

# CARTOGRAPHIE DU RISQUE POUR MALADIE CONTAGIEUSE EN UTILISANT LA LOI BINOMIALE NÉGATIVE

David Abrial & Myriam Charras-Garrido

*Unité d'Epidémiologie Animale, UR346, Centre INRA de Clermont-Ferrand-Theix,  
63122, Saint-Genès-Champanelle, France.*

*myriam.charras-garrido@clermont.inra.fr, david.abrial@clermont.inra.fr*

**Résumé.** La cartographie du risque en épidémiologie permet de localiser des aires avec un risque fort ou un risque faible de maladie, et produit une mesure de la différence de risque entre ces régions. Les modèles actuels de cartographie du risque pour données groupées par unité géographique utilisés par les épidémiologistes sont basés sur un modèle de Poisson mixte log-linéaire intégrant un champ de Markov continu latent. Ce modèle comprend un lissage spatial par champ gaussien auto-régressif et fait partie des méthodes bayésiennes hiérarchiques. Il est adapté à des maladies non contagieuses car il utilise la loi de Poisson qui est sans mémoire pour modéliser les cas de maladie dans chaque unité géographique. Pour adapter cette méthode aux maladies contagieuses, nous proposons de remplacer la loi de Poisson par une loi binomiale négative. En effet, la loi binomiale négative prend en compte une surdispersion par rapport à la loi de Poisson. Il apparaît que ce nouveau modèle améliore le lissage spatial comparé au modèle usuel de Poisson car il prend en compte une surdispersion locale au niveau de chaque unité géographique, en complément des échelles plus larges de surdispersion prises en compte par le modèle log-linéaire mixte.

**Mots-clés.** Estimation bayésienne, Maladie transmissible ou contagieuse, surdispersion, MCMC

**Abstract.** Disease mapping in epidemiology enables to identify the location of areas with low or high risk of disease, and provides a measure of risk differences between these regions. Current disease mapping models for pooled data used by epidemiologists are based on Poisson log-linear mixed model with latent intrinsic continuous Markov random field. Designed for the risk estimation in each geographical unit, this model is built on a gaussian auto-regressive spatial smoothing and hierarchical Bayesian approaches. It is adapted for non contagious diseases since it uses the memoryless Poisson distribution to model disease counts, that imply an independence of the observed cases. In order to adapt this method to contagious diseases, we propose to replace this Poisson distribution by a negative binomial. Indeed, the negative binomial distribution takes into account an overdispersion compared to the Poisson distribution. It appears to improve the spatial smoothing in disease mapping compared to the usual Poisson model taking into account a

local overdispersion at the level of each geographical unit, in addition to the larger scales of overdispersion incorporated in the log-linear mixed model.

**Keywords.** Disease Mapping, Negative binomial distribution, Bayesian estimation, Transmissible or Contagious disease, Overdispersion, MCMC

La cartographie du risque est un outil largement utilisé pour l'étude des maladies non contagieuses humaines (par exemple en cancérologie [12]). Elle permet l'étude de la répartition spatiale, et plus récemment de l'évolution spatio-temporelle, de ces maladies, ainsi que la mise en évidence de facteurs épidémiologiques. La cartographie du risque a été récemment introduite en épidémiologie animale [1,2,5,6]. Dans ce contexte, elle a montré un potentiel intéressant d'une part en terme de compréhension de la dynamique des épidémies et d'autre part en terme d'aide à la décision en tant que support pour la mise en œuvre de mesures de protections spatialement délimitées comme la vaccination, la restriction de circulation ou l'abattage du cheptel. Cependant son champ d'application actuel est restreint aux maladies non contagieuses. Ceci limite fortement son développement dans le cadre vétérinaire car les maladies non contagieuses sont peu fréquemment étudiées chez l'animal. Cette limitation existe également, mais n'a pas la même importance, en épidémiologie humaine où les maladies non contagieuses sont plus étudiées. L'approche de cartographie du risque pour maladie contagieuse que nous proposons apporte un nouveau cadre méthodologique pour l'étude pratique des maladies contagieuses chez l'animal, comme chez l'homme.

Les méthodes statistiques de cartographie sont souvent nécessaires pour la représentation spatiale interprétable de données réparties aléatoirement sur un territoire. En particulier, une carte des risques épidémiologiques basée sur le rapport cas population dans chaque unité géographique est généralement très bruitée. Pour pallier ce bruit, un lissage est nécessaire pour obtenir des cartes interprétables, c'est-à-dire faisant ressortir des zones à plus ou moins fort risque. Ce lissage est effectué en modélisant un lien entre unités géographiques voisines. Les modèles actuels de cartographie du risque, en particulier le modèle classique basé sur un modèle mixte log-linéaire de Poisson avec un lissage spatial auto-régressif gaussien [2,3,4,8], sont développés pour des maladies non contagieuses ou non transmissibles. En effet, le nombre de cas dans une unité géographique est modélisé par une loi de Poisson qui suppose une indépendance des différents événements à l'intérieur de l'unité géographique. Cette indépendance n'est pas vérifiée dans le cas d'une maladie contagieuse où certains cas entraînent d'autres. Cependant, l'intérêt de disposer de carte de risque pour les maladies contagieuses est tel que de récentes études appliquent à des maladies transmissibles [9] les méthodes actuelles sans adaptation particulière.

Nous proposons une adaptation des méthodes spatiales actuelles de cartographie du risque pour des maladies contagieuses ou transmissibles. Notre modèle s'inspire de la

modélisation d'une hétérogénéité non spatiale proposée dans de récents articles [7,10,11]. Ce type de modèles intègre à la fois un modèle log-normal pour modéliser l'hétérogénéité spatiale et un modèle Gamma pour l'hétérogénéité non corrélée, issue dans notre cas de la contagion. Il est à noter que le mélange d'une loi de Poisson avec une loi Gamma ainsi introduit revient à utiliser une loi binomiale négative. Or la loi binomiale négative est généralement utilisée, entre autres, pour la modélisation d'une maladie contagieuse, afin de prendre en compte une sur-dispersion par rapport à la loi de Poisson.

Nous allons utiliser la formulation de la loi binomiale négative sous forme de mélange Poisson-Gamma (PG). Les nombres de cas  $Y_i$  dans chaque unité géographique  $i$  sont modélisés par des variables aléatoires de lois de Poisson [3] :

$$Y_i \sim \text{Poisson}(\lambda_i) \text{ pour } i = 1, \dots, n, \quad (1)$$

Un second niveau prend en compte la surdispersion locale, liée à la transmission, avec une loi de mélange Gamma :

$$\lambda_i \sim \text{Gamma}(r_i, 1/\theta_i). \quad (2)$$

Le mélange Poisson-Gamma correspond à  $Y_i$  issu d'une loi binomiale négative  $NB(r_i, p_i)$  avec  $p_i = 1/(\theta_i + 1)$ , La surdispersion, reliée au taux de transmission de la maladie, est contrôlée par le paramètre  $\theta_i$ . Le troisième niveau prend en compte un lien entre aires géographiques à travers un modèle mixte log-linéaire, qui, comme pour les modèles usuels de cartographie du risque [2,3,4,8], est introduit au niveau des logarithmes des risques relatifs. Nous posons

$$\log(\mu_i) = \log(e_i) + u_i + v_i, \text{ pour } i = 1, \dots, n$$

où les  $u_i$  sont modélisés avec un a priori ICAR (*intrinsic gaussian conditional autoregressive*) dont la distribution est

$$U_i | U_j, j \neq i \sim \mathcal{N} \left( \bar{u}_i, \frac{\sigma_u^2}{m_i} \right), \quad (3)$$

Un *a priori* Gamma non-informatif est choisi pour les paramètres  $\tau_u = 1/\sigma_u^2$  et  $\tau_v = 1/\sigma_v^2$ , plus précisément une loi *Gamma*(0.01, 0.01). Comme pour les modèles classiques basés sur une loi de Poisson, le lissage spatial est introduit sur la moyenne  $\mu_i$  de la distribution des données. Dans les modèles usuels,  $\mu_i$  est directement relié au paramètre de la loi de Poisson. Mais dans notre modèle,  $\mu_i$  est relié aux deux paramètres de la loi binomiale négative. Nous avons choisi de faire agir le lissage spatial sur  $\theta_i = \mu_i/r_i$ . Nous avons modélisé le paramètre  $r_i$  restant par une loi Gamma, en testant différentes valeurs pour ses paramètres.

Le logiciel Bugs (Bayesian inference Using Gibbs Sampling, voir le site <http://www.mrc-bsu.cam.ac.uk/bugs/>) a été utilisé pour l'implémentation et l'estimation de nos modèles.

Les propriétés du modèle actuel à base d'un mélange d'une loi de Poisson par une loi Gamma, c'est-à-dire faisant intervenir une loi binomiale négative, ont été étudiées par des simulations intensives, et comparés au modèle classique [2,3,4,8] basé sur une loi de Poisson. Nos simulations, ainsi que l'application sur données réelles montrent les apports de notre modèle pour la modélisation de maladies transmissibles. En particulier, il produit des estimations du risque relatif mieux lissées et plus réalistes. Les valeurs les plus extrêmes, irréalistes et difficilement interprétables, sont minorées par notre modèle binomial négatif par rapport au classique modèle de Poisson.

## Bibliographie

- [1] F. Forbes, M. Charras-Garrido, L. Azizi, S. Doyle, D. Abrial (2013). Spatial risk mapping for rare disease with hidden Markov fields and variational EM. *Annals of Applied Statistics*, 7(2), 1192-1216.
- [2] D. Abrial, D. Calavas, N. Jarrige, C. Ducrot (2005). Poultry, pig and the risk of BSE following the feed ban in France - a spatial analysis. *Veterinary Research*, 36(4), 615-628.
- [3] J. Besag, J. York, A. Mollié (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1), 1-20.
- [4] N. Best, S. Richardson, A. Thomson (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14(1), 35-59.
- [5] M. Charras-Garrido, L. Azizi, F. Forbes, S. Doyle, N. Peyrard, D. Abrial (2013). On the difficulty to delimit disease risk hot spots. *International Journal of Applied Earth Observation and Geoinformation*, 22, 99-105. doi:10.1016/j.jag.2012.04.005
- [6] M. Charras-Garrido, D. Abrial, J. De Gor, S. Dachian, N. Peyrard (2012). Classification method for disease risk mapping based on discrete hidden Markov random fields. *Biostatistics*, 13(2), 241-255. doi:10.1093/biostatistics/kxr043
- [7] S. Gschlöbl, C. Czado (2006). Modelling count data with overdispersion and spatial effects. *Statistical Papers*, 49(3), 531-552.
- [8] D. Lee (2011). A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial and Spatio-temporal Epidemiology*, 2(2), 79-89.
- [9] B.Lowe, T. Bailey, D. Stephenson, R. Graham, C. Coelho, M. Sà Carvalho, C. Barcellos (2010). Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Computers and Geosciences*, 37(3), 371-381.
- [10] T. Neyens, C. Faes, G. Molenberghs (2011). A Generalized Poisson-Gamma Model for Spatially Overdispersed Data. *Spatial and Spatio-temporal Epidemiology*, 3, 185-194.
- [11] M. Paul, S. Tavoranpanich, D. Abrial, P. Gasqui, M. Charras-Garrido, W. Thanapongtharm, X. Xiao, et al. (2010). Anthropogenic factors and the risk of highly pathogenic avian influenza H5N1: prospects from a spatial-based model. *Veterinary research*, 41(3).
- [12] L. Waller, B. Carlin, H. Xia, A. Gelfand (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92(438), 607-617.