

*Inférence bayésienne sur un modèle de mélange
à interaction spatiale*

Lionel Cucala et Jean-Michel Marin
Université Montpellier 2

Jeudi 10 Mai 2012

Plan de l'exposé

1-Motivations

2-Modélisation et inférence

3-Sélection du nombre de composantes

4-Conclusion

Plan de l'exposé

1-Motivations

2-Modélisation et inférence

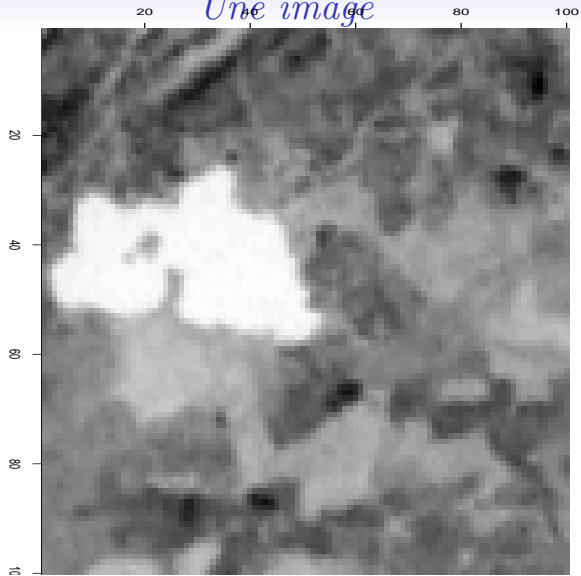
3-Sélection du nombre de composantes

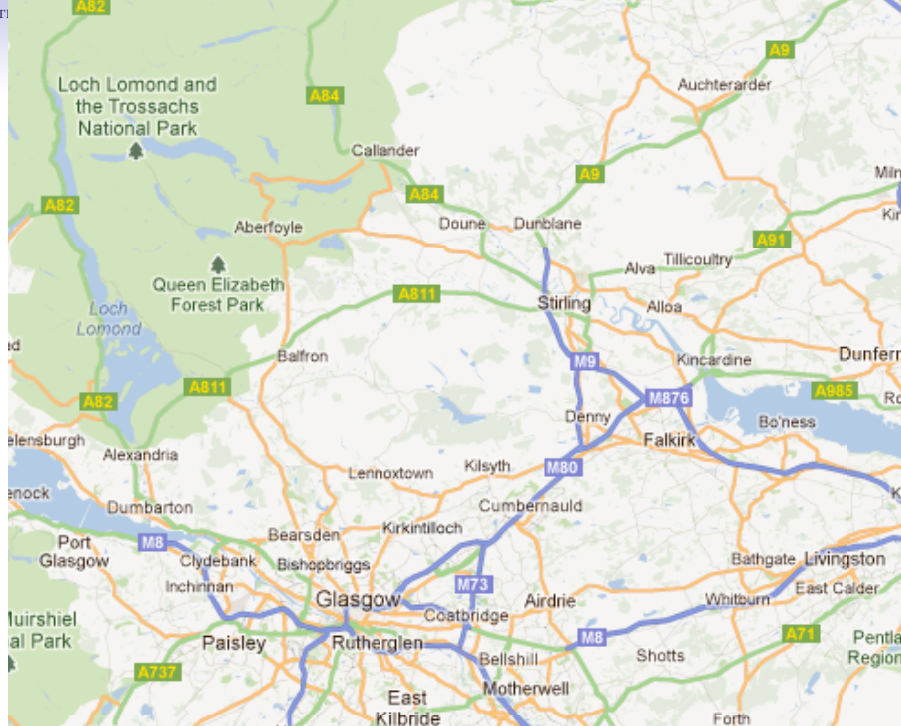
4-Conclusion

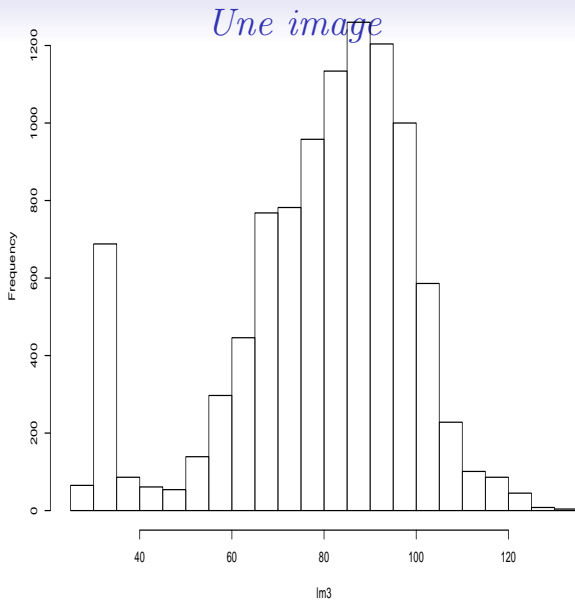
Données

- Images satellite d'une région lacustre d'Ecosse : Menteith lake.
- 1 image : 100×100 pixels codés entre 1 et 256.
- Objectif : identifier les zones aquatiques, marécageuses, terrestres...

Une image







Plan de l'exposé

1-Motivations

2-Modélisation et inférence

3-Sélection du nombre de composantes

4-Conclusion

Plan de l'exposé

1-Motivations

2-Modélisation et inférence

3-Sélection du nombre de composantes

4-Conclusion

Objectif

- ➡ Segmentation ou Clustering : séparer les observations en composantes homogènes.

- ➡ Nombre de composantes : k .

Premier modèle : le mélange gaussien

- Niveaux de gris observés : $\mathbf{y} = (y_1, \dots, y_N)$, avec $y_i \in \{1, 256\}$.
- Composantes inconnues : $\mathbf{z} = (z_1, \dots, z_N)$, avec $z_i \in \{1, k\}$.
- z_1, \dots, z_N i.i.d. $\sim \mathcal{M}(p_1, \dots, p_k)$
avec $p_1 + \dots + p_k = 1$.
- $(y_i | z_i = j) \sim \mathcal{N}(\mu_j, \sigma^2)$.

Premier modèle : le mélange gaussien

- $\mathbf{z} \in \{1, k\}^N$ donc k^N segmentations possibles.
- Approche bayésienne : on cherche

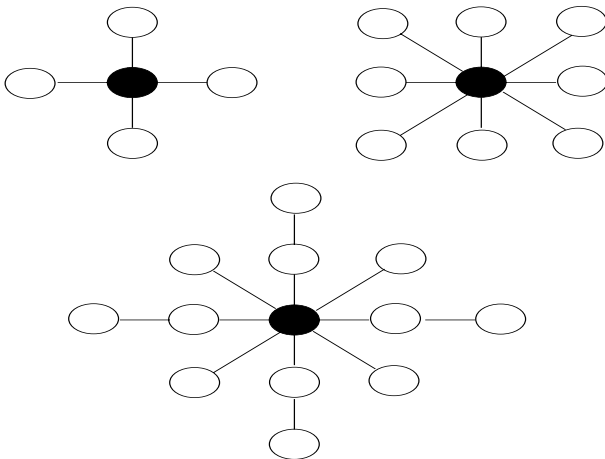
$$\begin{aligned}\Pi(\mathbf{z}|\mathbf{y}) &= \frac{\Pi(\mathbf{z}, \mathbf{y})}{\Pi(\mathbf{y})} = \frac{\Pi(\mathbf{z})\Pi(\mathbf{y}|\mathbf{z})}{\sum_{\mathbf{z}} \Pi(\mathbf{z})\Pi(\mathbf{y}|\mathbf{z})} \\ &= \frac{\int \int \int \Pi(\mathbf{z}|\mathbf{p})\Pi(\mathbf{y}|\mathbf{z}, \mu, \sigma^2)\Pi(\mathbf{p}, \mu, \sigma^2)d\mathbf{p} d\mu d\sigma^2}{\sum_{\mathbf{z}} \int \int \int \Pi(\mathbf{z}|\mathbf{p})\Pi(\mathbf{y}|\mathbf{z}, \mu, \sigma^2)\Pi(\mathbf{p}, \mu, \sigma^2)d\mathbf{p} d\mu d\sigma^2}\end{aligned}$$

- Remarque : info spatiale ignorée !

Modèle à dépendance spatiale

- Composantes spatialement dépendantes.
- $(z_1, \dots, z_N) \sim \text{Potts}(\beta)$.
- Modèle de champ markovien : $z_i | z_{-i}$ ne dépend que du voisinage.

Le système de voisinage



Le modèle de Potts

- $\sum_{j \sim i}$: somme sur les pixels voisins de i
- Densité :

$$f(\mathbf{z}|\beta) = \frac{\exp \left\{ \beta \sum_{i=1}^N \sum_{j \sim i} \delta_{z_j}(z_i) \right\}}{Z(\beta)} = \frac{\exp \left\{ \beta S(\mathbf{z}) \right\}}{Z(\beta)},$$

où $Z(\beta) = \sum_{\mathbf{z} \in \{1, \dots, k\}^N} \exp \left\{ \beta S(\mathbf{z}) \right\}$.

- Densité conditionnelle :

$$f(z_i | \mathbf{z}_{-i}) \propto \exp \left\{ \beta \left(\sum_{j \sim i} \delta_{z_j}(z_i) \right) \right\}.$$

Simulation : échantillonneur de Gibbs

➤ Méthode MCMC.

➤ Vecteur d'origine : $(z_1^{(0)}, \dots, z_N^{(0)})$.

➤

$$1. \quad z_1^{(t+1)} \sim f(z_1 | z_2^{(t)}, \dots, z_N^{(t)});$$

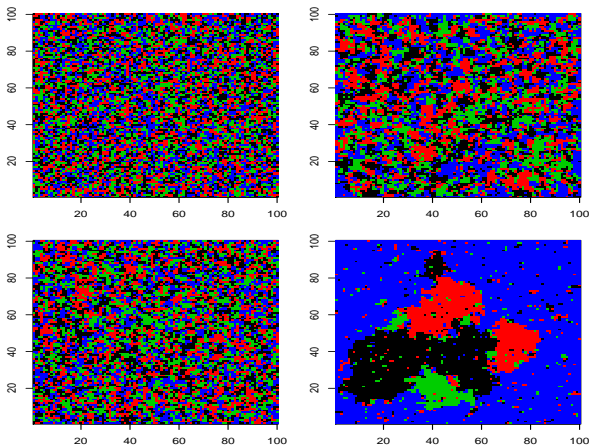
\vdots

\vdots

$$N. \quad z_N^{(t+1)} \sim f(z_N | z_1^{(t+1)}, \dots, z_{N-1}^{(t+1)});$$

➤ On considère que $(z_1^{(t)}, \dots, z_N^{(t)})$, $t > T$, suit à peu près la loi de probabilité du modèle.

Simulation : échantillonneur de Gibbs



Résolution

- On cherche la loi a posteriori de \mathbf{z} .
- Lois a priori
- Echantillonneur de Gibbs en 3 étapes \Rightarrow réalisations de

$$\mu, \sigma^2, \beta, \mathbf{z}.$$

Lois a priori

- $\sigma^2 \sim \mathcal{IG}(a, b)$,
- $\mu = (\mu_1, \dots, \mu_k) \sim \mathcal{U}(0 \leq \mu_1 \leq \dots \leq \mu_k \leq 256)$,
- $\beta \sim \mathcal{U}([0, 3])$.

Etape 1

- $p(z_i|z_{\delta i}, y, \beta, \mu, \sigma^2) \propto p(y_i|z_i, \mu, \sigma^2)p(z_i|z_{\delta i}, \beta)$.
- On a $p(z_i|z_{\delta i}, y, \beta, \mu, \sigma^2) \propto \text{Multinomial}(1; \omega_{i1}, \dots, \omega_{ik})$
avec :



$$\omega_{ij} = \frac{(\sqrt{2\pi}\sigma)^{-1} \exp[-\frac{1}{2}(\frac{y_i - \mu_j}{\sigma})^2 + \beta U(z_{\delta i}, j)]}{\sum_{j=1}^k \{(\sqrt{2\pi}\sigma)^{-1} \exp[-\frac{1}{2}(\frac{y_i - \mu_j}{\sigma})^2 + \beta U(z_{\delta i}, j)]\}}$$

où $U(z_{\delta i}, j) = \sum_{l \sim i} \mathbb{1}(z_l = j)$.

Etape 2

- $p(\mu, \sigma^2 | \beta, z, y) \propto [\prod_{i=1}^N p(y_i | z_i, \mu, \sigma^2)] p(\mu, \sigma^2)$.
- $\sigma^2 \sim \mathcal{IG}(\frac{N}{2} + a, \frac{\sum_{i=1}^N (y_i - \mu_{z_i})^2}{2} + b)$;
- $\mu_j \sim \mathcal{N}(\bar{y}_j, \frac{\sigma^2}{n_j}) \mathbf{1}(\mu_{j-1} \leq \mu_j \leq \mu_{j+1}), \quad \forall j = 1, \dots, k$;

Etape 3

- $p(\beta|\mu, \sigma^2, z, y) \propto p(z|\beta)p(\beta)$.
- Problème : $p(z|\beta)$ inconnue (constante de normalisation).
- Algorithme de Murray.

Algorithme de Murray

- Metropolis-Hastings avec variable auxiliaire.
- Variables courantes : β et z .
- Proposition : $\beta' \sim Q(.|\beta)$ et $z' \sim f(.|\beta')$.
- Proba d'acceptation explicite :

$$\frac{f(z|\beta')\pi(\beta')Q(\beta|\beta')f(z'|\beta)}{f(z|\beta)\pi(\beta)Q(\beta|\beta')f(z'|\beta')}.$$

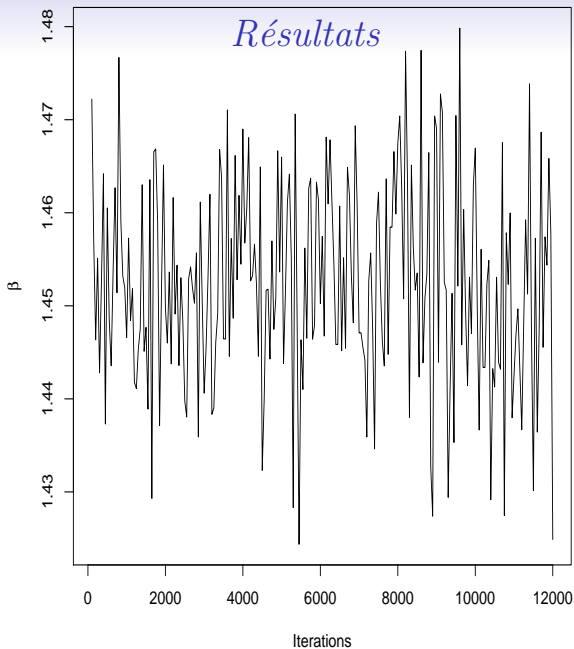
Finalemment

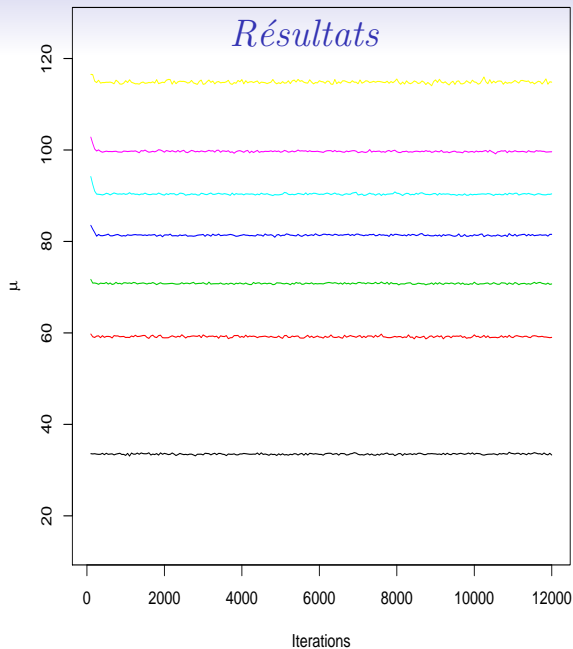
- On laisse passer une période de chauffe
- On obtient

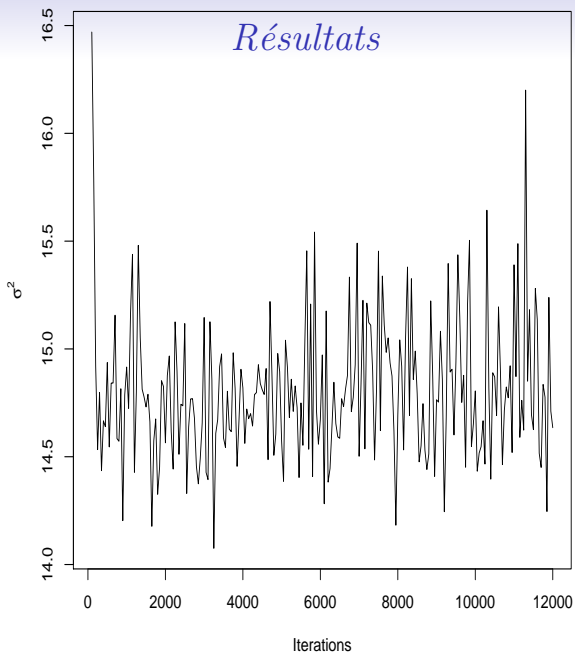
$$(\mu^{(t)}, \sigma^{2(t)}, \beta^{(t)}, z^{(t)}), t = 1, \dots, T.$$

- On estime toutes les lois a posteriori.
-

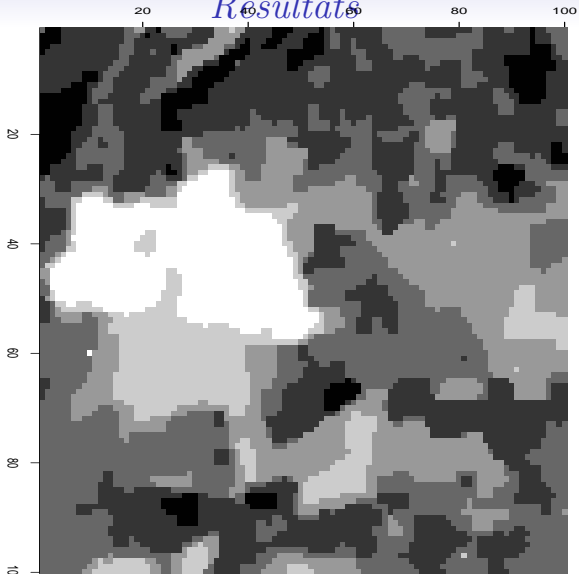
$$\hat{z}_i = \arg \max_{j=1, \dots, k} \sum_{t=1, \dots, T} \mathbb{1}(z_i^{(t)} = j).$$







Résultats



Plan de l'exposé

1-Motivations

2-Modélisation et inférence

3-Sélection du nombre de composantes

4-Conclusion

Plan de l'exposé

1-Motivations

2-Modélisation et inférence

3-Sélection du nombre de composantes

4-Conclusion

Comment choisir k ?

➤ $\theta_k = (\mu_{1,k}, \dots, \mu_{k,k}, \sigma_k, \beta_k) \in \mathbb{R}^{\otimes k+2}$.

➤ Loi a priori uniforme.

➤

$$\mathbb{P}(K = k) = \pi(k|\mathbf{y}) \propto m_k(\mathbf{y}) = \int f(\mathbf{y}|\theta_k)\pi(\theta_k)d\theta_k.$$

➤ Maximisation de la vraisemblance intégrée :

$$m_k(\mathbf{y}) = \int \sum_{\mathbf{z}_k} f(\mathbf{z}_k|\beta_k) \prod_{i=1}^N f(y_i|z_{i,k}, \theta_k) d\theta_k$$

Comment choisir k ?

- ➡ Vraisemblance complète intégrée :

$$\int f(\mathbf{y}, \mathbf{z}_k | \theta_k) \pi(\theta_k) d\theta_k$$

- ➡ Approximation de Laplace :

$$ICL(k) = -2 \log f(\mathbf{y}, \mathbf{z}_k^* | \theta_k^*) \pi(\theta_k^*) + (k + 2) \log n$$

où

$$f(\mathbf{y}, \mathbf{z}_k^* | \theta_k^*) \pi(\theta_k^*) = \max_{\mathbf{z}_k, \theta_k} f(\mathbf{y}, \mathbf{z}_k | \theta_k) \pi(\theta_k)$$

Critère ICL

- ➡ On rappelle que

$$f(\mathbf{y}, \mathbf{z}_k | \theta_k) = f(\mathbf{y} | \mathbf{z}_k, \theta_k) f(\mathbf{z}_k | \beta_k)$$

- ➡ Il faut approximer

$$f(\mathbf{z}_k | \beta_k) = \exp \left[\beta_k \sum_{l \sim i} \mathbb{1}(z_{i,k} = z_{l,k}) \right] / Z(\beta_k)$$

Critère ICL

➡ On a

$$f(\mathbf{z}_k | \beta_k) = \frac{f(S(\mathbf{z}_k) | \beta_k)}{N(S(\mathbf{z}_k))}$$

avec

$$N(s) = \sum_{\mathbf{z}_k} \mathbb{1}(S(\mathbf{z}_k) = \mathbf{s})$$

Critère ICL

Estimation du numérateur : $f(S(\mathbf{z}_k)|\beta_k)$ en deux étapes

- Simulation de $z^{(1)}, \dots, z^{(T)} \sim f(\cdot|\beta_k)$
- Estimation à noyau de la densité à partir de

$$S(z^{(1)}), \dots, S(z^{(T)})$$

Critère ICL

Estimation du dénominateur :

$$N(s) = \sum_{\mathbf{z}_k} \mathbb{1}(S(\mathbf{z}_k) = s) = k^N \mathbb{P}_{\beta_k=0}(S(\mathbf{z}_k) = s)$$

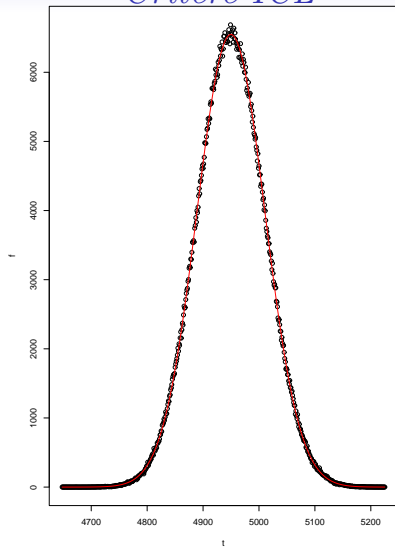
- ➡ k^N : nombre total de champs
- ➡ $\mathbb{P}_{\beta_k=0}(\cdot)$: probabilité lorsque les pixels sont i.i.d. uniformément sur $\{1, \dots, k\}$

Critère ICL

Estimation de $\mathbb{P}_{\beta_k=0}(S(\mathbf{z}_k) = s)$

- Solution naïve : simulation de $z^{(1)}, \dots, z^{(T)} \sim f(\cdot | \beta_k = 0)$
- Inefficace pour des valeurs de s trop grandes.

Critère ICL



Critère ICL

- On note N_{nei} le nombre de couples de pixels voisins.
- On associe à chaque couple $i \sim j$:

$$v_{i,j} = (z_{j,k} - z_{i,k}) \text{ modulo } k$$



$$S(\mathbf{z}_k) = \sum_{i \sim j} \mathbb{1}(v_{i,j} = 0)$$

- Si $\beta_k = 0$, $v_{i,j} \sim U(\{0, \dots, k-1\})$.

Critère ICL

- Connaître $N - 1$ arêtes non bouclantes, c'est les connaître toutes
- Simuler $S(\mathbf{z}_k)$ c'est simuler la valeur de $N - 1$ arêtes non bouclantes.
- Soit N_0 le nombre de ces arêtes non bouclantes à valeur nulle.
Pour $\beta_k = 0$,

$$N_0 \sim \mathcal{B}(N - 1, 1/k)$$

- De plus,

$$\mathbb{P}_{\beta_k=0}(S(\mathbf{z}_k) = s) = \sum_{i=0}^n \mathbb{P}_{\beta_k=0}(N_0 = i) \mathbb{P}_{\beta_k=0}(S(\mathbf{z}_k) = s | N_0 = i)$$

Critère ICL

Reste à estimer $\mathbb{P}_0(S(\mathbf{z}_k) = s | N_0 = i)$.

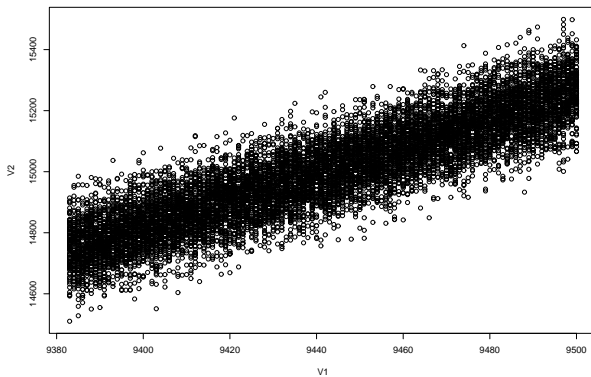
Pour chaque valeur de N_0 :

- tirage aléatoire des $N - 1$ arêtes non bouclantes,
- tirage des N_0 parmi eux fixés à 0,
- pour chacun des $N - 1 - N_0$ restants, simulation de sa valeur uniforme sur $\{1, \dots, k - 1\}$,
- calcul de la statistique exhaustive $S(\mathbf{z}_k)$.

Critère ICL

- La proba conditionnelle $\mathbb{P}_0(S(\mathbf{z}_k)|N_0)$ estimée uniquement pour les valeurs de $S(\mathbf{z}_k)$ obtenues après convergence. Pour $k = 10$, $S(\mathbf{z}_k) \in [14947, 15245]$ après convergence.
- Estimation de la probabilité par la fréquence empirique sur l'échantillon simulé.

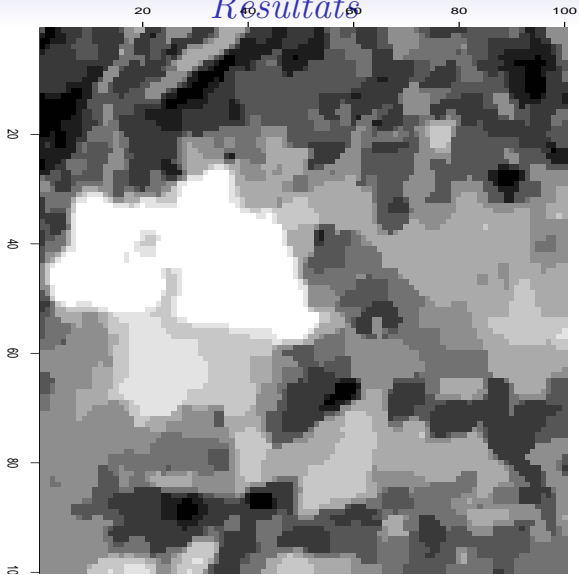
Critère ICL



Résultats

k	5	6	7	8	9	10	11
ICL	64174	61620	59874	59699	59580	58401	58515

Résultats



Plan de l'exposé

1-Motivations

2-Modélisation et inférence

3-Sélection du nombre de composantes

4-Conclusion

Plan de l'exposé

1-Motivations

2-Modélisation et inférence

3-Sélection du nombre de composantes

4-Conclusion

Conclusion

- ➡ Introduction de l'aspect spatial dans le modèle de mélange
- ➡ Algo de Murray : plus précis que pseudo-vraisemblance, plus rapide que estimation de la constante.
- ➡ Choix du nombre de composantes basé sur une estimation du critère ICL