Statistiques spatiales

Juste Raimbault[1]

2025-2026

[1]LaSTIG, IGN-ENSG-UGE

Non-stationnarité : définition

X_t processus stochastique avec loi de probabilité jointe F_X

Stationnarité stricte : pour tous τ et $t_1, \ldots, t_n$

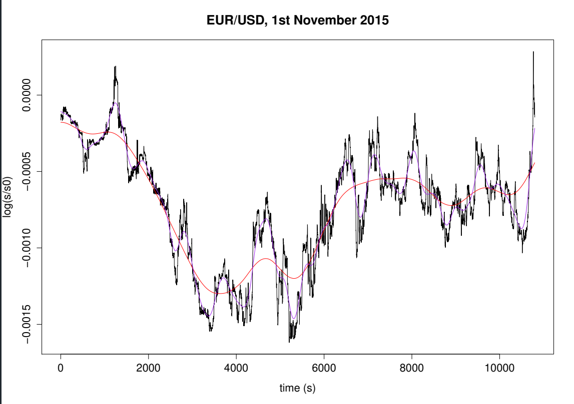
$$F_X(x_{t_1+\tau}, \ldots, x_{t_n+\tau} = F_X(x_{t_1}, \ldots, x_{t_n})$$

Stationnarité faible : pour tous $t, \tau, t_1, t2$, et X_t de variance finie,

$$\mathbb{E}[X_t] = \mathbb{E}[X_{t+\tau}]$$

$$\mathrm{Cov}[X_{t_1}, Xt_2] = \mathrm{Cov}[X_{t_1-t_2}, X0]$$

Non-stationnarité temporelle : exemple



Séries temporelles financières avec moyenne et autocovariance variable dans le temps [Raimbault, 2019]

Non-stationnarité spatiale : exemples

- Tout processus ponctuel dont les moyennes agrégées à d'autres niveaux géographiques varient dans l'espace
- Tout processus de moyenne constante mais dont la fonction d'autocorrélation varie dans l'espace ou n'est pas définie
- En fait la quasi totalité des processus impliquant des structures ou formes spatiales
- Même pour des processus pouvant être stationnaires (géographie physique, climat), celle-ci est locale

Tests de non-stationnarité spatiale

Méthodes variées dépendant de l'approche prise et du contexte :

- test de comparaison des moyennes entre zones géographiques
- variation du spectre [Fuentes, 2005]
- significativité statistique de la variation des coefficients d'une régression géographique pondérée [Leung et al., 2000]
- ...

Auto-corrélation spatiale : définition et tests

$\rightarrow$ corrélation entre valeurs voisines d'un processus spatial :
$\mathrm{Cov}[X_{x_1}, X_{x_2}] \neq 0$

Tests statistiques sur les résidus d'un modèle linéaire :

- Test de Moran
- Tests des multiplicateurs de Lagrange

Méthodes

Extensions spatiales des modèles statistiques :

- Régression géographique pondérée
- Auto-régressions spatiales
- Régression multi-niveaux

Méthodes avancées permettant de gérer la **non-stationnarité spatiale** (GWR et multi-niveau), **l'aspect multi-échelle** (multi-niveaux) et **l'autocorrélation spatiale** (GWR et auto-régressions)

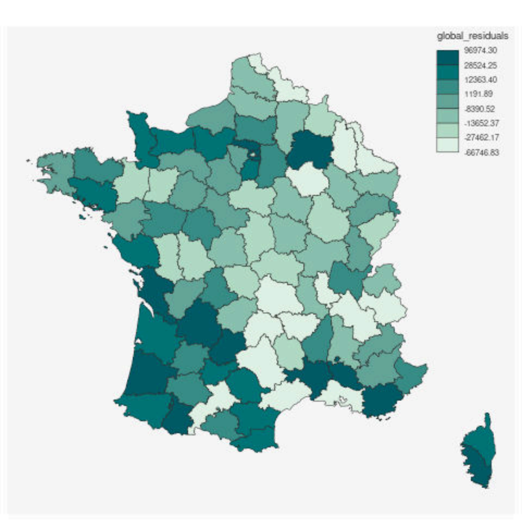
Préparation des données

Première partie du TP en R :

→ construction d'une base des prix immobiliers au niveau départemental, à partir de la base DVF et INSEE

→ premières explorations, cartes et diagnostics

Exemple de résidus globaux structurés



Régression géographique pondérée

Comment inclure des effets de voisinage et prendre en compte la non-stationnarité spatiale dans des modèles statistiques ?
[Fotheringham et al., 2003]

Modèle GWR basique pour les variables y_i aux positions $\vec{u_i}$ et variables explicatives x_{ik}

$$y_i = \beta_0(\vec{u_i}) + \sum_k \beta_k(\vec{u_i}) x_{ik} + \varepsilon_i$$

avec les observations pondérées par un poids spatial $w_i(r)$ en fonction de la distance à $\vec{u_i}$

- Moindres-carrés pondérés, estimés à chaque localisation avec des poids spatiaux variables
- Différents kernels pour les poids spatiaux (gaussien, exponentiel, puissance, bisquare)
- Estimation de la taille de kernel optimale par optimisation de l'AIC par exemple

AIC : définition

Pour comparer des modèles statistiques ajustés sur le même jeu de données, le **Critère d'Information d'Akaike** permet de prendre en compte le nombre de paramètres :

$$AIC = 2k - 2\ln L$$

pour un modèle de vraisemblance L et paramètres k

Correction pour les échantillons de petite taille (n observations) :

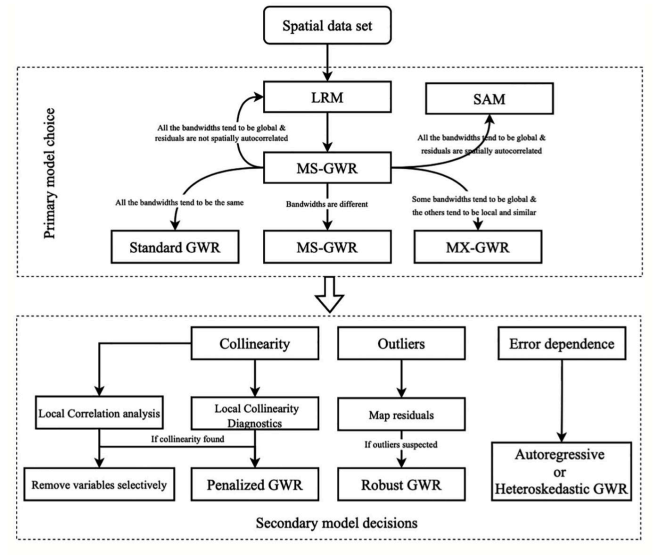
$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

Distance optimale et sélection de modèles

→ sélection de distance optimale par algorithme d'optimisation GSS et minimisation de l'AICc (ou d'un critère de validation croisée) : `bw.gwr`

→ sélection de modèle par méthode directe en minimisant l'AICc : `gwr.model.selection`

Pour aller plus loin : cadre global pour GWR



[Comber et al., 2021]

Application

Deuxième partie du TP en R : analyse GWR

Données : DVF agrégées au niveau départemental construites précédemment

Modèles d'auto-régression spatiale

Ajout d'un terme auto-régressif dans le modèle linéaire :

$$y = \rho Wy + \beta X + \varepsilon$$

- Introduction d'effets de voisinages : prise en compte de l'autocorrélation spatiale (*spatial Durbin model*)
- Correlation avec l'erreur par le terme Wy
- Estimation par Maximum de Vraisemblance : `spatialreg::lagsarlm`
- Version avec décalage sur la variable X : modèle *SLX*, `spatialreg::lmSLX`

Modèles à erreur spatiale

Prise en compte de la structure spatiale dans le terme d'erreur :

$$y = \beta X + \varepsilon$$

$$\varepsilon = \lambda W \varepsilon + u$$

- Erreur auto-régressive dans l'espace
- Test des multiplieurs de Lagrange pour savoir si préférable à un modèle d'auto-régression
- Estimation par Maximum de Vraisemblance : `spatialreg::errorsarlm`

Modèles spatiaux “emboîtés” : modèle de Durbin spatial

$\rightarrow$ combinaison d'un décalage sur Y et sur X

$$y = \rho W y + \beta X + W X \Theta + \varepsilon$$

- Estimation : `spatialreg::lagsarlm(type="mixed")`
- Existe aussi avec le terme d'erreur (modèle de Durbin spatial à erreur)

Test des multiplicateurs de Lagrange

Test générique pour des modèles statistiques avec maximum de vraisemblance $\mathcal{L}(\vec{x}, \theta)$, avec le score défini par :

$$s(\theta) = \frac{\partial \log \mathcal{L}(\vec{x}, \theta)}{\partial \theta}$$

Alors $\sqrt{s(\theta_0)/I(\theta_0)}$ avec $I(\theta_0)$ information de Fisher, suit une distribution normale.

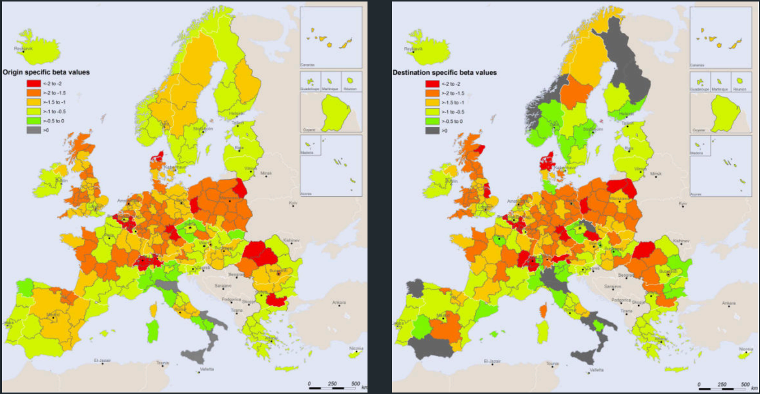
$\rightarrow$ application aux différents modèles spatiaux via leur maximum de vraisemblance

Application

Troisième partie du TP en R : auto-régressions spatiales

Données : DVF agrégées au niveau départemental construites précédemment

Régression multi-niveau : exemple



Paramètres de l'effet de la distance pour un modèle de migration, spécifiques aux origines et destinations, calculés par une régression multi-niveaux [Dennett and Wilson, 2013]

Régression multi-niveau

Lorsque des données peuvent être groupées par une variable catégorielle, une estimation au sein de chaque groupe est incluse dans la régression (“effets fixes”) :

$$y = \alpha + \beta X + \sum_j (\alpha_j + \beta_j X_j + \varepsilon_j)$$

- *Random intercepts :* constantes α_j uniquement
- *Random slopes :* coefficients β_j

Application aux données géographiques

- Groupement des observations selon des niveaux géographiques supérieurs, potentiellement plusieurs : régions, pays, . . .
- non-stationnarité prise en compte avec des coefficients variables
- pas de matrice de distance ou de poids spatiaux : niveau géographique exogène; mais prise en compte du caractère multi-échelle dans le cas de plusieurs niveaux
- pas de prise en compte de l'autocorrélation spatiale

Application

Quatrième partie du TP en R : régressions multi-niveaux spatiales

Données : DVF agrégées au niveau départemental, groupement par régions

Résumé des méthodes

	Non-stationnarité	Auto-corrélation	Multi-échelles
Régression géographique pondérée	✓	✓	✗
Auto-régression spatiale	✗	✓	✗
Régression multi-niveau	✓	✗	✓

Synthèse

$\rightarrow$ méthodes statistiques adaptées à différents aspects des processus spatiaux

$\rightarrow$ méthodes complémentaires, à appliquer selon le contexte et les propriétés des données (tests d'auto-corrélation, de non-stationnarité)

$\rightarrow$ méthodes allant d'une application basique à un cadre complet plus avancé

A retenir : concepts et principes d'application, utilisation basique en R

- Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Lü, Y., Murakami, D., Nakaya, T., Wang, Y., et al. (2021).
  A route map for successful applications of geographically weighted regression.
  Geographical Analysis.
- Dennett, A. and Wilson, A. (2013).
  A multilevel spatial interaction modelling framework for estimating interregional migration in europe.
  Environment and Planning A, 45(6):1491–1507.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2003).
Geographically weighted regression: the analysis of spatially varying relationships.
John Wiley & Sons.

Fuentes, M. (2005).
A formal test for nonstationarity of spatial stochastic processes.
Journal of Multivariate Analysis, 96(1):30–54.

Leung, Y., Mei, C.-L., and Zhang, W.-X. (2000).
Statistical tests for spatial nonstationarity based on the geographically weighted regression model.
Environment and Planning A, 32(1):9–32.

Raimbault, J. (2019).
Second-order control of complex systems with correlated synthetic data.
Complex Adaptive Systems Modeling, 7(1):1–19.