RAPPORT PROJET ECONOMETRIE SPATIALE

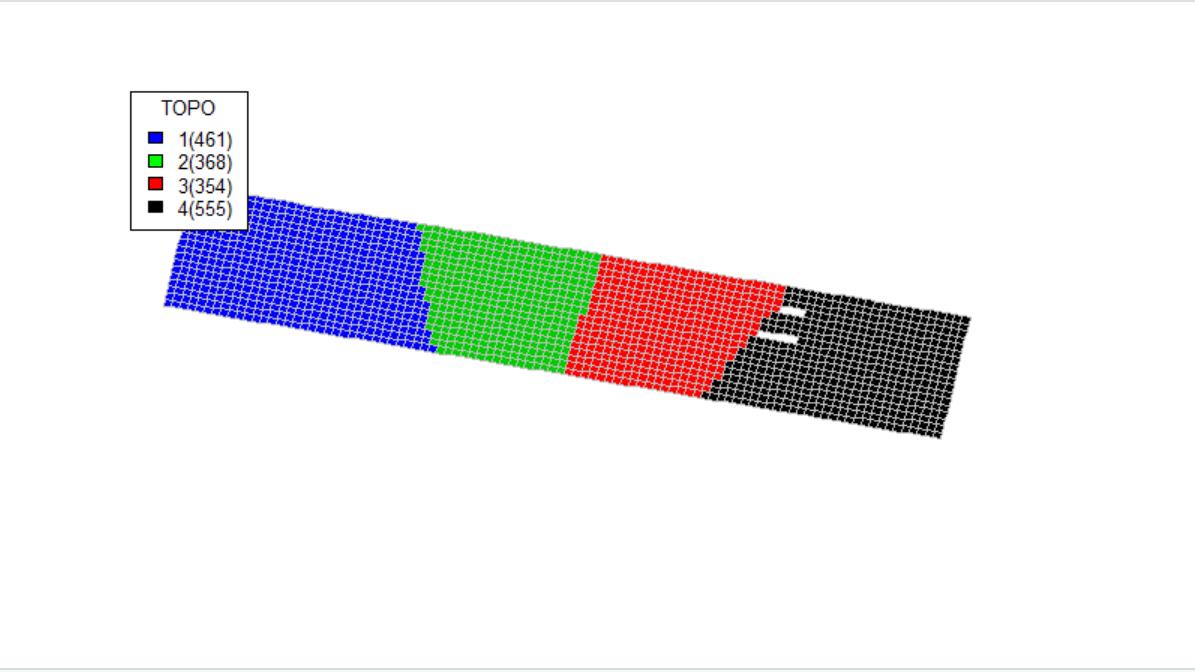
**INTRODUCTION**

La statistique spatiale connaît un développement important du fait de son utilisation dans de nombreux domaines: sciences de la terre, environnement et climatologie, épidémiologie, économétrie, analyse d'image, etc…

L’économétrie spatiale est un ensemble de méthodes et techniques permettant de prendre en compte explicitement la dimension spatiale des activités économiques dans les modèles économétriques. Les méthodes de l’économétrie spatiale visent à traiter les deux particularités des données spatiales: l’autucorrélation spatiale qui se refère à l’absence d’indépendance des observations géographiques et l’hétérogéneité spatiale qui est liée à la différentiation dans l’espace des variables et des comportements.

**PRESENTATION DES DONNEES et PROBLEMATIQUE de L’ETUDE**

Notre base conserne un champs de mais en Argentine sur quatres types de champs (Slope W, Hillope*,* Slope E et Low E). Elle contient les indicateurs sur la production de maïs de 1738 parcelles de culture de maïs. Le carte ci-dessous donne une representation des quatres types de champs.



L’objectif de notre étude est d’analyser l’impact de l’utilisation de nitrogène comme engrais sur les rendements du maïs (mesurés en nombre de quintaux par hectare). Pour ce faire nous utilisons le logiciel R et le logiciel Geoda

**ANALYSE EXPLORATOIRE DES DONNEES SPATIALES**

**1- MATRICE PONDERATION**

La matrice de pondération spatiale est le fruit de la transformation de la matrice de distance afin d’exprimer la proximité spatiale entre observations. La matrice de contiguïté à l’ordre 1 est une matrice carrée symétrique ayant autant de lignes que de colonnes entre les zones géographique Pour notre étude on s’intéresse à la matrice de pondération basée sur la contiguïté au sens de la reine à l’ordre 1. Le tableau ci-dessous donne les résultats de la matrice de pondération effectué avec le logiciel R.

|  |  |  |  |
| --- | --- | --- | --- |
| **LSTES DE VOISINS** | | | |
| Nombre de régions | Nombre de liens non nul | Pourcentage de poids non nuls | Nombre de liens moyen |
| 1738 | 13166 | 0.4358673 | 7.575374 |

Il y a 1738 parcelles dans les quatres champs. On note qu’il n’y a que 0.4358673 % de liaisons entre pacelles dans la matrice de contiguïté, ce qui signifie que celle -ci est vide à 99%. Par ailleurs on observe que chaque parcerelle a en moyenne 7.575374 pacelles voisines. Si tous les parcelles étaient reliés entre eux, le nombre total de liens serait 1738x1738 = 3020644. En réalité, on dénombre 13166 liens non nuls avec une pondération uniforme (1/3020644).

**2- AUTOCORRELATION SPATIALE GLOBALE**

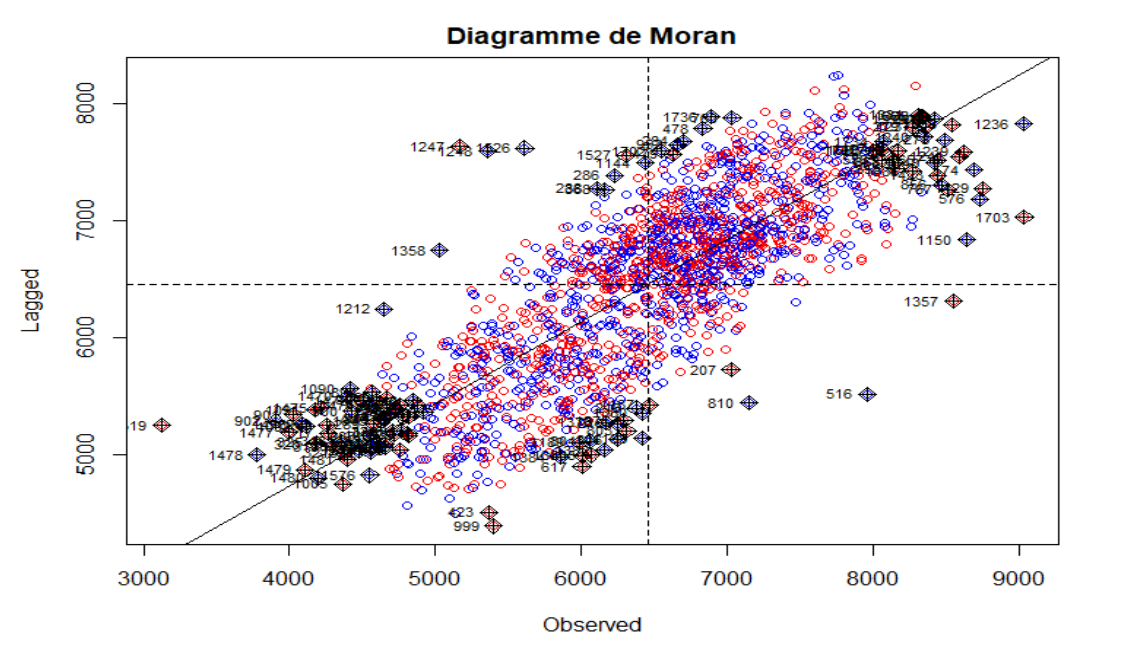
L’autocorrélation spatiale est définie par l’absence d’indépendance entre observations géographiques. En présence d’autocorrélation spatiale, la répartition spatiale des valeurs n’est pas  
aléatoire: il y a un phénomène de ressemblance des valeurs en fonction de la localisation géographique. On parle d’autocorrélation spatiale positivelorsque les observations proches ont plus tendance à se ressembler que les observations plus éloignées, et d’autocorrélation spatiale négativedans le cas inverse de dissemblance des observations proches

Le but de cette partie est de comprendre la dépendance spatiale globale sur les rendements du maïs dans les parcelles en se limitant à celui qui est le plus utilisé: l’indice de Moran. Pour faire ce test on doit dans un premier temps standardiser la matrice de contiguïté. Le tableau ci-dessous donne les résultats du test de I Moran globale.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Moran I statistic standard deviate** | **p-value** | **Moran I statistic** | **Expectation** | **Variance** |
| 56.503 | 2.2e-16 | 0.7008644728 | -0.0005757052 | 0.0001541122 |

Le coefficient d’autocorrélation spatiale de I Moran est égal à 0.7008644728. La p\_value de ce test est inférieur au seuil de significativité de 5%, donc l’hypothèse d’absence d’autocorrélation spatiale H₀ est rejetée. Il existe bien une autocorrélation spatiale globale sur les rendements du maïs en kilogramme par hectare dans les 1738 parcelles. De plus le coefficient de I Moran étant proche de 1, les données présentent une autocorrélation spatiale globale positive: les valeurs de YIELDS100 semblables ont tendance à se regrouper spatialement.

De méme les résultats de test de Geary confirme aussi l’existence d’autocorrélation spatiale positive globale. Car l’indice de Geary C=0.2986872738 est positif et la p\_value de ce test strictement inférieur à 5%. Le graphe ci-dessous est le diagramme de Moran.



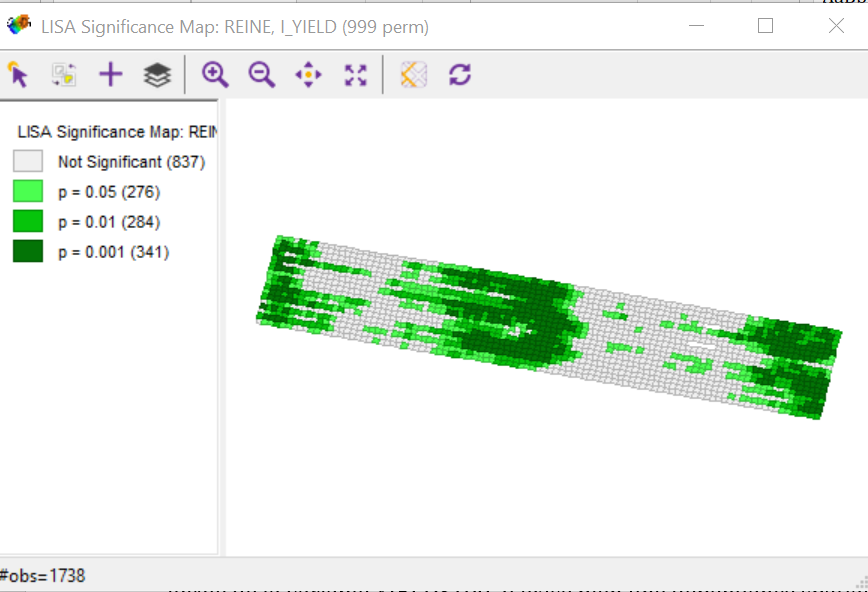
**3- INDICATEUR D’AUTOCORRELATION SPATIALE LOCALE (LISA)**

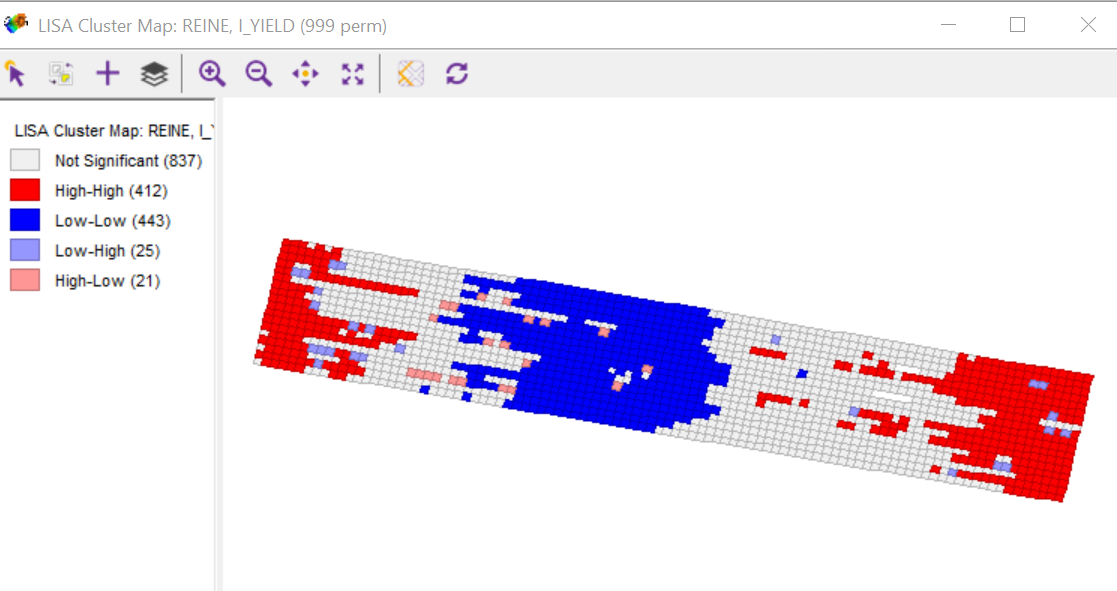
* **Analyse autocorrélation spatiale de la variable YIELD100 à l’aide du test de I Moran local**

Dans cette partie on cherche à analyses la présence d’autocorrélation spatiale locale dans la variable YIELDS100 à l’aide due test de I Moran locaux. On a effectué ce test avec le logiciel GeoDa.

La moyenne des indices de I Moran locaux effectué par R est égale à l’indice de I Moran globale I=0.7008645.

Les graphes ci-dessous représentent les résultats du test de I Moran locaux de la variable YIELD effectué sous GeoDa.



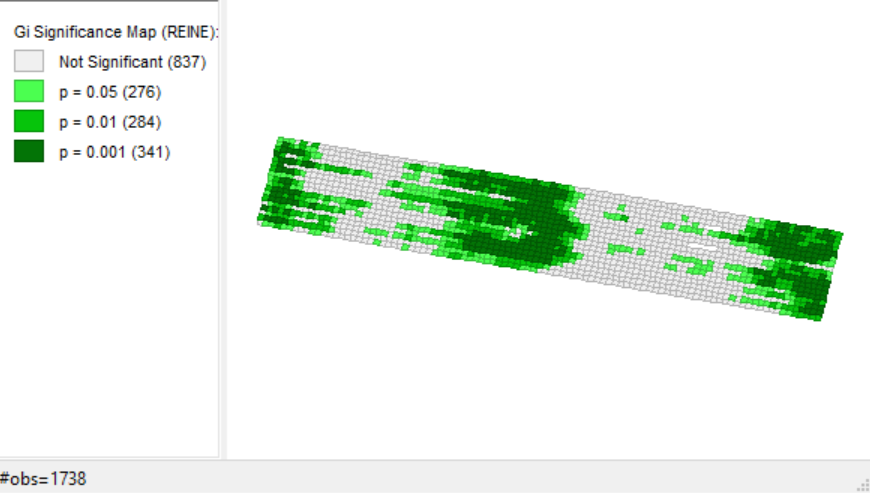


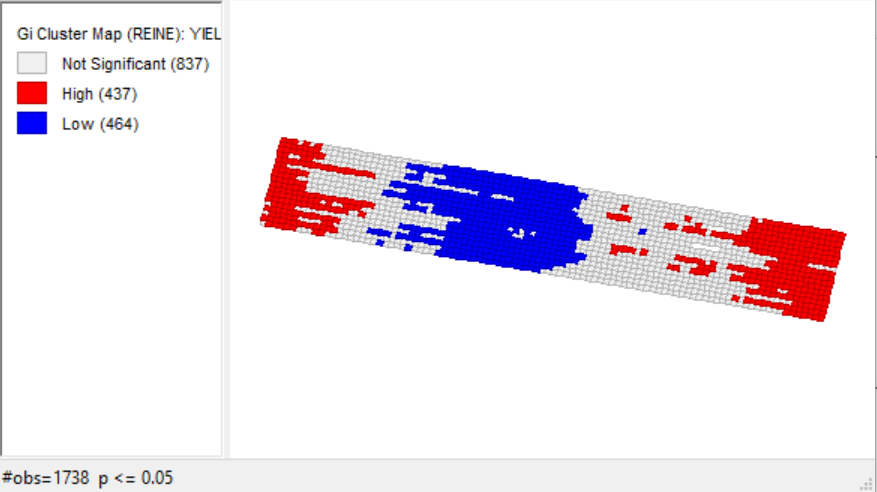
On remarque qu’il y’a 276 parcelles voisines dont leurs coefficients de Moran locaux ont une p\_value inférieur au seuil de 5%. Ceci indique qu’il existe 276 parcelles dont le rendement du mais se regroupent de façon significative.

De plus nous constatons d’après le graphe de I Moran tout les valeurs voisines semblent à ce regrouper dans les deux quadrants Low-Low **(443)** et High-High **(412)**. Donc il y’a **412** parcelles dont leurs valeurs similiaires tendent à ce regrouper faiblement et 443 parcelles dont leurs valeurs similiaires se regroupent de façon plus élevés.

* **Analyse d’autocorrélation spatiale locale de la variable YIELD à l’aide des statistiques de Getis locales (Local G).**

Les graphes ci-dessous résultats de la statisitque de Geatis local





Il existe 276 parcelles dont leurs reudements du maïs similiaires tendent à se regrouper de façon significative (5%). Nous remarquons aussi 437 parcelles ont une autocorrélation forte et 464 une autocorrélation faible.

Les résultats du test de I Moran local sont a peu prés sont à peu près identiques à celui du test de la statistique de Geatis. Il ont les méme nombre de parcelles similiaires et significatives.

###### **ECONOMETRIE SPATIALE**

### L’économétrie spatiale propose un ensemble de modèles statistiques destinés à modéliser les phénomènes en considérant la présence d’autocorrélation. Pour ce faire, une des démarches préconisées consiste à estimer un modèle par moindres carrés ordinaires (MCO) et à tester statistiquement la présence d’autocorrélation spatiale entre les résidus du modèle statistique afin de valider si les hypothèses d’indépendance du modèle de régression linéaire multiple sont respectées. Une seconde approche vise à estimer un modèle autorégressif complet intégrant toutes les formes de dépendances spatiales possibles afin d’en tester statistiquement la significativité.

### CONSTRUCTION DE LA SPÉCIFICATION ÉCONOMÉTRIQUE PRESENTEE EN (1)

Le modèle linéaire classique s’écrit en (1):

Voici ci-dessous les résultats du modèle MCO estimé

## Call:  
## lm(formula = YIELDS100 ~ N + N2 + TOP2 + TOP3 + TOP4, data = shp1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2212.93 -374.39 8.52 364.02 2827.22   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.597e+03 3.786e+01 174.254 < 2e-16 \*\*\*  
## N 1.144e+01 1.067e+00 10.722 < 2e-16 \*\*\*  
## N2 -3.510e-02 7.627e-03 -4.602 4.49e-06 \*\*\*  
## TOP2 -6.316e+02 3.922e+01 -16.105 < 2e-16 \*\*\*  
## TOP3 -1.769e+03 3.965e+01 -44.604 < 2e-16 \*\*\*  
## TOP4 -5.265e+02 3.535e+01 -14.893 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 560.9 on 1732 degrees of freedom  
## Multiple R-squared: 0.6004, Adjusted R-squared: 0.5993   
## F-statistic: 520.5 on 5 and 1732 DF, p-value: < 2.2e-16

La p\_value de la statistique LR est inférieure au seuil 10%, donc le modèle MCO est globalement significatif. On constate aussi que toutes les variables semblent avoir un effet sur le rendement du maïs car leurs coefficients estimés sont tous significatifs.

1. **METHODE DU SPHESIPHIQUE AU GENERAL**

#### Le tableau ci-dessous résume les résultats des différents tests éffectué sous R.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **SARMA** | **LM\_lag** | **LM\_err** | **RLM\_lag** | **RLM\_err** |
| **Test** | 2028 | 1410.5 | 2027.4 | 0.60218 | 617.5 |
| **df** | 2 | 1 | 1 | 1 | 1 |
| **P\_value** | < 2.2e-16 | <2.2e-16 | <2.2e-16 | 0.4377 | <2.2e-16 |

La probabilité critique du test joint est inférieur au seuil de significativité de 10%, on rejette l’hypothèse de nulleté des coefficients autorégressifs ρ et λ. Donc ce modèle est significatif. D’où la nécessicité d’effectuer les tests LM\_lag et LM\_err. On peut remarquer que la statistique de test pour une alternative SEM est supérieure à celle correspondant à une alternative SAR. Pour conclure de façon plus crédible, on regarde les tests robustes à la présence de la spécification alternative de l’autocorrélation spatiale. Il s’agit pour le RLMlag de tester l’absence de terme autorégressif spatial. Lorsque le modèle contient déjà un terme autorégressif spatial dans les erreurs (RLMlag), ou inversement pour RLMerr de tester l’absence de terme autorégressif spatial dans les erreurs lorsque le modèle contient un terme autorégressif spatial. La version robuste RLMerr est fortement significative alors que RLMlag ne l’est pas. Nous estimons donc un modèle à effets fixes avec un processus autorégressif spatial dans les erreurs. C’est à dire un modèle SEM.

On estime donc le modèle SEM.

## Call:errorsarlm(formula = MLS, data = shp1, listw = mbcs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2083.120 -225.327 -10.447 208.213 2718.356   
##   
## Type: error   
## Coefficients: (asymptotic standard errors)   
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 6.3341e+03 9.2915e+01 68.1711 < 2.2e-16  
## N 1.0864e+01 6.4931e-01 16.7317 < 2.2e-16  
## N2 -2.2646e-02 4.6429e-03 -4.8776 1.074e-06  
## TOP2 -3.8945e+02 1.0539e+02 -3.6954 0.0002195  
## TOP3 -9.8135e+02 1.2164e+02 -8.0674 6.661e-16  
## TOP4 -4.7784e+02 1.2367e+02 -3.8639 0.0001116  
##   
## Lambda: 0.83544, LR test value: 1133.8, p-value: < 2.22e-16  
## Asymptotic standard error: 0.016774  
## z-value: 49.806, p-value: < 2.22e-16  
## Wald statistic: 2480.6, p-value: < 2.22e-16  
##   
## Log likelihood: -12897.09 for error model  
## ML residual variance (sigma squared): 141060, (sigma: 375.59)  
## Number of observations: 1738   
## Number of parameters estimated: 8   
## AIC: 25810, (AIC for lm: 26942)

Le coefficient autorégressif dans les erreurs est égal à λ=0.83544 et il est significativement different de 0.

**Fonction de vraisemblance du modèle SEM**

OU

W=mbcs est la matrice de contiguïté standardisée

1. **INTERPRÉTATION L’IMPACT D’UNE VARIATION DE LA QUANTITÉ D’ENGRAIS NUTROGÈNE, ÉVALUÉE À LA VALEUR MOYENNE DE L’ÉCHANTILLON, SUR LE RENDEMENT DU MAÏS,**

**Impact**=

La formule de l’impact ci-dessus nous permet de faire l’analyse suivant: lorsque la quantité d’engrais augmente d’une unité le rendement moyen du maïs va augmenter de 10.81871. On peut conclure que le rendement de mais augmente la quantité d’engrais

1. **ESTIMATION DU MODÈLE (1) EN INTÉGRANT LE DÉCALAGE SPATIAL DE LA VARIABLE DÉPENDANTE.**

## Call:lagsarlm(formula = YIELDS100 ~ N + N2 + TOP2 + TOP3 + TOP4, data = shp1,   
## listw = mbcs, type = "lag", method = "eigen")  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2208.8874 -251.2361 -8.3889 252.5305 2719.2941   
##   
## Type: lag   
## Coefficients: (asymptotic standard errors)   
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 1.7102e+03 1.4042e+02 12.1790 < 2.2e-16  
## N 1.1226e+01 7.9234e-01 14.1686 < 2.2e-16  
## N2 -3.1553e-02 5.6635e-03 -5.5712 2.530e-08  
## TOP2 -1.8694e+02 3.1646e+01 -5.9070 3.483e-09  
## TOP3 -5.5014e+02 4.4683e+01 -12.3120 < 2.2e-16  
## TOP4 -1.6321e+02 2.8343e+01 -5.7584 8.490e-09  
##   
## Rho: 0.68496, LR test value: 884.38, p-value: < 2.22e-16  
## Asymptotic standard error: 0.019902  
## z-value: 34.417, p-value: < 2.22e-16  
## Wald statistic: 1184.5, p-value: < 2.22e-16  
##   
## Log likelihood: -13021.78 for lag model  
## ML residual variance (sigma squared): 173130, (sigma: 416.09)  
## Number of observations: 1738   
## Number of parameters estimated: 8   
## AIC: 26060, (AIC for lm: 26942)  
## LM test for residual autocorrelation  
## test value: 14.599, p-value: 0.00013297

* **Valeur, significativité du coefficient autorégressif et autocorrélation des résidus**

La coefficient de la variable autorégressif est égale à . La p\_value du test LR est égale à celle du test basé de sur la matrice de variance asymptotique **(=2.22.e-16**) et cette probabilité critique est strictement inférieure au seuil de significativité de 10%, donc on rejette et on conclut que ce coefficient autorégressif est significativement différent de 0. Ceci indique que le rendement d maïs d’une parcelle depend fortement de celui de ces voisins. D’après le test de Multiplicateur de Lagrange, on rejette l’hypothèse de nulleté de l’autocorrélation spatiale des résidus car la p\_value de ce test est égale à 0.00013297 et cette dernière est inférieure à 0.1. Donc il reste toujours de l’autocorrélation spatiale résiduelle dans les erreurs.

Le tableau ci-dessous donne les résultats du test de I Moran globale des résidus

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Moran I statistic standard deviate** | **p-value** | **Moran I statistic** | **Expectation** | **Variance** |
| 2.61 | 0.0045 | 0.0317967436 | -0.0005757052 | 0.0001541122 |

D’après les résultats du test de I Moran des ci-dessus, on en déduis aussi qu’il reste toujours de l’autocorrélation spatiale des résidus

## Comparaisons les résultats avec ceux du MCO

L’estimateur de l’effet d‟interaction ρ̂ est significativement positive, cela implique que les  
estimateurs du modèle des moindres carrés sont biaisés et inefficients et par conséquent  
l‟effet du rendement de mais estimé ̂ est biaisé. De plus l’AIC du modèle SAR est égale à 26060 est inférieur à l’AIC du modèle MCO 26942. D’apràs le critère de AIC le modèle SAR est préférable à celui du modèle MCO.

### 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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parcelle** | 87 | 181 | 1534 | 1628 |
| **Valeur maximale effet direct** | 13,27355 | 13,27355 | 13,27355 | 13,27355 |

L’effet direct est maximal est égal 13.27355 et elle est dans les parcelles 87, 181, 1534 et 1628 et celui-ci est à. La moyenne des effets directs est égale à 12.43853 et l’effet maximale est égale à 13.27355.

### Au regard des résultats de l’estimation du modèle avec variable endogène spatialement

Le log-likehood pour SEM vaut **-12897.09** est plus élevé que pour le modèle SAR **-13021.78**. Ce qui témoingne une bonne amélioration du modèle SEM sur le modèle SAR. L’AIC vaut **25810** pour SEM, **26060** pour SAR (et **26942** pour LM), ce qui montre que SEM est meilleur que SAR.

On conclut l’estimation du modèle autorégressif avec erreurs est plus adéquat pour expliquer l’effet du rendement du maïs sur ces variables explicatives.