



*Les temps du territoire*

*Jeudi 21 mai 2015*

# Géohistoire du lien réseau/territoire en France entre 1860 et 1930

*Une approche géographiquement pondérée*

Christophe MIMEUR<sup>1</sup>

Thomas THEVENIN<sup>1</sup>

Gilles VUIDEL<sup>2</sup>

Ludovic GRANJON<sup>3</sup>

Robert SCHWARTZ<sup>4</sup>

<sup>1</sup> TheMA, UMR 6049 CNRS, Dijon, France

<sup>2</sup> TheMA, UMR 6049 CNRS, Besançon, France

<sup>3</sup> MSH, USR 3516 CNRS, Dijon, France

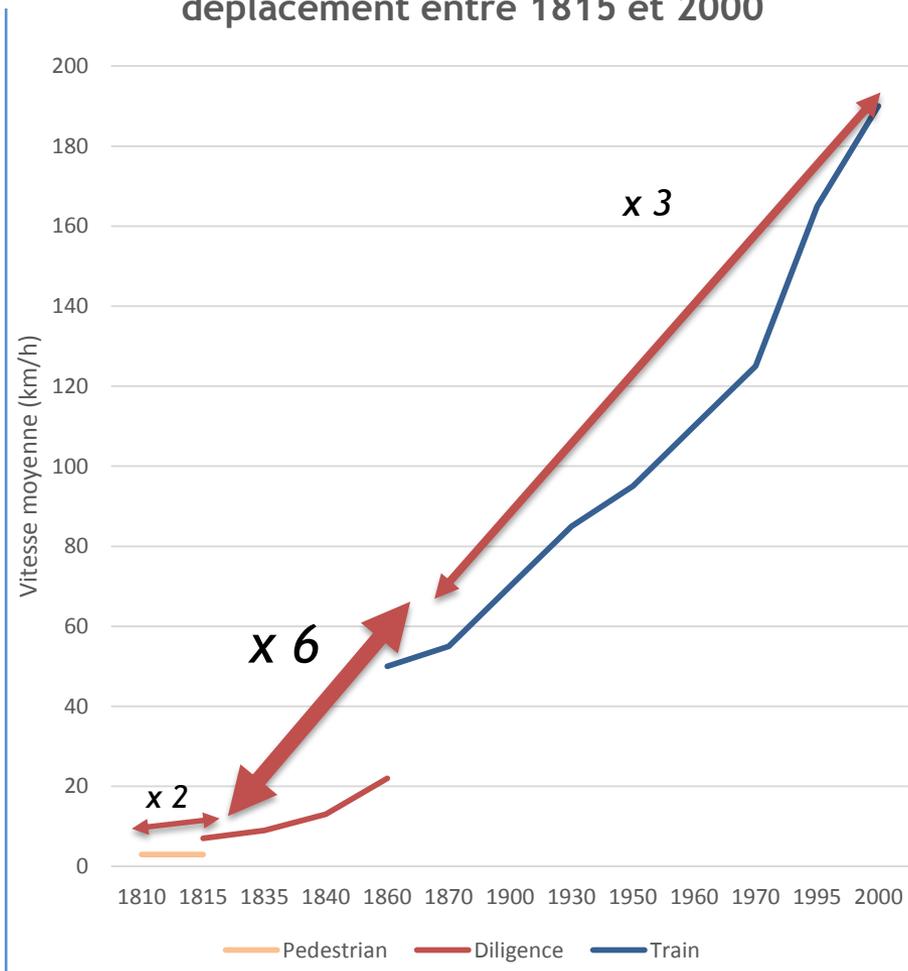
<sup>4</sup> Mount Holyoke College, Massachusetts, United States

- ▶ Envisager le lien réseau/territoire dans la longue durée
- ▶ La structuration des données
- ▶ Les apports de l'approche géographiquement pondérée



# « Un vertigineux rétrécissement spatio-temporel »

### Evolution de la vitesse moyenne de déplacement entre 1815 et 2000



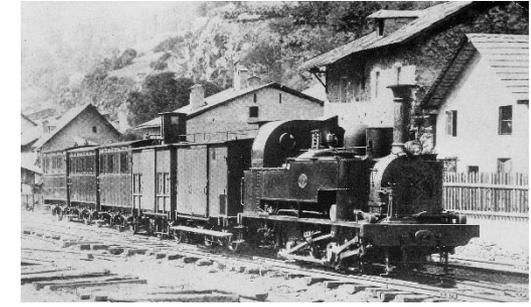
De l' « immense piétinement des Hommes » (Bavoux et al., 2005) ...



... à la traction hippomobile ...



... à la « révolution de la vitesse » (Studeny, 1995)



La notion de « grande vitesse » est relative



## Nos deux objets de recherche

- Question de recherche  
Structuration des données  
Résultats

### RESEAU



*Transport system*

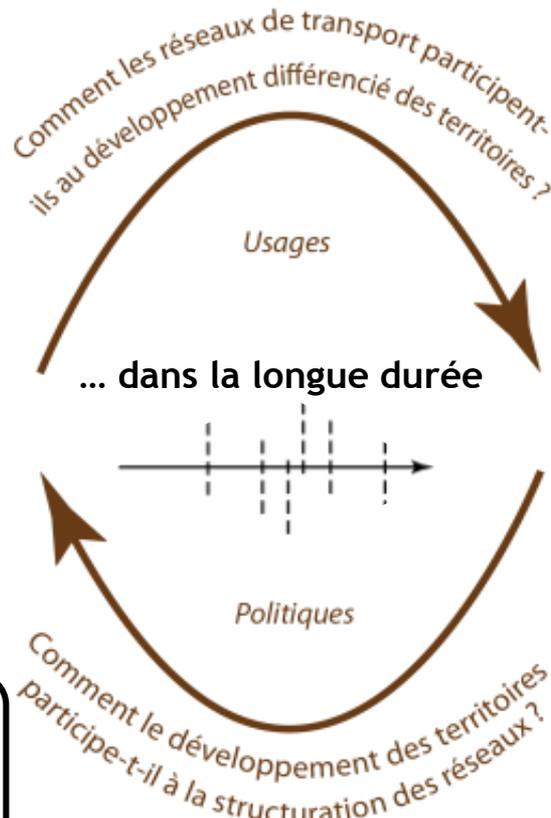
L'évolution du réseau ferroviaire peut être considéré comme un « marqueur des traces du temps » (DURAND-DASTES, 1998)

### TERRITORY



*Demographic dynamics*

« La taille démographique est un excellent (et le meilleur) résumé de très nombreuses propriétés fonctionnelles » (Pumain, 1997)



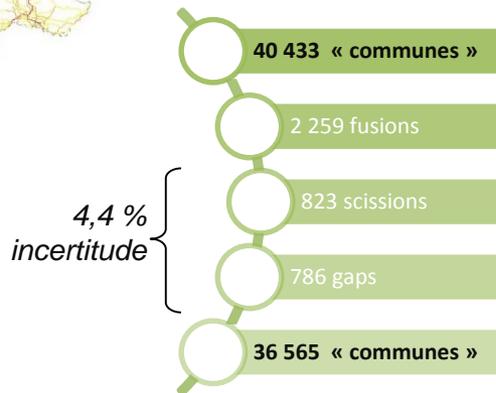
### Inscription dans le champ géohistorique, qui suppose deux conditions :

- Saisir le « recul et la prudence de l'histoire » (BRETAGNOLLE, 2014)
- Une « exigence épistémologique de contextualisation de l'explication » (PASSERON, 1991 ; PUMAIN, 2014)

# Du SIG-H aux indicateurs d'accessibilité ...

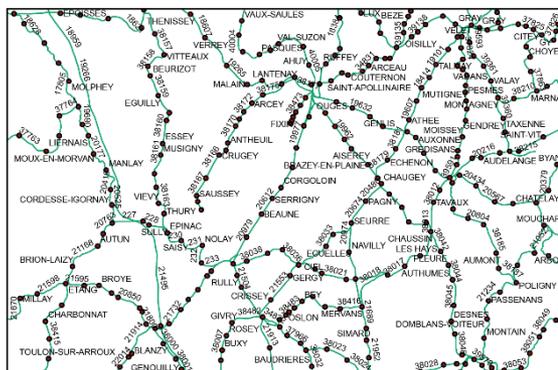


## La donnée démographique



MOTTE Claude et al. (2003). *Communes d'hier, communes d'aujourd'hui*, INED, Paris

## La donnée réseau



Une diversité des sources historiques pour reconstruire près de 43 000 tronçons : textes, plans, statistiques, ...

### L'information « gare »

● Points

Toponyme
Commune
Nature
Ouverture
Fermeture

### L'information « tronçon »

▣ Lignes

Nature
Largeur
Ouverture
Fermeture
Réouverture
Electrification
Vitesse maximale
Vitesse moyenne
Impédance
Classement
Opérateur

■ Une information temporelle continue  
■ Une information temporelle par décennie

## La construction des indicateurs d'accessibilité

Un accès au réseau forcément discontinu (STATHOPOULOS, 1997)



Origin



Station 1

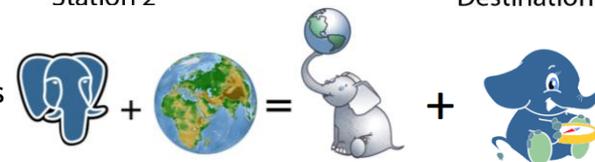


Station 2



Destination

→ Relocalisation du SIG-H dans un SGBD : changement dans la structuration des données



→ Quelles performances ?

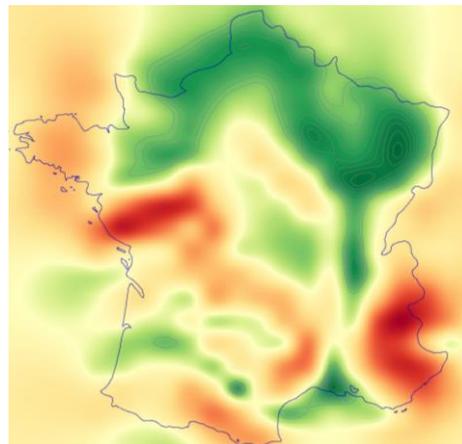
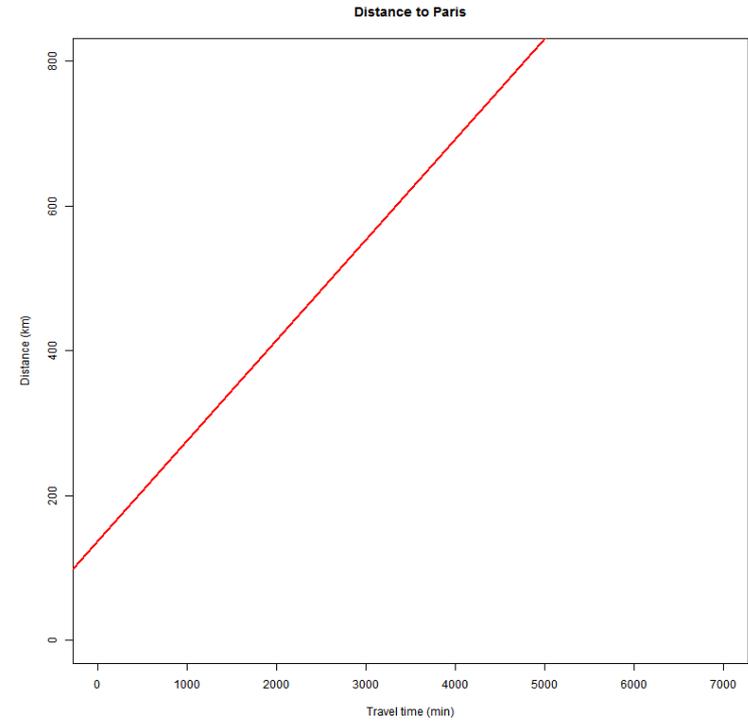
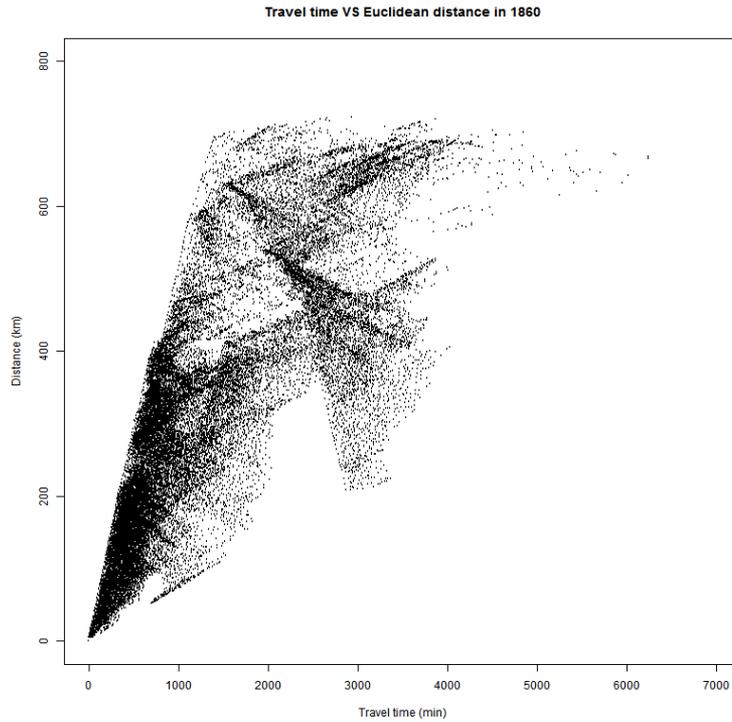
- Matrice origine/destination géante
- Temps de calcul et stockage difficilement gérable

Faire des choix dans la construction des indicateurs d'accessibilité

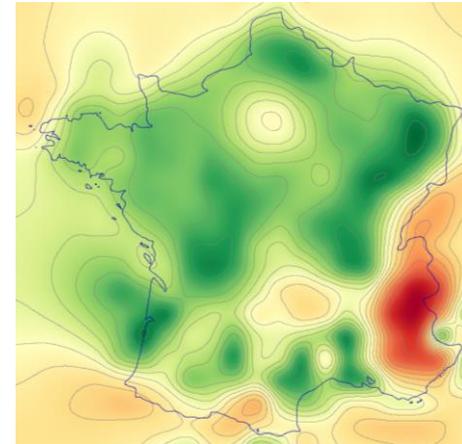


# L'essor d'une rapidité différenciée ...

► **Résultats**



1860

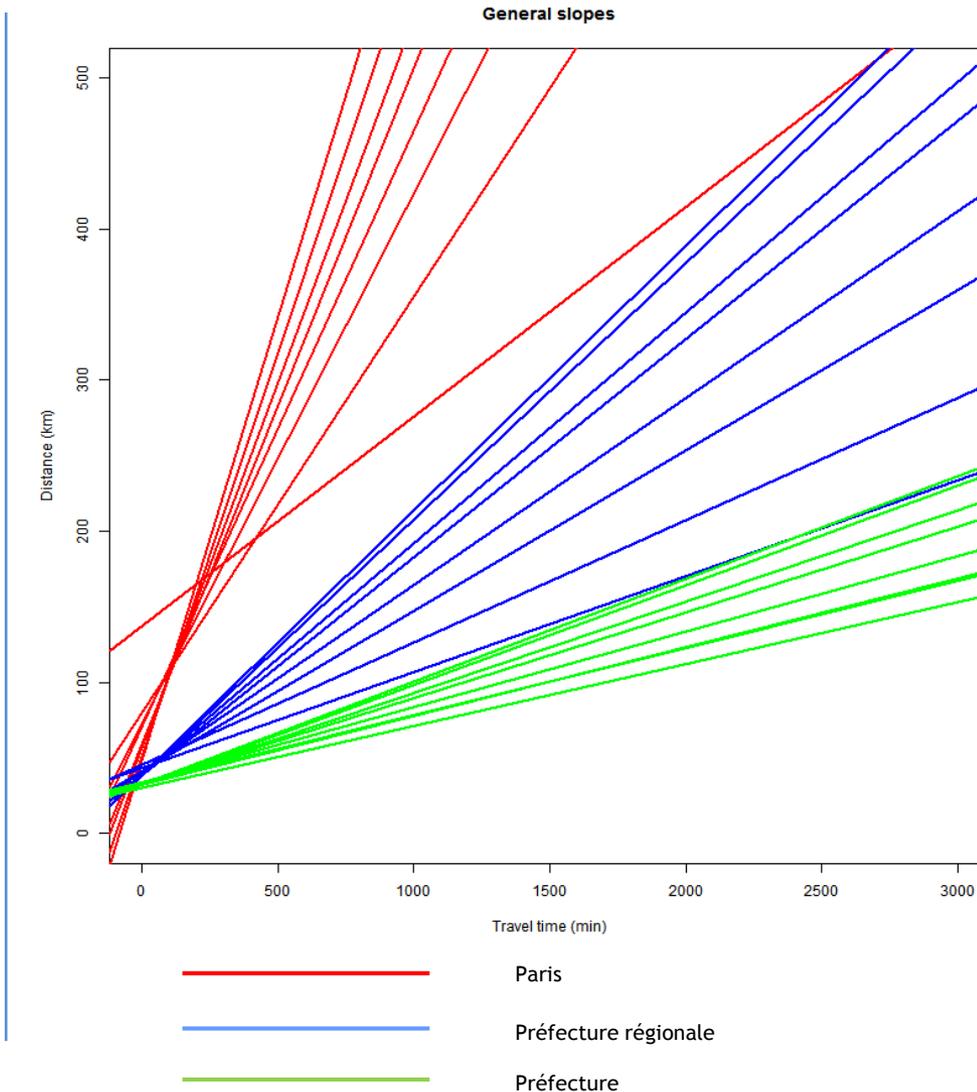


1930



... à tous les échelons ...

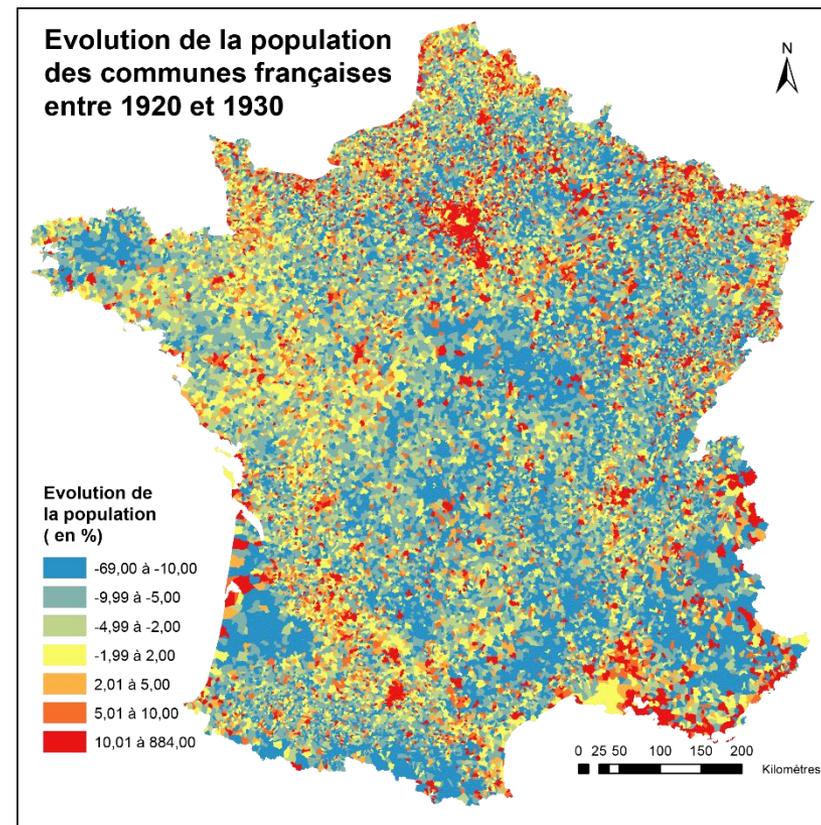
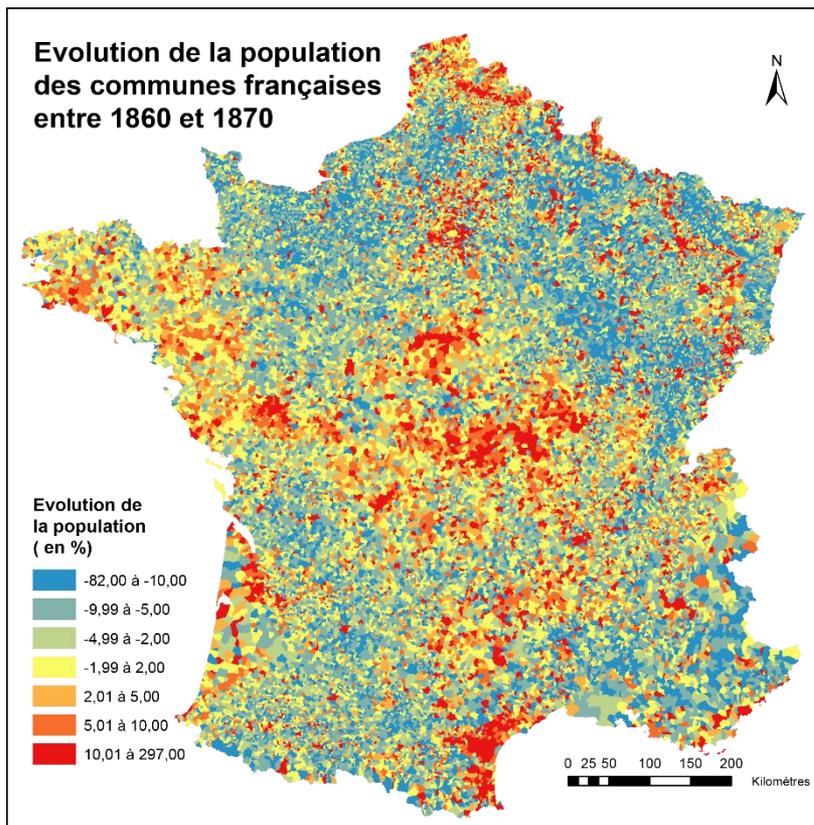
**Le lien distance euclidienne / distance-temps  
aux différents échelons administratifs**



- **Différentiels de pente selon l'échelon administratif**
  - L'accessibilité à Paris est plus sensible à la vitesse le long du réseau.
- **Les écarts sont décalés dans le temps :**
  - Rejoindre Paris coûte que coûte
  - Accélérer le réseau principal
  - Densifier le réseau



*Est-ce que cette accélération des vitesses sur le réseau ferroviaire explique les dynamiques de population / les différentiels de dynamiques de population ?*



*Une OLS très peu satisfaisante*

	1860-1870	1920-1930
R <sup>2</sup>	0,0402	0,0553

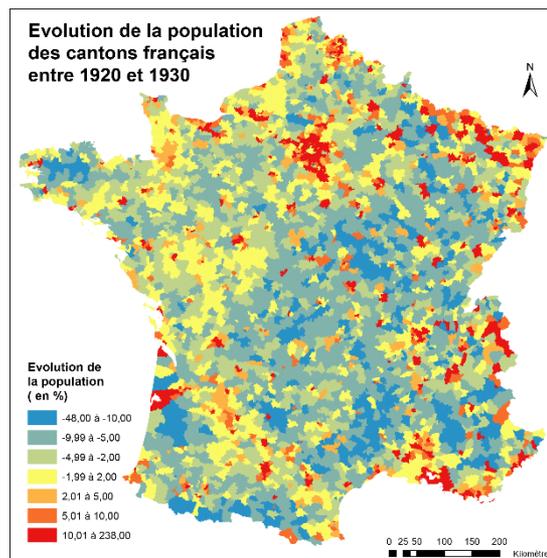
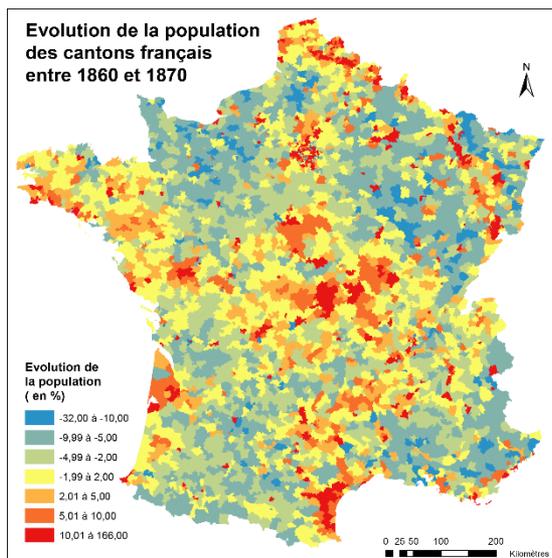
L'hétérogénéité spatiale, couplée à une autocorrélation spatiale, n'aboutissent qu'à des modèles de régression globaux très peu significatifs.



## Confronter les dynamiques de population ...

### ► Résultats

### Agrégation des données au canton : de 36 655 communes à 3 642 cantons



### Quelles variables explicatives ?

- Le temps d'accès à la gare la plus proche
- Le temps de parcours jusqu'au centre le plus proche : la préfecture, la préfecture de région, la capitale ...
- La place de l'unité spatiale dans la hiérarchie du réseau

### Résultats des OLS :

- Des  $R^2$  inférieurs à 0,05.
- Des résidus très insatisfaisants.
- Test Koenker significatif

- **Hétéroscédasticité** : les unités spatiales sont très hétérogènes
- **Instabilité des paramètres** : les paramètres de régression ne sont pas les mêmes selon la localisation



**En quoi la Régression Géographiquement Pondérée  
(Geographically Weighted Regression) peut répondre aux enjeux du modèle ?**

$$y_i = a_0 + \sum_l a_k x_{ik} + \varepsilon_i$$

Modèle de régression global

$$y_i = a_0(u_i, v_i) + \sum_l a_k(u_i, v_i) x_{ik} + \varepsilon_i$$

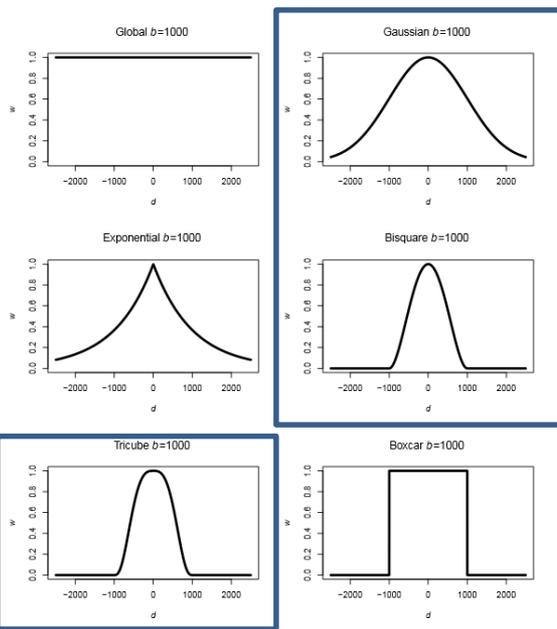
Modèle de régression géographiquement pondérée

### Quels avantages ?

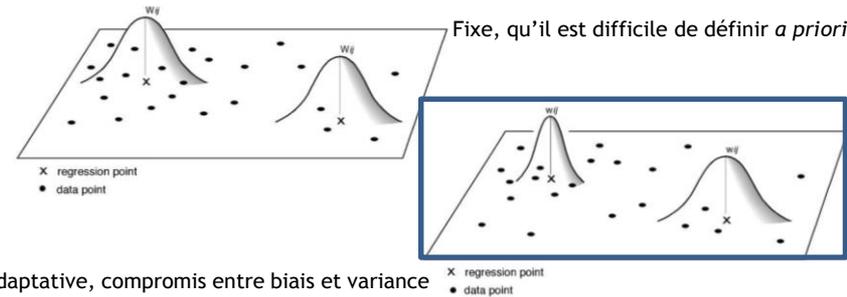
- Une prise en compte directe de la non-stationnarité spatiale.
- Une équation de régression est obtenue pour chaque entité spatiale.
- De la qualité de la régression à la valeur des estimateurs : tous les résultats sont cartographiables ...

### Deux choix à effectuer ...

#### La fonction de pondération



#### La bande passante



Adaptative, compromis entre biais et variance

#### Les choix adoptés :

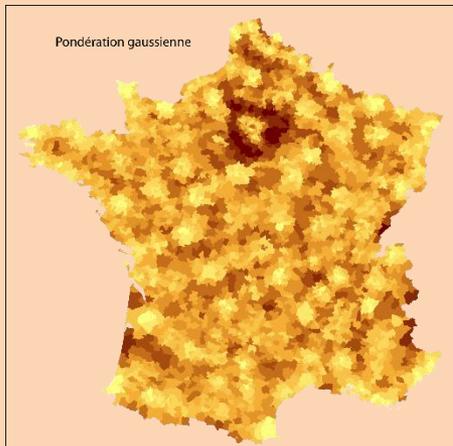
- 3 fonctions de pondération testées :
  - Gauss
  - Bisquare
  - Tricube
- Une bande passante optimisée par validation croisée



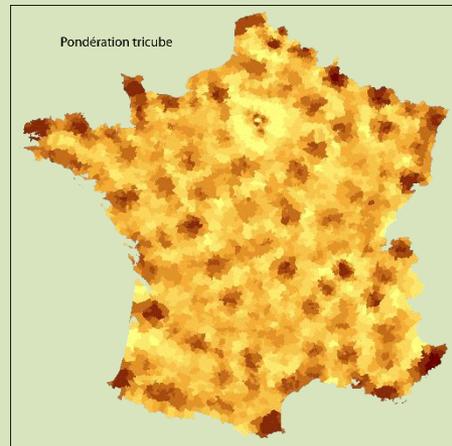
# Parmi les 63 sorties de modèles ...

## Les signatures des $R^2$

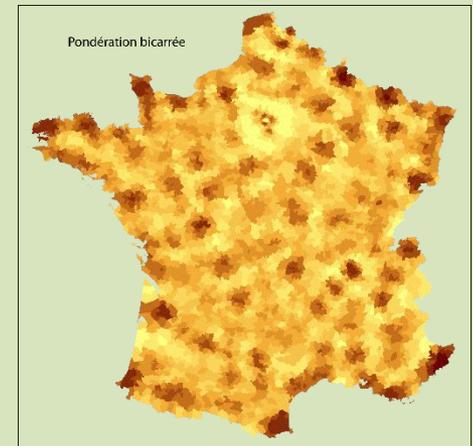
### $R^2$ avec une pondération gaussienne



### $R^2$ avec une pondération tricube



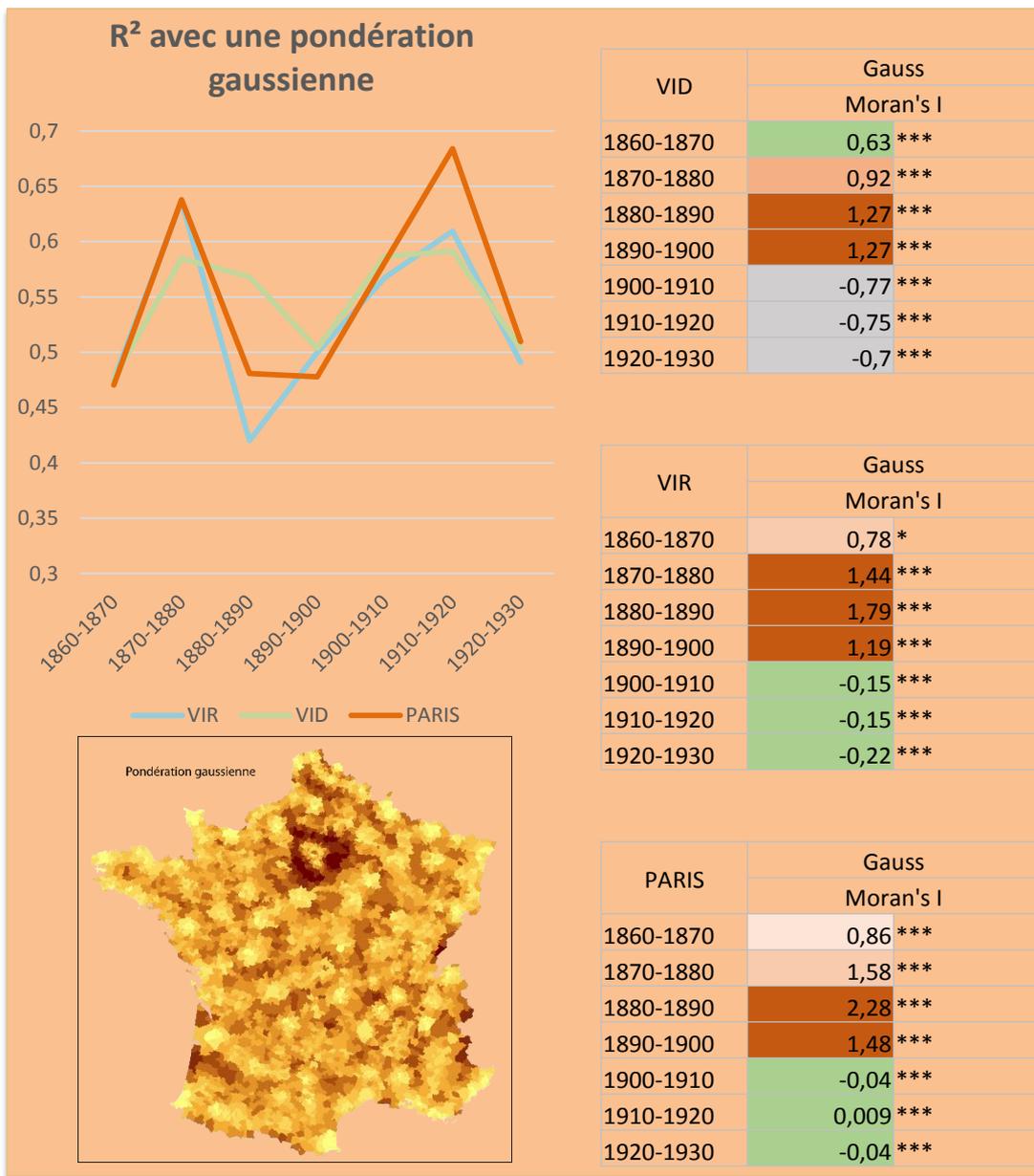
### $R^2$ avec une pondération bicarrée





## Parmi les 63 sorties de modèles ...

### Confronter l'autocorrélation des résidus



### Rejet du modèle gaussien

- Prépondérance parisienne
- Peu de structures sur le reste du territoire
- Des résidus largement autocorrélés



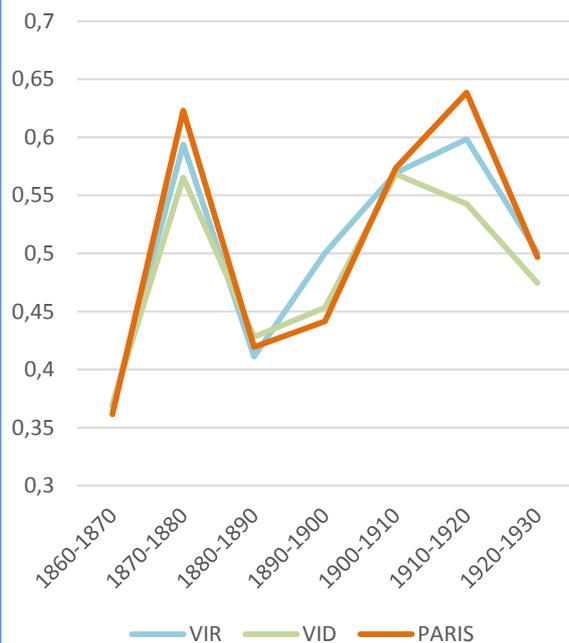
## Parmi les 63 sorties de modèles ...

### ► Résultats

*Peu de différences entre les modèles « bisquare » et « tricube »*

*→ Choisir entre les 3 approches de l'accessibilité : Paris, région, département ...*

R<sup>2</sup> avec une pondération  
tricube

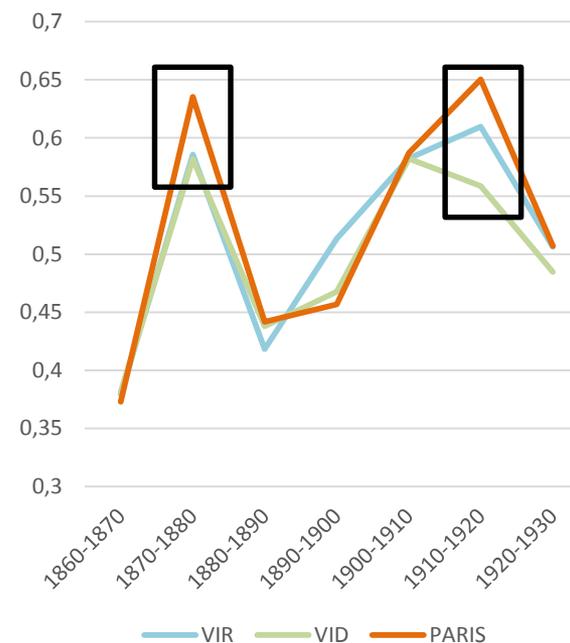


PARIS	Bisquare	Tricube
	Moran's I	Moran's I
1860-1870	0,4 ***	0,41 ***
1870-1880	0,73 **	0,72 ***
1880-1890	0,81 ***	0,82 ***
1890-1900	0,8 **	0,84 **
1900-1910	-0,27 ***	-0,25 ***
1910-1920	-0,29 ***	-0,26 ***
1920-1930	-0,31 ***	-0,31 ***

VIR	Bisquare	Tricube
	Moran's I	Moran's I
1860-1870	0,34 ***	0,35 ***
1870-1880	0,44 ***	0,4 ***
1880-1890	0,46 ***	0,44 ***
1890-1900	0,63 ***	0,63 ***
1900-1910	-0,34 ***	-0,31 ***
1910-1920	-0,39 ***	-0,39 ***
1920-1930	-0,37 ***	-0,39 ***

VID	Bisquare	Tricube
	Moran's I	Moran's I
1860-1870	0,24 ***	0,24 ***
1870-1880	0,32 **	0,33 ***
1880-1890	0,49 ***	0,46 ***
1890-1900	0,58 ***	0,58 ***
1900-1910	-0,95 ***	-0,97 **
1910-1920	-0,83 ***	-0,83 **
1920-1930	-0,77 **	-0,78 *

R<sup>2</sup> avec une pondération  
bicarrée





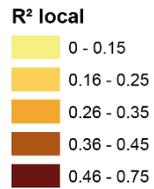
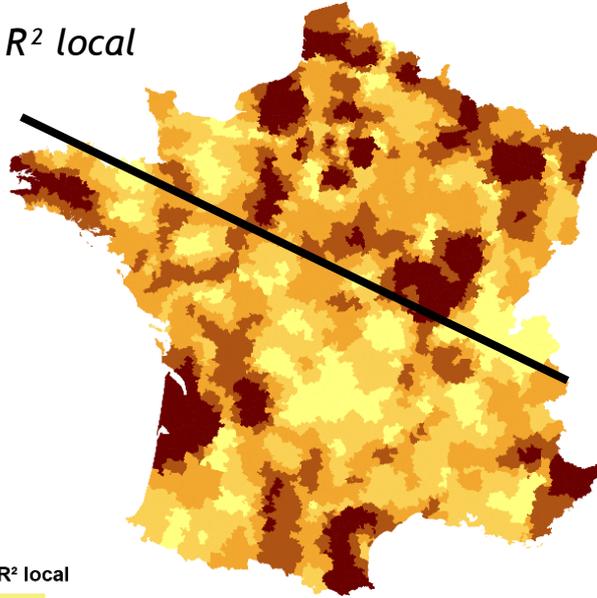
# De la significativité au rôle de la desserte ...

Question de recherche  
Structuration des données

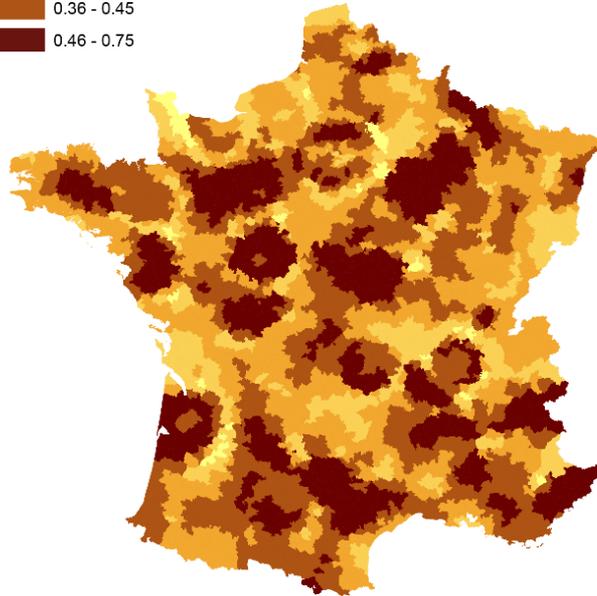
► Résultats

1870 - 1880

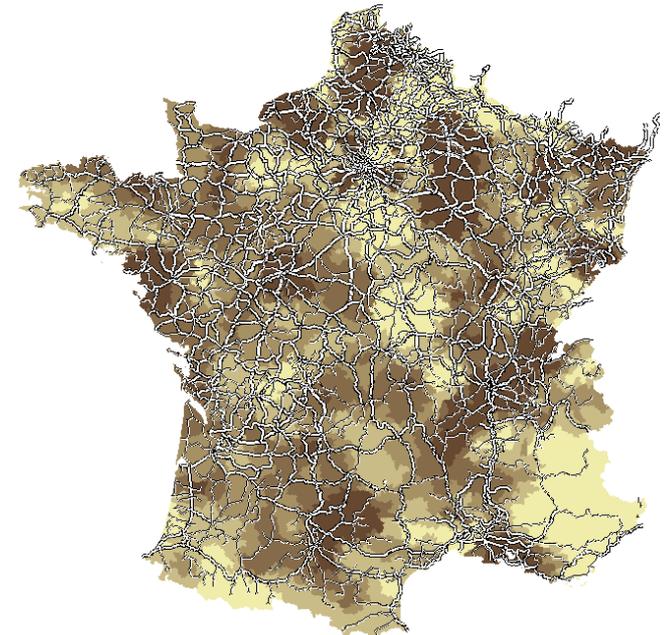
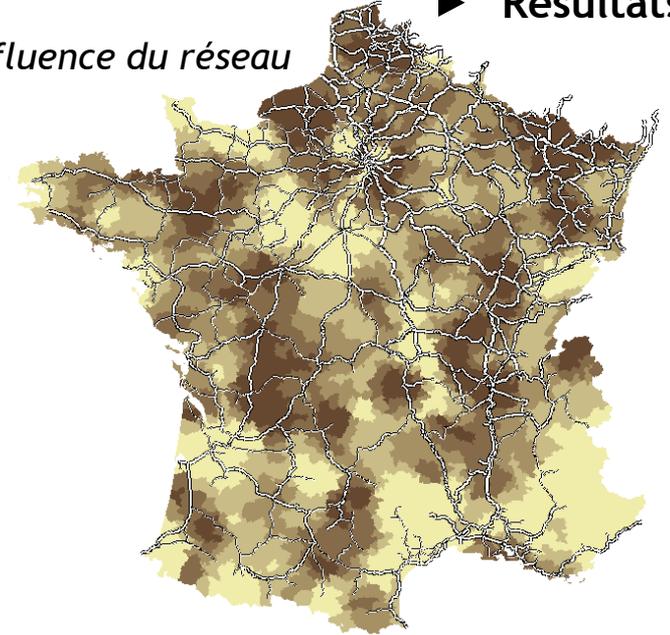
$R^2$  local



1910 - 1920



Influence du réseau



## Conclusion et perspectives ...

---

### Une première démarche exploratoire

- Une relative dépendance à la fonction de pondération : Gauss VS Bisquare et Tricube
- Une bande passante définie par le modèle pour davantage de significativité.

### Quelles perspectives de la GWR ?

- Tester le modèle avec des bandes passantes définies a priori :
  - Quels critères pour la définition de la bande passante ?
- Les modèles GWR sur l'ensemble du territoire font davantage ressortir des structures urbaines ...
  - Créer des sous-ensembles pour évaluer la stabilité/instabilité du modèle
- Développement en cours de la « Localised Flexible Bandwidth Geographically Weighted Regression » (FBGWR)
  - Comment le processus opère à différentes échelles spatiales ?
  - Bande passante spécifique à chaque point
- Quelles possibilités de visualisation pour mieux interpréter les paramètres locaux, dans une perspective temporelle ?

### Des perspectives plus larges ?

- Des enjeux géohistoriques ...
  - Prendre en compte des « histoires locales »
  - Introduire de nouvelles variables
- Opportunité d'intégrer des données socio-économiques
  - Quelle continuité temporelle ?
  - Quelle méthode d'agrégation / désagrégation ?





*Les temps du territoire*

*Jeudi 21 mai 2015*

# Géohistoire du lien réseau/territoire en France entre 1860 et 1930

*Une approche géographiquement pondérée*

Christophe MIMEUR<sup>1</sup>

Thomas THEVENIN<sup>1</sup>

Gilles VUIDEL<sup>2</sup>

Ludovic GRANJON<sup>3</sup>

Robert SCHWARTZ<sup>4</sup>

<sup>1</sup> TheMA, UMR 6049 CNRS, Dijon, France

<sup>2</sup> TheMA, UMR 6049 CNRS, Besançon, France

<sup>3</sup> MSH, USR 3516 CNRS, Dijon, France

<sup>4</sup> Mount Holyoke College, Massachusetts, United States

- ▶ Envisager le lien réseau/territoire dans la longue durée
- ▶ La structuration des données
- ▶ Les apports de l'approche géographiquement pondérée