Projet VITAMINES

Nicolas Debarsy[1], Laurent Delsol[2], Cécile Louchet[2]

[1]LEO, Université d'Orléans [2]MAPMO, Université d'Orléans

19èmes Journée CaSciModOT
26 Novembre 2013

Introduction

- VITAMINES : Validation et études des Interactions, Techniques statistiques et Analyse de Modèles appliqués aux Images Naturelles et à l'Econométrie Spatiale

Introduction

- VITAMINES : Validation et études des Interactions, Techniques statistiques et Analyse de Modèles appliqués aux Images Naturelles et à l'Econométrie Spatiale
- Présentation aux 17èmes journée CaSciModOT de *Statistiques fonctionnelles pour l'imagerie hyperspectrale*

Introduction

- VITAMINES : Validation et études des Interactions, Techniques statistiques et Analyse de Modèles appliqués aux Images Naturelles et à l'Econométrie Spatiale
- Présentation aux 17èmes journée CaSciModOT de *Statistiques fonctionnelles pour l'imagerie hyperspectrale*
- PEPS HuMaIn

Introduction

- VITAMINES : Validation et études des Interactions, Techniques statistiques et Analyse de Modèles appliqués aux Images Naturelles et à l'Econométrie Spatiale
- Présentation aux 17èmes journée CaSciModOT de *Statistiques fonctionnelles pour l'imagerie hyperspectrale*
- PEPS HuMaIn
- Lien entre traitement d'images et économétrie spatiale : Les données traitées ne sont pas considérées comme indépendantes les unes par rapport aux autres

Introduction

- VITAMINES : Validation et études des Interactions, Techniques statistiques et Analyse de Modèles appliqués aux Images Naturelles et à l'Econométrie Spatiale
- Présentation aux 17èmes journée CaSciModOT de *Statistiques fonctionnelles pour l'imagerie hyperspectrale*
- PEPS HuMaIn
- Lien entre traitement d'images et économétrie spatiale : Les données traitées ne sont pas considérées comme indépendantes les unes par rapport aux autres
  $\Rightarrow$ Nécessité de modéliser ces interdépendances

Interdépendances en traitement d'image

- Débruitage d'images
  - Régularité de l'image provient de la répétition de certains motifs dans les niveaux de gris des images
    $\Rightarrow$ Nécessité de pouvoir détecter des pixels correspondant à des motifs semblables.
    $\Rightarrow$ Définition d'une mesure d'interaction entre pixels, fonction de la proximité géographique et similarité photométrique (Buades et al. 2005).
  - L'estimation de l'image débruitée par MV dépend de la qualité de ces mesures d'interaction.

Interdépendances en économie

- Croissance économique
  - Etude de la croissance économique (au niveau international ou au niveau régional).
    Plusieurs articles (Ertur et Koch, 2007, 2011 ; Pfaffermayr 2009) ont montré l'importance de la prise en compte de l'autocorrélation spatiale entre pays ou régions pour expliquer leur croissance (diffusion technologique, commerce international,...)

Interdépendances en économie

- Croissance économique
  - Etude de la croissance économique (au niveau international ou au niveau régional).
    Plusieurs articles (Ertur et Koch, 2007, 2011 ; Pfaffermayr 2009) ont montré l'importance de la prise en compte de l'autocorrélation spatiale entre pays ou régions pour expliquer leur croissance (diffusion technologique, commerce international,...)
  - Nécessité de prendre en compte ces interactions dans les modèles économiques et économétriques.
    $\Rightarrow$ Recours aux outils de l'économétrie spatiale

Interdépendances en économie

- Croissance économique
  - Etude de la croissance économique (au niveau international ou au niveau régional).
    Plusieurs articles (Ertur et Koch, 2007, 2011 ; Pfaffermayr 2009) ont montré l'importance de la prise en compte de l'autocorrélation spatiale entre pays ou régions pour expliquer leur croissance (diffusion technologique, commerce international,...)
  - Nécessité de prendre en compte ces interactions dans les modèles économiques et économétriques.
    $\Rightarrow$ Recours aux outils de l'économétrie spatiale
  - Modèle économétrique : Confrontation du modèle économique aux données.

$$C = \alpha + \beta Y + \epsilon$$

Econométrie spatiale

Outils destinés à prendre explicitement en compte la présence de l'espace dans le modèle

1. Autocorrélation spatiale : Comportement d'un individu est influencé par le comportement d'autres individus.

$$g_t = \lambda W g_t + X\beta + \epsilon$$

- W est la matrice d'interactions résumant le schéma d'interaction entre observations
- X est une matrice de variables explicatives
- λ coefficient capturant l'intensité des interactions

Econométrie spatiale

Outils destinés à prendre explicitement en compte la présence de l'espace dans le modèle

1. Autocorrélation spatiale : Comportement d'un individu est influencé par le comportement d'autres individus.

$$g_t = \lambda W g_t + X\beta + \epsilon$$

   - W est la matrice d'interactions résumant le schéma d'interaction entre observations
   - X est une matrice de variables explicatives
   - λ coefficient capturant l'intensité des interactions

2. Hétérogénéité spatiale : Comportement différencié en fonction de la localisation spatiale. Ex de la croissance éco pour les régions européennes : Nord-Sud, Est-Ouest.

Projet VITAMINES

- Projet interdisciplinaire car utilisation des outils appliqués en traitement d'images, des statistiques (test de mélange, stat fonctionnelle, stat bayésiennes) et de l'économétrie spatiale afin de répondre à la question de la prise en compte des interactions.
- Différents axes de recherches
  1. Détection de la présence de concentration de valeurs similaires de variables en utilisant des distances entre observations basées sur différents espaces (géographique, institutionnel, économique, sociologique, ...)
  2. Déterminer la meilleure manière de prendre en compte les interactions entre observations pour une problématique donnée

Projet VITAMINES

1. Amélioration des outils de l'économétrie spatiale :
   - Déterminer la matrice d'interaction optimale (par rapport à un critère statistique : Validation croisée ou statistiques bayésiennes)
   - Capter la présence d'hétérogénéité spatiale (détection de clusterisation spatiale)
2. Amélioration des méthodes de traitement d'image bas-niveau (segmentation, débruitage) grâce à une prise en compte plus précise des interactions entre (groupes de) pixels

Projet VITAMINES

1. Amélioration des outils de l'économétrie spatiale :
   - Déterminer la matrice d'interaction optimale (par rapport à un critère statistique : Validation croisée ou statistiques bayésiennes)
   - Capter la présence d'hétérogénéité spatiale (détection de clusterisation spatiale)
2. Amélioration des méthodes de traitement d'image bas-niveau (segmentation, débruitage) grâce à une prise en compte plus précise des interactions entre (groupes de) pixels

Projet a débuté sur l'étude de l'hétérogénéité spatiale en économie.

Méthodologie

Cadre

- I : Ensemble des observations ($\#I < \infty$)
- $\forall i, j \in I$, $d(i, j) = d(j, i)$, mesure de distance. Ex : distance géographique
- y_i : variable associée à chaque localisation (scalaire ou vecteur) : Croissance éco, criminalité, réussite scolaire

Méthodologie

Cadre

- I : Ensemble des observations ($\#I < \infty$)
- $\forall i, j \in I$, $d(i,j) = d(j,i)$, mesure de distance. Ex : distance géographique
- y_i : variable associée à chaque localisation (scalaire ou vecteur) : Croissance éco, criminalité, réussite scolaire

Objectif : Classer les observations en k clubs (ou clusters)
Une segmentation est définie comme

$$x : I \mapsto \{1, \ldots, k\}.$$

Ainsi $x(i)$ est le club de la région i.

Classification basée sur la variance

Objectif : Minimisation de la variance intra-cluster.

1. Clusterisation sur la valeur de variable.

$$E_{\text{éco}}(x) = \sum_{i \in I} \sum_{j \in I} (y_i - y_j)^2 \mathbb{1}_{x(i)=x(j)}.$$

où $E_{\text{éco}}$ est l'énergie de la variance intra-groupe.

Classification basée sur la variance

Objectif : Minimisation de la variance intra-cluster.

1. Clusterisation sur la valeur de variable.

$$E_{\text{éco}}(x) = \sum_{i \in I} \sum_{j \in I} (y_i - y_j)^2 \mathbb{1}_{x(i)=x(j)}.$$

où $E_{\text{éco}}$ est l'énergie de la variance intra-groupe.
Interprétation pour $k = 2$,

- tracer l'histogramme des y_i
- chercher une abscisse Y qui coupe en 2 l'histogramme
- on prend Y qui minimise la variance intra (qui donc maximimise la variance inter). $\Rightarrow$ Méthode d'Otsu.

Classification basée sur la variance

Objectif : Minimisation de la variance intra-cluster.

1. Clusterisation sur la valeur de variable.

$$E_{\text{éco}}(x) = \sum_{i \in I} \sum_{j \in I} (y_i - y_j)^2 \mathbb{1}_{x(i)=x(j)}.$$

où $E_{\text{éco}}$ est l'énergie de la variance intra-groupe.
Interprétation pour $k = 2$,

- tracer l'histogramme des y_i
- chercher une abscisse Y qui coupe en 2 l'histogramme
- on prend Y qui minimise la variance intra (qui donc maximimise la variance inter). $\Rightarrow$ Méthode d'Otsu.

2. Au lieu de considérer les valeurs y_i et y_j pour classifier, utilisation de la distance les séparant.
   La distance peut-être géographique, institutionnelle, sociologique, linguistique.

$$E_{\text{géo}}(x) = \sum_{i \in I} \sum_{j \in I} d(i,j)^2 \mathbb{1}_{x(i)=x(j)}.$$

MAP Gaussien

- Clusterisation sur la variable d'intérêt avec une régularisation par la distance
  $\Rightarrow$ On minimise :

$$\begin{aligned} E_{\text{mixte}}(x) &= E_{\text{éco}}(x) + \lambda E_{\text{géo}}(x) \\ &= \sum_{i \in I} \sum_{j \in I} (y_i - y_j)^2 \mathbb{1}_{x(i)=x(j)} + \lambda \sum_{i \in I} \sum_{j \in I} d(i,j)^2 \mathbb{1}_{x(i)=x(j)}, \end{aligned}$$

où β est un paramètre de régularisation.

MAP Gaussien

- Clusterisation sur la variable d'intérêt avec une régularisation par la distance
  $\Rightarrow$ On minimise :

$$E_{\text{mixte}}(x) = E_{\text{éco}}(x) + \lambda E_{\text{géo}}(x)$$
$$= \sum_{i \in I} \sum_{j \in I} (y_i - y_j)^2 \mathbb{1}_{x(i)=x(j)} + \lambda \sum_{i \in I} \sum_{j \in I} d(i,j)^2 \mathbb{1}_{x(i)=x(j)},$$

où β est un paramètre de régularisation.

Si $\lambda \approx 0$, on minimise $E_{\text{éco}}(x)$, si $\lambda \to \infty$, on minimise $E_{\text{géo}}(x)$

Application

- Application à la $\beta-$convergence économique
  - Théorie économique basée sur les rendements marginaux décroissants.
    $\Rightarrow$ Plus le niveau de richesse initial est élevé, moins le taux de croissance éco est grand.

$$g_t = \alpha + \beta PIB_{t0} + \epsilon$$

Application

- Application à la $\beta-$convergence économique
  - Théorie économique basée sur les rendements marginaux décroissants.
    $\Rightarrow$ Plus le niveau de richesse initial est élevé, moins le taux de croissance éco est grand.

$$g_t = \alpha + \beta PIB_{t0} + \epsilon$$

  - $\beta-$convergence absolue car on suppose que toutes les éco convergent vers le même équilibre de long terme.
  - Théorie indique $\beta < 0$

Application

- Application à la $\beta-$convergence économique
  - Théorie économique basée sur les rendements marginaux décroissants.
    $\Rightarrow$ Plus le niveau de richesse initial est élevé, moins le taux de croissance éco est grand.

$$g_t = \alpha + \beta PIB_{t0} + \epsilon$$

  - $\beta-$convergence absolue car on suppose que toutes les éco convergent vers le même équilibre de long terme.
  - Théorie indique $\beta < 0$
  - Généralisable à la $\beta-$convergence conditionnelle si addition de variables explicatives supplémentaires.

Application

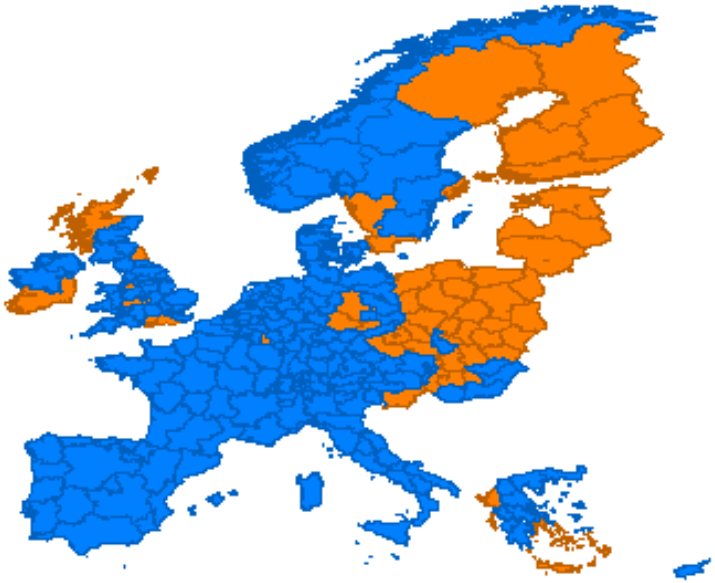
- Application à la $\beta-$convergence économique
  - Théorie économique basée sur les rendements marginaux décroissants.
    $\Rightarrow$ Plus le niveau de richesse initial est élevé, moins le taux de croissance éco est grand.

$$g_t = \alpha + \beta PIB_{t0} + \epsilon$$

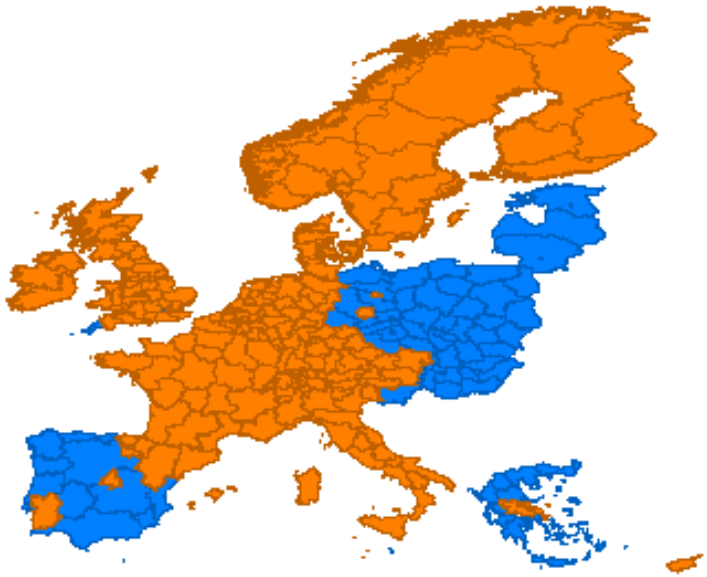
  - $\beta-$convergence absolue car on suppose que toutes les éco convergent vers le même équilibre de long terme.
  - Théorie indique $\beta < 0$
  - Généralisable à la $\beta-$convergence conditionnelle si addition de variables explicatives supplémentaires.
  - Etude de la croissance des régions européennes de 1995 à 2009. N = 262 (N=$\#I$).

Illustration Croissance et richesse

k-means sans régularisation spatiale ($\beta = 0$)



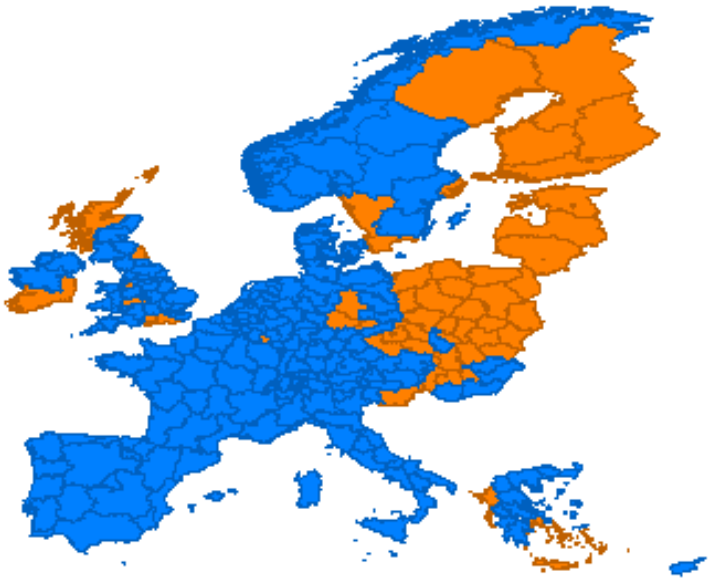
Taux de croissance annuel moyen 95-09



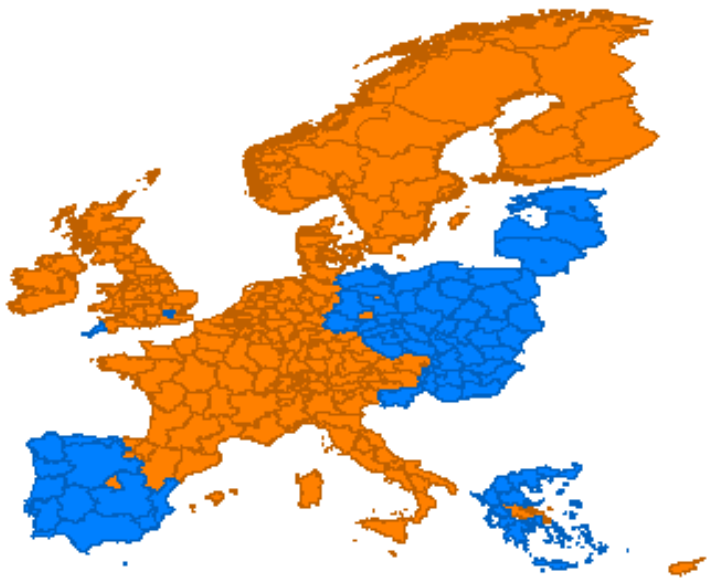
Richesse par tête initiale en 95

Illustration Croissance et richesse

MAP Gaussien ($\lambda = 10^{-10}$)



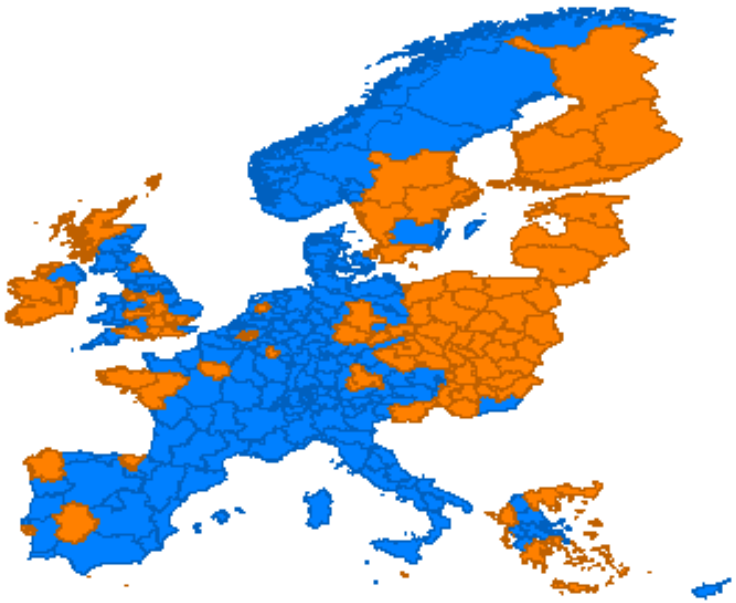
Taux de croissance annuel moyen 95-09



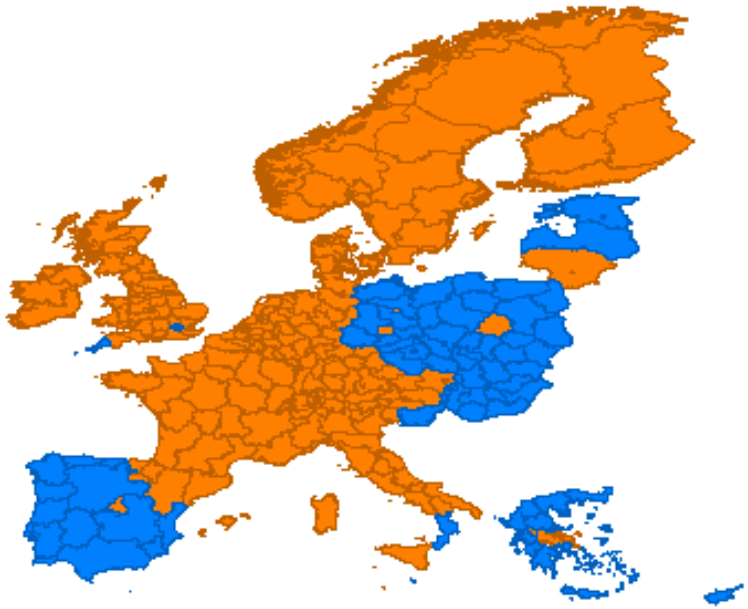
Richesse par tête initiale en 95

Illustration Croissance et richesse

MAP Gaussien ($\lambda = 10^{-7}$)



Taux de croissance annuel moyen 95-09



Richesse par tête initiale en 95

Estimation du modèle de $\beta-$ convergence

- Estimation du modèle sans autocorrélation spatiale :

$$g_t = \alpha + ln(PIB\_pc)_{t0}\beta + \epsilon$$

g_t le taux de croissance annuel moyen sur la période du revenu par tête
PIB_pc le revenu par tête à la période initiale

Estimation du modèle de $\beta-$ convergence

- Estimation du modèle sans autocorrélation spatiale :

$$g_t = \alpha + ln(PIB\_pc)_{t0}\beta + \epsilon$$

g_t le taux de croissance annuel moyen sur la période du revenu par tête
PIB_pc le revenu par tête à la période initiale

- Estimation du modèle avec autocorrélation spatiale (SAR) :

$$g_t = \alpha + ln(PIB\_pc)_{t0}\beta + \rho W g_t + \epsilon$$

avec *W* la matrice d'interactions.
Matrice considérée : 15 plus proches voisins.

Résultats de $\beta-$convergence

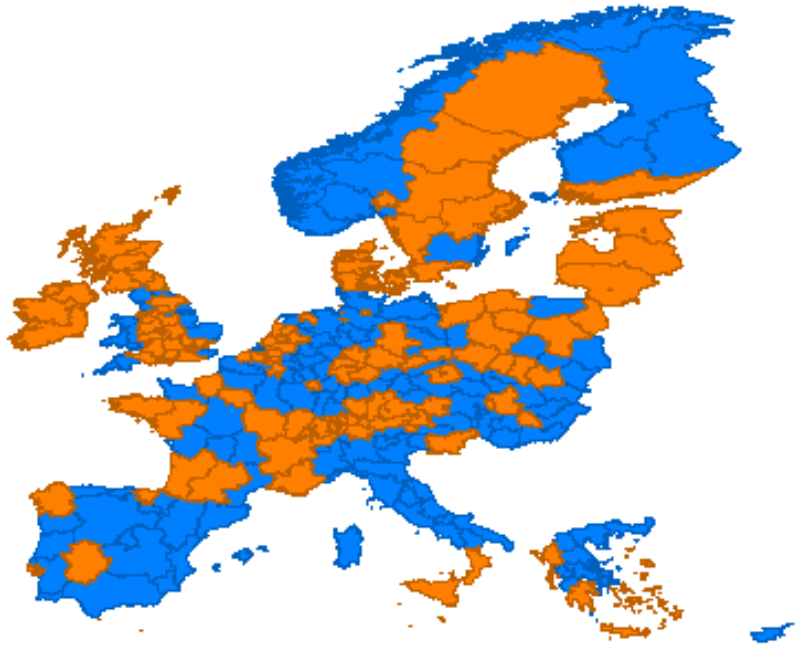
	Variable dépendante	g_t
Var. Exp.	OLS	SAR
α	0.055	0.026
	(18.751)	(5.977)
β	-0.012	-0.006
	(-15.113)	(-5.966)
ρ	-	0.632
		(8.649)
N=262		
R^2=0.49		

t-stats entre parenthèses

$\beta < 0 \Rightarrow$ convergence (lente) des économies.
Autocorrélation spatiale positive et significative

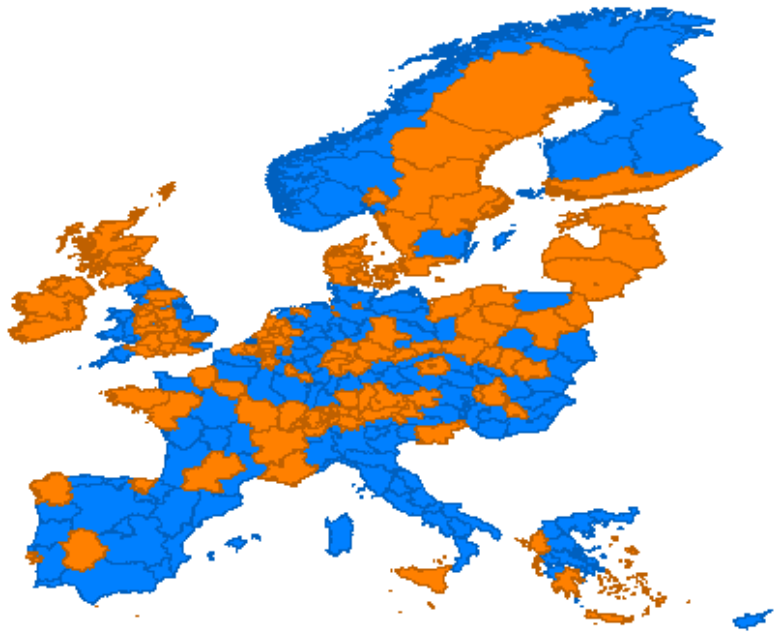
Hétérogénéité ou autocorrélation spatiale ?

FIGURE : *k*-means sans régularisation



Hétérogénéité ou autocorrélation spatiale ?

FIGURE : MAP Gaussien ($\lambda = 10^{-8}$)



Hétérogénéité ou autocorrélation spatiale ?

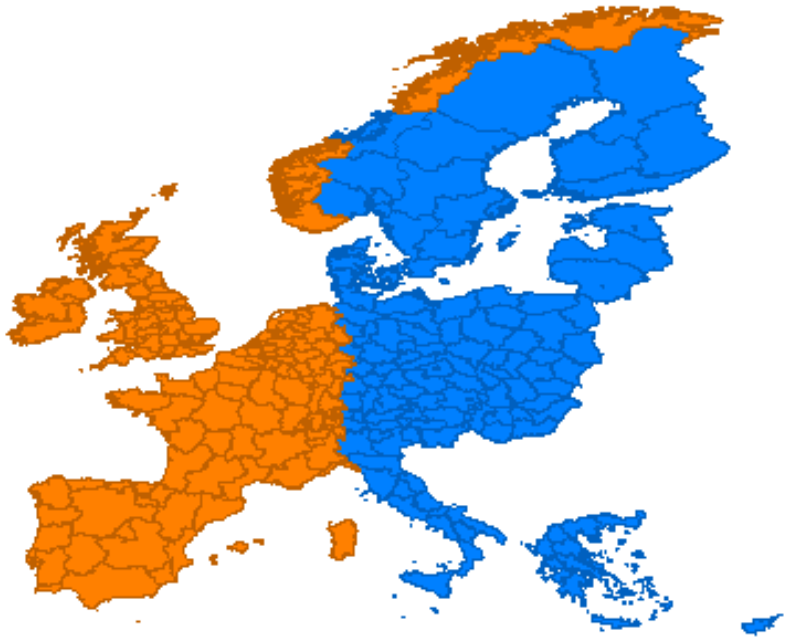
- Il semble que les deux soient conjointement présents.
- Des clusters de résidus similaires semblent émerger.

Hétérogénéité ou autocorrélation spatiale ?

- Il semble que les deux soient conjointement présents.
- Des clusters de résidus similaires semblent émerger.
- Cependant ...

Hétérogénéité ou autocorrélation spatiale ?

FIGURE : MAP Gaussien ($\lambda = 10^{-4}$)



Recherches futures

1. Concernant l'hétérogénéité spatiale
   - Développement de statistiques pour tester la significativité des clusters (test de mélange, tests à contrario)
   - Généralisation à d'autres distances que la distance géographique (institutionnelle, linguistique, ...)

Recherches futures

1. Concernant l'hétérogénéité spatiale
   - Développement de statistiques pour tester la significativité des clusters (test de mélange, tests à contrario)
   - Généralisation à d'autres distances que la distance géographique (institutionnelle, linguistique, ...)
2. Concernant les interactions
   - Développement d'outils statistiques permettant de choisir la matrice d'interactions la plus pertinente pour une problématique donnée (méthodes de validation croisée, stat. bayésiennes)
   - En traitement d'image, définition d'un graphe sous-jacent suffisamment efficace pour améliorer les méthodes de segmentation, filtrage, décomposition ..