

Inférence bayésienne d'un modèle de Poisson bivarié à effets aléatoires pour la modélisation des variations spatiales de données de comptage corrélées

Édouard Chatignoux[†], Isabelle Albert*, Laurent Remontet[‡], Marc Colonna[§], Zoé Uhry^{†‡}

[†]Santé Publique France, *UMR518 AgroParisTech/INRAE,

[‡]Service de biostatistique des Hospices civils de Lyon,

[§]Réseau français des registres de cancer Francim



INRAE



Hospices Civils de Lyon



AppliBUGS, 13 juin 2023

Déroulé de la présentation

Contexte

Extension de la modélisation jointe à des données corrélées

Performance de la méthode

Application

Discussion

Déroulé de la présentation

Contexte

- Disease mapping en épidémiologie
- Modélisation jointe de plusieurs pathologies
- Données de comptage corrélées

Extension de la modélisation jointe à des données corrélées

Performance de la méthode

Application

Discussion

Déroulé de la présentation

Contexte

Disease mapping en épidémiologie

Modélisation jointe de plusieurs pathologies

Données de comptage corrélées

Extension de la modélisation jointe à des données corrélées

Performance de la méthode

Application

Discussion

Modélisation des variations spatiales en épidémiologie

Cartographie de l'incidence

Disease mapping

- ▶ Description des variations spatiales d'une pathologie
 - ▶ Zones de sur/sous incidence
 - ▶ Allocation des ressources
- ▶ Générer des hypothèses sur les facteurs entraînant une variation du risque
 - ▶ Unité spatiale = proxy de facteurs de risques inégalement distribués sur le territoire
 - ▶ Régression écologique

Données disponibles

- ▶ Données de comptage agrégées sur les unités spatiales (communes/départements...) $\rightarrow N_j$
- ▶ Effectifs attendus (standardisation indirecte age/sex) $\rightarrow E_j$

$$\hookrightarrow N_j \sim \mathcal{P}(E_j \lambda_j)$$

Objectif

Estimer λ_j

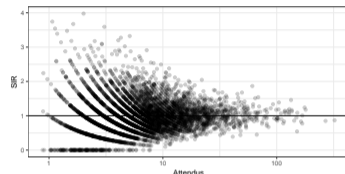
Modélisation des variations spatiales en épidémiologie

Modélisation des risques

Maximum de vraisemblance : $\lambda_j = N_j/E_j$

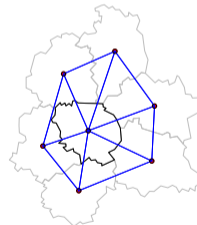
Instabilité des SIR

- ▶ Variabilité de Poisson des observations
- ▶ SIR d'autant plus variables que l'unité géographique est faiblement peuplée
- ▶ Peu utile pour caractériser les variations du risque



Utilisation de la structure spatiale des observations

- ▶ Besoin de lissage
- ▶ Proximité spatiale des observations
 - Risque varie de façon continue dans l'espace
 - Utilisation de distances/voisinages
- ▶ Modèles Bayésiens avec des effets latents par unité géographique
 - Prior autoregressifs conditionnels (CAR)^a



a. Julian BESAG et al. "Bayesian image restoration, with two applications in spatial statistics". *Annals of the Institute of Statistical Mathematics* 43.1 (1991), p. 1-20.

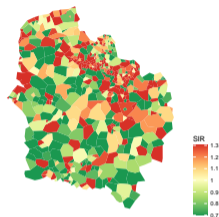
Modélisation des variations spatiales en épidémiologie

Modélisation des risques

Maximum de vraisemblance : $\lambda_j = N_j/E_j$

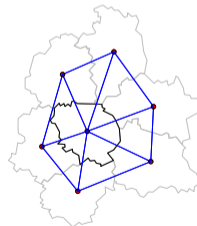
Instabilité des SIR

- ▶ Variabilité de Poisson des observations
- ▶ SIR d'autant plus variables que l'unité géographique est faiblement peuplée
- ▶ Peu utile pour caractériser les variations du risque



Utilisation de la structure spatiale des observations

- ▶ Besoin de lissage
- ▶ Proximité spatiale des observations
 - Risque varie de façon continue dans l'espace
 - Utilisation de distances/voisinages
- ▶ Modèles Bayésiens avec des effets latents par unité géographique
 - Prior autoregressifs conditionnels (CAR)^a



a. Julian BESAG et al. "Bayesian image restoration, with two applications in spatial statistics". *Annals of the Institute of Statistical Mathematics* 43.1 (1991), p. 1-20.

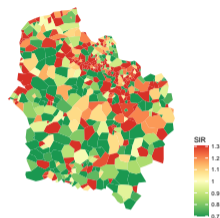
Modélisation des variations spatiales en épidémiologie

Modélisation des risques

Maximum de vraisemblance : $\lambda_j = N_j/E_j$

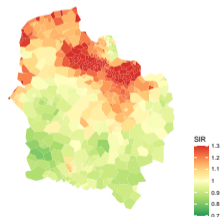
Instabilité des SIR

- ▶ Variabilité de Poisson des observations
- ▶ SIR d'autant plus variables que l'unité géographique est faiblement peuplée
- ▶ Peu utile pour caractériser les variations du risque



Utilisation de la structure spatiale des observations

- ▶ Besoin de lissage
- ▶ Proximité spatiale des observations
 - Risque varie de façon continue dans l'espace
 - Utilisation de distances/voisinages
- ▶ Modèles Bayésiens avec des effets latents par unité géographique
 - Prior autoregressifs conditionnels (CAR)^a



a. Julian BESAG et al. "Bayesian image restoration, with two applications in spatial statistics". *Annals of the Institute of Statistical Mathematics* 43.1 (1991), p. 1-20.

Modèle BYM

► $N_j \sim \mathcal{P}(E_j \exp(c_j))$

Modèle BYM

- ▶ $N_j \sim \mathcal{P}(E_j \exp(c_j))$
- ▶ $c_j = v_j + u_j$



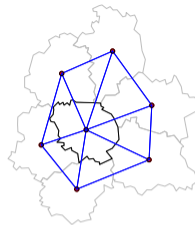
Modèle BYM

- ▶ $N_j \sim \mathcal{P}(E_j \exp(c_j))$
- ▶ $c_j = v_j + u_j$
- ▶ v_j : iid, sans structure spatial



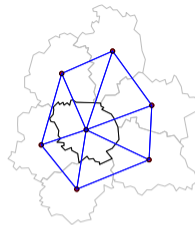
Modèle BYM

- ▶ $N_j \sim \mathcal{P}(E_j \exp(c_j))$
- ▶ $c_j = v_j + u_j$
- ▶ v_j : iid, sans structure spatiale
- ▶ u_j : dépend des voisins (voisinage $\delta_j =$ frontière commune) \rightarrow modèle ICAR
 $u_j | \mathbf{u}_j \sim \mathcal{N}\left(\frac{1}{n_{\delta_j}} \sum_{l \in \delta_j} u_l, \frac{1}{n_{\delta_j} \tau}\right)$



Modèle BYM

- ▶ $N_j \sim \mathcal{P}(E_j \exp(c_j))$
- ▶ $c_j = v_j + u_j$
- ▶ v_j : iid, sans structure spatiale
- ▶ u_j : dépend des voisins (voisinage $\delta_j =$ frontière commune) \rightarrow modèle ICAR
 $u_j | \mathbf{u}_j \sim \mathcal{N}\left(\frac{1}{n_{\delta_j}} \sum_{l \in \delta_j} u_l, \frac{1}{n_{\delta_j} \tau}\right)$



Paramétrisation BYM2¹

$$\begin{cases} c_j = \sigma_k(\sqrt{\phi}u_j + \sqrt{1-\phi}v_j) \\ \mathbf{u} \sim \text{ICAR}^* \\ v_j \sim \mathcal{N}(0, 1) \end{cases} \rightarrow \text{ICAR normé pour que } V(u_j) \simeq 1$$

- ▶ ϕ mesure la part de variance spatialement structurée
- ▶ σ_k mesure la variance totale des effets aléatoires

1. Andrea RIEBLER et al. "An intuitive Bayesian spatial model for disease mapping that accounts for scaling". *Statistical Methods in Medical Research* 25.2 (août 2016), p. 1-24. arXiv : 1601.01180.

Déroulé de la présentation

Contexte

Disease mapping en épidémiologie

Modélisation jointe de plusieurs pathologies

Données de comptage corrélées

Extension de la modélisation jointe à des données corrélées

Performance de la méthode

Application

Discussion

Distributions spatiales similaires pour plusieurs pathologies (i.e. facteurs de risques communs) → modélisation jointe des incidences

Statistical Methods in Medical Research 2005; **14**: 61–82

- ▶ Identification des variations communes
 - ↪ Partage d'effets aléatoires
- ▶ Identification des variations spécifiques
- ▶ Amélioration de la précision des estimations
 - ↪ Renforcement de l'information

Towards joint disease mapping

Leonhard Held Department of Statistics, University of Munich, Germany, **Isabel Natário** CEAUL, Faculdade de Ciências da Universidade de Lisboa and Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal, **Sarah Elaine Fenton** Department of Veterinary Clinical Sciences and Department of Mathematical Sciences, University of Liverpool, UK, **Håvard Rue** Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway and **Nikolaus Becker** German Cancer Research Center, Heidelberg, Germany

Développement des méthodes/applications

- ▶ Nombreuses méthodes, cadre général commun (*multidimensional disease mapping*)¹
- ▶ Applications régulières en épidémiologie depuis 10 ans²
 - Similarités dans la distribution des risques
 - Amélioration des estimateurs
 - Combinaison/calibration de plusieurs proxy

1. Miguel A. MARTINEZ-BENEITO et al. "Towards a Multidimensional Approach to Bayesian Disease Mapping". *Bayesian Analysis* 12.1 (mars 2017), p. 239-259; Ying C. MACNAB. "Bayesian disease mapping : Past, present, and future". en. *Spatial Statistics. Special Issue : The Impact of Spatial Statistics* 50 (août 2022), p. 100593.

2. Getayeneh Antehunegn TESEMA et al. "A Systematic Review of Joint Spatial and Spatiotemporal Models in Health Research". en. *International Journal of Environmental Research and Public Health* 20.7 (jan. 2023), p. 5295.

Modélisation jointe

Shared Component Model

Cas de deux sources de données

- ▶ Deux données de comptage (N_{1j}, N_{2j})
 - $(N_j|\mu_{ij}) \sim \mathcal{P}(E_{ij}\mu_{ij})$
- ▶ λ_{1j} dépend d'une variable latente partagée c_j et d'une spécifique ϵ_{ij}
 - $\log(\mu_{ij}) = \delta_i c_j + \epsilon_{ij}$
 - ▶ c_j capte les covariations de N_1 et de N_2
 - $\delta_i =$ facteurs d'échelle
 - ▶ ϵ_{ij} capte les variations spécifiques de N_1 et de N_2
- ▶ c_j et ϵ_{ij} suivent des GMRF
 - BYM, ICAR, \mathcal{N}
 - $\mathbb{V}(c_j) = 1$; $\mathbb{V}(\epsilon_{ij}) = \sigma_i^2$

Hypothèse

- ▶ $N_{1j}|\mu_{1j} \perp N_{2j}|\mu_{2j}$
- ▶ Toute la dépendance entre N_1 et N_2 est prise en compte dans c_j
- ▶ Quid de situations où N_1 et N_2 partagent des cas communs?

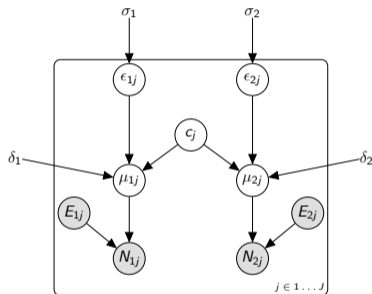


Figure – Shared component model

Déroulé de la présentation

Contexte

Disease mapping en épidémiologie

Modélisation jointe de plusieurs pathologies

Données de comptage corrélées

Extension de la modélisation jointe à des données corrélées

Performance de la méthode

Application

Discussion

Exemples

- ▶ Co-infection à plusieurs pathologies¹
- ▶ Covariation incidence/mortalité²
- ▶ Modélisation jointe de plusieurs proxys

Distribution de Poisson-Bivariée³

Si N_{1j} et N_{2j} partagent des cas communs en nombre X_{3j} , on peut réécrire $N_{1j} = X_{1j} + X_{3j}$ et $N_{2j} = X_{2j} + X_{3j}$.
Sous l'hypothèse que $X_{ij} | \lambda_{ij} \sim \mathcal{P}(\lambda_{ij})$, (N_{1j}, N_{2j}) suit une loi de Poisson-bivariée : $(N_{1j}, N_{2j}) \sim \mathcal{BP}(\lambda_{1j}, \lambda_{2j}, \lambda_{3j})$

$$P(N_1 = x, N_2 = y | \lambda_1, \lambda_2, \lambda_3) = e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \frac{\lambda_1^x}{x!} \frac{\lambda_2^y}{y!} \sum_{i=0}^{\min(x,y)} \binom{x}{i} \binom{y}{i} i! \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right)^i \quad (1)$$

1. Nadine SCHUR et al. "Modelling the geographical distribution of co-infection risk from single-disease surveys". en. *Statistics in Medicine* 30.14 (2011), p. 1761-1776.

2. J. ETXEBERRIA et al. "Joint modelling of brain cancer incidence and mortality using Bayesian age- and gender-specific shared component models". en. *Stochastic Environmental Research and Risk Assessment* 32.10 (oct. 2018), p. 2951-2969.

3. P. HOLGATE. "Estimation for the Bivariate Poisson Distribution". *Biometrika* 51.1/2 (1964), p. 241-245.

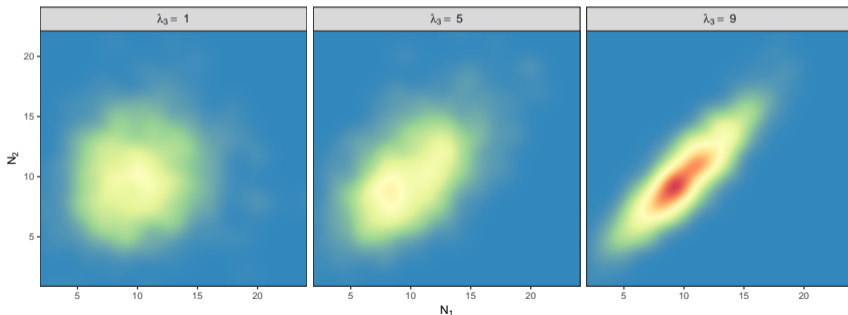
Distribution Poisson-bivariée

Propriétés

- ▶ Espérances et variances marginales inchangées
 - ▶ $\mathbb{E}(N_i | \lambda_i, \lambda_3) = \lambda_i + \lambda_3 = \mu_i$
 - ▶ $\mathbb{V}(N_i | \lambda_i, \lambda_3) = \lambda_i + \lambda_3 = \mu_i$
- ▶ Covariance positive, irréductible
 - ▶ $\text{Cov}(N_1, N_2 | \lambda_1, \lambda_2, \lambda_3) = \lambda_3$

Distribution données \mathcal{BP}

$n=1000$, $\lambda_1 = \lambda_2 = 10 - \lambda_3$



Déroulé de la présentation

Contexte

Extension de la modélisation jointe à des données corrélées

Problématique

Formulation du modèle

Performance de la méthode

Application

Discussion

Déroulé de la présentation

Contexte

Extension de la modélisation jointe à des données corrélées

Problématique

Formulation du modèle

Performance de la méthode

Application

Discussion

Violation de l'hypothèse d'indépendance

Conséquences

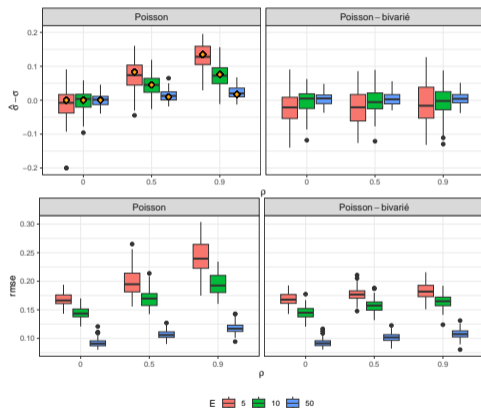
- ▶ Inflation de la co-variance
 - ▶ $\text{Cov}(N_1, N_2) = E_1 E_2 \text{Cov}(\mu_1, \mu_2) + \lambda_3$
 - ▶ v.s. $E_1 E_2 \text{Cov}(\mu_1, \mu_2)$ si $N_1 \perp N_2$,
- ⇒ Surestimation de la variance partagée
 - ▶ D'autant plus que λ_3 grand devant $E_1 E_2 \text{Cov}(\mu_1, \mu_2)$

Illustration - simulations

- ▶ $(N_{1j}, N_{2j}) \sim \mathcal{BP}(E \exp(c_j), E \exp(c_j), \rho E \exp(c_j))$
- ▶ $c_j \sim \mathcal{N}(0, \sigma^2)$
- ▶ Simulations
 - ▶ 100 unités géo, 100 simulations
 - ▶ $E = 5, 10, 50, \rho = 0, 0.5, 0.9, \sigma = 0.2$

Sous l'hypothèse $N_1 \perp N_2$

- ▶ Variance des c_j sur-estimée de $\frac{\rho}{2E}$
- ▶ Dégradation des RMSE



Déroulé de la présentation

Contexte

Extension de la modélisation jointe à des données corrélées

Problématique

Formulation du modèle

Performance de la méthode

Application

Discussion

Formulation d'un modèle SCM avec distribution Poisson-bivarée - I

	$N_1 \perp N_2$	$N_1 \not\perp N_2$
Distribution	$N_{ij} \sim \mathcal{P}(E_{ij}\mu_{ij})$	$(N_{1j}, N_{2j}) \sim \mathcal{BP}(A_{1j}\lambda_{1j}, A_{2j}\lambda_{2j}, A_{3j}\lambda_{3j})$
Variations spatiales	Deux sources μ_{1j} et μ_{2j}	Trois sources <ul style="list-style-type: none"> ▶ Cas partagés λ_{3j} ▶ Cas spécifiques λ_{1j} et λ_{2j}
Covariations	Entre les espérances μ_{1j} et μ_{2j}	Deux sources de co-variation <ul style="list-style-type: none"> ▶ Entre les espérances λ_{ij} ▶ Entre les observations <ul style="list-style-type: none"> ▶ Cas communs X_{3j}
Effectifs attendus	E_{ij} observés	<ul style="list-style-type: none"> ▶ $A_{1j} = E_{1j} - A_{3j}$, $A_{2j} = E_{2j} - A_{3j}$ ▶ A_{3j} non observés

Formulation d'un modèle SCM avec distribution Poisson-bivarée - II

Estimation des cas attendus

- ▶ $E_{ij} = \sum_a T_{ai} p a_{aj}$
 - ▶ T_{ai} : taux d'incidence pour la classe d'âge a
 - ▶ $p a_{aj}$: population d'âge a dans j
- ▶ $A_{3j} < \min(E_{1j}, E_{2j})$
 - ⇒ $A_{3j} = \rho \min(E_{1j}, E_{2j})$, $\rho \in [0 - 1]$
 - Taux d'incidence par âge pour $X_3 \simeq \rho T_{ai}$

Covariations entre les λ_{ij}

- ▶ Covariations entre λ_{3j} et $\lambda_{1j}, \lambda_{2j}$
 - ▶ Facteurs de risques communs aux cas communs X_3 et aux cas spécifiques X_1 et X_2
 - ▶ Assez probable : facteur de risque des cas communs également facteur de risque des cas spécifiques
 - ▶ Ex. : Obésité est un facteur du diabète, de l'hypertension, et de diabète + hypertension
- ▶ Covariations entre λ_{1j} et λ_{2j} indépendantes de λ_{3j}
 - ▶ Moins probable
 - ▶ Facteurs de risques des cas spécifiques qui ne serait pas un FDR des cas communs...
 - ▶ Peut arriver si λ_{1j} et λ_{2j} contiennent des erreurs corrélées (e.g. erreurs de notifications pour la modélisation de plusieurs proxys d'une maladie)
 - ▶ Formulation des λ_{ij}
 - ▶ $\lambda_{3j} = \exp(\delta_3 c_j)$
 - ▶ $\lambda_{ij} = \exp(\delta_i c_j + \epsilon_{ij})$, $i = 1, 2$

Formulation d'un modèle SCM avec distribution Poisson-bivarée - III

Distribution

- ▶ $(N_{1j}, N_{2j} | \lambda_{ij}) \sim \mathcal{BP}(A_{1j}\lambda_{1j}, A_{2j}\lambda_{2j}, A_{3j}\lambda_{3j})$

Cas attendus

- ▶ $A_{3j} = \rho \min(E_{1j}, E_{2j}), \rho \in [0, 1]$
- ▶ $A_{ij} = E_{ij} - \rho E_{1j}, i = 1, 2$
 - $\rho =$ proportion moyenne de cas partagés

Ratios d'incidence λ_{ij}

- ▶ $\log(\lambda_{3j}) = \delta_3 c_j$
- ▶ $\log(\lambda_{ij}) = \alpha_i + \delta_i c_j + \epsilon_{ij}, i = 1, 2$
- ▶ c_j et $\epsilon_{ij} \sim \text{GMRF}$
 - BYM, ICAR, \mathcal{N}
 - $\mathbb{V}(c_j) = 1; \mathbb{V}(\epsilon_{ij}) = \sigma_j^2$

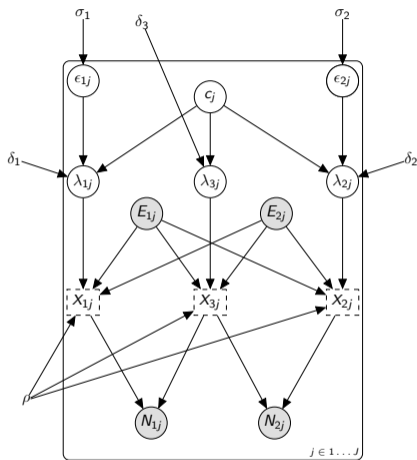
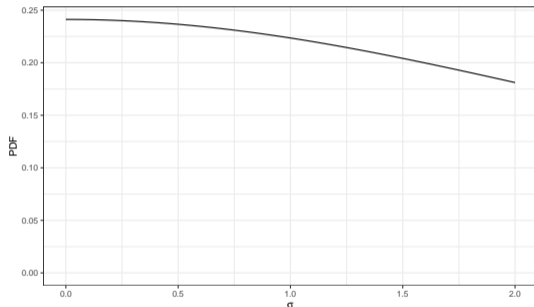


Figure – Shared component model pour données de Poisson-bivarie

Formulation d'un modèle SCM avec distribution Poisson-bivarée - IV

Prior

- ▶ Écart types (δ_i, σ_i) : Half student, de paramètre d'échelle 2.5 et de position 3 (défaut `brms`), restreints à l'intervalle 0-2
- ▶ Proportion de variance spatialement structurée (BYM) ϕ et coef ρ : Uniforme 0-1
- ▶ Autres constantes : gaussienne vague $\mathcal{N}(0, 5^2)$



Stan

- ▶ 2000 itérations, 1000 warmup
- ▶ Diagnostique des modèles : \hat{R} , ESS et transitions divergentes
- ▶ Implémentation ICAR, BYM selon Morris¹

1. Mitzi MORRIS et al. "Bayesian hierarchical spatial models : Implementing the Besag York Mollie model in stan". en. *Spatial and Spatio-temporal Epidemiology* 31 (nov. 2019), p. 100301.

Déroulé de la présentation

Contexte

Extension de la modélisation jointe à des données corrélées

Performance de la méthode

Application

Discussion

Simulations

Modèle de génération des données

$$(N_{1j}, N_{2j}) \sim \mathcal{PB}(A_{1j}\lambda_{1j}, A_{2j}\lambda_{2j}, A_{3j}\lambda_{3j})$$

▶ Cas attendus

$$\text{▶ } A_{3j} = \rho E_j$$

$$\text{▶ } A_{ij} = (1 - \rho)E_j, \quad i = 1, 2$$

→ E_j : Effectifs attendus pour hospitalisations pour cancer du Pomoun, $\bar{E}_j = 10$

▶ Ratios d'incidence

$$\text{▶ } \log(\lambda_{3j}) = \delta_3 u_j$$

$$\text{▶ } \log(\lambda_{ij}) = \delta_i u_j + \epsilon_{ij}, \quad i = 1, 2$$

$$\text{▶ } u_j \sim \text{BYM2}(\phi, 1), \quad \phi = 0.7$$

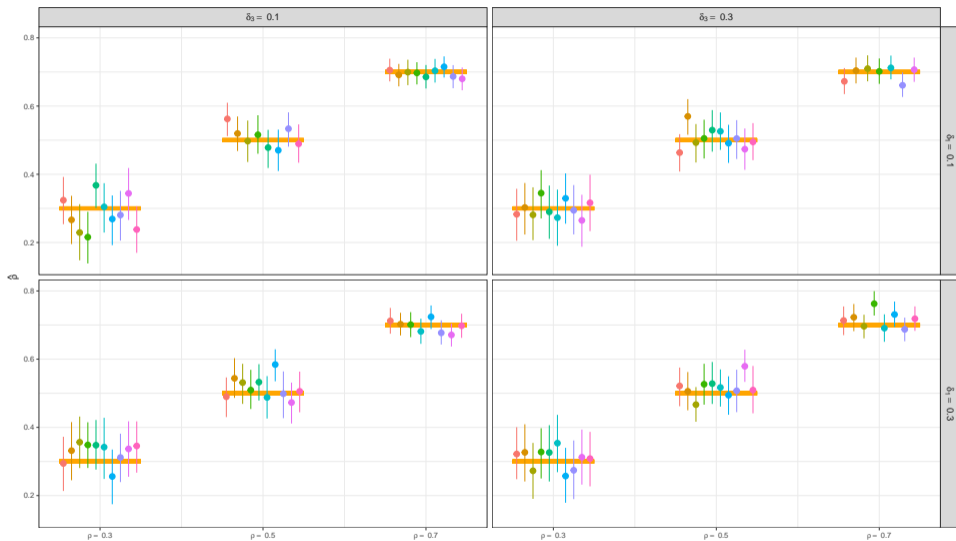
$$\text{▶ } \epsilon_{ij} \sim \mathcal{N}(\sigma_i^2), \quad i = 1, 2, \quad \sigma_{1j} = 0.1, \quad \sigma_{2j} = 0.2$$

Simulations

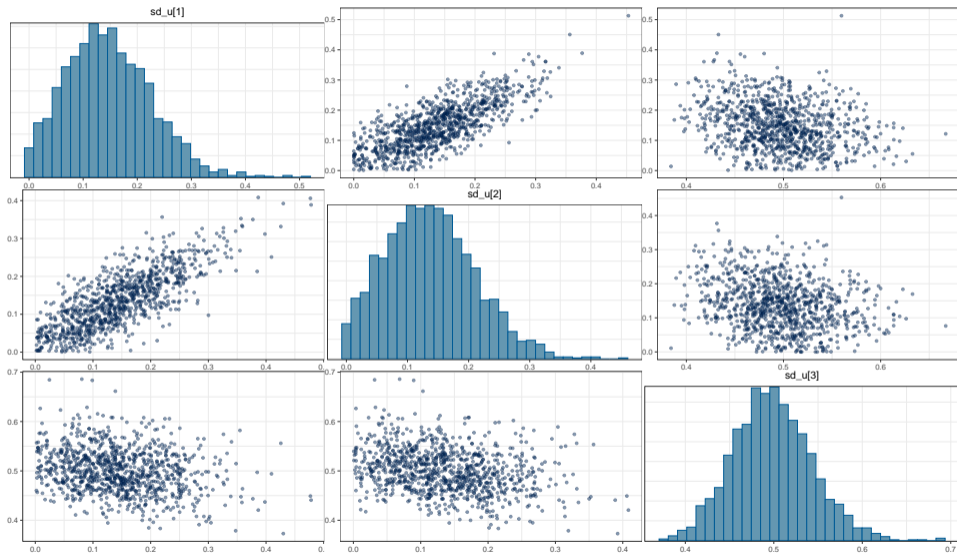
Pour $\rho = (0.3, 0.5, 0.7)$, $\delta_3 = (0.1, 0.3)$, $\delta_1 = (0.1, 0.3)$, $\delta_2 = 0.2$, $J = 1000$ unité géographique
 $s = 1 \dots 10$ simulations

1. Génération des cas selon le modèle de génération
2. Estimation des paramètres par
 - ▶ Modèle de génération des données
 - ▶ Modèles SCM équivalent (hypothèse $N_1 \perp N_2$)

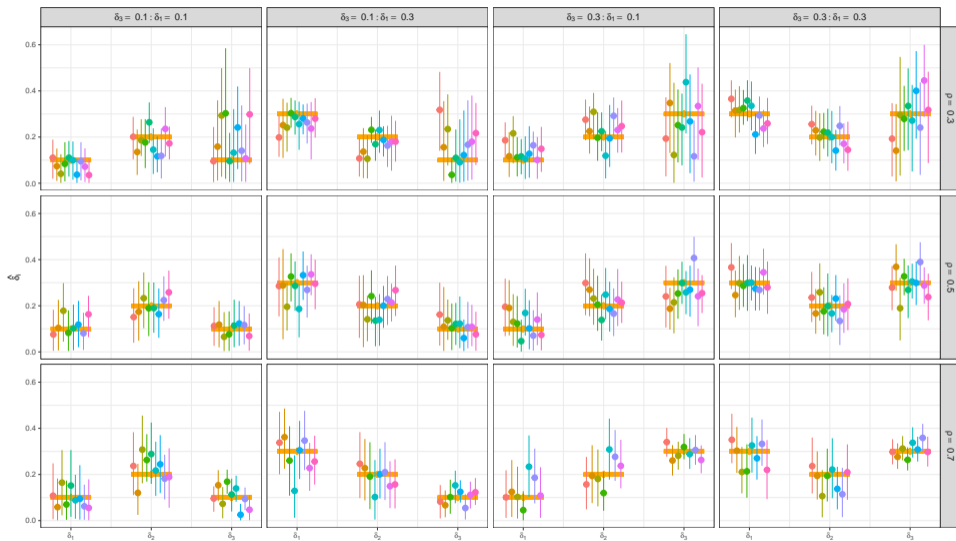
Estimation de ρ



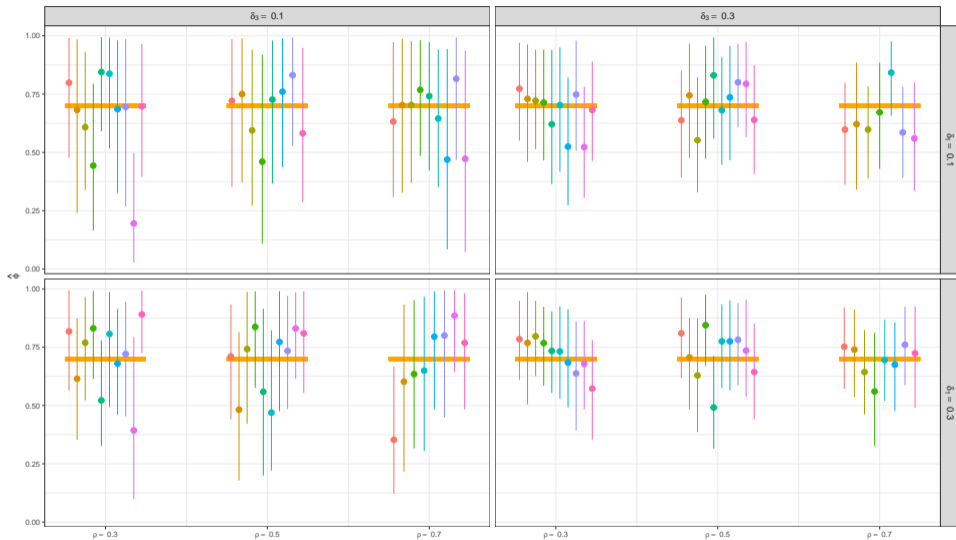
Estimation des δ_j - I



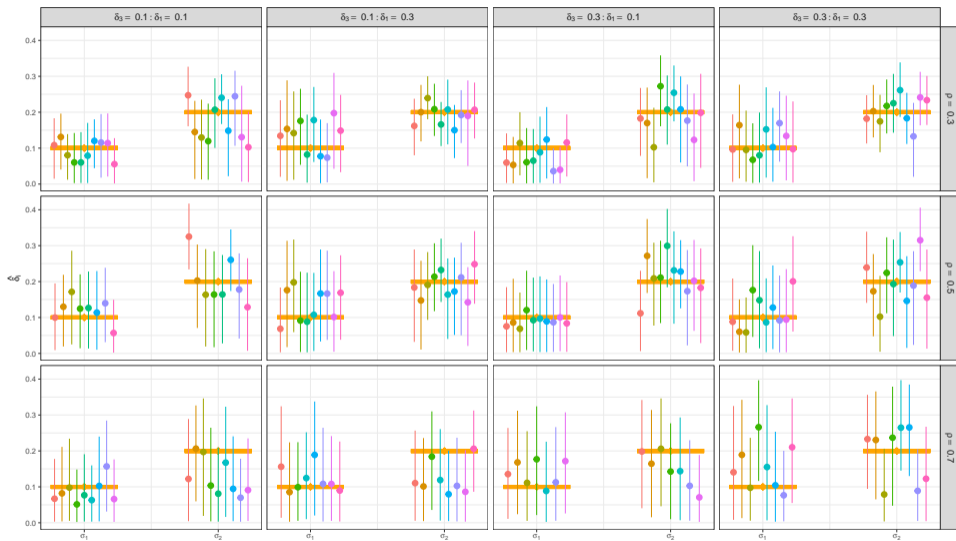
Estimation des δ_j - II



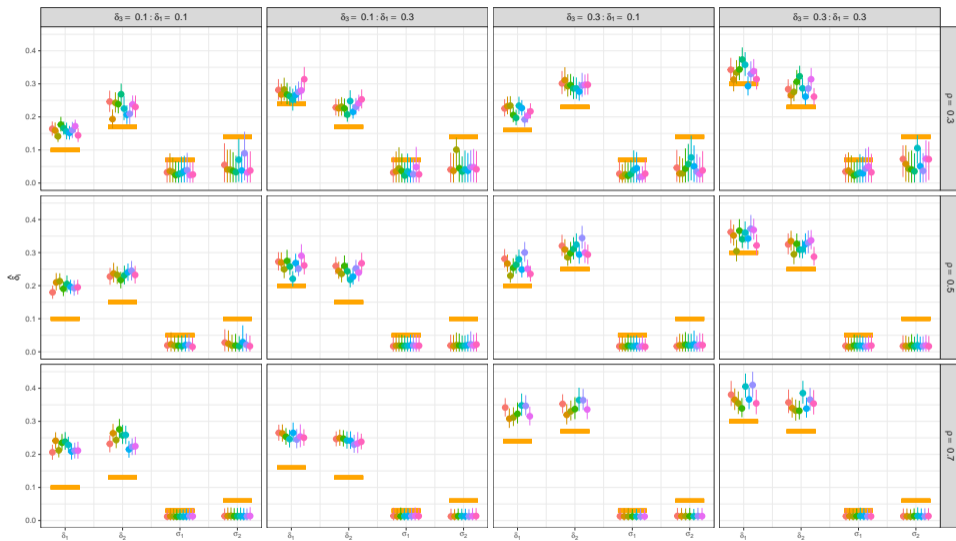
Estimation de ϕ



Estimation des σ_i



Comparaison avec SCM Poisson



Bilan des simulations

Déroulé de la présentation

Contexte

Extension de la modélisation jointe à des données corrélées

Performance de la méthode

Application

Discussion

Proxys de l'incidence des cancers

- ▶ Nombre d'hospitalisations et de mises en affection longue durée pour cancer
- ▶ Période : 2013-2019
- ▶ Échelle des codes postaux (n=5533 unités géographiques)
- ▶ Données SNDS

Prévalence de l'hypertension et du diabète

- ▶ Nombre de cas de diabète traité et d'hypertendus, hommes de 50 ans et moins
- ▶ Période : 2020
- ▶ Échelle des codes communes, nord de la France (n=15600 unités géographiques)
- ▶ Données SNDS (algorithmes SPF)

Modèles

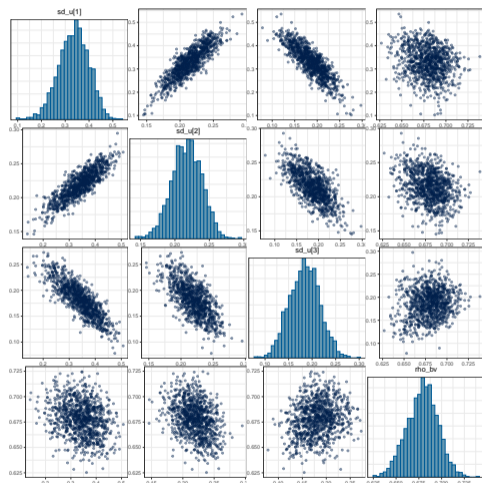
- ▶ Modèle BYM sur comptages pris isoléments
 - ▶ Cas totaux dans chaque source
 - ▶ Cas partagés et cas spécifiques
- ▶ Modèles SCM
 - Formulations \mathcal{P} et \mathcal{PB}

Source	P_{25}	P_{50}	P_{75}	Moy.	Somme
Foie					
A	1	3	7	5.4	29,920
H	3	6	12	10.1	55,955
H&A	1	2	4	3.1	17,333
SNC					
A	1	2	5	4.2	23,498
H	1	3	6	5.3	29,065
H&A	0	1	3	2.4	13,336
HTN-dia					
Dia	0	0	1	1.7	27,226
HTN	0	1	3	4.2	65,130
D&H	0	0	0	0.3	4,822

Cancer du foie, hommes, 2013-2019 - I

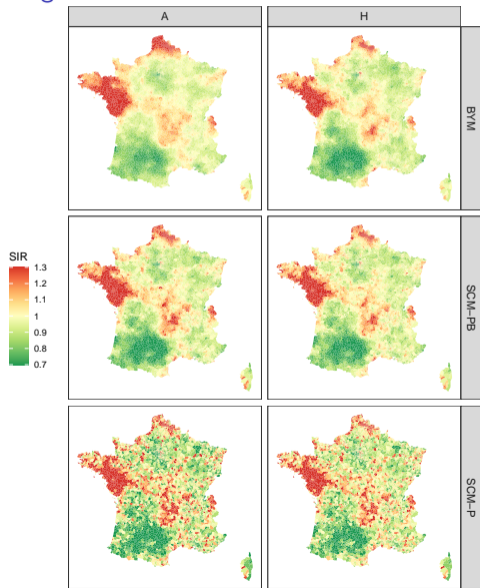
Paramètres

	SCM - \mathcal{PB}			SCM - \mathcal{P}		
	mean	$P_{2.5}$	$P_{97.5}$	mean	$P_{2.5}$	$P_{97.5}$
δ_1	0.34	0.21	0.45	0.34	0.32	0.36
δ_2	0.22	0.17	0.26	0.28	0.26	0.30
δ_3	0.18	0.12	0.25	0.31	0.29	0.33
σ_1	0.19	0.03	0.32	0.01	0.00	0.03
σ_2	0.13	0.09	0.16	0.01	0.00	0.02
ϕ	0.97	0.91	1.00	0.67	0.58	0.76
ρ	0.68	0.64	0.71			

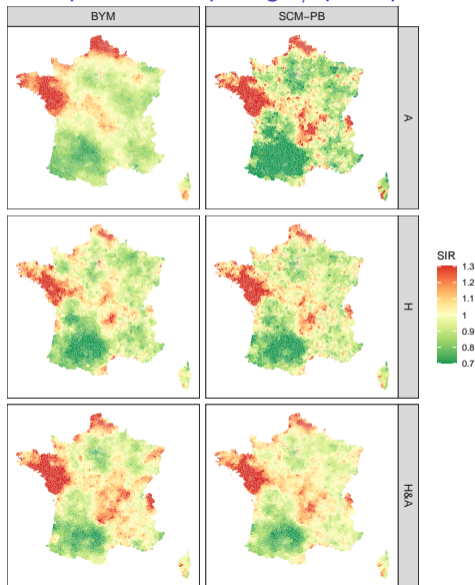


Cancer du foie, hommes, 2013-2019 - II

Marges



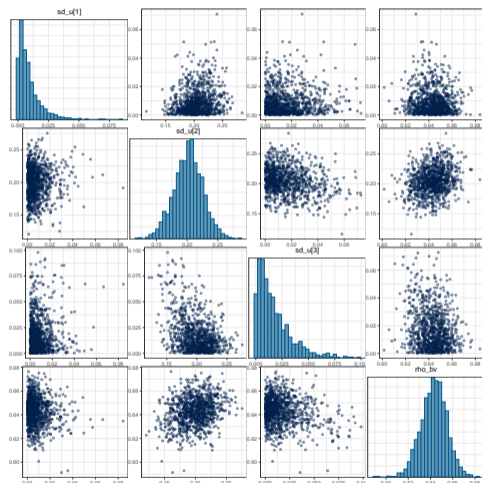
Décomposition cas partagés/spécifiques



Cancer du SNC, hommes, 2013-2019 - I

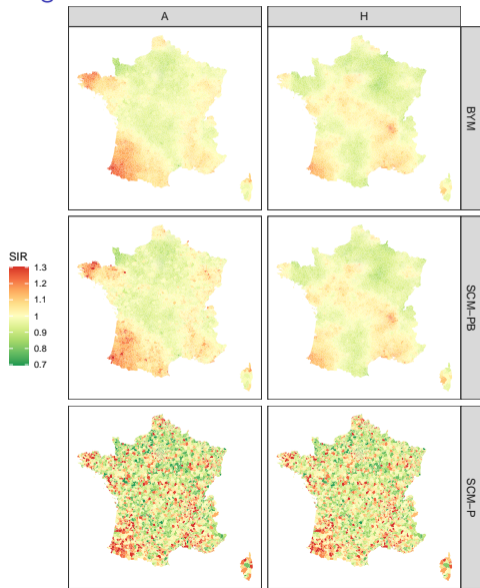
Paramètres

	SCM - \mathcal{PB}			SCM - \mathcal{P}		
	mean	$P_{2.5}$	$P_{97.5}$	mean	$P_{2.5}$	$P_{97.5}$
δ_1	0.03	0.00	0.12	0.28	0.26	0.30
δ_2	0.06	0.00	0.15	0.24	0.23	0.26
δ_3	0.09	0.04	0.14	0.26	0.24	0.28
σ_1	0.32	0.23	0.39	0.01	0.00	0.03
σ_2	0.17	0.07	0.23	0.01	0.00	0.03
ϕ	0.75	0.33	0.99	0.14	0.07	0.23
ρ	0.64	0.61	0.66			

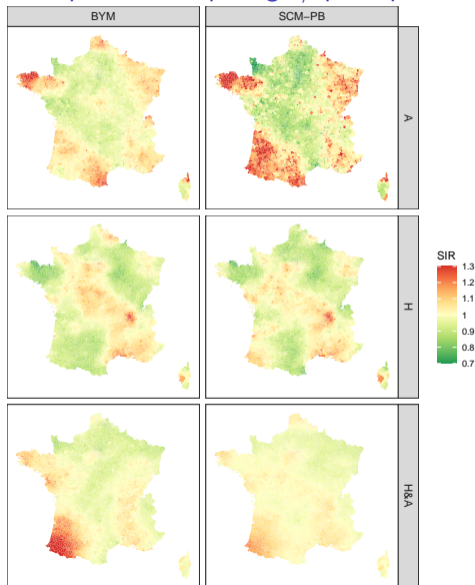


Cancer du SNC, hommes, 2013-2019 - II

Marges



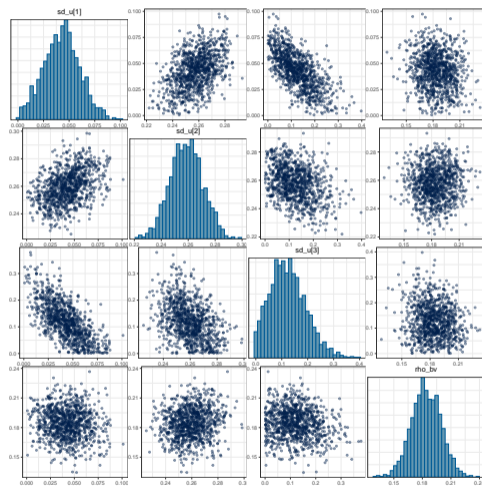
Décomposition cas partagés/spécifiques



Diabète et hypertension, 2019 - I

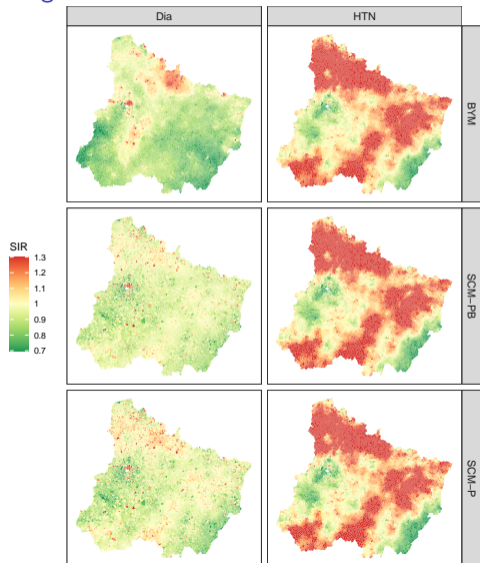
Paramètres

	SCM - \mathcal{PB}			SCM - \mathcal{P}		
	mean	$P_{2.5}$	$P_{97.5}$	mean	$P_{2.5}$	$P_{97.5}$
δ_1	0.04	0.01	0.08	0.09	0.08	0.10
δ_2	0.26	0.24	0.28	0.26	0.23	0.30
δ_3	0.12	0.01	0.27	0.16	0.13	0.18
σ_1	0.33	0.30	0.35	0.27	0.25	0.29
σ_2	0.03	0.00	0.07	0.02	0.00	0.05
ϕ	0.94	0.85	1.00	0.93	0.86	0.99
ρ	0.18	0.15	0.21			

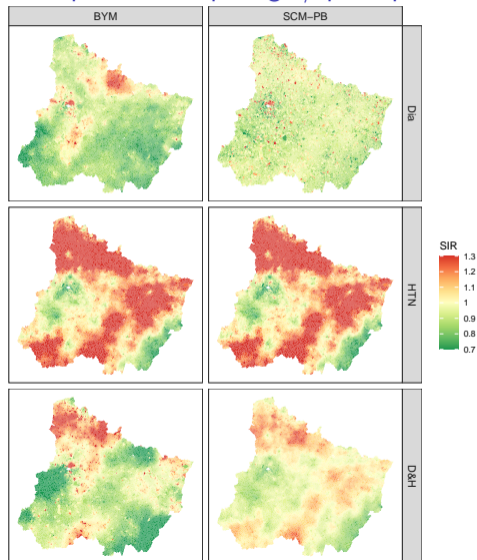


Diabète et hypertension, 2019 - II

Marges



Décomposition cas partagés/spécifiques



Déroulé de la présentation

Contexte

Extension de la modélisation jointe à des données corrélées

Performance de la méthode

Application

Discussion

Modélisation de données de comptage corrélées

Méthodes

- ▶ Pluralité de méthodes¹
 - Copules, Poisson log-normale, distribution de Poisson généralisée...
- ▶ Pas toujours de distinction claire entre dépendance des espérances et dépendance des observations conditionnellement aux espérances

Distribution de Poisson-bivariée

- ▶ Distribution assez peu utilisée en pratique
 - ▶ Buts inscrits en match de foot (GORGI et al. 2023)
 - ▶ Consultation/prescription (ALMUHAYFITH et al. 2016)
 - ▶ Quelques extensions mixtes (KARLIS et al. 2005a)
 - ▶ Paramètre aléatoire commun aux $\lambda_i \rightarrow \alpha \lambda_i$: surdispersion...
 - ▶ Distribution jointe ($\lambda_1, \lambda_2, \lambda_3$)
- ▶ Inférence
 - ▶ Approches EM (KARLIS et al. 2005b)
 - ▶ Bayésiennes (KARLIS et al. 2008 ; AL-WAHSH et al. 2020)
 - ▶ Autres (KOCHERLAKOTA et al. 2001 ; JOHNSON 1997)

1. David I. INOUE et al. "A review of multivariate distributions for count data derived from the Poisson distribution". en. *WIREs Computational Statistics* 9.3 (2017), e1398.

SCM et données de comptage corrélées

Enjeu des cas partagés

Si non pris en compte :

- ▶ Surestimation de la puissance disponible
- ▶ Transfert de la corrélation de Poisson vers la corrélation des espérances
→ $\text{Cov}(N_1, N_2) = E_1 E_2 \text{Cov}(\mu_1, \mu_2) + \lambda_3$
- ▶ D'autant plus problématique que $\frac{\lambda_3}{E_1 E_2} \simeq \frac{\rho}{\max(E_1, E_2)}$ grand devant $\text{Cov}(\mu_1, \mu_2)$
 - ▶ Proportion de cas partagé important
 - ▶ Nombre de cas faible

Extension SCM au cas Poisson-bivarié

- ▶ Permet de corriger le transfert de corrélation
- ▶ Performances
 - ▶ Bonne estimation de ρ
 - ▶ Dépendance importante entre les paramètres de variance
→ ↗ incertitudes sur les paramètres
- ▶ Application
 - ▶ Permet de retrouver les grands contrastes partagés et spécifiques
 - ▶ Estimation supplémentaire de la proportion de cas partagés

Approche bayésienne

- ▶ Utilisation “facile” d'une distribution non-standard
- ▶ Utilité +++ des diagnostics (divergences, mixage)
- ▶ Temps + ressources de calcul
 - 12-24h pour les modèles proxys

Perspectives

- ▶ Article en cours de rédaction
- ▶ Application pour la prédiction de l'incidence des cancers (calibration)
 - ▶ Incidence des cancers issues registres, disponibles sur 20% du territoire
 - ▶ Proxys de l'incidence (e.g. hospitalisation, ALD) disponibles sur tout le territoire

Références I

- [1] Julian BESAG et al. "Bayesian image restoration, with two applications in spatial statistics". *Annals of the Institute of Statistical Mathematics* 43.1 (1991), p. 1-20.
- [2] Andrea RIEBLER et al. "An intuitive Bayesian spatial model for disease mapping that accounts for scaling". *Statistical Methods in Medical Research* 25.2 (août 2016), p. 1-24. arXiv : 1601.01180.
- [3] Miguel A. MARTINEZ-BENEITO et al. "Towards a Multidimensional Approach to Bayesian Disease Mapping". *Bayesian Analysis* 12.1 (mars 2017), p. 239-259.
- [4] Ying C. MACNAB. "Bayesian disease mapping : Past, present, and future". en. *Spatial Statistics. Special Issue : The Impact of Spatial Statistics* 50 (août 2022), p. 100593.
- [5] Getayeneh Antehuneqn TESEMA et al. "A Systematic Review of Joint Spatial and Spatiotemporal Models in Health Research". en. *International Journal of Environmental Research and Public Health* 20.7 (jan. 2023), p. 5295.
- [6] Nadine SCHUR et al. "Modelling the geographical distribution of co-infection risk from single-disease surveys". en. *Statistics in Medicine* 30.14 (2011), p. 1761-1776.
- [7] J. ETXEBERRIA et al. "Joint modelling of brain cancer incidence and mortality using Bayesian age- and gender-specific shared component models". en. *Stochastic Environmental Research and Risk Assessment* 32.10 (oct. 2018), p. 2951-2969.
- [8] P. HOLGATE. "Estimation for the Bivariate Poisson Distribution". *Biometrika* 51.1/2 (1964), p. 241-245.
- [9] Mitzi MORRIS et al. "Bayesian hierarchical spatial models : Implementing the Besag York Mollié model in stan". en. *Spatial and Spatio-temporal Epidemiology* 31 (nov. 2019), p. 100301.

Références II

- [10] David I. INOUYE et al. "A review of multivariate distributions for count data derived from the Poisson distribution". en. *WIREs Computational Statistics* 9.3 (2017), e1398.
- [11] P. GORGI et al. "Estimation of final standings in football competitions with a premature ending : the case of COVID-19". en. *AStA Advances in Statistical Analysis* 107.1 (mars 2023), p. 233-250.
- [12] Fatimah E. ALMUHAYFITH et al. "On bivariate Poisson regression models". en. *Journal of King Saud University - Science* 28.2 (avr. 2016), p. 178-189.
- [13] Dimitris KARLIS et al. "Mixed Poisson Distributions". en. *International Statistical Review* 73.1 (2005), p. 35-58.
- [14] Dimitris KARLIS et al. "Bivariate Poisson and Diagonal Inflated Bivariate Poisson Regression Models in R". en. *Journal of Statistical Software* 14.1 (sept. 2005), p. 1-36.
- [15] Dimitris KARLIS et al. "Exact Bayesian modeling for bivariate Poisson data and extensions". en. *Statistics and Computing* 18.1 (mars 2008), p. 27-40.
- [16] Huda AL-WAHSH et al. "A bivariate autoregressive Poisson model and its application to asthma-related emergency room visits". eng. *Statistics in Medicine* 39.23 (oct. 2020), p. 3184-3194.
- [17] Subrahmaniam KOCHERLAKOTA et al. "Regression in the Bivariate Poisson Distribution". *Communications in Statistics - Theory and Methods* 30.5 (avr. 2001), p. 815-825.
- [18] JOHNSON. *Discrete Multivariate Distributions*. en-us. John Wiley & Sons, 24 jan. 1997. 332 p.