Benchmark technique : Lissage spatial

Mame Diarra FAYE, Manager en Actuariat Produit
Chaymae GRAOUII, Consultante en Actuariat Produit
Nexialog Consulting, Paris, France

7 mai 2025

Résumé

Le lissage spatial est une des méthodes les plus pertinentes et adaptées dans le cadre d'une représentation de dynamiques géographiques. En effet, certains acteurs du marché de l'assurance santé ont recours à ce dernier pour répondre à certaines limites relatives au zonier traditionnel. Ce travail propose un benchmark technique des principales méthodes de lissage existantes (KDE, KNN, krigeage, méthodes bayésiennes, MCMC, quantiles...). Ces différentes méthodes seront analysées de par leur principe, les avantages et les limites relatifs à leur application ainsi que leur cas d'usage. Un des objectifs de cette note sera par ailleurs de valider la pertinence des choix méthodologiques en fonction du type de données disponibles, des besoins opérationnels et du niveau de granularité recherché. Le découpage géographique par bassin de vie sera aussi introduit en alternative aux frontières administratives traditionnelles.

Mots-clés : lissage spatial, zonier santé, bassin de vie, krigeage, KDE, KNN, MCMC, modèles bayésiens, interpolation, assurance santé, tarification

Table des matières

Introduction

En assurance santé les modèles de segmentation géographique reposent traditionnellement sur des zonages administratifs. Ces découpages, bien que pratiques, ne reflètent pas fidèlement le risque lié à l'appartenance géographique des assurés. Ils ne tiennent pas forcément compte des flux inter-zones ni des conditions d'accès aux soins médicaux tels que la disponibilité des services, la distance à parcourir pour accéder à un service médical et les besoins spécifiques des assurés. Tous ces facteurs vont générer des écarts tarifaires entre zones voisines, ou au contraire masquer des disparités réelles entre territoires éloignés.

Le lissage spatial s'impose alors comme étant une des alternatives les plus adaptées pour améliorer la lisibilité des phénomènes géographiques, atténuer le bruit statistique, et renforcer la robustesse des analyses spatiales. Son objectif est de produire une représentation plus continue et interprétable des données en tenant compte à la fois de la proximité géographique et des structures locales.

Ces techniques, historiquement utilisées en épidémiologie et en environnement, s'appliquent aujourd'hui de plus en plus dans le cadre assurantiel.

Ce document se présente comme suit :

- Une première partie sur l'analyse de l'existant, avec l'exploration d'un panel de cas d'usage sur le lissage spatial appliqué dans différents domaines (santé, environnement, assurance...).
- Afin d'identifier les approches les plus pertinentes dans le contexte de l'assurance santé, une seconde partie sera proposée sur la présentation et la comparaison des méthodes de lissage. Cette section permettra ainsi de mieux jauger l'efficacité des différentes méthodes en prenant en compte les réalités médicales relatives à la zone géographique.

1 Les enjeux du lissage spatial en tarification santé

Le marché, de plus en plus compétitif, combiné à l'explosion de la quantité de données santé disponible oblige les acteurs du marché à avoir recours à des méthodologies, jusqu'ici très peu utilisées en santé, afin de proposer des tarifs plus justes qui répondent au mieux aux profils de risque des assurés couverts.

La modélisation du risque lié au lieu de résidence d'un assuré constitue une étape clé dans le processus de tarification. La méthodologie la plus répandue consiste à mettre en place un zonier qui, par expérience métier, est constitué de 5 zones au maximum fluctuant généralement sur un intervalle de [90% - 115%].

La création d'un zonier peut se faire soit de manière empirique, soit avec une modélisation de type GLM. Avec la méthode GLM, les résidus de la modélisation de la prime pure sont généralement utilisés, avec comme loi de la famille exponentielle celle de Tweedie.

Notons par ailleurs que le zonier peut être modélisé avec d'autres indicateurs tels que la fréquence et le coût moyen afin de capter séparément le risque géographique lié à ces derniers. Cette méthode reste très peu utilisée en santé pour des raisons pratiques, par souci de rigueur mathématique, mais également pour des questions de pertinence vu la faible volatilité du coût moyen en assurance santé.

Contrairement à la non-vie, où la tarification peut se faire à l'adresse près, le découpage géographique retenu pour la création d'un zonier en assurance santé se fait très souvent à la maille départementale. Cette méthode de découpage et de création de zonier a deux limites :

- Les risques liés au lieu de résidence ne sont pas homogènes par département.
- Deux assurés habitant dans des départements différents peuvent avoir le même profil de risque du fait de leur proximité géographique (pollution, météo, épidémie, accès aux services médicaux, ...).

Voici un exemple pour illustrer ce propos.

Cas 1 :

Appliquer le même coefficient de zonier à un habitant du 19^{e} et un habitant du 15^{e} (département de Paris), ainsi qu'un même coefficient à un habitant de Pantin et un habitant de Villepinte (département de Seine-Saint-Denis).

Cas 2 :

Appliquer des coefficients évolutifs en fonction de la proximité géographique (un habitant du 19^{e} et un habitant de Pantin auront des coefficients de zonier plus proches).

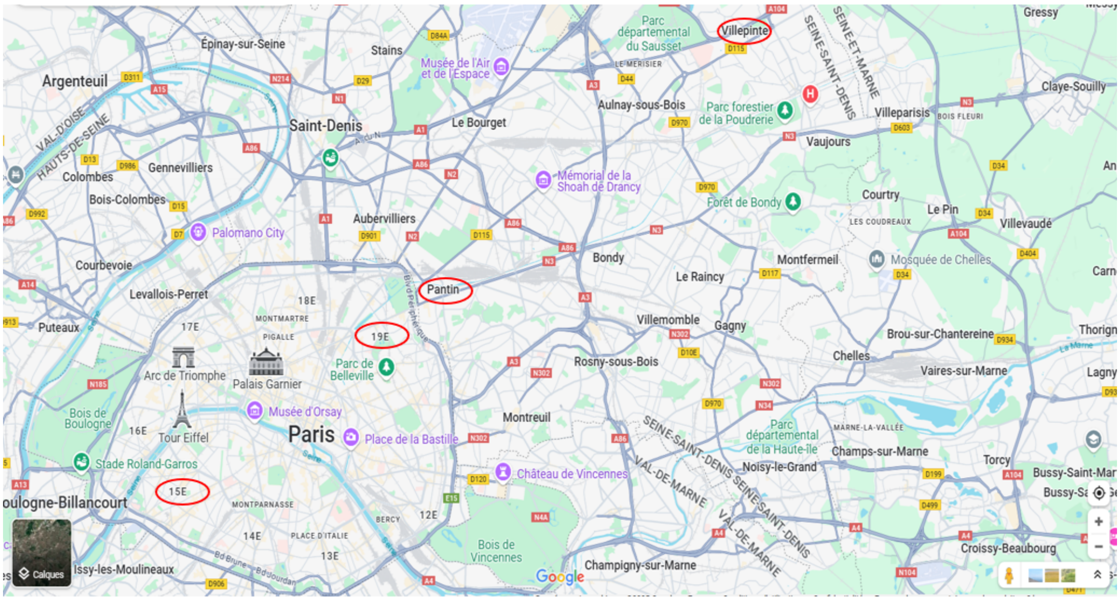


Figure 1 Carte illustrant les limites du zonier départemental.

La logique voudrait que le cas 2 soit appliqué en tarification, **c'est le principe du lissage spatial**.

2 Introduction au lissage spatial

Le lissage spatial est une technique souvent utilisée en analyse géostatistique permettant d'obtenir une représentation plus fluide et interprétable d'un phénomène sur un territoire. Il atténue les variations aléatoires dues à l'échantillonnage ou aux découpages administratifs arbitraires [1]. Cela permet d'éviter des effets de seuil non justifiés, tout en mettant en évidence les tendances structurelles du phénomène étudié.

Un des intérêts majeurs du lissage est de contourner le problème des unités spatiales modifiables (MAUP) [2] lorsque les regroupements sont artificiels (IRIS [3], communes, départements...). Le lissage apporte alors une lecture plus fidèle des dynamiques territoriales.

La notion de bassin de vie, bien que très pertinente de par sa définition, reste très peu connue dans le monde de la tarification santé.

1. Le lissage spatial est largement utilisé en épidémiologie et en statistique spatiale pour détecter des tendances cachées.
2. Le MAUP (*Modifiable Areal Unit Problem*) désigne l'influence des regroupements géographiques arbitraires sur les résultats statistiques.
3. IRIS (Ilots Regroupés pour l'Information Statistique) : une unité géographique utilisée dans l'analyse statistique en France pour l'organisation spatiale.

À quoi sert le lissage par bassin de vie ?

Le **lissage spatial par bassin de vie** [a] propose une alternative : modéliser les zones non pas selon des divisions administratives figées, mais en fonction des flux de soins, de la densité médicale et des usages réels du système de santé.

a. Selon l'INSEE, un **bassin de vie** est "le plus petit territoire sur lequel les habitants ont accès aux équipements et services les plus courants." https://www.insee.fr/fr/metadonnees/definition/c2060

2.1 Enjeux du découpage par bassin de vie

Le recours au bassin de vie comme unité de lissage soulève plusieurs enjeux structurants, nous pouvons citer entre autres :

- **L'équité territoriale** : réduction des disparités entre zones voisines ; harmonisation des tarifs avec l'accès réel aux soins.
- **La cohérence géographique** : rapprochement de territoires éloignés mais présentant des caractéristiques médicales similaires ; justification de la distinction entre zones proches mais médicalement distinctes.
- **La qualité actuarielle** : atténuation du bruit dans les données, amélioration de la stabilité des coefficients, et prévision plus précise du risque.
- **L'acceptabilité par les assurés** : meilleure compréhension et acceptation des tarifs lorsque ceux-ci sont alignés avec la réalité de l'environnement médical.

2.2 Impact sur les méthodes de lissage

Le découpage par bassin de vie n'entraîne pas un changement significatif dans les méthodes classiques de lissage. Cependant il fonctionne comme un "mini-lissage" car, de par sa définition, on s'attend à avoir une certaine homogénéité au sein du même bassin de vie face à certains comportements liés à la situation géographique. En effet, à travers ce découpage, la plupart des risques liés au lieu de résidence peuvent s'aligner parfaitement(désert médical, zones urbaines/ zones rurales, temps de trajet pour des services médicaux, épidémie...)

Le lissage par bassin de vie complète ce découpage en affinant les profils de risque et en permettant de mieux gérer les effets de bord et les frontières entre les zones. Cela évite les écarts tarifaires importants souvent observés aux frontières administratives.

2.3 Variables de référence : une dépendance au périmètre et à l'usage

Le choix des variables utilisées comme inputs dans le lissage spatial dépend de plusieurs facteurs :

- **Le périmètre d'analyse** : à maille fine, on observe des disparités fortes et un besoin de données fiables ; à maille plus large, les variables agrégées (ex. taux de pathologies chroniques ou dépenses moyennes) sont plus pertinentes et robustes.
- **Le cas d'usage visé** :

  Pour des objectifs tarifaires, on privilégiera des variables directement liées au risque assurantiel, comme le coût moyen des soins, la fréquence des actes ou encore le niveau de consommation médicale.

  Pour des objectifs de pilotage ou d'aménagement du territoire, on mobilisera davantage des variables comme la densité médicale, les temps d'accès aux soins ou les flux inter-zones.

Notons par ailleurs que certaines variables jouent un rôle structurant car fortement corrélées à la consommation médicale. Elles sont donc utilisées dans tous les cas d'usage.

Un lissage spatial efficace repose autant sur la méthode choisie que sur la sélection pertinente des variables d'entrée, en cohérence avec le périmètre étudié et les finalités du modèle.

À retenir

Passer d'un découpage géographique classique à un découpage par **bassin de vie** permet d'obtenir :

- une représentation fidèle des pratiques médicales et des dynamiques locales ;
- un zonier plus juste, adapté à la réalité fonctionnelle des territoires ;
- une adaptation des méthodes de lissage à des zones irrégulières ;
- une amélioration de la performance tarifaire.

3 Analyse de l'existant : panorama des travaux examinés

Cette section vise à présenter une revue structurée des travaux existants portant sur le lissage spatial, en mettant en lumière leur contexte d'application, leur contribution méthodologique, ainsi que leur pertinence pour le secteur de l'assurance santé en France.

L'application des méthodes de lissage géospatial à la tarification santé en France reste encore peu explorée dans la littérature scientifique. En revanche, ces techniques sont bien établies dans d'autres domaines, tels que :

- L'assurance habitation ou auto : pour la segmentation géographique des risques.
- L'épidémiologie : pour la cartographie des maladies (ex. VIH).

Ces constats font du lissage géospatial une opportunité innovante pour optimiser la segmentation des risques santé.

3.1 Revues scientifiques sur le lissage spatial : focus sur le secteur assurantiel

Les recherches académiques portant sur le lissage spatial en assurance restent encore limitées. Deux publications notables peuvent néanmoins être citées :

- **Kopczyk, D. (2024)** explore des techniques d'interpolation spatiale multivariée afin d'améliorer la modélisation du risque en assurance non-vie.
- **Clemente, G. P., Della Corte, F. et Zappa, D. (2024)** proposent un modèle hiérarchique de réseau spatial appliqué à l'analyse du risque routier.
- **Faller, D. (2022)** modélise statistiquement la fréquence des sinistres à l'aide de MCMC avec diffusion de Langevin riemannienne continue.

3.2 Autres publications relatives au lissage spatial

Plusieurs travaux de recherche en santé publique ont recours au lissage spatial. Parmi eux :

- **Larmarange, J., Vallo, R., Yaro, S., Msellati, P. et Méda, N. (2011)** appliquent le krigeage et les noyaux adaptatifs à la cartographie de la prévalence du VIH en Afrique.
- **Held, L. et Rue, H. (2016)** proposent une régression de Poisson pondérée géographiquement pour la cartographie des maladies.
- **Kauhl, B., Schweikart, J., Krafft, T., Keste, A. et Moskwyn, M. (2016)** analysent les facteurs de risque du diabète de type 2 à l'aide de l'estimation par noyau et de la régression pondérée.
- **Casper, M., Kramer, M. R., Peacock, J. M. et Vaughan, A. S. (2019)** illustrent l'intégration de la dimension spatiale dans les politiques de santé publique.

Dans le domaine environnemental :

- **Lemarchand, O. et Jeannée, N. (2009)** explorent les méthodes de lissage géostatistique pour cartographier la pollution au dioxyde d'azote en Alsace, notamment via le krigeage.

Certaines ressources proposent des outils méthodologiques dédiés à la représentation spatiale :

- **Rose, T. (2021) et Genebes, L., Renaud, A. et Sémécurbe, F. (2018)** présentent des approches par carroyage et par noyaux pour la cartographie thématique.
- **Boisseau, M. (2021)** décrit les paramètres de lissage dans "Cartes & Données", incluant noyaux gaussiens et quadratiques.
- **Viry, M. et Giraud, T. (2020)** proposent une plateforme interactive de cartographie thématique nommée Magrit, développée au sein de l'UMS RIATE (Unité Mixte de Service dédiée à l'analyse spatiale et à la géovisualisation, soutenue par le CNRS).

Synthèse des travaux existants

L'analyse des publications révèle une prédominance des applications du lissage spatial dans les domaines sanitaire, environnemental et territorial, en particulier pour la cartographie des maladies et la modélisation de la pollution. À l'inverse, les usages dans le secteur assurantiel restent très limités.

Les travaux recensés dans les mémoires d'actuariat sont plutôt des démarches exploratoires, sans mise en œuvre à grande échelle. Les méthodes les plus utilisées sont le krigeage, le lissage par noyau (KDE), les simulations MCMC et certaines approches bayésiennes. Ces outils permettent de concilier la précision locale et la stabilité globale, mais leur application en assurance santé reste encore embryonnaire. Cela constitue un axe d'innovation prometteur, sous réserve de démontrer leur robustesse dans un cadre réglementaire strict.

4 Méthodes de lissage spatial

L'ensemble des méthodes présentées sont regroupées en quatre familles : les méthodes locales basées sur la proximité spatiale, les méthodes d'interpolation mathématique, les méthodes probabilistes et bayésiennes, ainsi que les méthodes basées sur la distance et les réseaux.

Mode de lecture

Dans cette section, le système de notation appliqué pour évaluer les avantages et les limites de chaque méthode est détaillé dans le tableau 12 en annexe.

4.1 Méthodes locales basées sur la proximité spatiale

Les méthodes locales sont basées sur l'idée que l'influence d'un point de donnée diminue avec la distance parcourue [4].

4.1.1 Méthode à noyau (KDE : Kernel Density Estimation)

Le lissage spatial par noyau est une technique d'estimation non paramétrique utilisée pour analyser des phénomènes spatiaux en attribuant une pondération aux observations en fonction de leur proximité [5]. Il repose sur l'utilisation d'une fonction noyau K_h qui diminue avec la distance, permettant ainsi une estimation fluide de l'intensité locale des observations.

4. Ce principe est un des fondements de la première loi de la géographie de Tobler, qui stipule que « tout est lié à tout, mais les objets proches le sont plus que ceux éloignés ».

5. La méthode KDE est notamment utilisée en criminologie, en épidémiologie et en écologie pour cartographier des densités d'événements.

L'intensité lissée estimée en un point x est définie par :

$$\hat{\lambda}_h(x) = \sum_i K_h(x - x_i) \tag{1}$$

où :

- $K_h(x)$ est une fonction de pondération décrivant l'influence des observations en fonction de leur distance à x.
- h est la bande passante qui contrôle l'étendue du lissage [6].

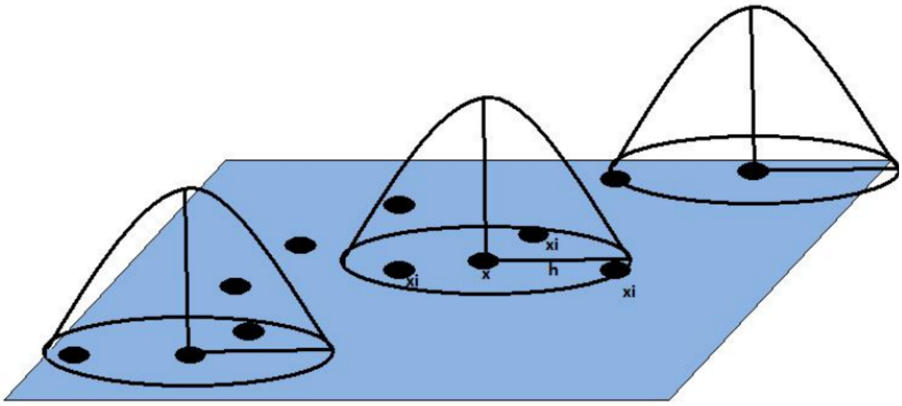


Figure 2 Illustration du lissage spatial par la méthode de noyau.

Cette méthode permet une estimation continue des phénomènes spatiaux sans imposer de structure rigide sur les données. La figure 2 illustre comment chaque point d'estimation est associé à une fonction noyau : la valeur de cette fonction est maximale au niveau du point et décroît au fur et à mesure que l'on s'en éloigne [7]. Ce principe permet de prendre en compte l'influence des observations voisines tout en atténuant l'effet des points éloignés.

a) **Types de noyaux**

Les noyaux sont des fonctions de pondération utilisées dans le lissage spatial pour attribuer des poids aux observations en fonction de leur distance à un point donné [8]. Dans ce cadre, différents types de noyaux peuvent être mobilisés selon les priorités de l'analyse. On distingue notamment les deux formes suivantes, parmi les plus utilisées :

- **Noyau gaussien** : ce noyau est défini par :

$$K_h^N(x) = \frac{1}{2\pi} e^{-\|x\|^2} \tag{2}$$

Il accorde une importance plus grande aux points proches tout en prenant en compte l'ensemble des observations de la zone d'étude.

- **Noyau quadratique (ou Epanechnikov)** : ce noyau est défini par :

$$K_h^B(x) = \frac{9}{16} 1_{\|x\|<h} \left(1 - \frac{\|x\|^2}{h^2}\right)^2 \tag{3}$$

6. Un choix de bande passante trop grand donne comme résultat un modèle avec un effet de sur-lissage des données, masquant les structures locales. À l'inverse, une bande passante trop faible crée un effet de bruit.
7. La décroissance de la pondération suit généralement une loi gaussienne, mais d'autres noyaux peuvent être utilisés selon les besoins de l'étude.
8. En pratique, le choix du noyau a souvent un effet mineur sur le résultat final, contrairement au choix de la bande passante qui influence significativement le degré de lissage.

Il attribue des poids plus élevés aux points les plus proches et s'annule au-delà du rayon de lissage [9].

Le choix du noyau a un impact limité sur l'estimation, mais le choix de la bande passante h est déterminant [10]. Le noyau quadratique favorise les observations proches en réduisant l'influence des points éloignés, tandis que le noyau gaussien prend en compte l'ensemble des points de la zone d'étude.

b) **Correction des effets de bord**

Dans les études spatiales, l'estimation de densité par noyau est souvent biaisée aux frontières de la région d'étude, car les points situés en bordure ont moins de voisins contribuant à leur estimation [11]. Ce phénomène est appelé effet de bord et peut entraîner une sous-estimation systématique de l'intensité dans ces zones.

Afin de prendre en compte ces effets et d'améliorer l'estimation de l'intensité spatiale, deux méthodes de correction sont couramment utilisées :

- **Correction uniforme** : supposant une continuité de l'intensité au-delà des frontières, elle ajuste l'estimation en normalisant les valeurs en fonction de la masse du noyau contenue dans la fenêtre d'étude :
$$\hat{\lambda}_h^U(x) = \frac{1}{e_h(x)} \sum_i K_h(x - x_i)$$
avec
$$e_h(x) = \int_W K_h(x - v)dv.$$
- **Correction de Diggle** : plus conservatrice, cette approche considère que l'intensité est nulle en dehors de la zone d'étude. Chaque point est corrigé individuellement en fonction de la proportion du noyau incluse dans la fenêtre :
$$\hat{\lambda}_h^D(x) = \sum_i \frac{1}{e_h(x_i)} K_h(x - x_i).$$

Si la correction uniforme est plus simple et rapide à implémenter, la correction de Diggle est préférable lorsque l'on suspecte une réelle discontinuité aux frontières, bien qu'elle soit plus coûteuse en calcul.

La figure 3 permet d'illustrer la forme typique de fonctions noyaux utilisées dans les méthodes de lissage spatial.

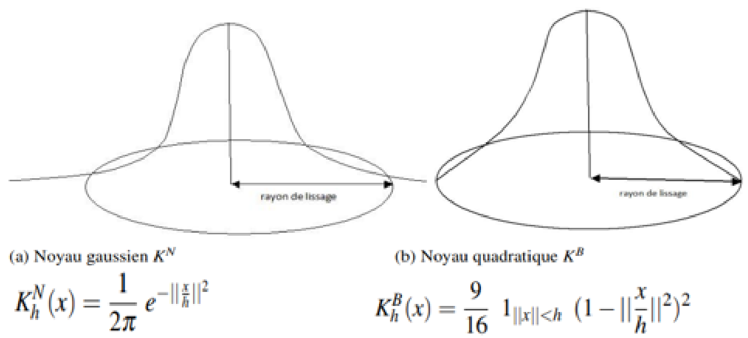


Figure 3 Illustration de deux fonctions noyaux utilisées pour le lissage spatial.

9. Le rayon de lissage correspond à une distance autour de chaque territoire dans laquelle les valeurs voisines sont prises en compte pour corriger la sienne. Il est exprimé dans l'unité des coordonnées du document de travail (par exemple : mètres ou kilomètres).

10. En statistique spatiale, la sélection optimale de la bande passante peut être effectuée via des critères tels que la validation croisée ou l'approximation de la vraisemblance maximale.

11. Cet effet est particulièrement marqué dans les zones à faible densité d'observations et peut fausser les résultats si aucune correction n'est appliquée.

c) **Avantages, limites et notations de la méthode à noyau**

Le tableau ci-après résume les principaux avantages, limites ainsi que les notations associées aux méthodes de lissage par noyau. Ces éléments permettent d'évaluer leur pertinence selon les contextes d'analyse spatiale.

Avantages		Limites	
Notations	Commentaires	Notations	Commentaires
(+++)	**Efficacité sur les grandes bases de données**, permettant d'éliminer le bruit	(--)	**Sensibilité au choix du noyau et de la bande passante**, influençant fortement la qualité de l'estimation
(++)	**Flexibilité**, car la méthode peut être appliquée avec différents types de noyaux adaptés à divers contextes	(--)	**Effet de bord**, pouvant biaiser les estimations aux frontières si aucune correction n'est appliquée
(++)	**Approche non paramétrique**, ne nécessitant pas d'hypothèses fortes sur la distribution des données	(--)	**Perte de précision en cas de sur-lissage**, pouvant masquer des structures fines ou des discontinuités importantes
(+)	**Visualisation intuitive**, facilitant l'analyse spatiale et l'interprétation des tendances des données	(-)	**Complexité computationnelle**, notamment pour les grands ensembles de données où le calcul de chaque estimation peut être coûteux

Table 1 Synthèse des avantages, limites et notations des méthodes de lissage par noyau.

4.1.2 K-plus proches voisins (K-NN)

La méthode des k-plus proches voisins (K-NN) est une technique non paramétrique utilisée pour estimer la densité spatiale en se basant sur les k observations les plus proches d'un point donné [12]. Contrairement à la méthode à noyau, qui utilise une fonction de pondération décroissante avec la distance, le k-NN considère uniquement les k voisins les plus proches avec une pondération égale.

En analyse spatiale, la méthode k-NN est utilisée pour estimer la densité d'événements ou de points d'intérêt [13]. En attribuant à chaque point de l'espace une valeur basée sur les k observations les plus proches, on obtient une surface lissée représentant la distribution spatiale des phénomènes étudiés.

L'estimateur k-NN de l'intensité en un point x est défini par :

$$\hat{\lambda}_k(x) = \frac{k}{n \cdot V_k(x)} \tag{4}$$

où :

- n est le nombre total d'observations.
- $V_k(x)$ est le volume de la région englobant les k plus proches voisins du point x.

Cette approche adapte la taille de la région de lissage en fonction de la densité locale des observations : dans les zones denses, $V_k(x)$ sera petit, alors que dans les zones moins denses, il sera plus grand [14].

12. Introduite en 1951 par Fix et Hodges, la méthode k-NN a d'abord été utilisée pour la reconnaissance de formes avant d'être appliquée en géostatistique et analyse spatiale.
13. Elle est notamment utilisée en épidémiologie pour cartographier les foyers de maladies et en urbanisme pour identifier les zones denses d'activités.
14. Cette adaptabilité permet de mieux représenter les phénomènes avec des concentrations inégales de points, contrairement aux méthodes à bande passante fixe.

a) **Pseudo-code de l'algorithme K-NN**

Le processus de classification ou d'estimation par k-NN peut être résumé en quelques étapes :

Pseudo-code de l'algorithme k-NN

```
Entrée :
 • Un jeu de données d'entraînement étiquetées
 • Une observation inconnue à classer
 • Un entier k (nombre de voisins à considérer)

1. Pour l'observation inconnue :
   a. Calculer les distances avec toutes les observations connues.
   b. Sélectionner les k plus proches voisins.

2. Si la tâche est une classification :
   • Assigner l'étiquette majoritaire parmi les k voisins.

3. Si la tâche est une régression :
   • Calculer la moyenne des valeurs des k voisins.

4. Retourner la valeur prédite.
```

b) **Choix du paramètre K**

Le choix du nombre de voisins K est crucial pour la performance de la méthode K-NN. Un petit K peut rendre l'estimation sensible au bruit local (forte variance, faible biais), tandis qu'un grand K peut lisser excessivement les variations locales (fort biais, faible variance).

c) **Métriques de distance utilisées en K-NN**

Pour identifier les K voisins les plus proches, l'algorithme K-NN repose sur différentes métriques de distance [15] :

Distance Euclidienne : Elle mesure la distance en ligne droite entre deux points. Cette métrique est adaptée aux données continues et aux contextes où les variations sont similaires dans toutes les directions. Elle peut cependant être sensible aux valeurs extrêmes (outliers).

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{5}$$

Distance de Manhattan : Elle mesure la distance entre deux points en ne considérant que les déplacements horizontaux et verticaux. Elle est préférable dans des contextes où les déplacements suivent des axes fixes, comme les trajets en ville ou sur un quadrillage.

$$d(x, y) = \sum_{i=1}^{n}|x_i - y_i| \tag{6}$$

Distance de Minkowski : C'est une généralisation des distances Euclidienne et de Manhattan. En ajustant le paramètre p, on peut adapter la distance aux spécificités des données (par exemple, $p = 1$ donne Manhattan, $p = 2$ donne Euclidienne).

$$d(x, y) = \left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{\frac{1}{p}} \tag{7}$$

d) **Avantages et limites de la méthode des k plus proches voisins (KNN)**

Le tableau suivant présente les principaux avantages et limites de la méthode des k plus proches voisins (KNN), en soulignant les éléments essentiels à prendre en compte pour un lissage spatial pertinent.

15. Le choix de la métrique influence directement la performance du modèle. Par exemple, la distance Euclidienne est efficace pour des données continues, tandis que la distance de Manhattan est plus adaptée lorsque les déplacements s'effectuent selon des axes fixes, comme dans un plan urbain structuré.

Avantages		Limites	
Notations	**Commentaires**	**Notations**	**Commentaires**
(+++)	**Simple à utiliser et fonctionne bien** avec peu de réglages	(---)	**Sensibilité au choix de K** : un K trop petit peut être trop sensible au bruit
(++)	**Ajustable** : le paramètre K permet de contrôler le degré de lissage	(--)	**Problèmes aux frontières** : estimations biaisées dans certaines zones peu denses
(++)	**Non paramétrique** : ne nécessite pas d'hypothèses sur la distribution des données	(--)	**Calculs intensifs sur grands ensembles** : nécessite de nombreuses comparaisons de distances
(+)	Capable de gérer des relations non linéaires	(-)	**Coût computationnel élevé** : calcul des distances coûteux pour de grands ensembles

Table 2 Synthèse des avantages, limites et notations des k plus proches voisins (KNN).

4.1.3 Moyennes mobiles spatiales

Le lissage par moyennes mobiles spatiales est une méthode permettant d'obtenir une estimation lissée d'un phénomène spatial en remplaçant la valeur d'un point par la moyenne des valeurs de ses voisins. Elle est couramment utilisée en géostatistique, cartographie et analyse environnementale pour atténuer les variations locales et mettre en avant les tendances globales [16].

L'idée principale est que les points proches ont souvent des valeurs similaires. Appliquer une moyenne locale permet donc de réduire les variations brusques et le bruit, et d'obtenir une carte plus homogène et interprétable [17]. Cette approche est particulièrement utile pour éviter les artefacts liés aux découpages administratifs arbitraires.

a) **Rappel de la méthodologie**

A. Définir une fenêtre de voisinage (ex. carré de 3×3 ou rayon de 5 km [18]).
B. Calculer la moyenne des valeurs des points situés dans cette fenêtre.
C. Remplacer la valeur du point central par cette moyenne selon la formule suivante :

$$\hat{Z}(s) = \frac{1}{|N(s)|} \sum_{i \in N(s)} Z_i \tag{8}$$

où :

- $\hat{Z}(s)$ est la valeur estimée au point s après lissage.
- $N(s)$ représente l'ensemble des points voisins définis par la fenêtre de voisinage.
- Z_i est la valeur mesurée aux points voisins.

b) **Variantes de la moyenne mobile**

Il existe plusieurs variantes du lissage par moyennes mobiles spatiales [19] :

La **moyenne mobile simple** consiste à remplacer chaque point par la moyenne des valeurs de ses voisins. Elle est facile à mettre en œuvre et permet de lisser efficacement

16. Le lissage par moyennes mobiles spatiales est une technique utilisée depuis longtemps en analyse de données spatiales, notamment pour la visualisation de cartes météorologiques et d'indices de pollution.
17. Cette technique repose sur l'hypothèse de continuité spatiale, qui suppose que les variations locales sont souvent progressives plutôt que brutales.
18. Le choix de la fenêtre de voisinage influence directement le niveau de lissage : une fenêtre trop petite capte trop de détails locaux, tandis qu'une fenêtre trop grande lisse excessivement les variations.
19. Les descriptions des méthodes sont basées sur les explications disponibles sur Wikipédia (https://fr.wikipedia.org/wiki/Moyenne_mobile) et le blog HubSpot pour le lissage exponentiel (https://blog.hubspot.fr/sales/lissage-exponentiel).

les données, mais elle présente l'inconvénient de réduire fortement les variations locales, ce qui peut masquer des ruptures importantes dans les données.

La **moyenne mobile pondérée**, quant à elle, introduit un poids décroissant selon la distance entre les points : les plus proches ont plus d'influence sur la moyenne que les plus éloignés. Cette méthode offre une meilleure prise en compte de la structure locale et limite l'effet de dilution des zones hétérogènes. En revanche, elle nécessite de définir une fonction de pondération adaptée, ce qui ajoute un paramétrage supplémentaire à maîtriser.

Enfin, la **moyenne mobile exponentielle** applique une pondération qui décroît de façon exponentielle avec l'éloignement dans l'espace ou dans le temps, en accordant une plus grande importance aux données récentes ou proches. Cette méthode est particulièrement efficace pour suivre des évolutions rapides, mais elle peut introduire un biais si les anciennes valeurs restent significatives, et elle peut aussi lisser de manière excessive certaines tendances de fond.

c) **Avantages et limites de la moyenne mobile spatiale**

Le tableau suivant présente les principaux avantages et limites de la méthode de moyenne mobile spatiale, en mettant en évidence les aspects à considérer pour son utilisation dans un contexte d'analyse spatiale.

Avantages		Limites	
Notations	Commentaires	Notations	Commentaires
(+++)	Facile à comprendre et rapide à appliquer	(--)	Peut entraîner un sur-lissage, effaçant des détails importants
(++)	Permet une meilleure visualisation des tendances spatiales	(--)	Ne prend pas en compte la structure spatiale réelle des données
(++)	Particulièrement adapté aux applications spatiales telles que la cartographie, la météorologie et l'analyse environnementale	(--)	N'est pas adapté aux zones avec des discontinuités fortes
(+)	Atténue les variations extrêmes et réduit le bruit	(-)	Peut générer des anomalies en bordure si la fenêtre est mal définie

Table 3 Synthèse des avantages, limites et notations de la méthode de moyenne mobile spatiale.

4.2 Méthodes d'interpolation mathématique

L'interpolation spatiale est une technique permettant d'estimer des valeurs inconnues en se basant sur des observations existantes. Son objectif principal est de produire une surface continue des données mesurées en tenant compte des relations spatiales entre les points.

Principe : Les phénomènes spatiaux présentent souvent une continuité géographique. L'interpolation permet de prédire la valeur en un point non observé en exploitant l'information des points environnants [20].

4.2.1 Krigeage : Une méthode d'interpolation géostatistique

Le krigeage est une technique avancée d'interpolation spatiale qui repose sur la corrélation spatiale des données. Contrairement aux méthodes classiques de lissage, il modélise la variabilité spatiale des observations et minimise l'erreur d'estimation [21].

20. L'interpolation spatiale est largement utilisée en climatologie, en hydrologie et en cartographie pour combler les lacunes dans les données mesurées.
21. Le krigeage a été développé par le statisticien sud-africain Danie Krige dans les années 1950 pour optimiser l'exploitation minière.

a) **Formulation mathématique du krigeage**

Le krigeage repose sur une moyenne pondérée des observations environnantes. L'estimation $\hat{Z}(s_0)$ au point s_0 est donnée par :

$$\hat{Z}(s_0) = \sum_{i=1}^{N} w_i Z(s_i) \tag{9}$$

où :

- $Z(s_i)$ est la valeur mesurée en un point voisin s_i.
- w_i sont des poids optimisés en fonction de la covariance spatiale.
- N est le nombre de points pris en compte pour l'interpolation.

Les poids w_i sont déterminés par le variogramme, une fonction qui quantifie la dépendance spatiale en fonction de la distance [22] :

$$\gamma(h) = \frac{1}{2}\mathbb{E}\left[Z(s) - Z(s+h)\right]^2 \tag{10}$$

où h est la distance entre deux points.

La figure 4 montre un exemple classique d'un modèle de krigeage. Le modèle interpole les données connues et prédit les valeurs inconnues en minimisant l'erreur et en ajustant la variance. On observe que la variance est nulle aux points d'observation et augmente entre ces points.

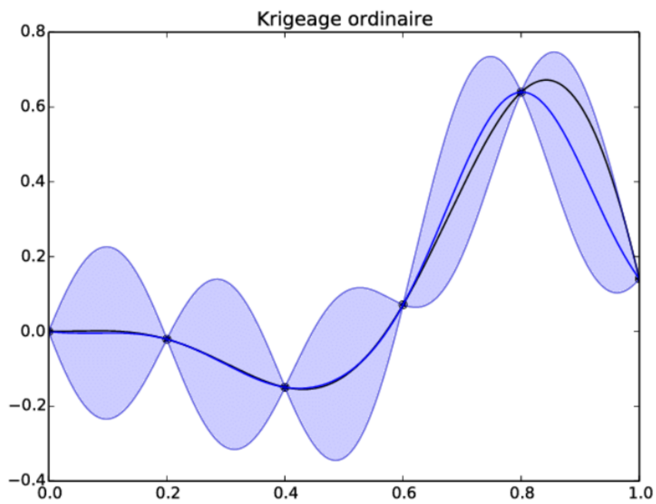


Figure 4 Illustration d'un modèle classique de krigeage.

22. Le variogramme est un outil clé en géostatistique, permettant de modéliser la structure spatiale des données et d'optimiser l'interpolation.

b) **Avantages et limites du krigeage**

Ce tableau fournit une synthèse des avantages et des inconvénients associés au krigeage, permettant d'apprécier sa pertinence selon les objectifs de modélisation spatiale.

Avantages		Limites	
Notations	**Commentaires**	**Notations**	**Commentaires**
(+++)	**Précision élevée** : exploite la relation spatiale pour une estimation fiable	**(--)**	**Complexité de mise en œuvre** : requiert une bonne connaissance des modèles géostatistiques
(++)	**Interpolation fiable** : permet d'estimer les valeurs dans les zones sans données	**(--)**	**Ajustement du variogramme** : une mauvaise modélisation fausse les résultats
(++)	**Prise en compte de l'incertitude** : fournit une mesure de confiance sur chaque estimation	**(--)**	**Coût computationnel élevé** : calculs lourds sur de grands jeux de données
(++)	**Adapté pour les données continues** : couramment utilisé en climatologie, hydrologie et cartographie	**(-)**	**Sensibilité aux données d'entrée** : des mesures imprécises peuvent fausser les résultats

Table 4 Synthèse des avantages, limites et notations de la méthode de krigeage.

4.2.2 Loess spatial : lissage basé sur une régression locale pondérée

Le loess spatial est une méthode de lissage qui permet d'adoucir les variations des données sans faire d'hypothèses fortes sur leur répartition [23]. Contrairement au krigeage, qui suppose une structure spatiale précise, le loess ajuste une courbe locale en s'adaptant dynamiquement à chaque point.

a) **Principe mathématique du loess spatial**

Le loess ajuste localement une régression pondérée pour estimer la tendance des données. L'estimateur du loess au point x_i est défini par :

$$\hat{y}_i = \sum_{j=1}^{N} w_j(x_i)\beta_j$$

où :

- N est le nombre de points voisins pris en compte,
- β_j sont les coefficients de la régression locale ajustée par moindres carrés pondérés,
- $w_j(x_i)$ est une fonction de pondération attribuant plus de poids aux points proches de x_i.

La fonction de pondération courante est la fonction tricube [24] :

$$w_j(x_i) = \begin{cases} \left(1 - \left(\frac{|x_j - x_i|}{h}\right)^3\right)^3, & \text{si } |x_j - x_i| < h, \\ 0, & \text{sinon.} \end{cases}$$

Avec :

- h : paramètre de bande passante, contrôlant la taille du voisinage [25].
- $|x_j - x_i|$: distance entre le point d'intérêt x_i et un point voisin x_j.

23. Le loess a été introduit dans les années 1970 par Cleveland et Devlin pour le lissage adaptatif des séries temporelles et des données spatiales.
24. La fonction tricube est l'une des fonctions de pondération les plus utilisées en loess, mais d'autres fonctions, comme la gaussienne, peuvent être employées selon les besoins.
25. Le choix du paramètre de bande passante h est crucial : une valeur trop petite capture trop de détails (sur-ajustement), tandis qu'une valeur trop grande lisse excessivement la structure locale.

b) **Étapes de l'algorithme**

Les principales étapes de mise en œuvre de cet algorithme sont décrites ci-dessous de manière simplifiée :

A. **Définition du voisinage** : On choisit une zone autour de chaque point (ex. un rayon de 10 km).
B. **Attribution des poids aux voisins** :
   - Plus un point est proche, plus son poids est important.
   - Plus un point est éloigné, moins il influence l'estimation.
C. **Ajustement d'une régression polynomiale locale** sur le voisinage [26].
D. **Remplacement de la valeur du point** par cette estimation.

c) **Avantages et limites du loess spatial**

Le tableau ci-après fournit une synthèse des avantages et des inconvénients associés à la méthode loess appliquée à un contexte spatial, permettant d'évaluer sa pertinence selon les besoins d'analyse locale et de modélisation souple.

Avantages		Limites	
Notations	Commentaires	Notations	Commentaires
(+++)	**S'adapte aux variations locales**	(--)	**Sensibilité au choix du voisinage,** impactant la qualité du lissage
(++)	**Ne suppose pas une structure spatiale fixe** contrairement au krigeage	(--)	Peut être difficile à calibrer en présence de bruit important
(++)	Effectue un lissage tout en préservant les tendances générales	(-)	**Plus lent qu'une simple moyenne mobile**, car nécessite plusieurs calculs locaux

Table 5 Synthèse des avantages, limites et notations de la méthode loess spatiale.

En résumé, le loess constitue une approche flexible et adaptative qui permet de lisser les données tout en préservant leurs tendances locales [27]. Contrairement aux méthodes basées sur un modèle spatial fixe, il s'ajuste dynamiquement en fonction du voisinage. Toutefois, son efficacité repose sur un choix judicieux des paramètres, notamment la bande passante et la taille du voisinage, qui influencent directement la qualité de l'estimation.

4.2.3 Lissage par quantile

Le lissage par quantile est une méthode robuste permettant d'analyser les données spatiales en conservant les structures locales tout en réduisant l'influence des valeurs extrêmes [28]. Contrairement au lissage par noyau qui estime une moyenne lissée, cette approche consiste à calculer un quantile donné (Q10, Q50, Q90, etc.) pour chaque zone, ce qui permet de mieux capter les disparités locales sans être biaisé par des valeurs aberrantes [29].

a) **Définition mathématique d'un quantile**

Soit une variable aléatoire X d'une distribution de fonction de répartition $F(x)$. Le quantile d'ordre q (avec $0 < q < 1$) est défini comme :

$$Q_q = \inf\{x \in \mathbb{R} \mid F(x) \geq q\}. \tag{11}$$

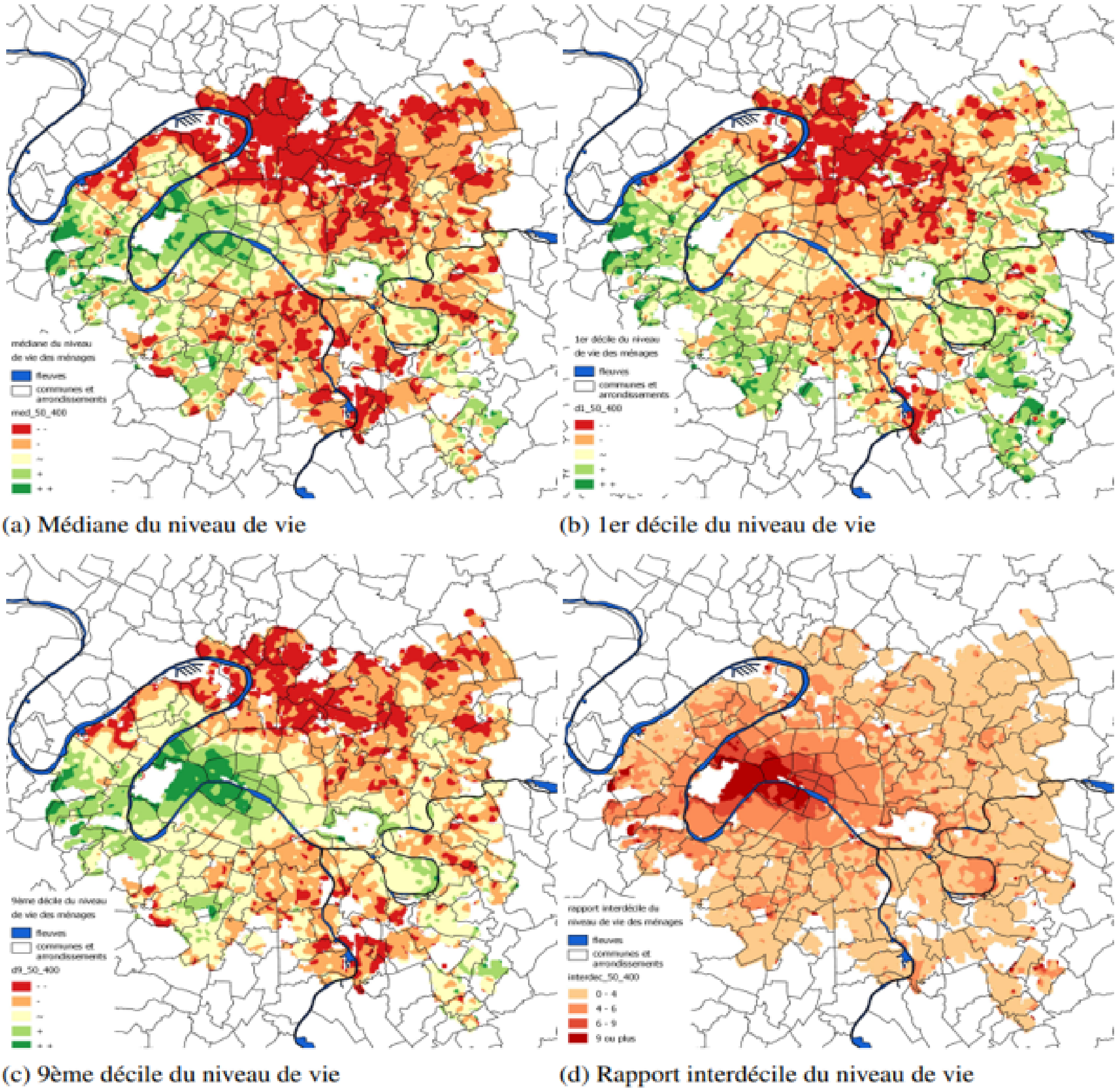
26. Le degré du polynôme ajusté est généralement 1 (régression linéaire locale) ou 2 (régression quadratique locale) pour éviter le sur-ajustement.
27. Le loess est souvent utilisé en statistique exploratoire, en cartographie environnementale et en traitement d'images pour réduire le bruit et lisser les valeurs aberrantes.
28. Contrairement aux moyennes, les quantiles sont insensibles aux valeurs aberrantes et sont largement utilisés en économétrie et en épidémiologie.
29. Par exemple, dans l'analyse des revenus, la médiane est souvent privilégiée à la moyenne pour éviter l'influence des très hauts revenus.

Cela signifie que Q_q est la plus petite valeur pour laquelle la fonction de répartition atteint au moins q.

Par exemple :

- Le premier quartile (Q1) correspond à $q = 0.25$,
- La médiane correspond à $q = 0.50$,
- Le troisième quartile (Q3) correspond à $q = 0.75$.

Prenons l'exemple de la répartition des niveaux de vie en Île-de-France (figure 5) :



(a) Médiane du niveau de vie

(b) 1er décile du niveau de vie

(c) 9ème décile du niveau de vie

(d) Rapport interdécile du niveau de vie

Figure 5 Exemple de lissage quantile appliqué aux niveaux de vie en Île-de-France.[30]

Dans la figure ci-dessus, les cartes sont numérotées de (a) à (d) :

- La carte (a) représente la **médiane du niveau de vie**, illustrant une vue d'ensemble des revenus par zone.
- La carte (b) montre le **premier décile du niveau de vie** ($D1$, valeur en dessous de laquelle se situent 10% des observations), identifiant les zones où les revenus sont les plus bas.

30. Source : Insee-DGFiP-Cnaf-Cnav-CCMSA, Fichier localisé social et fiscal 2012.

- La carte (c) représente le **neuvième décile du niveau de vie** ($D9$, valeur en dessous de laquelle se situent 90% des observations), mettant en évidence les zones les plus riches.
- La carte (d) illustre le **rapport interdécile**, qui met en lumière les inégalités économiques entre les différentes zones [31].

Quelques observations majeures :

- **Fortes inégalités spatiales** : L'ouest parisien affiche des niveaux de vie plus élevés tandis que certaines zones du nord-est et de la banlieue restent économiquement en difficulté [32].
- **Disparités locales bien captées** : Contrairement à une moyenne unique, l'utilisation des quantiles permet d'identifier les zones précises de précarité et de richesse.
- **Marquage des écarts riches/pauvres** : L'image du rapport interdécile (d) révèle des écarts considérables, surtout dans et autour de Paris.

b) **Avantages et limites du lissage par quantile**

Ce tableau présente une synthèse des principaux avantages et limites du lissage par quantile, en soulignant ses apports spécifiques dans un contexte de lissage spatial.

Avantages		Limites	
Notations	Commentaires	Notations	Commentaires
(+++)	**Moins sensible aux valeurs aberrantes**	(--)	**Nécessite un grand échantillon** pour être statistiquement valide
(++)	**Adapté aux distributions asymétriques** comme celles du revenu ou de la pollution	(--)	**Interprétation moins intuitive** que celle d'un lissage classique
(++)	Facilite l'analyse de la dispersion spatiale via la comparaison de plusieurs quantiles	(-)	**Sensibilité au choix du quantile** : un mauvais paramétrage peut masquer les tendances

Table 6 Synthèse des avantages, limites et notations du lissage par quantile.

Ainsi, le lissage par quantile est une méthode robuste adaptée aux données asymétriques, permettant d'explorer la dispersion spatiale, mais nécessitant un grand échantillon et une interprétation plus délicate.

4.3 Méthodes probabilistes et bayésiennes

Les méthodes probabilistes et bayésiennes [33] sont utilisées pour estimer des valeurs inconnues en prenant en compte l'incertitude et les distributions de probabilité. Contrairement aux approches déterministes comme la moyenne mobile spatiale ou le krigeage, elles reposent sur des modèles statistiques qui intègrent des informations préalables (a priori) et les mettent à jour avec de nouvelles données (a posteriori).

Ces méthodes permettent notamment :

- D'intégrer l'incertitude dans les estimations.
- D'exploiter des distributions de probabilité pour mieux représenter la variabilité des données.
- De faire des prédictions en l'absence de nouvelles observations grâce à des modèles bayésiens.

31. Les rapports interdéciles sont largement utilisés en sociologie et en urbanisme pour mesurer les écarts de richesse entre territoires.
32. Ces disparités spatiales sont souvent renforcées par les dynamiques du marché immobilier et des infrastructures de transport.
33. Introduites par Thomas Bayes au XVIIIe siècle, elles sont restées marginales jusqu'au développement des algorithmes MCMC dans les années 1950. Aujourd'hui, elles sont largement utilisées en géostatistique, en intelligence artificielle et en finance.

4.3.1 Méthode bayésienne avec simulation MCMC

Cette méthode appartient aux techniques probabilistes et bayésiennes de lissage spatial. Contrairement aux approches déterministes, elle repose sur des modèles statistiques qui prennent en compte l'incertitude et les dépendances spatiales.

a) **Fondements de la méthode bayésienne**

Les méthodes bayésiennes sont utilisées pour estimer des valeurs inconnues dans l'espace en combinant :

- **Les données observées (réelles)**.
- **Une connaissance a priori** (hypothèses basées sur des tendances connues).
- **Une mise à jour progressive** grâce aux nouvelles données.

Selon la disponibilité des données, deux cas peuvent se présenter :

A. **Cas 1 : Mise à jour classique** : lorsqu'une nouvelle donnée est disponible, le modèle met à jour ses estimations. Plus il y a d'observations, plus les prédictions sont précises.

B. **Cas 2 : Simulation Monte-Carlo par Chaînes de Markov (MCMC)** : lorsqu'aucune donnée supplémentaire n'est disponible, la simulation MCMC est utilisée pour approximer une distribution de probabilité à partir des observations existantes.

b) **Théorème de Bayes et mise à jour des estimations**

La mise à jour des estimations bayésiennes repose sur le **théorème de Bayes** :

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Avec :

- $P(\theta)$ est la distribution a priori (avant d'avoir les données).
- $P(D \mid \theta)$ est la probabilité des nouvelles données sous notre hypothèse actuelle.
- $P(D)$ est le facteur de normalisation.
- $P(\theta \mid D)$ est la distribution a posteriori (mise à jour avec les nouvelles données).

c) **Simulation MCMC et échantillonnage de Gibbs**

Dans le cadre des méthodes MCMC, la chaîne de Markov est un outil algorithmique qui permet d'explorer l'espace des paramètres inconnus. Elle évolue dans un espace multidimensionnel, où chaque dimension peut correspondre à une zone géographique ou un paramètre spatial à estimer. Cette chaîne est dite Markovienne car chaque état ne dépend que du précédent, ce qui permet une exploration progressive de la distribution a posteriori.

Il est important de souligner que cette chaîne ne modélise pas les interactions entre unités géographiques voisines : elle n'intègre pas de structure spatiale explicite. Elle sert uniquement à simuler des échantillons dans un cadre bayésien. Les dépendances spatiales explicites seront introduites dans la section dédiée aux chaînes de Markov spatiales.

Les simulations MCMC permettent d'approximer des distributions complexes en générant un grand nombre d'échantillons successifs. Une des méthodes les plus utilisées est l'échantillonnage de Gibbs, qui permet d'obtenir des valeurs échantillonnées selon une loi conditionnelle.

Principe de l'échantillonnage de Gibbs :

A. On initialise aléatoirement les paramètres inconnus.
B. À chaque itération, on met à jour un paramètre en fonction des valeurs actuelles des autres paramètres.
C. Après un certain nombre d'itérations, les *samples* convergent vers la distribution cible.

Ainsi, l'échantillonnage de Gibbs est particulièrement utile pour les modèles bayésiens où la distribution conjointe est difficile à estimer directement, mais où les distributions conditionnelles sont plus faciles à manipuler.

d) **Application : estimation du coût moyen des soins en assurance santé**

Supposons qu'un assureur veuille estimer le coût moyen des soins pour une ville C en utilisant les données des villes voisines A et B.

A. **Estimation a priori** : On suppose que le coût des soins suit une distribution normale :

$$P(\theta) = \mathcal{N}(240, 10^2)$$

Cela signifie que la moyenne estimée est de 240€, avec une variance initiale de $10^2 = 100$.

B. **Nouvelle observation** : Un hôpital de la ville C rapporte un coût observé de 245€, supposé suivre une distribution normale avec une incertitude de 5€, soit :

$$P(D|\theta) = \mathcal{N}(245, 5^2)$$

Cette information permet d'affiner l'estimation initiale en combinant la distribution a priori et la distribution des nouvelles données.

C. **Mise à jour bayésienne** :

En combinant les distributions a priori et des observations, on obtient une nouvelle estimation :

$$\mu_{\text{post}} = \frac{\mu_0\sigma_2^2 + \mu_2\sigma_0^2}{\sigma_0^2 + \sigma_2^2}$$

$$\sigma_{\text{post}}^2 = \frac{\sigma_0^2\sigma_2^2}{\sigma_0^2 + \sigma_2^2}$$

où :

- $\mu_0 = 240$ et $\sigma_0^2 = 100$ sont les paramètres de la distribution a priori.
- $\mu_2 = 245$ et $\sigma_2^2 = 25$ sont les paramètres de la distribution des observations.

D. **Résultat final** :

$$\mu_{\text{post}} = \frac{240 \times 25 + 245 \times 100}{125} = 243$$

$$\sigma_{\text{post}}^2 = \frac{100 \times 25}{125} = 5^2$$

Ainsi, la nouvelle distribution est :

$$P(\theta|D) = \mathcal{N}(243, 5^2)$$

La moyenne passe de 240 € à 243 €, l'incertitude diminue de 10 € à 5 €. Cela montre comment une nouvelle observation peut améliorer une estimation initiale.

e) **Avantages et limites des méthodes MCMC**

Ce tableau synthétise les principaux avantages et limites des méthodes de simulation MCMC.

Avantages		Limites	
Notations	**Commentaires**	**Notations**	**Commentaires**
(+++)	**Intègre l'incertitude** et génère une distribution de valeurs possibles	(---)	**Nécessite énormément de ressource**, en particulier sur des modèles complexes
(++)	**Adaptée aux zones avec peu de données** grâce aux hypothèses a priori	(---)	Peut entraîner des temps de calcul importants pour obtenir une estimation stable
(++)	Flexibilité et capacité de mise à jour progressive avec de nouvelles données	(--)	**Paramétrage complexe** notamment pour définir des distributions cohérentes avec les données
(+)	Utilisation de l'échantillonnage de Gibbs pour simplifier le calcul des distributions complexes	(--)	**Convergence lente** si les distributions sont fortement corrélées

Table 7 Synthèse des avantages, limites et notations des méthodes MCMC.

En définitive, les méthodes MCMC, notamment l'échantillonnage de Gibbs, constituent des approches puissantes pour le lissage spatial. Elles permettent d'estimer des valeurs inconnues tout en intégrant progressivement les observations et en quantifiant l'incertitude, ce qui les rend particulièrement pertinentes lorsque les données spatiales sont incomplètes.

4.3.2 Modèles bayésiens hiérarchiques

a) **Principe de la méthode**

Les modèles bayésiens hiérarchiques [34] constituent une méthode avancée d'estimation, qui structure les données en plusieurs niveaux d'information afin de mieux capturer à la fois les tendances globales et les particularités locales.

Contrairement aux approches classiques fondées uniquement sur les données locales, ces modèles procèdent de manière hiérarchique :

- Ils commencent par une estimation globale, fondée sur les données disponibles à un niveau supérieur (national ou régional).
- Puis, ils introduisent des niveaux intermédiaires (région, département) pour renforcer la précision.
- Enfin, ils affinent l'estimation à l'aide des **observations locales** disponibles.

b) **Exemple : estimation du coût moyen des soins**

Un assureur souhaite estimer le coût moyen des soins dans une ville C, mais dispose de peu de données locales.

Cas 1 : Utilisation de MCMC uniquement

Avec la méthode MCMC, on utilise uniquement les distributions a priori et les quelques données disponibles.

- **Avantage** : Permet une estimation du coût dans C même en l'absence de données locales.
- **Limite** : L'estimation peut être biaisée si les données sont trop rares.

34. Introduits dans les années 1970 en statistique appliquée, ces modèles ont été largement popularisés par Gelman et Hill (2006). Ils sont aujourd'hui utilisés en économie, en biologie et en intelligence artificielle.

Cas 2 : Utilisation d'un modèle bayésien hiérarchique

Avec un modèle hiérarchique, on intègre plusieurs niveaux d'information :

- Niveau national : Le coût moyen des soins en France est 250 €.
- Niveau régional : Dans la région de C, les coûts sont généralement de 270 €.
- Niveau local : Quelques données de C indiquent 260 €.

Résultat : Plutôt que d'utiliser uniquement la moyenne nationale (250 €) ou seulement les données locales (260 €), le modèle trouve un équilibre et estime le coût des soins à environ 265 €.

c) **Avantages et limites du modèle bayésien hiérarchique**

Ce tableau présente une synthèse des principaux avantages et limites de ce modèle.

Avantages		Limites	
Notations	**Commentaires**	**Notations**	**Commentaires**
(+++)	**Meilleure gestion des données rares**, en utilisant des tendances globales pour ajuster les estimations locales	(--)	**Plus complexe à implémenter**, nécessitant une modélisation fine des relations hiérarchiques
(+++)	**Intègre les dépendances spatiales** entre zones voisines et niveaux d'information	(--)	**Exige plus de puissance de calcul**, car repose sur des mises à jour dynamiques et des distributions multiples
(++)	**Réduction de l'incertitude** grâce à l'intégration des tendances régionales et nationales	(--)	**Nécessite des données fiables à différents niveaux**, sinon il y a un risque de biais
(++)	**Mise à jour dynamique des estimations**, contrairement aux modèles statiques traditionnels	(-)	**Moins réactif que certaines méthodes bayésiennes** (ex. MCMC) en cas de forte variabilité locale

Table 8 Synthèse des avantages, limites et notations du modèle bayésien hiérarchique.

En résumé, le modèle bayésien hiérarchique permet une estimation plus robuste et précise en intégrant plusieurs niveaux d'information. Contrairement à une approche purement locale, il ajuste les estimations en tenant compte des tendances globales, réduisant ainsi l'incertitude et le risque de sur-ajustement. Il est particulièrement adapté aux situations où les données locales sont rares ou insuffisantes, en équilibrant les informations issues des niveaux nationaux, régionaux et locaux. Son principal avantage est d'offrir une vision plus complète et plus fiable des estimations, bien que son implémentation soit plus complexe et gourmande en calculs.

4.4 Méthodes basées sur la distance et les réseaux

Les méthodes de lissage spatial basées sur la distance exploitent les relations de voisinage direct pour estimer une valeur inconnue en fonction des valeurs observées aux alentours. Ces approches considèrent que la proximité spatiale est un facteur déterminant dans la continuité des données spatiales.

4.4.1 Méthode de pondération inverse à la distance (IDW)

La méthode *Inverse Distance Weighting* (IDW) repose sur l'idée que plus une observation est proche du point à estimer, plus son influence doit être importante. Elle attribue ainsi à chaque donnée un poids inversement proportionnel à sa distance au point cible.

Les principes fondamentaux de cette méthode sont :

- La contribution d'un point diminue avec l'augmentation de la distance.
- L'estimation spatiale s'effectue par une pondération des observations disponibles selon leur proximité.

a) **Formulation mathématique**

La valeur estimée $Z(x)$ en un point cible x est obtenue par une moyenne pondérée des valeurs des points voisins Z_i :

$$Z(x) = \frac{\sum_{i=1}^{N} w_i Z_i}{\sum_{i=1}^{N} w_i}$$

où le poids w_i attribué à chaque point i est défini par :

$$w_i = \frac{1}{d_i^p}$$

Avec :

- d_i : la distance entre le point cible x et le point voisin i.
- p : un paramètre de pondération (souvent $p = 2$).

Interprétation : Plus la distance d_i est petite (point proche), plus le poids w_i est grand (forte influence).

b) **Avantages et limites de l'IDW** AZ Ce tableau présente les principaux avantages et limites dans un cadre d'analyse spatiale.

Avantages		Limites	
Notations	Commentaires	Notations	Commentaires
(+++)	**Simple et rapide** : calcul facile à implémenter	(---)	Ne prend en compte que la distance, sans considérer la structure spatiale globale
(++)	Idéal pour des données bien réparties	(--)	**Sensible aux valeurs extrêmes** qui peuvent fausser l'estimation
(++)	**Méthode intuitive** : les points proches ont une influence plus forte dans l'estimation	(--)	Moins efficace si les données sont très hétérogènes

Table 9 Synthèse des avantages, limites et notations de la méthode IDW.

Ainsi, l'IDW est une méthode efficace mais basique, qui fonctionne bien si les relations spatiales sont simples et que les points sont bien distribués. Elle est souvent utilisée en climatologie et en analyse environnementale [35].

4.4.2 Les chaînes de Markov spatiales

Les chaînes de Markov spatiales sont une approche probabiliste avancée qui repose sur la modélisation des interactions spatiales entre zones voisines. Contrairement aux méthodes basées uniquement sur la distance (ex. IDW, Krigeage), elles prennent en compte les relations dynamiques entre les zones.

a) **Idée principale**

- Chaque zone dépend des zones voisines selon une dynamique probabiliste.
- L'état d'une zone est influencé par les états des zones adjacentes via un processus markovien.
- Le lissage est effectué par mise à jour itérative des valeurs selon les probabilités de transition entre zones.

35. L'IDW est fréquemment employée pour interpoler les précipitations ou la température à partir de stations météorologiques.

b) **Formulation mathématique**

Le modèle repose sur la propriété markovienne d'ordre 1 :

$$P(Z_i \mid Z_{i-1}, Z_{i-2}, \dots) = P(Z_i \mid Z_{i-1}) \tag{12}$$

Cette relation exprime que l'état d'une zone ne dépend que de son état précédent proche, et non de l'ensemble des états passés ni des points éloignés : c'est la définition d'un processus de Markov. Cette propriété est valable que le processus soit homogène ou non.

Une chaîne de Markov homogène est un cas particulier dans lequel les probabilités de transition ne varient pas dans le temps (ou l'espace). Cela signifie que :

$$\forall n, \quad P(Z_{n+1} = j \mid Z_n = i) = P(Z_1 = j \mid Z_0 = i) \tag{13}$$

Dans le cas non homogène, ces probabilités peuvent dépendre de l'indice n, c'est-à-dire de la position ou du moment. Ainsi, la propriété markovienne est plus générale : elle définit une dépendance locale, tandis que l'homogénéité concerne la stabilité des règles de transition.

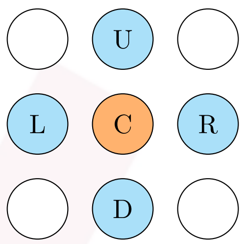
Dans un cadre multidimensionnel, comme en analyse spatiale, la propriété markovienne se généralise en supposant que l'état d'un point $Z_{\boldsymbol{s}}$, identifié par sa position $\boldsymbol{s} = (s_1, s_2, \dots, s_d)$ dans un espace à d dimensions, dépend uniquement d'un ensemble limité de points voisins, noté $\mathcal{N}(\boldsymbol{s})$. Cela signifie qu'on suppose une dépendance locale entre chaque point et son environnement immédiat, selon la relation :

$$P(Z_{\boldsymbol{s}} \mid Z_{\boldsymbol{t}}, \boldsymbol{t} \notin \mathcal{N}(\boldsymbol{s})) = P(Z_{\boldsymbol{s}} \mid \mathbf{Z}_{\mathcal{N}(\boldsymbol{s})}) \tag{14}$$

Dans le cas particulier du plan (c'est-à-dire en deux dimensions), chaque cellule est repérée par ses coordonnées (i, j). On définit alors un voisinage spatial autour de cette cellule pour modéliser les interactions locales. Deux structures de voisinage sont classiquement utilisées :

- **Voisinage de Von Neumann** : il correspond aux 4 cellules adjacentes à la cellule centrale, situées au nord, au sud, à l'est et à l'ouest :

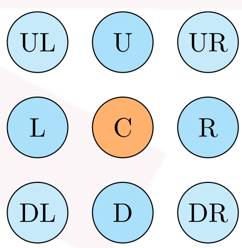
$$\mathcal{N}_{VN}(i, j) = \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\}$$



Le voisinage de Von Neumann capture uniquement les dépendances dans les directions cardinales (nord, sud, est, ouest), ce qui est pertinent pour modéliser des interactions directionnelles, par exemple dans le cas de flux suivant un axe ou de processus de diffusion anisotropes.

- **Voisinage de Moore** : il inclut non seulement les 4 cellules adjacentes, mais aussi les 4 situées en diagonale autour de la cellule centrale, soit 8 voisines au total :

$$\mathcal{N}_M(i, j) = \{(i+a, j+b) \mid a, b \in \{-1, 0, 1\}, \ (a, b) \neq (0, 0)\}$$



Le voisinage de Moore prend également en compte les cellules diagonales, permettant ainsi de modéliser des interactions plus isotropes autour de la cellule centrale. Il est particulièrement adapté lorsque les dépendances spatiales ne sont pas limitées à une direction donnée, comme dans certains phénomènes naturels homogènes.

La cellule centrale est repérée par ses coordonnées (i, j) dans une grille régulière bidimensionnelle. Les cellules voisines sont notées selon leur position relative :

- **C** (Center) : cellule centrale, de coordonnées (i, j).
- **U** (Up) : cellule au nord, $(i-1, j)$.
- **D** (Down) : cellule au sud, $(i+1, j)$.
- **L** (Left) : cellule à l'ouest, $(i, j-1)$.
- **R** (Right) : cellule à l'est, $(i, j+1)$.
- **UL**, **UR**, **DL**, **DR** : cellules en diagonale (haut-gauche, haut-droite, bas-gauche, bas-droite).

Ces structures de voisinage permettent de formaliser les interactions locales dans un modèle de chaîne de Markov spatiale, en adaptant finement la portée spatiale de la dépendance entre cellules.

c) **Étapes du modèle**

Le fonctionnement du modèle repose sur une série d'étapes structurées, décrites ci-dessous :

A. Définition des zones voisines et des probabilités de transition.
B. Calcul des probabilités conditionnelles en fonction des observations et hypothèses.
C. Mise à jour itérative des valeurs pour assurer un lissage progressif.

d) **Avantages et limites des chaînes de Markov spatiales**

Le tableau suivant résume les principaux avantages et limites dans un contexte d'analyse spatiale.

Avantages		Limites	
Notations	**Commentaires**	**Notations**	**Commentaires**
(+++)	**Modélise la dépendance spatiale** de manière plus réaliste	(--)	**Plus complexe à implémenter** qu'un IDW ou un KNN
(++)	Meilleure prise en compte des relations structurelles entre zones	(--)	Requiert une calibration des probabilités de transition
(++)	Modèles flexibles adaptés à la simulation et à la prévision spatiale Permettent une évolution itérative des valeurs, tout en prenant en compte les dépendances locales	(--)	**Convergence parfois lente**, nécessitant plusieurs itérations
(+)	Applicable à des domaines variés comme la cartographie, la géostatistique ou la modélisation environnementale	(-)	**Problèmes aux frontières des cartes** lorsque certaines zones manquent de voisins

Table 10 Synthèse des avantages, limites et notations des chaînes de Markov spatiales.

Les chaînes de Markov spatiales sont particulièrement utiles pour modéliser la continuité spatiale et sont largement utilisées en géostatistique, en modélisation de risques et en simulation d'évolution de territoires [36].

36. Ces méthodes sont notamment employées en écologie pour modéliser la répartition des espèces et en urbanisme pour prévoir l'expansion des zones urbaines.

5 Résumé

Le tableau ci-après présente une synthèse comparative des principales méthodes de lissage spatial. Chaque méthode est évaluée selon six critères : la facilité d'implémentation, la robustesse, le temps d'exécution, l'adaptabilité aux grandes bases de données, la capacité à modéliser l'incertitude et le traitement des valeurs aberrantes.

	Facilité d'implémentation	Robustesse	Temps d'exécution	Adaptation aux grosses bases de données	Modélisation de l'incertitude	Traitement des valeurs aberrantes
Bayésien hiérarchique	(--)	(+++)	(---)	(++)	(++)	(++)
Krigeage	(--)	(+++)	(--)	(--)	(++)	(–)
MCMC bayésien	(--)	(+++)	(---)	(-)	(+++)	(–)
Chaînes de Markov spatiales	(--)	(+++)	(--)	(++)	(++)	(+)
Méthodes à noyau	(++)	(++)	(-)	(++)	(-)	(--)
Lissage par quantile	(++)	(+++)	(--)	(++)	(+)	(+++)
LOESS	(++)	(++)	(-)	(-)	(-)	(--)
K-Nearest Neighbors (K-NN)	(+++)	(++)	(-)	(--)	(--)	(--)
Moyenne mobile spatiale	(+++)	(++)	(++)	(++)	(--)	(+)
Pondération Inverse à la Distance (IDW)	(+++)	(+)	(+++)	(++)	(---)	(--)

Table 11 Résumé comparatif des méthodes de lissage spatial.

Les méthodes sont classifiées en cinq niveaux de pertinence :

- Niveau 1 (vert foncé) : méthodes expertes (Krigeage, Bayésien hiérarchique), très robustes, intégrant l'incertitude et adaptées à des structures spatiales complexes.
- Niveau 2 (vert moyen) : méthodes avancées (MCMC, Markov spatiales), puissantes mais techniquement exigeantes.
- Niveau 3 (doré) : méthodes intermédiaires (KDE, Quantile, loess), souples et accessibles, mais sensibles aux paramètres ou instables.
- Niveau 4 (orange) : méthodes simples (KNN, Moyenne mobile), faciles à mettre en œuvre, mais limitées dans la modélisation spatiale.
- Niveau 5 (rouge foncé) : méthode exploratoire (IDW), rapide et intuitive, mais peu fiable pour des analyses approfondies.

Il convient de souligner qu'il ne s'agit pas d'une hiérarchie absolue : chaque méthode peut se révéler pertinente en fonction du contexte d'application, des objectifs analytiques visés et des caractéristiques propres aux données disponibles. Cette classification vise à orienter le choix des méthodes en fonction des contraintes méthodologiques et des exigences opérationnelles.

6 Packages R pour le lissage spatial

Plusieurs packages R permettent de réaliser du lissage spatial selon différentes approches. Voici une sélection des plus utilisés :

6.1 spatstat

Développé comme un package de référence pour l'analyse des processus ponctuels spatiaux, spatstat offre un large éventail de fonctionnalités pour la modélisation, l'estimation et la visualisation des distributions de points [a]

- Permet de lisser une densité spatiale à partir d'un jeu de points avec la fonction density.ppp().
- Propose des méthodes avancées de sélection de bande passante (bw.diggle, bw.frac, etc.).
- Inclut plusieurs fonctions d'estimation et d'analyse spatiale adaptées aux distributions de points.

a. Voir la documentation du package spatstat : https://spatstat.org/.

6.2 btb (Beyond The Border)

Développé par les agents de l'Insee depuis 2018, le package btb implémente une méthode de lissage par noyau quadratique [a].

- Prend en compte les effets de bord, ce qui est essentiel pour l'analyse de zones limitées géographiquement.
- Permet un lissage quantile, particulièrement adapté aux variables comme les revenus.
- Utile pour la production de cartes de densité spatiale sur des découpages administratifs précis.

a. Voir le package btb sur GitHub : https://github.com/InseeFr/btb

6.3 ggplot2 et sf

Les packages ggplot2 et sf permettent une visualisation avancée et une manipulation efficace des données spatiales [a].

- ggplot2 permet de visualiser des cartes lissées avec des gradients de couleur grâce aux fonctions geom_sf() et stat_density2d(), qui permettent respectivement de représenter des objets géographiques et d'afficher la densité des points sur une carte.
- sf facilite la manipulation et la conversion des données spatiales, et s'intègre efficacement avec d'autres packages du tidyverse.

a. Le package ggplot2 est décrit ici : https://ggplot2.tidyverse.org/

6.4 KernelKnn

Le package KernelKnn est conçu pour appliquer la méthode des k-Plus Proches Voisins (k-NN) en lissage spatial [a].

- Implémente une estimation locale en fonction des k voisins les plus proches.
- Idéal pour les zones hétérogènes, où l'on veut adapter la taille du voisinage à la densité des points.
- Compatible avec des approches non paramétriques et fonctionne bien pour la prévision spatiale.

a. Voir la documentation de KernelKnn : https://cran.r-project.org/web/packages/KernelKnn/

Le choix du bon package pour le lissage spatial dépend du type de données et de l'objectif de l'analyse.

- Pour une analyse de densité ponctuelle : spatstat est recommandé.
- Pour un lissage à l'échelle administrative : btb est plus adapté.
- Pour la visualisation avancée : ggplot2 et sf sont les meilleurs outils.
- Pour une approche k-NN : KernelKnn permet un lissage flexible et adaptatif.

Ces outils permettent une meilleure compréhension des phénomènes spatiaux et facilitent la prise de décision basée sur des analyses rigoureuses.

Conclusion

Le lissage spatial reste jusqu'ici la méthode la plus pertinente pour prendre en compte la proximité géographique dans la création du zonier. Toutefois, son efficacité dépend fortement du support géographique utilisé.

L'approche classique, basée sur des zonages administratifs (IRIS, communes), présente des limites : elle ne reflète pas les parcours de soins réels ni la densité médicale locale. À l'inverse, lisser les données par bassin de vie c'est-à-dire par des zones définies en fonction des flux de soins et des comportements de recours, permet de mieux capter la réalité fonctionnelle des territoires.

Le choix de la méthode de lissage doit donc s'adapter aux spécificités des territoires et des données disponibles :

- En zones denses avec données abondantes : les méthodes par noyau (KDE) ou le krigeage sont recommandées, car elles offrent un bon équilibre entre précision locale et stabilité globale.
- En zones peu denses ou avec incertitudes fortes : les modèles bayésiens hiérarchiques ou les simulations MCMC permettent de lisser tout en intégrant l'incertitude, avec une meilleure robustesse dans les zones mal renseignées.
- Pour des applications rapides et opérationnelles : les moyennes mobiles spatiales ou les k-NN (k plus proches voisins) sont simples à mettre en œuvre et conviennent bien à des cas d'usage pragmatiques.
- En présence de valeurs extrêmes ou de disparités fortes : les méthodes de lissage par quantile permettent de limiter l'impact des outliers et de mieux refléter les disparités locales.
- Lorsque les relations de voisinage sont complexes ou structurelles : les approches par les chaînes de Markov spatiales sont les plus adaptées pour modéliser finement les interdépendances spatiales.

NEXIALOG CONSULTING

References

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician 46*(3), 175–185. https://sci-hub.se/10.1080/00031305.1992.10475879.

Banerjee, S., Carlin, B. P. et Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data* (2 ed.). https://api.pageplace.de/preview/DT0400.9781439819180_A37871102/preview-9781439819180_A37871102.pdf.

Boisseau, M. (2021). Module lissage : mettez en évidence les tendances spatiales de vos phénomènes. *Articque by ChapsVision*. https://www.articque.com/solutions/cartes-et-donnees/des-fonctionnalites-multiples/lissage/.

Brunsdon, C., otheringham, A. S. et Charlton, M. E. (1996). Geographically weighted regression : A method for exploring spatial nonstationarity. *International Journal of Geographical Information Science 10*(5), 281–298. https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1538-4632.1996.tb00936.x.

Casper, M., Kramer, M. R., Peacock, J. M. et Vaughan, A. S. (2019). Population health, place, and space : Spatial perspectives in chronic disease research and practice. *Preventing Chronic Disease 16*, E123. https://www.cdc.gov/pcd/issues/2019/19_0237.htm.

Cissé, D. (2022). Construction de micro-zoniers en assurance mrh à l'aide d'outils de data science. *Mémoire de fin d'études, ENSAE Paris*. https://www.institutdesactuaires.com.

Clemente, G. P., Della Corte, F. et Zappa, D. (2024). Hierarchical spatial network models for road accident risk assessment. *Annals of Operations Research*. https://link.springer.com/article/10.1007/s10479-024-06049-7.

Faller, D. (2022). Spatial statistical modelling of insurance claim frequency using markov chain monte carlo based inference with riemannian langevin diffusion and continuous spatial dependence. *Master's Thesis, Lund University*. https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=9073261&fileOId=9073270.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. et Rubin, D. B. (2021). *Bayesian Data Analysis* (3 ed.). https://sites.stat.columbia.edu/gelman/book/BDA3.pdf.

Genebes, L. and Renaud, A. and Sémécurbe, F. (2018). Lissage spatial. *Manuel d'analyse spatiale*, INSEE, n° 131, chapitre 8. pp. 1–27, https://www.insee.fr/fr/statistiques/fichier/3635442/imet131-l-chapitre-8.pdf.

Kauhl, B., Schweikart, J. Krafft, T., Keste, A. et Moskwyn, M. (2016). Do the risk factors for type 2 diabetes mellitus vary by location ? a spatial analysis of health insurance claims in northeastern germany using kernel density estimation and geographically weighted regression. *International Journal of Health Geographics 15*(38). https://doi.org/10.1186/s12942-016-0068-2.

Kopczyk, D. (2024). Geography modelling in insurance pricing. *Quantee Blog*. https://www.quantee.ai/resources/geography-modelling-in-insurance-pricing.

Larmarange, J., Vallo, R., Yaro, S. et Msellati, P. and Méda, N. (2011). Méthodes pour cartographier les tendances régionales de la prévalence du vih à partir des enquêtes démographiques et de santé (eds). *Cybergeo : European Journal of Geography* (539). https://journals.openedition.org/cybergeo/23782.

Lemarchand, O. et Jeannée, N. (2009). Méthodes de cartographie et approche géostatistique. la cartographie de la pollution au dioxyde d'azote en alsace. *Cahier des thèmes transversaux ArScAn*, IX. pp. 203–214, https://hal.science/hal-02264055v1/document.

Marlier, A. (2018). Zonier sur un risque à événements rares : l'inondation. *Mémoire de fin d'études, ISFA - Lyon*. https://www.ressources-actuarielles.net/EXT/ISFA/1226-02.nsf/0/76b964cd9eff098cc12581dd0072127e/$FILE/Memoire_Actuariat_MARLIER_Aurelia.pdf.

Mathis, J. (2009). Élaboration d'un zonier en assurance de véhicules par des méthodes de lissage spatial basées sur des simulations mcmc. *Mémoire de fin d'études, Université Louis Pasteur de Strasbourg*. https://www.ressources-actuarielles.net/C12574E200674F5B/0/D7313694D94D10BAC125783A005364A2.

Nakaya, T., Fotheringham, A. S., Brunsdon, C. et Charlton, M. (2005). Geographically weighted poisson regression for disease association mapping. *Statistics in Medicine 24* (17), 2695–2717. https://mural.maynoothuniversity.ie/id/eprint/5955/1/MC_Poisson.pdf.

Rose, T. (2021). Méthodologie et lissage des cartes. *Agreste Grand Est – Dossier numéro 3.* pp. 82–84, https://draaf.grand-est.agriculture.gouv.fr/IMG/pdf/2021-03-atlas2020-p82a84_methodo_et_lissage_des_cartes_cle893c8d.pdf.

Sepulveda, C. (2017). Modélisation du risque géographique en santé, pour la création d'un nouveau zonier. comparaison de deux méthodes de lissage spatial. *Mémoire de fin d'études, ISUP – Paris.* https://www.ressources-actuarielles.net/EXT/ISFA/1226-02.nsf/0/68c1024b306d64a8c12580b4005fb59c//$FILE/SEPULVEDA.pdf.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London : Chapman and Hall. pp. 1–175, https://doi.org/10.1201/9781315140919.

Statistics Easily (2021). Qu'est-ce que c'est : le lissage du voisin le plus proche (k-nearest neighbor smoothing). *Statistics Easily – Glossaire.* https://fr.statisticseasily.com/glossario/what-is-k-nearest-neighbor-smoothing/.

Viry, M. et Giraud, T. (2020). Magrit : Cartographie thématique. *UAR 2414 RIATE – Centre pour l'analyse spatiale et la géovisualisation.* https://github.com/riatelab/magrit.

Annexe

Notation	Commentaire
(+++)	Très favorable : Avantage-clé, structurant pour la méthode, indispensable pour la performance.
(++)	Favorable : Bénéfice important mais non exclusif, améliore les résultats sans être indispensable.
(+)	Légèrement favorable : Atout secondaire, utile dans des cas spécifiques mais non déterminant.
(−)	Légèrement défavorable : Inconvénient modéré, à prendre en compte dans certains cas.
(− −)	Défavorable : Limitation importante à prendre en compte lors du choix ou de l'adaptation de la méthode.
(− − −)	Très défavorable : Handicap majeur, empêche l'utilisation de la méthode dans certains cas.

Table 12 Notation d'échelle utilisée pour les méthodes de lissage spatial.

Nexialog Consulting est un cabinet de conseil spécialisé en Banque et en Assurance. Organisés autour de 3 domaines d'activité - Risques Bancaires, Financiers & Assurantiels - nous intervenons au sein des équipes métiers afin de les accompagner depuis le cadrage jusqu'à la mise en œuvre de leurs projets. Associant innovation et expertise, le savoir-faire de notre cabinet a permis de consolider notre positionnement sur ce segment et de bénéficier d'une croissance forte et régulière.

Les besoins de nos clients étant en constante évolution, nous nous adaptons continuellement pour proposer le meilleur accompagnement. Le département R&D de Nexialog Consulting se donne pour objectif de proposer des solutions innovantes à des problématiques métier ou d'actualité. Pour cela, nous nous appuyons sur des librairies internes et sur le travail de nos consultants. Le pôle R&D Nexialog a également pour mission de former les collaborateurs sur l'évolution des techniques et la réglementation en lien avec leur activité.

Site web du cabinet : https://www.nexialog.com
Publications : https://www.nexialog.com/publications/

Contacts

Ali BEHBAHANI
Associé, Fondateur
Tél : + 33 (0) 1 44 73 86 78
Email : abehbahani@nexialog.com

Christelle BONDOUX
Associée, Directrice commerciale
Tél : + 33 (0) 1 44 73 75 67
Email : cbondoux@nexialog.com

Hugo RAPIOR
Responsable de programme R&D
Email : hrapior@nexialog.com

Areski COUSIN
Directeur scientifique R&D
Email : acousin@nexialog.com