

Champs de Markov caché pour la modélisation spatio-temporelle du risque épidémiologique

Lamiae Azizi⁽¹⁾⁽²⁾

⁽¹⁾Florence Forbes, Stéphane Girard
INRIA, Grenoble

⁽²⁾Myriam Garrido, David Abrial
Unité d'Epidémiologie Animale

1 Décembre 2009

Plan de la présentation

1. Présentation du sujet
2. Modèles Spatiaux
3. Algorithmes
4. Jeux de données en Épidémiologie Animale
5. Pistes explorées
6. Futur : Modèles Spatio-Temporels
7. Questions

« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Présentation du sujet

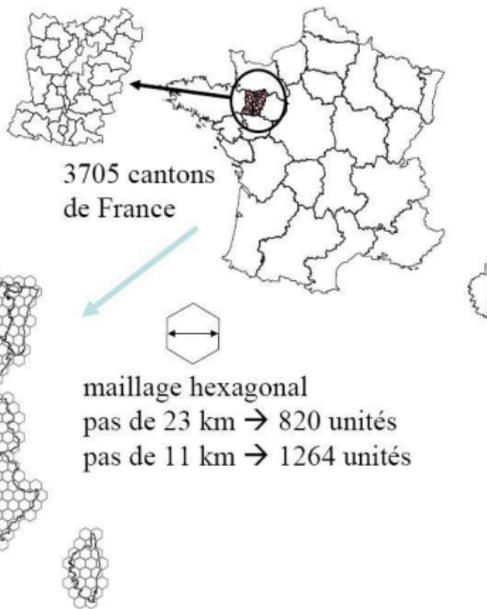
« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Cartographie du risque

Données disponibles

Changement d'unités géographiques



Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

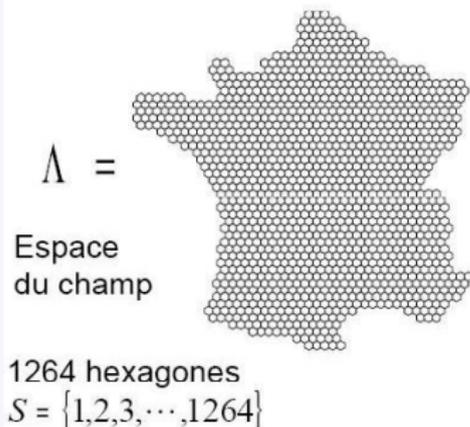
Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Présentation du sujet

Cartographie du risque



Données disponibles

y_i nombre de cas observés
 n_i effectifs de la population

$r = \{r_1, \dots, r_i, \dots, r_{1264}\} \in \Omega^{1264}$

$i \in S$

variables de risque



Classification des variables de risque R en 5 modalités par exemple

$\Omega = \{\text{'Très faible'}, \text{'Faible'}, \text{'Moyen'}, \text{'Fort'}, \text{'Très fort'}\}$



« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

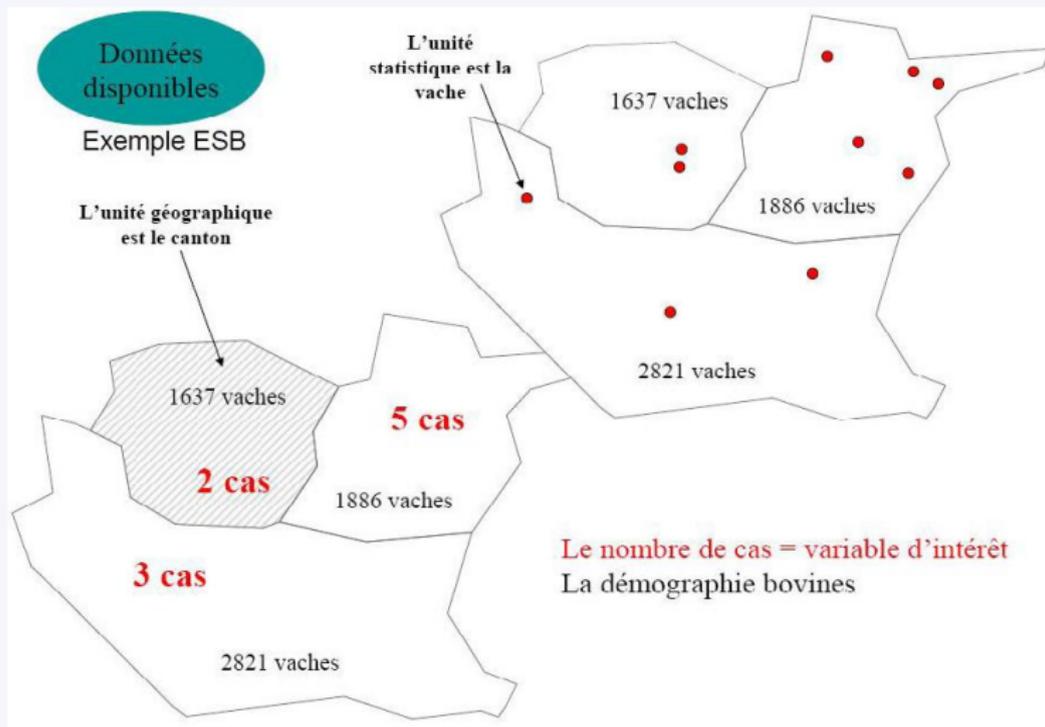
Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Présentation du sujet

Cartographie du risque



« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Modèle autorégressif conditionnel Gaussien(CAR)

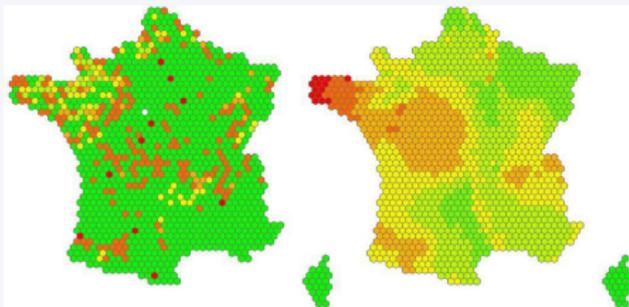
- ▶ $u_i | u_{-i}, \kappa_i \sim \mathcal{N}(\sum_{j \neq i} \beta_{ij} u_j, \kappa_i)$ avec : $i, j = 1, \dots, N$
- ▶ N représente le nombre d'unités géographiques.
- ▶ β_{ij} contrôlent la force d'interactions entre les unités.
- ▶ Besag et al (1991) ont proposé d'utiliser ce modèle pour la cartographie du risque en imposant des contraintes sur les β_{ij} et les variances κ_i : **Modèle autorégressif intrinsèque Gaussien (Besag york Mollié)**.

Présentation du sujet

Modèle de Besag York Mollié (BYM)

- ▶ $y_i \sim \text{Poisson}(e_i r_i)$ (resp. $n_i \lambda_i$)
- ▶ $\log(r_i) = u_i + v_i$
 1. e_i : nombre de cas attendus
 2. r_i : risque relatif
 3. u_i, v_i : composantes spatiales
- ▶ $u_i | u_{-i} \sim \mathcal{N}(\sum_{i \neq j} \beta_{ij} u_j, \kappa_i)$ (CAR)
- ▶ $v_i \sim N(\mu, \tau_v)$ (v_i indépendants)

Cartographie du risque



« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

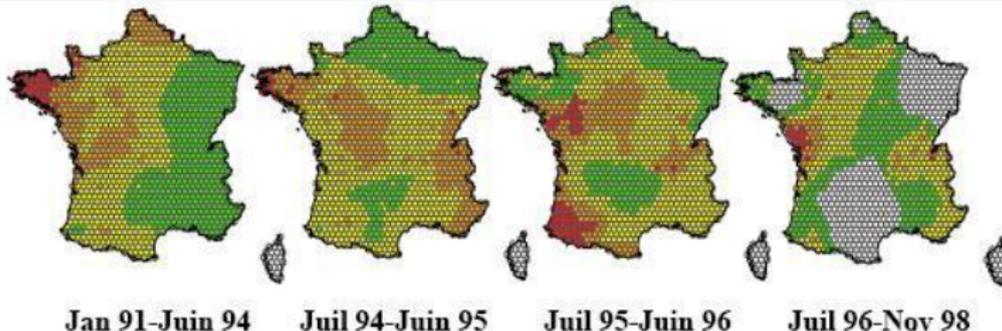
Questions

Présentation du sujet

Objectif de la thèse

- ▶ Prise en compte de la composante temporelle.
- ▶ Modélisation spatio-temporelle du risque épidémique sous forme de cartes de risque.

Cartographie du risque



« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Plan de la présentation

1. Présentation du sujet
2. **Modèles Spatiaux**
3. Algorithmes
4. Jeux de données en Épidémiologie Animale
5. Pistes explorées
6. Futur : Modèles Spatio-Temporels
7. Questions

« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Les Modèles

Hypothèse de base commune à (tous) les modèles :

- ▶ Observations : $y = \{y_1, \dots, y_n\} = \{y_i, i \in S\}$
- ▶ risque : $r = \{r_1, \dots, r_n\}$
- ▶ $\forall i \in S, y_i | r_i \sim \text{Poisson}(e_i r_i)$ (resp. $\text{Poisson}(n_i \lambda_i)$)
 1. e_i : nombre de cas attendu
 2. r_i : risque relatif
 3. n_i : population dans l'unité i
 4. λ_i : risque absolu
- ▶ Les modèles sont différents selon le choix de la loi sur r . (resp. λ)

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Modèles spatiaux

Modèles continus : Le Modèle de Besag York Mollié (BYM)

$\log(r_i) = u_i + v_i$ avec :

- ▶ $u_i, i \in S \sim \text{CAR Gaussien}$
- ▶ $v_i, i \in S \sim \text{Normale}$

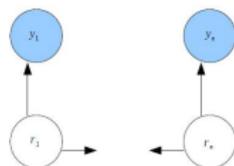
J.Besag (1986, 1991).

Le modèle Gaussien Multivarié avec Covariance Spatiale

Modèle avec $\{\log r_i, i \in S\}$ Gaussienne multivariée.

Best et al., Wakefield et al..

Modèle hiérarchique avec spatial au niveau "1"



« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

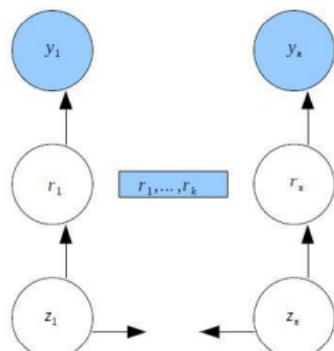
Modèles spatiaux

Modèles discrets

- ▶ $\forall r_i \in r_1, \dots, r_k \sim \mathcal{M}(\pi_1, \dots, \pi_k) \iff r_i = r_{z_i}$ et $z_i \in 1, \dots, K$ (indépendants) \sim Finite Poisson Mixture.
- ▶ $r_i = r_{z_i}$ et $Z = \{z_i, i \in S\} \sim$ Potts (et Extensions).

Green et Richardson (2002), Alfo et al (2009), Garrido et al (en cours).

Modèle hiérarchique avec Le spatial au niveau "2"



« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Couplage de continu et discret

$r_i = r(X_i, Z_i)$ avec : X_i continue et Z_i discrète.

► $r_i = r_{z_i} \exp(u_i + v_i)$

Ici, $X_i = (u_i, v_i)$

► Cas plus général :

1. $r_i = r_{i,z_i}$ avec $\forall k \in \{1, \dots, K\}$,
 $\{\log(r_{i,k}); i \in S\} \sim CAR_k$

2. $r_{i,k} = \exp(u_{i,k})$ avec $\{u_{i,k}, i \in S\} \sim CAR_k$

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

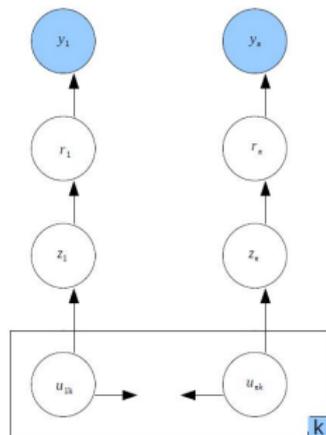
Modèles spatiaux

Approche hiérarchique Niveau "3"

- ▶ $\pi_{ik} = P(z_i = k)$
- ▶ Modèle logistique : $\pi_{ik} = \frac{\exp(u_{ik}|\phi)}{\sum_k^K \exp(u_{ik}|\phi)}$ Avec
 $u_k = \{u_{ik}, i \in S\} \rightsquigarrow$ CAR gaussien (indépendants).

Fernandez et Green 2002.

Modèle hiérarchique avec Le spatial au niveau "3": Logistique



« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

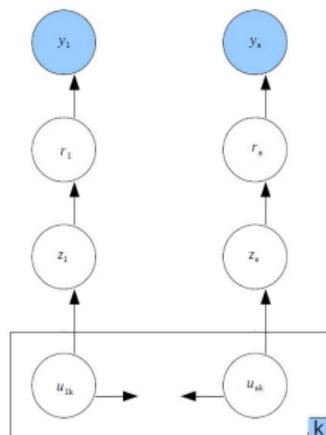
Modèles spatiaux

Approche hiérarchique Niveau "3"

- ▶ $w_{ik} = P(z_i = k)$
- ▶ $w_{i.} = \{w_{i1}, \dots, w_{ik}\} \sim \text{Dir}(\alpha_{i,1}, \dots, \alpha_{i,k})$
- ▶ $\alpha_{.k} = \{\alpha_{i,k}, i \in S\} = \{\exp(u_{i,k}), i \in S\} \sim \text{CAR}_k$

Durand et Francois(2009).

Modèle hiérarchique avec Le spatial au niveau "3": Dirichlet



« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Approches avec covariables

« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Modèle avec covariables

Si on a J covariables $\forall i$, r_i remplacé par :

$r_i \exp\langle C_i, \beta \rangle$ avec, $\langle C_i, \beta \rangle = \sum_{j=1}^J C_{ij} \beta_j$

- ▶ Cas discret : $r_{z_i} \exp\langle C_i, \beta_{z_i} \rangle$
- ▶ Autre cas : $r_{z_i} \exp(u_i + v_i) \exp\langle c_i, \beta_{z_i} \rangle$

Modèle avec covariables

- ▶ A essayer : Introduction de covariables géographiques ?

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Plan de la présentation

1. Présentation du sujet
2. Modèles Spatiaux
3. Algorithmes
4. Jeux de données en Épidémiologie Animale.
5. Pistes explorées
6. Futur : Modèles Spatio-Temporels
7. Questions

« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Algorithmes existants

- ▶ Gibbs et MCMC : Les plus courants dans la littérature
- ▶ L'algorithme Expectation-Maximisation (EM) et extensions
- ▶ L'Expectation-Maximisation Variationnel Bayésien (VEM)
- ▶ Les algorithmes hybrides :
 1. Monte-Carlo Expectation Maximisation (MCEM)
 2. Monte Carlo Variationnel Expectation Maximisation (MCVEM)

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

L'algorithme EM

- ▶ L'algorithme EM : Estimation de la solution du maximum de vraisemblance :

$$\hat{\theta} = \arg \max_{\theta \in \Theta} p(y|\theta)$$

- ▶ L'algorithme EM Bayésien : Estimation du maximum **a posteriori** (MAP) :

$$\hat{\theta} = \arg \max_{\theta \in \Theta} p(\theta|y)$$

- ▶ Interprétation fonctionnelle de EM utilisée pour estimer le MAP revient à :

$$F_{MAP}(q, \theta) = \sum_{z \in Z} \log(p(y, z|\theta))q(z) + \log(p(\theta)) + I[q]$$

1. q : Une famille de distributions
2. θ : les paramètres régissant la distribution jointe
3. $I[q] = -E_q[\log(q(Z))]$

EM Variationnel Bayésien

- ▶ L'estimation est réalisée via la maximisation alternée selon q et θ de F_{MAP}
- ▶ Valeur Initiale ($q^{(r)}, \theta^{(r)}$)
- ▶ Mise à jour itérative selon :
 1. Etape E :

$$\begin{aligned}q^{(r+1)} &= \arg \max_{q \in \mathcal{D}} F_{MAP}(q, \theta^{(r)}) \\ &= \arg \max_{q \in \mathcal{D}} \sum_{z \in Z} \log(p(z|y, \theta^r)) q(z) + I[q]\end{aligned}$$

2. Etape M :

$$\begin{aligned}\theta^{(r+1)} &= \arg \max_{\theta \in \Theta} F_{MAP}(q^{(r+1)}, \theta) \\ &= \arg \max_{\theta \in \Theta} \sum_{z \in Z} \log(p(\theta|y, z)) q^{(r+1)}(z)\end{aligned}$$

Plan de la présentation

1. Présentation du sujet
2. Modèles Spatiaux
3. Algorithmes
4. Jeux de données en Épidémiologie Animale
5. Pistes explorées
6. Futur : Modèles Spatio-Temporels
7. Questions

« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Données Épidémiologiques

Spécificité par rapport à l'épidémiologie Humaine :

- ▶ Les observations y_i sont petites (avec beaucoup de 0)
- ▶ Les populations n_i sont très variables et parfois aussi petits (voire nulles aussi)
- ▶ Autres ?
- ▶ \implies Besoin de modèles spécifiques ?

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Épidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Plan de la présentation

1. Présentation du sujet
2. Modèles Spatiaux
3. Algorithmes
4. Jeux de données en Épidémiologie Animale
5. Pistes explorées
6. Futur : Modèles Spatio-Temporels
7. Questions

« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Travail effectué

- ▶ Utilisation du logiciel SpaCEM3 : EM pour une classification floue + Champ de Markov pour modélisation des dépendances.
- ▶ Modèle testé : Modèle discret de Potts (et extensions)
- ▶ Application de différents algorithmes EM sur des données simulées épidémiologiques :
 1. Mean-Field Expectation-Maximisation
 2. Modal-Field Expectation-Maximisation
 3. Simulated-Field Expectation-Maximisation
- ▶ Différentes techniques d'initialisation
- ▶ résultats pas très satisfaisants :
 1. L'initialisation n'est pas très bonne ; problème de 0 ?
 2. Problème de classes vides
 3. Estimation des paramètres raisonnables en pratique

Plan de la présentation

1. Présentation du sujet
2. Modèles Spatiaux
3. Algorithmes
4. Jeux de données en Épidémiologie Animale
5. Pistes explorées
6. **Futur : Modèles Spatio-Temporels**
7. Questions

« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Épidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Modèles spatio-temporels

Modèles spatiaux avec incorporation de la dimension temporelle

- ▶ **Champ aléatoire de Markov caché (HMRF), 3D, vectoriel ou fonctionnel**
- ▶ **Modélisation Bayésienne**

Champ aléatoire de Markov caché (HMRF) vectoriel ou fonctionnel

- ▶ Les observations du modèle spatiale sont des vecteurs dont les composantes correspondent à différents temps ou des fonctions du temps.
 1. Cas vectoriel : covariances entre composantes des vecteurs ne dépendent pas de la **distance** entre elles.
 2. Cas fonctionnel.
- ▶ Carte 2D représente des groupes homogènes spatialement ayant des comportements similaires dans le temps.

« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Modèles spatio-temporels

Champ aléatoire de Markov caché (HMRF) 3D

- ▶ Graphe 3D ayant la dimension temporelle.
- ▶ Même modèle qu'en 2D mais paramètres différents :
interaction spatiale $\beta_s \neq$ interaction temporelle β_t
- ▶ Notion de voisinage temporelle à définir selon objectifs.

Modélisation Bayésienne

- ▶ Un modèle 2D spatial paramétrique (HMRF) dépendant d'un paramètre θ
- ▶ Poser une loi à priori $P(\theta)$
- ▶ Intégrer la dynamique temporelle par ce biais :
 1. Chercher une contrainte globale temporelle que l'on traduit sur une loi à priori globale $P(\theta_1, \dots, \theta_t)$
 2. Définir $P(\theta)$ à partir de l'estimation de $\hat{\theta}_{t-1}$ à $t - 1$
 3. Autre manière à explorer

Intérêt : Intégration de connaissances **à priori** d'experts, 27./29

« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Plan de la présentation

1. Présentation du sujet
2. Modèles Spatiaux
3. Algorithmes
4. Jeux de données en Épidémiologie Animale
5. Pistes explorées
6. Futur : Modèles Spatio-Temporels
7. Questions

« Champs de Markov cachés pour la modélisation spatio-temporelle du risque épidémiologique

L. Azizi

Présentation du sujet

Modèles spatiaux

Algorithmes

Jeux de données en Epidémiologie Animale

Pistes Explorées

Futur : Modèles Spatio-temporels

Questions

Problèmes pratique

- ▶ Application de l'algorithme EM : Problème ?
- ▶ Problème du "Label switching" ?
- ▶ Données avec beaucoup de 0 : Problème ?

Interrogations Méthodologiques

- ▶ L'intérêt pour les épidémiologistes d'avoir des cartes spatio-temporelles ? A quel type de questions peut répondre ce genre de cartes ?
- ▶ **Approximation par la loi de Poisson** : Adaptée quand on a peu de données ? jusqu'à quel point ?
- ▶ **Approches hybrides ?** : Mélanger le discret et le continu, quel apport ?
- ▶ Introduction des covariables géographiques ?