



HAL
open science

Compte-rendu: Loonis V., Bellefon M.-P. (dir.), 2018 : Manuel d'analyse spatiale. Théorie et mise en œuvre pratique avec R, Insee Méthodes n° 131, Insee, Eurostat, 392 p

Thierry Feuillet

► **To cite this version:**

Thierry Feuillet. Compte-rendu: Loonis V., Bellefon M.-P. (dir.), 2018 : Manuel d'analyse spatiale. Théorie et mise en œuvre pratique avec R, Insee Méthodes n° 131, Insee, Eurostat, 392 p. Cybergeog : Revue européenne de géographie / European journal of geography, 2018. <halshs-01972198>

HAL Id: halshs-01972198

<https://shs.hal.science/halshs-01972198v1>

Submitted on 7 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/329453717>

Loonis V., Bellefon M.-P. (dir.), 2018, Manuel d'analyse spatiale. Théorie et mise en oeuvre pratique avec R, Insee Méthodes n° 131, Insee, Eurostat, 392 p.

Article in *CyberGeo* · December 2018

CITATIONS

0

READS

15

1 author:



Thierry Feuillet

Université de Vincennes - Paris 8

84 PUBLICATIONS 445 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Comparative analysis of Icelandic and Martian glacial and paraglacial morphological systems [View project](#)



Arctic facing Climate Change: analysis of paraglacial geomorphological crisis (2009-2014 - IUF grants) [View project](#)



Loonis V., Bellefon M.-P. (dir.), 2018, *Manuel d'analyse spatiale. Théorie et mise en œuvre pratique avec R*, Insee Méthodes n° 131, Insee, Eurostat, 392 p.

Thierry Feuillet



Édition électronique

URL : <http://journals.openedition.org/cybergeography/29853>

ISSN : 1278-3366

Éditeur

UMR 8504 Géographie-cités

Ce document vous est offert par Université Paris 8



Référence électronique

Thierry Feuillet, « Loonis V., Bellefon M.-P. (dir.), 2018, *Manuel d'analyse spatiale. Théorie et mise en œuvre pratique avec R*, Insee Méthodes n° 131, Insee, Eurostat, 392 p. », *Cybergeography : European Journal of Geography* [En ligne], Revue de livres, mis en ligne le 05 décembre 2018, consulté le 06 décembre 2018. URL : <http://journals.openedition.org/cybergeography/29853>

Ce document a été généré automatiquement le 6 décembre 2018.

© CNRS-UMR Géographie-cités 8504

Loonis V., Bellefon M.-P. (dir.), 2018,
*Manuel d'analyse spatiale. Théorie et
mise en œuvre pratique avec R*,
Insee Méthodes n° 131, Insee,
Eurostat, 392 p.

Thierry Feuillet

- 1 Cet imposant manuel collaboratif (392 p., incluant 25 auteurs dont 18 appartenant à l'Insee) affiche clairement l'ambition de constituer une référence à la fois théorique et pratique pour l'analyse statistique des données géolocalisées, en particulier (mais pas seulement) au sein des instituts statistiques nationaux. Pour le resituer dans son contexte, cet ouvrage a été proposé par la division des méthodes et référentiels géographiques de l'Insee à Eurostat et au forum européen de statistique et de géographie (EFGS) – qui l'ont donc accepté – et s'inscrit dans le cadre du développement d'un système d'information statistique géoréférencée au sein de l'Insee. Le manuel est disponible en ligne (en français et bientôt en anglais) sur les sites d'Eurostat et de l'EFGS.
- 2 Comme l'expliquent Vincent Loonis et Marie-Pierre de Bellefon – respectivement directeur et coordinatrice de l'ouvrage – en avant-propos, le point fort du manuel est clairement de combiner équitablement théorie et pratique, ce qui n'est pas si fréquent dans ce type de littérature. Certains ouvrages sont en effet trop théoriques pour les non-statisticiens (par ex. Cressie, 1993 ; LeSage et Pace, 2009), d'autres presque exclusivement pratiques (par ex. Bivand *et al.*, 2008 ; Brunson *et al.*, 2015). Chaque chapitre débute par une mise au point théorique (entendu ici comme formalisation mathématique), parfois également conceptuelle, suivie d'une mise en application en R. Cette logique efficace, couplée au choix très pertinent de l'outil (rappelons que R est libre et communautaire, et de plus en plus adapté à l'analyse de données spatiales) et à l'explicitation souvent détaillée des équations, rendent la lecture très agréable et comblera des lecteurs aux attentes variées, si tant est qu'ils disposent d'un bagage statistique relativement

important. L'ensemble comporte également un nombre important de figures originales pertinentes et souvent pédagogiques, même si bon nombre de cartes ne respectent pas certaines règles basiques de sémiologie (absence quasi systématique d'échelles, mauvais choix de couleurs (par ex. figure 8.11), illisibilité partielle de l'information (par ex. figures 5.22, 9.1, 9.2, 9.6), absence de légende (par ex. figures 5.2, 5.13, 5.20, 9.12, 9.13, 11.1, 11.8). Son autre point fort est d'embrasser un très large spectre de méthodes de statistiques spatiales et de géostatistiques, dont certaines sont rarement présentées dans des manuels en français, telles que la régression géographiquement pondérée et le krigeage. La rigueur et la fiabilité des formalisations mathématiques sont de mise, et il en revient à l'excellence des auteurs, dont certain-es sont mondialement reconnu-es dans leur domaine de spécialité.

- 3 Il convient cependant de noter que ce manuel n'a pas été écrit par, ni pour, des géographes en particulier. En effet, et quoi de plus normal pour un manuel de l'Insee, l'économétrie et les applications économiques prédominent sur le reste. Ce n'est bien entendu par un reproche – le mot géographie n'apparaissant pas dans le titre et l'ouvrage n'ayant donc pas cette prétention – mais cela s'en ressent parfois sur le caractère partiel de la définition de certains concepts fondamentaux de la géographie théorique. L'hétérogénéité spatiale n'est par exemple définie ici qu'à travers l'hétéroscédasticité ou l'instabilité structurelle des paramètres d'une régression, c'est-à-dire sous un angle purement économétrique. On ne trouvera non plus presque aucune définition ou mention des grandes lois ou modèles théoriques de l'analyse spatiale tel que ceux établis par les géographes, exceptée la première loi de Tobler bien entendu, et une rapide mention de la théorie des lieux centraux et des modèles gravitaires (p. 340). Le manuel ne s'adresse donc clairement pas aux étudiants de géographie d'un niveau inférieur au master, voire au doctorat, mais davantage aux étudiants économistes/économètres/statisticiens et à tous les chercheurs, ingénieurs et chargé(e)s d'études ayant à mobiliser des méthodes de statistiques spatiales dans leurs analyses. Il constituera pour ce public, à n'en pas douter, un précieux compagnon pour de nombreuses années.
- 4 Le manuel comprend 14 chapitres découpés en quatre parties. La 1^{ère} partie, « Décrire les données géolocalisées », donne d'abord l'occasion de définir les notions et les outils de base en analyse spatiale descriptive (chap. 1), de façon efficace. Cela comprend la définition des types de données spatiales, la présentation des principales règles de sémiologie graphique (qui n'ont pas toujours été respectées dans le corps même du manuel, comme mentionné plus haut, et qui auraient par ailleurs mérité de se référer au récent et excellent *manuel de cartographie*, de Lambert et Zanin [2016]) et la présentation de scripts de cartographie et de géotraitements simples sous R (y compris à l'aide du récent package *sf* (pour *simple features*), qualifié à juste titre d'avenir du traitement de données spatiales sous R). Les auteur-es présentent ensuite de façon complète et très bien illustrée les différentes manières de codifier les structures de voisinage et de définir les pondérations spatiales.
- 5 La 2^{ème} partie, *Mesurer l'importance des effets spatiaux*, s'intéresse aux indices globaux et locaux d'autocorrélation spatiale (chap. 3), aux configurations de points (chap. 4) et à la géostatistique (chap. 5), en suivant donc la typologie classique des données spatiales : surfaciques, ponctuelles et continues, respectivement. Le chapitre 3 présente les indices habituels d'autocorrélation spatiale globale et locale pour les variables quantitatives (Moran, Geary, Getis-Ord), mais aussi pour les variables qualitatives (statistiques des *join count*). Le chapitre suivant offre une présentation particulièrement fournie et très claire

de l'analyse des semis de points, incorporant la question importante des effets de bord d'un point de vue théorique. Les auteurs y présentent ensuite de très nombreux indices de quantification des configurations ponctuelles (fonction K de Ripley et ses variantes, ainsi que d'autres indicateurs récents encore peu connus, dont certains développés par les auteurs) et des exemples de modélisation complexe de processus générateurs de distributions (processus par ailleurs utilement listés en annexe). À noter que les auteurs ont ici choisi un exemple d'application en écologie, et non en économie. Le chapitre 5 est quant à lui dédié à la géostatistique tout au long de ses 36 pages. L'auteur y définit les notions importantes et présente les principes et les modèles d'ajustement du variogramme. Le krigeage ordinaire et ses extensions sont ensuite déclinés, avec des applications en géographie physique, domaine qui s'y prête particulièrement bien. Si l'auteur de ce chapitre n'est assurément pas cartographe, il est en revanche très bon pédagogue !

- 6 La 3^{ème} partie, « *Prendre en compte les effets spatiaux* », constitue le gros bloc du manuel (184 pages, 7 chapitres). Les deux premiers chapitres sont consacrés aux modèles d'économétrie spatiale, d'abord sur données transversales (chap. 6) puis sur données de panel (chap. 7). Rappelons que ces modèles sont des régressions qui prennent explicitement en compte la dimension spatiale dans leurs équations (via une matrice de pondération spatiale), mais dont les coefficients sont estimés en moyenne sur tout le territoire étudié. Le chapitre 6 reprend, en plus détaillé, une excellente synthèse déjà publiée par les auteurs (Le Saout et Floch, 2016). Ils y présentent les principaux modèles (modèle autorégressif, modèle à erreurs spatiales, etc.), leurs critères de sélection et la façon dont il convient d'interpréter les coefficients (ce qui est loin d'être simple en raison des rétroactions spatiales entre l'observation et son voisinage, comme l'expliquent justement les auteurs). Le chapitre 7 sur les données de panel est utile mais fait un peu redondance dans sa première partie et aurait peut-être mérité d'être mieux harmonisé avec le chapitre précédent, notamment dans la notation mathématique. Ce chapitre paraît également moins accessible que les autres à qui ne dispose pas de solides acquis en statistique. Le chapitre 8 est dédié au lissage spatial et est judicieusement (d'un point de vue pédagogique) placé juste avant le chapitre sur la régression géographiquement pondérée. Le chapitre débute par un paragraphe très clair sur l'origine du lissage dans les analyses géographiques. Si les auteurs présentent le lissage comme un moyen de limiter en partie le MAUP (c'est le cas pour ce qui concerne les effets de zonage), on pourrait aussi le considérer comme une manière de le révéler : les effets statistiques de la variation du rayon de lissage (bande passante) en sont justement une illustration (effets d'échelle). Les auteurs poursuivent par le paramétrage de la méthode, et notamment du choix des noyaux. La figure 8.5 part d'une bonne intention mais ne permet malheureusement pas de visualiser les différences d'estimation en fonction du noyau choisi. En conclusion, les auteurs ont raison de rappeler que la méthode du lissage ne doit en théorie s'appliquer qu'à des « phénomènes dont l'ordre de grandeur est très supérieur au rayon de lissage » (p. 238).
- 7 Le chapitre 8 embraye logiquement sur la régression géographiquement pondérée (GWR en anglais). Cette méthode, développée par des géographes et de plus en plus populaire en raison de sa relative facilité d'accès, est ici très clairement présentée, malgré la mauvaise qualité des cartes. Des méthodes d'estimation robustes aux valeurs aberrantes sont également proposées, ce qui montre encore une fois la rigueur des auteurs pour tout ce qui relève du domaine statistique. Les applications en R sont également très complètes,

y compris en ce qui concerne les fonctions de diagnostic des modèles. Le paragraphe sur les causes possibles de la non-stationnarité (p. 263) est en revanche très court, malgré l'intérêt de la question. Un détail enfin : la GWR a été proposée en 1996 et non en 1998, contrairement à ce qui est dit en conclusion.

- 8 Le chapitre 10 a pour titre « Echantillonnage spatial ». En réalité, seule une partie de ce problème classique en géographie est considérée ici, puisque le chapitre ne traite que de l'échantillonnage spatial d'enquêté-es dans le cadre d'un plan de sondage. Une comparaison d'une dizaine de méthodes est proposée. Le chapitre n'est pas simple à lire pour des non-statisticiens, car la dimension mathématique est très présente. Le très pertinent chapitre 11 s'intéresse au problème des données non exhaustives en économétrie spatiale, avec un focus sur les données d'enquête. Comme le rappellent clairement les auteurs, la non-exhaustivité des données spatialisées a, en théorie, pour effet de biaiser l'estimation du paramètre spatial d'un modèle de type SAR (modèle spatial autorégressif) et les effets spatiaux en général. Les auteurs le prouvent d'ailleurs de façon très pédagogique à partir d'un jeu de données simulées à l'échelle des NUTS3 européens. Ils en ont tiré des échantillons selon différentes méthodes (aléatoirement, par grappes, etc.) et ont comparé les paramètres des échantillons avec les paramètres réels de modèles SAR. En plus d'un effet d'échantillonnage spatial, ils prouvent ainsi statistiquement un « effet taille » (plus l'échantillon est grand, plus on tend vers la vraie valeur du paramètre). La figure 11.4 est très parlante à ce propos, même s'il est dommage qu'elle ne dispose pas d'une légende. Les auteurs proposent ensuite deux « pistes de résolution » : l'agrégation spatiale des données à un niveau supérieur et leur imputation (méthode consistant à estimer des données manquantes), mises en application sur des données réelles à travers deux exemples (le code R n'est ici pas partagé). Ceux ou celles qui ne travaillent pas ou peu sur les données de recensement de l'Insee ne seront pas nécessairement en phase avec cette phrase de conclusion selon laquelle « l'estimation de modèles spatiaux sur des données échantillonnées reste rare » (p. 310). Au contraire, bon nombre de modèles spatiaux dans la littérature concernent des échantillons et non des populations (en écologie ou en épidémiologie spatiale, entre autres), ce qui n'en rend que plus utile la lecture attentive de ce chapitre. La 3^{ème} partie se clôt sur le chapitre 12, dédié au problème des petits échantillons et au traitement de variables dépendantes qualitatives (modèles linéaires généralisés).
- 9 La dernière partie, « Prolongements », comprend deux chapitres. Le chapitre 13 est presque double, en traitant des analyses de graphes, mais presque en fait uniquement comme moyen de partitionner l'espace. Or, graphes et partitionnement ne se subsument pas. Peut-être aurait-il été plus judicieux d'en faire deux chapitres ? Quoiqu'il en soit, le texte est clair et, une fois n'est pas coutume, des travaux de géographes francophones sont cités ! Néanmoins, retour à la case départ avec la première phrase du chapitre suivant : « La profusion récente de sources de données géolocalisées [...] offre de nombreux champs aux économistes, démographes ou sociologues » (p. 361). En effet, quelle drôle d'idée que des géographes travaillant sur des données géolocalisées ? Trêve d'ironie, ce chapitre original nous place dans la peau de « l'expert en confidentialité statistique » qui dispose de données individuelles exhaustives devant être diffusées à un niveau géographique fin, mais sans que quiconque puisse ensuite y identifier un individu (ou un ménage). Ce problème est bien entendu commun dans les instituts de statistique publique. Les auteurs présentent les principales méthodes de gestion du risque de divulgation : méthodes pré-tabulées et post-tabulées, imputation locale, agrégation

spatiale. Cette dernière a été utilisée par l'Insee en 2013 pour publier des statistiques de revenus dans les carreaux de 200 mètres. Ces carreaux ont été agrégés en rectangles quand ils comprenaient moins de 11 ménages fiscaux, seuil réglementaire à respecter. Ces méthodes sont comparées en termes de conservation des corrélations spatiales et de perturbations des observations.

- 10 Pour conclure, ce *manuel d'analyse spatiale* présente de façon claire un nombre conséquent de méthodes de statistique spatiale, des plus communes aux plus originales, à la fois théoriquement et empiriquement (à l'aide de R). L'ouvrage s'adresse en priorité à des chercheurs ou à des chargés d'études souhaitant mettre en œuvre des méthodes spécifiques dans l'analyse de leurs données spatiales, ou à des étudiants en économie ou en statistiques. Les connaissances statistiques assez poussées nécessaires à sa compréhension le rendront difficilement accessible aux étudiants de géographie avant le 3^{ème} cycle. En cela, le très prochain *manuel de géographie quantitative* (Feuillet *et al.*, 2019) apparaîtra complémentaire et plus adapté à ce public, en compensant la moindre présence des formules mathématiques par une approche à la fois plus conceptuelle et plus attachée aux racines de la géographie théorique.

BIBLIOGRAPHIE

Bivand R., Pebesma E., Gomez-Rubio V., 2008, *Applied spatial data analysis with R*, New York, Springer.

Brunsdon C., Comber L., 2015, *An introduction to R for spatial analysis and mapping*, Londres, Sage.

Cressie N., 1993, *Statistics for spatial data*, New York, John Wiley & Sons.

Feuillet T., Cossart E., Commenges H., 2019. *Manuel de géographie quantitative. Concepts, outils, méthodes*, Paris, A. Colin.

Le Saout R., Floch J.-M., 2016. *Économétrie spatiale : une introduction pratique*, Insee.

Lambert N., Zanin C., 2016, *Manuel de cartographie. Principes, méthodes, applications*, Paris, A. Colin.

LeSage J., Pace R., 2009, *Introduction to spatial econometrics*, Londres, Chapman & Hall/CRC.

AUTEUR

THIERRY FEUILLET

Université Paris 8 Vincennes-Saint-Denis

LADYSS (UMR 7533), France

thierry.feuillet@univ-paris8.fr