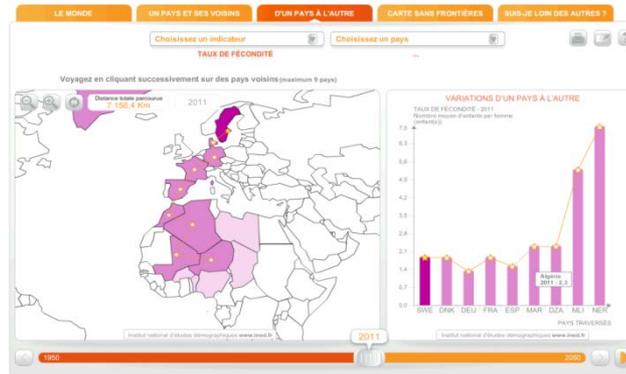
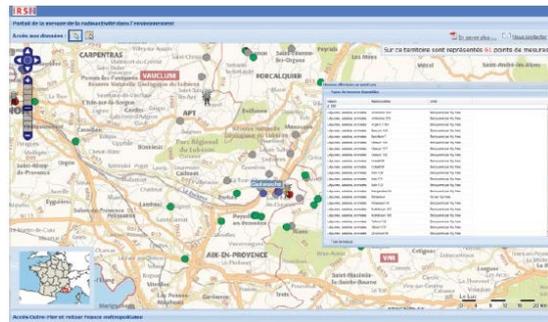
Cartographie

Cartographie interactive

géovisualisation

« geovisual analytics »

Statistique



- Questions de géo-référencement et d'analyse des organisations



Localisation des lieux nommés dans le corpus de romans (traitement partiel, visualisation indicative)

Et les big data !?!

- Questions de géo-référencement et d'analyse des organisations 😊
- Questions de contenus 😮

Exemple des tweets sur l'alimentaire à Saint-Chamond

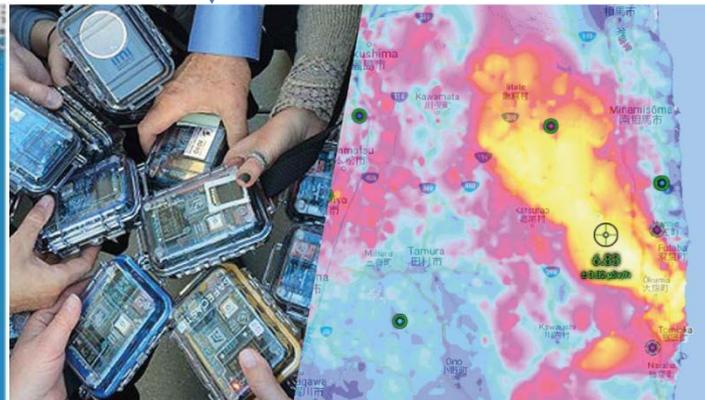
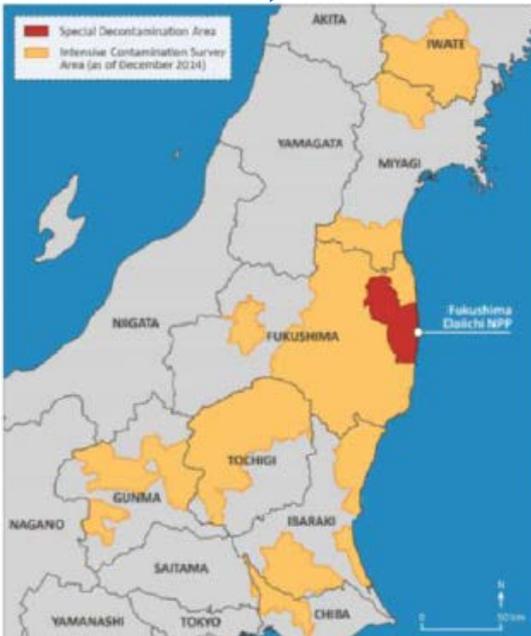


- Questions de référencement et d'analyse des organisations 😊
- Questions de contenus 😮
- Questions d'intégration des données 😏

Officielle

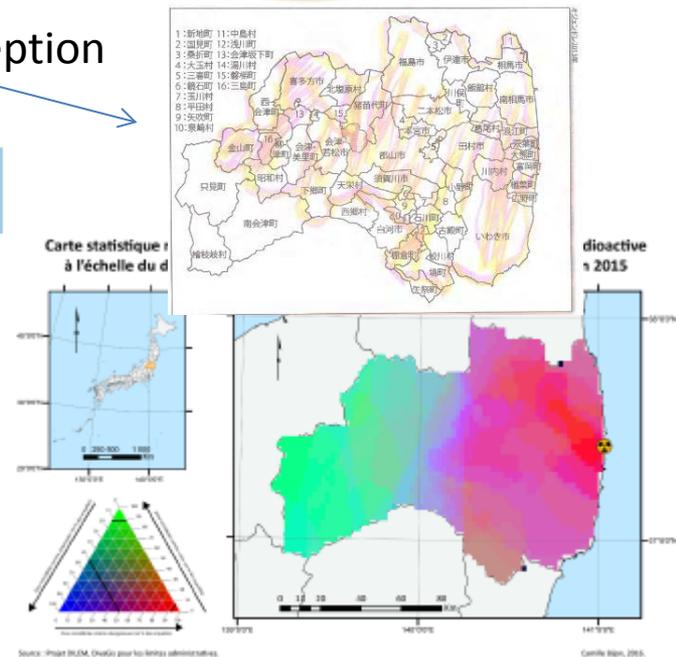
Participatives

Perception



Source : Projet Safecast, 2016.

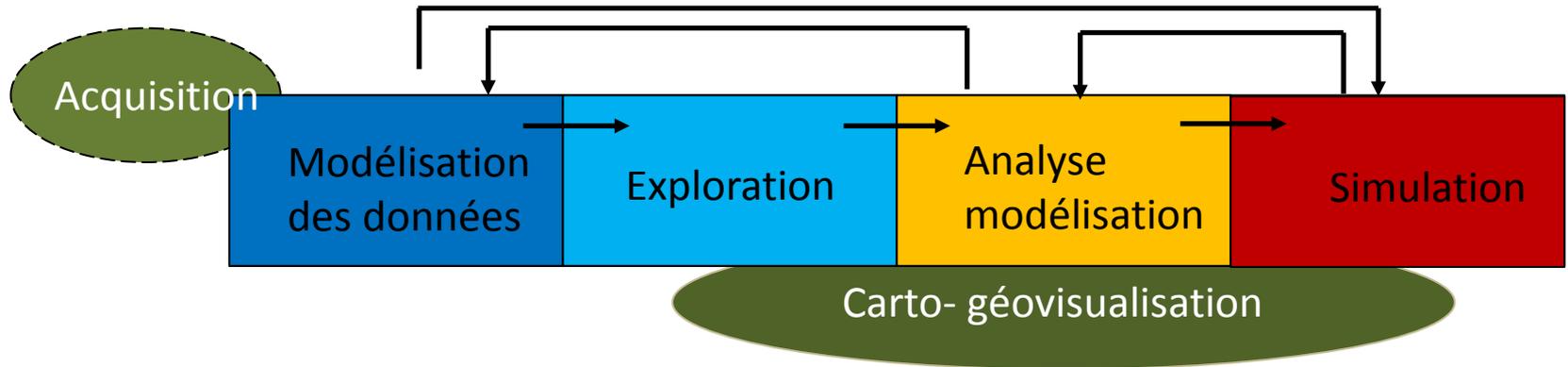
Camille Bijon, 201



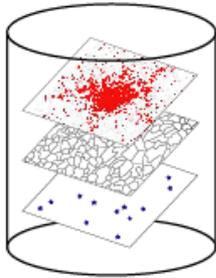
Source : Projet IJUM, Densho pour les entités administratives.

Camille Bijon, 2015.

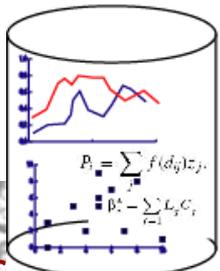
Evolution méthodologique et évolutions technologiques:



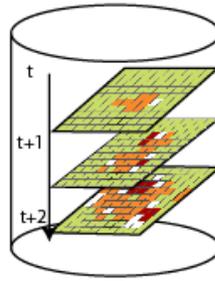
Système d'Information Géographique



Système d'analyse statistique



Modèle informatique

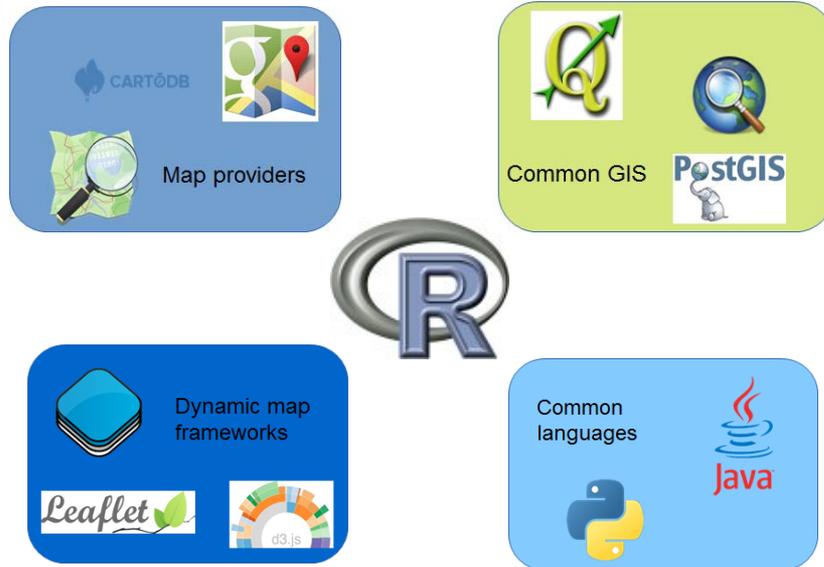


Approches qui mobilisent un couplage de méthodes et d'outils => Renouveau

Un environnement intégrateur technologique et méthodologique

Potentialité de transfert ?

Technologique pour les géographes

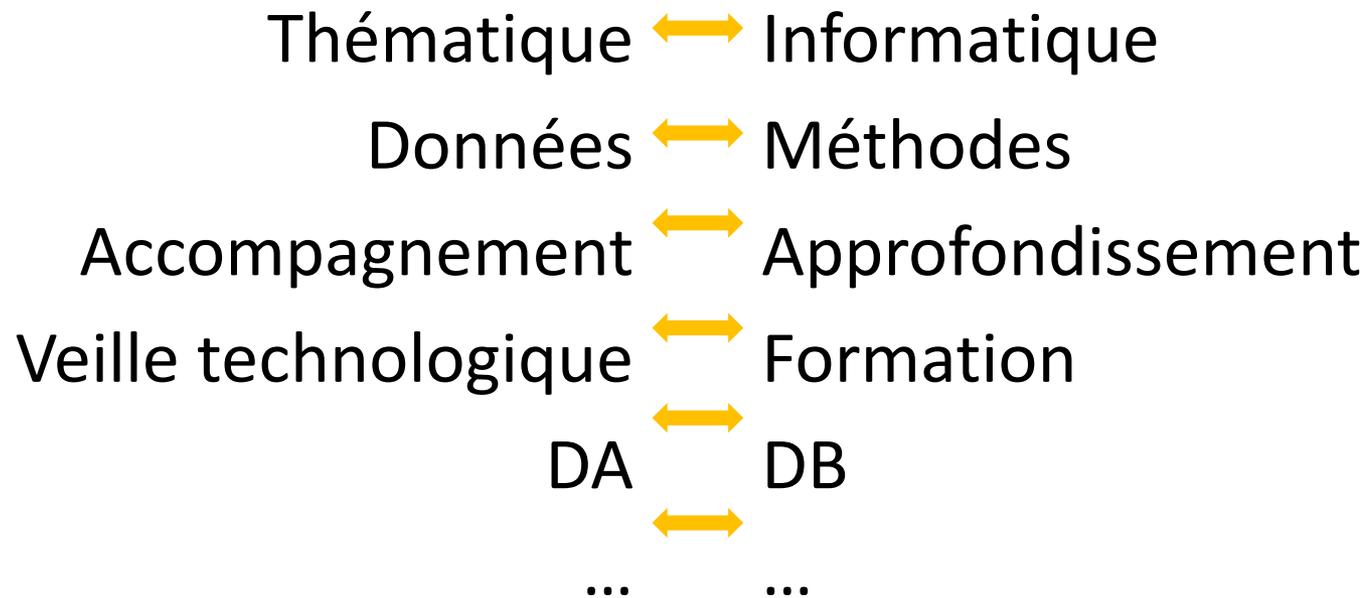


Méthodologique pour les autres disciplines

Packages de gestion de données spatiales et cartographie sous R

- maptools
- sp
- Rgdal
- Rgeos
- mapproj
- Rcartography
-

Retour sur le métier





« passeurs » - « facilitateur »

Thématique – Méthode - Technique

Ingénieur - chercheur

Entre domaines

...

Statistiques et géographie, Analyse spatiale, Modèles d'analyse spatiale

- **Béguin M., Pumain D.** 1994: La représentation des données géographiques, *Cursus, Armand Colin*.
- **Charre J.**, 1995, Statistique et territoire, *coll. espaces modes d'emploi, RECLUS*
- **Fotheringham A.S, Brudson C, Charlton M**, 2000, Quantitative geography, SAGE London
- **Hagget P.**, 1973, L'analyse spatiale en géographie, Paris, *A.Colin*.
- **Mathian H., Sanders L**, 2014, Approches spatio-temporelles. Objets géographiques et processus de changement, ISTE Edition, 178 p., col. Focus
- **Miller H.J., Han J**, 2009, Geographic Data Mining and Knowledge Discovery, Second Edition, CRC Press, 486 p-
- **O'Sullivan D. Unwin D.J**, 2010, Geographic information analysis, Wiley
- **Rogerson, P.A**, 2010, Statistical methods for geography : a students guide, SAGE
- **Saint-Julien T., Pumain D.** 1997 L'analyse spatiale: 1.localisation dans l'espace, *Cursus, Armand Colin*.
- **Saint-Julien T., Pumain D.** 2001 Les interactions spatiales, *Cursus, Armand Colin*.
- **Sanders L.** (dir), 2001, Modèles en analyse spatiale, Paris, Hermès lavoisier.
- **Sanders L.**, 1989, L'analyse statistique des données en géographie, *Alidade-RECLUS*

Revue,

- Revue internationale de géomatique (<http://rig.revuesonline.com/>)
- Cybergéo – (<https://cybergegeo.revues.org/>)
- M@ppemonde – (<http://mappemonde.mgm.fr/>)
- Applied Geography (<http://www.sciencedirect.com/science/journal/0143622>)