

Projet pour le cours de statistique spatiale

M2 ISN, Année Universitaire 2020–2021

Sophie Dabo et Azzouz Dermoune

Vous devez rendre à la fois un document reprenant les réponses aux questions posées et le code R utilisé pour obtenir les réponses.

Les deux fichiers sont à envoyer par email à déposer sous Moodle. Veuillez intituler vos fichiers en utilisant la forme suivante : `nom_prenom.docx` et `nom_prenom.R`

1 Géostatistique

1.1 Problématique d'étude et base de données

On souhaite réaliser une carte pluviométrique de l'état du Parana au Brésil durant la période sèche (Mai-Juin). Pour cela on dispose durant cette période de mesures (hauteur de pluie moyenne calculée en différentes années) des précipitations, effectuées en 143 stations.

Le fichier *parana.txt* en format ascii contient un tableau formé des colonnes suivantes :

- colonne 1 : abscisses des stations
- colonne 2 : ordonnées des stations
- colonne 3 : mesures aux stations

Le fichier *parana.borders.txt* contient les coordonnées géographiques des frontières de l'état du Parana.

Questions

1.1.1 Analyse exploratoire

1. Tracer la carte des données avec les frontières de l'état.
2. Faire une étude variographique des précipitations

1.1.2 Prédiction spatiale

1. Construire une grille (avec un pas de votre choix) sur la carte l'état du Parana
2. Cartographier les précipitations en chaque point d'une grille par krigeage (en utilisant les krigeages ordinaire et universelle adaptés, argumenter vos choix) à partir des mesures aux 143 stations.
3. Comparer les deux cartes obtenues par krigeage et sélectionner celle qui vous donne le meilleur modèle de prévision. Commenter vos résultats.

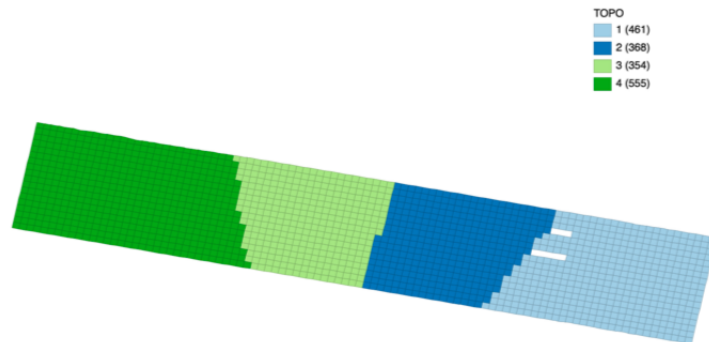


FIGURE 1 – Localisation des quatre types de terrains

2 Econométrie spatiale

2.1 Problématique d'étude

Ce projet vise à analyser l'impact de l'utilisation de nitrogène comme engrais sur les rendements du maïs (mesurés en nombre de quintaux par hectare). La base de données concerne un champ en Argentine qui comprend 4 types de sols différents (*Slope W*, *Hilltop*, *Slope E* et *Low E*), voir Figure 1.

Afin d'obtenir des résultats plus lisibles, veuillez tout d'abord multiplier la variable *YIELDS* par 100, pour obtenir un rendement exprimé en kg par hectare, plutôt qu'en quintaux par hectare. Nommez cette nouvelle variable *YIELDS100*

2.2 Base de données et modèle à estimer

La base de données pour effectuer ce devoir se trouve sur Moodle et est intitulée "yields" (la base est constituée du fichier .shp, .shx et .dbf). Si vous utilisez **Géoda** pour créer les matrices de pondération spatiale, la variable d'identification à utiliser est appelée **OBS**. Cette base de données contient des indications sur la production de maïs de 1738 parcelles de culture de maïs. La Table 1 présente les variables incluses dans la base.

TABLE 1 – Base de données

Libellé	Description
Id (01)	identifier
Top2	variable muette : 1 si parcelle est de type Slope E (terrain pentu orienté Est) et 0 sinon
Top3	variable muette : 1 si parcelle est de type Hilltop (terrain au sommet de la colline) et 0 sinon
Top4	variable muette : 1 si parcelle est de type Slope W (terrain pentu orienté Ouest) et 0 sinon
Nxtp2 à Nxtp4	interaction Nitrogène - zone topographique
Longitude	longitude
Latitude	latitude
Obs	numéro d'observation
Yield	rendement du maïs (quintal par hectare)
N	Quantité de nitrogène utilisée (kg par hectare)
N2	Quantité de nitrogène au carré
Topo	zone : Low E (1), Slope E (2), Hilltop (3), Slope W (4). Low E représente un terrain en vallée exposé à l'Est.
Bv	luminosité (proxy pour une faible teneur en matière organique)
Bv2	luminosité au carré
Nxbv	interaction Nitrogène - luminosité
Bvxt2 to Bvxt4	interaction luminosité - zone topographique
Bv2xt2 to Bv2xt4	interaction luminosité au carré et zone topographique
Sat	rayonnement (proxy pour une faible teneur en matière organique)
Sat2	rayonnement au carré
Nxsat	interaction Nitrogène - rayonnement
Satxt2 to Satxt4	interaction rayonnement - zone topographique
Sat2xt2 to Sat2xt4	interaction carré du rayonnement - zone topographique

2.3 Questions

2.3.1 Analyse exploratoire de données spatiales

1. Construisez une matrice de poids basée sur la contiguïté au sens de la reine à l'ordre 1
2. Analysez et interprétez l'autocorrélation spatiale globale dans la variable *YIELDS100*
3. Analysez la présence d'autocorrélation spatiale locale dans la variable *YIELDS100*. Effectuez l'analyse avec les I de Moran locaux
4. Analyser l'autocorrélation spatiale locale dans la variable *YIELDS100* à l'aide des statistiques de Getis locales (Local G). Quelles sont les différences avec l'analyse basée sur le I de Moran local ?

2.3.2 Econométrie spatiale

Pour répondre à ces questions, considérer un seuil de significativité à 10%.

1. Construisez la spécification économétrique présentée en (1)

$$YIELDS100 = \beta_0 + \beta_1 N + \beta_2 N^2 + \beta_3 TOP2 + \beta_4 TOP3 + \beta_5 TOP4 + \varepsilon \quad (1)$$

2. En utilisant la méthode du spécifique au général et la matrice de connectivité créée au point précédent, identifiez et écrivez la spécification spatiale la plus adéquate pour modéliser l'autocorrélation spatiale. Justifiez votre réponse. De plus, écrivez la fonction de vraisemblance associée au modèle sélectionné.
3. Interprétez l'impact d'une variation de la quantité d'engrais, évaluée à la valeur moyenne de l'échantillon, sur le rendement du maïs (attention à la forme fonctionnelle). Cet impact est-il significatif ?
4. Estimer maintenant le modèle (1) en intégrant le décalage spatial de la variable dépendante.
 - (a) Quelle est la valeur du coefficient autorégressif spatial ? Est-il significativement différent de 0 ? Reste-t-il de l'autocorrélation spatiale résiduelle dans les erreurs ?

- (b) Comparez les résultats obtenus avec ceux des MCO
- (c) Quels sont les effets direct et indirect moyens d'une variation de la quantité utilisée d'engrais, évaluée à la valeur moyenne de l'échantillon, sur le rendement du maïs ?
- (d) Pour quelle(s) parcelle(s) l'effet direct est-il maximal ? Et quelle est la valeur de l'effet ?
- (e) Au regard des résultats de l'estimation du modèle avec variable endogène spatialement décalée et du modèle avec autocorrélation spatiale des erreurs, quelle spécification préférez-vous ? Justifiez votre réponse.