HAL
open science

Spatial risk mapping for rare disease with hidden Markov fields and variational EM

Lamiae Azizi, Florence Forbes, Senan Doyle, Myriam Garrido, David Abrial

To cite this version:

Lamiae Azizi, Florence Forbes, Senan Doyle, Myriam Garrido, David Abrial. Spatial risk mapping for rare disease with hidden Markov fields and variational EM. [Research Report] RR-7572, INRIA. 2011. inria-00577793

HAL Id: inria-00577793

https://inria.hal.science/inria-00577793

Submitted on 18 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Spatial risk mapping for rare disease with hidden Markov fields and variational EM

Lamiae Azizi — Florence Forbes — Senan Doyle — Myriam Charras-Garrido — David Abrial

N° 7572

March 2011

Optimization, Learning and Statistical Methods

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7572--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE
INRIA
centre de recherche GRENOBLE - RHÔNE-ALPES

Spatial risk mapping for rare disease with hidden Markov fields and variational EM

Lamiae Azizi*† , Florence Forbes*, Senan Doyle* , Myriam Charras-Garrido† , David Abrial†

Theme : Optimization, Learning and Statistical Methods
Équipes-Projets Unité d'épidémiologie Animale, INRA Clermont-Ferrand-Theix et MISTIS

Rapport de recherche n° 7572 — March 2011 — 33 pages

Abstract: We recast the disease mapping issue of automatically classifying geographical units into risk classes as a clustering task using a discrete hidden Markov model and Poisson class-dependent distributions. The designed hidden Markov prior is non standard and consists of a variation of the Potts model where the interaction parameter can depend on the risk classes. The model parameters are estimated using an EM algorithm and the mean field approximation. This provides a way to face the intractability of the standard EM in this spatial context, with a computationally efficient alternative to more intensive simulation based Monte Carlo Markov Chain (MCMC) procedures. We then focus on the issue of dealing with very low risk values and small numbers of observed cases and population sizes. We address the problem of finding good initial parameter values in this context and develop a new initialization strategy appropriate for spatial Poisson mixtures in the case of not so well separated classes as encountered in animal disease risk analysis. Using both simulated and real data, we compare this strategy to other standard strategies and show that it performs well in a lot of situations.

Key-words: Classification; Discrete hidden Markov random field; Disease mapping; Poisson mixtures; Potts model; Variational EM.

* MISTIS team
† Unité d'épidémiologie Animale, INRA Clermont-Ferrand-Theix

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Variationnel EM pour la cartographie des maladies rares à l'aide de champs de Markov cachés

Résumé : Nous abordons la cartographie automatique d' unités géographiques en classes de risque comme un problème de clustering à l'aide de modèles de Markov cachés discrets et de modèles de mélange de Poisson. Le modèle de Markov caché proposé est une variante du modèle de Potts, où le paramètre d'interaction dépend des classes de risque. Afin d'estimer les paramètres du modèle, nous utilisons l'algorithme EM combiné à une approche variationnelle champ-moyen. Cette approche nous permet d'appliquer l'algorithme EM dans un cadre spatial et présente une alternative efficace aux méthodes d'estimation basées sur des simulations intensives de type Markov chain Monte Carlo (MCMC). Nous abordons également les problèmes d'initialisation, spécialement quand les taux de risque sont petits (cas des maladies animales). Nous proposons une nouvelle stratégie d'initialisation appropriée aux modèles de mélange de Poisson quand les classes sont mal séparées. Pour illustrer notre méthodolgie, nous présentons des résultats d'application sur des données épidémiologiques réelles et simulées et montrons la performance de la stratégie d'initialisation présentée en comparaison à celles utilisées usuellement.

Mots-clés : Classification, Champs aléatoires de Markov cachés discrets, Cartographie du risque, Mélanges de Poisson, Modèle de Potts, EM variationnel.

Contents

1 Introduction **4**

2 Designing hidden Markov fields for spatial disease mapping **6**
2.1 A spatial hidden structure model 7
2.2 An observation model for count data 9

3 Estimating disease maps using variational EM variants **10**
3.1 A tractable EM variant 10
3.2 The Search/Run/Select strategy 12

4 A search procedure for initializing EM **13**
4.1 Searching the risk parameters space using the EM trajectories properties 14
4.2 A full search procedure 15

5 Illustrations **16**
5.1 Simulated data sets 17
5.2 The BSE data set 20

6 Discussion **20**

1 Introduction

The analysis of the geographical variations of a disease and their representation on a map is an important step in epidemiology. The goal is to identify homogeneous regions in terms of disease risk and to gain better insights into the mechanisms underlying the spread of the disease. Traditionally, the region under study is partitioned into a number of areas on which the observed cases of a given disease are counted and compared to the population size in this area. It has long ago become clear that spatial dependencies between counts had to be taken into account when analyzing such location dependent data. The number of observed cases are usually modelled by Poisson distributions. Most statistical methods for risk mapping of aggregated data (e.g. Mollie, 1999; Richardson et al., 1995; Pascutto et al., 2000; Lawson et al., 2000) are based on a Poisson log-linear mixed model and follow the one proposed by Besag, York & Mollie (1991). The so-called BYM model introduced by Besag et al. (1991), extended by Clayton & Bernadinelli (1992) and called the convolution model by Mollie (1996), is one of the most popular approaches and used extensively in this context. This model corresponds to a Bayesian hierarchical modelling approach. It is based on an Hidden Markov Random Field (HMRF) model where the latent intrinsic risk field is modelled by a Markov field with continuous state space, namely a Gaussian Conditionally Auto-Regressive (CAR) model. Recent developments in this context concern in particular spatio-temporal mapping (Knorr-Held & Richardson, 2003; Robertson et al., 2010; Lawson & Song, 2010) and multivariate disease mapping (Knorr-Held et al., 2002; MacNab, 2010). For all these procedures, the model inference therefore results in a real-valued estimation of the risk at each location and one of the main reported limitations (e.g. by Green & Richardson, 2002) is that local discontinuities in the risk field are not modelled leading to potentially oversmoothed risk maps. Also, in some cases, coarser representations where areas with similar risk values are grouped are desirable (Abrial et al., 2005). Grouped representations have the advantage of providing clearly delimited areas for different risk levels, which is helpful for decision-makers to interpret the risk structure and determine protection measures. These areas at risk can be viewed as clusters as in (Knorr-Held & Rasser, 2000), but we prefer to interpret them as risk classes, as Green & Richardson (2002) and Alfo et al. (2009), since geographically separated areas can have similar risks and be grouped in the same class. In consequence, the classes can be less numerous than the clusters and their interpretation by decision-makers easier in terms of risk value. Using the BYM model, it is possible to derive from the output such a grouping, using either fixed risk ranges (usually difficult to choose in practice) or more automated clustering techniques (e.g. Fraley & Raftery, 2007). In any case this post-processing step is likely to be sub-optimal. By contrast, in this work, we investigate procedures that include such a risk classification.

There have been several attempts to take into account the presence of discontinuities in the spatial structure of the risk. Within hierarchical approaches, one possibility is to move the spatial dependence one level higher in the hierarchy. Green & Richardson (2002) proposed to replace the continuous risk field by a partition model involving the introduction of a finite number of risk levels and allocation variables to assign each area under study to one of these levels. Spatial dependencies are then taken into account by modelling the allocation

variables as a discrete state-space Markov field, namely a spatial Potts model. This results in a discrete HMRF modelling. The general effect is also to recast the disease mapping issue into a clustering task using spatial finite Poisson mixtures. In the same spirit, Fernandez & Green (2002) proposed another class of spatial mixture models, in which the spatial dependence is pushed yet one level higher. Of course, the higher the spatial dependencies in the hierarchy the more flexible the model but also the more difficult the parameter estimation. As regards inference, these various attempts have in common the use of MCMC techniques which can seriously limit even prevent from applying them to large data sets in a reasonable time.

Following the idea of using a discrete HMRF model for disease mapping, we propose to use for inference, as an alternative to simulation based techniques, an Expectation Maximization (EM) framework (Dempster et al., 1977; McLachlan & Peel, 2000). This framework is commonly used to solve clustering tasks (Fraley & Raftery, 2007) but leads to intractable computation when considering non trivial Markov dependencies. However, approximation techniques are available and among them, we propose to investigate variational approximations for their computational efficiency and good performance in practice. In particular, we consider the so-called mean field principle that provides a deterministic way to deal with intractable Markov Random Field (MRF) models (Celeux et al., 2003) and has proven to perform well in a number of applications, *e.g.* (Forbes et al., 2010; Vignes & Forbes, 2009; Blanchet & Forbes, 2008).

An attempt in this direction has been recently made by Alfo et al. (2009) but with a rather limited consideration for experimental validation and robustness of their setting. The approach in (Alfo et al., 2009) has been tested on a single data set regarding human heart disease. Human disease data usually has the particularity that the populations under consideration are large and the risk values relatively high. This is not fully representative of epidemiological studies, especially non contagious animal disease studies. When considering animal epidemiology, we may have to instead face low size populations and risk levels much smaller than 1, typically $1e^{-5}$ to $1e^{-3}$. Difficulties in applying techniques that work in the human case to data sets in the animal case have not been investigated. In addition, the authors in (Alfo et al., 2009) report no difficulties regarding initialization and model selection. This is far from being the case in all practical problems. In this paper we propose to go beyond the work of Alfo et al. (2009) and to address a number of related issues. More specifically, we investigate in more detail the model behavior. We pay special attention to one of the main inherent issues when using EM procedures, namely algorithm initialization. The model selection issue, *e.g.* the determination of the number of classes, is addressed using some previous work (Forbes & Peyrard, 2003) in which a mean field approximation of the Bayesian Information Criterion (BIC) is provided for HMRF models. The EM solution can depend highly on its starting position. We show that in contrast to the example in (Alfo et al., 2009), simple initializations do not always work, especially for rare disease for which the risks are small. We then propose and compare different initialization strategies in order to get a robust way of initializing for most situations arising in practice.

In addition, we build on the standard hidden Markov field model used in (Green & Richardson, 2002; Alfo et al., 2009) by considering a more general formulation that is able to encode more complex interactions than the standard

Potts model. In particular, we are able to encode the fact that risk levels in neighboring regions cannot be too different, whereas the standard Potts model penalizes neighboring risks equally, whatever the amplitude of their difference.

The paper is structured as follows. In Section 2, we present the HMRF setting and in particular a general Potts model formulation that allows us to define an instance of the model more appropriate for disease mapping than the standard version of (Green & Richardson, 2002; Alfo et al., 2009). In Section 3, we explain how disease mapping can be addressed in a more computationally efficient way using variational approximations within an EM framework for inference. Important issues related to the EM procedure are then addressed in Section 4. We propose a new efficient initialization strategy and a criterion for selecting the number of risk classes if necessary. Results are reported and compared in Section 5 on both simulated and real data sets. A discussion ends the paper in Section 6.

2 Designing hidden Markov fields for spatial disease mapping

Discrete HMRF have been widely used for a number of classification tasks. Most applications are related to image analysis but other examples include population genetics (Francois et al., 2006), bioinformatics (Vignes & Forbes, 2009), among others. In a clustering or classification context, it has the advantage to provide some insight and control on the clustering regularity through a meaningful and easy to understand parametric model. Hidden structure models and more specifically mixture models are among the most statistically mature methods of clustering. In this paper, we recast the disease mapping issue into a clustering task. A clustering or labelling problem is specified in terms of a set of sites S and a set of labels $\mathcal{L}$. A site often represents an item, a point or a region. A set of sites may be categorized in terms of their regularity. Sites on a lattice are considered as spatially regular (*e.g.* the pixels of a 2D image). Sites which do not present spatial regularity are considered as irregular. This is the usual case when sites represent geographic locations. A label is an event that may happen to a site. In our disease mapping problem, a typical event is the association of a region to a certain risk level. We will consider only the case where there are a finite number of such risk levels, *i.e.* the labels assume discrete values in a set of K labels. In the following, it is convenient to consider $\mathcal{L}$ as the set of K-dimensional indicator vectors $\mathcal{L} = \{e_1, \ldots, e_K\}$ where each e_k has all its components set to 0 except the kth which is 1.

Based on count data for a rare phenomenon observed in a predefined set S of N areas (*e.g.* geographical regions), the goal is to assign to each region a risk level among a finite set of K possible levels $\{\lambda_1, \ldots, \lambda_K\}$ when these risk levels are themselves unknown and need to be estimated. In addition, as already mentioned, we are interested in accounting for spatial dependencies between counts in various regions in order to get spatially consistent risk estimation. In general, risks are expected to be more similar in nearby areas than in areas that are far apart. The idea is to exploit the risk information from neighboring areas to provide more reliable risk estimates in each area. In each area, two values are usually available, the number y_i $(i = 1, \ldots, N)$ of observed cases of

the given disease and the population size n_i. A common assumption is that for an area indexed by $i \in S$, the number of cases y_i is a realization of a Poisson distribution whose parameter depends on the risk level assigned to the area. The unknown risk assignment and risk level compose respectively the hidden part and unknown parameters of the model.

Therefore, the data is naturally divided into observed data $\mathbf{y} = \{y_1, \ldots, y_N\}$ and unobserved or missing membership data $\mathbf{z} = \{z_1, \ldots, z_N\}$. The latter are considered as random variables denoted by $\mathbf{Z} = \{Z_1, \ldots, Z_N\}$. When the Z_i's are independent, the model reduces to a standard mixture model. When the Z_i's are not independent, the inter-relationship between sites can be maintained by a neighborhood system usually defined through a graph. Two neighboring sites correspond to two nodes of the graph linked by an edge.
For disease mapping, we usually consider the simplest possible graph structure connecting contiguous locations: regions i and j are neighbors if and only if they are spatially contiguous.

2.1 A spatial hidden structure model

The dependencies between neighboring Z_i's are then modelled by further assuming that the joint distribution of $\{Z_1, \ldots, Z_N\}$ is a discrete MRF on this specific graph:

$$P(\mathbf{z}; \beta) \quad = \quad W(\beta)^{-1} \exp(-H(\mathbf{z}; \beta)) \tag{1}$$

where β is a set of parameters, $W(\beta)$ is a normalizing constant and H is a function assumed to be of the following form (we restrict to pair-wise interactions),

$$H(\mathbf{z}; \beta) = \sum_{i \in S} V_i(z_i; \beta) + \sum_{\substack{i,j \\ i \sim j}} V_{ij}(z_i, z_j; \beta),$$

where the V_i's and V_{ij}'s are respectively functions referred to as singleton and pair-wise potentials. We write $i \sim j$ when areas i and j are neighbors, so that the second sum above is over neighboring areas.

The set of parameters β consists of two sets $\beta = (\alpha, \mathbb{B})$ where α and $\mathbb{B}$ are defined as follows. We can consider pair-wise potentials V_{ij} that depend on z_i and z_j but also possibly on i and j. Since the z_i's can only take a finite number K of values, for each i and j, we can define a $K \times K$ matrix $\mathbb{B}_{ij} = (\mathbb{B}_{ij}(k,l))_{1 \leq k,l \leq K}$ and write without loss of generality $V_{ij}(z_i, z_j; \beta) = -\mathbb{B}_{ij}(k,l)$ if $z_i = e_k$ and $z_j = e_l$. Using the indicator vector notation and denoting z_i^t the transpose of vector z_i, it is equivalent to write $V_{ij}(z_i, z_j; \beta) = -z_i^t \mathbb{B}_{ij} z_j$. This latter notation has the advantage to make sense also when the vectors are arbitrary and not necessarily indicators. This will be useful when describing the algorithms of Section 3.

Similarly we consider singleton potentials V_i that may depend on z_i and on i, so that denoting by α_i a $K-$dimensional vector, we can write $V_i(z_i, \beta) = -\alpha_i(k)$ if $z_i = e_k$, where $\alpha_i(k)$ is the kth component of α_i, or equivalently $V_i(z_i, \beta) = -z_i^t \alpha_i$. This vector α_i acts as weights for the different values of z_i. When α_i is zero, no risk level is favored, *i.e.* for a given area i, if no information on the neighboring areas is available, then all risk levels have the same probability. If in addition, for all i and j, $\mathbb{B}_{ij} = b \times I_K$ where b is a real scalar and I_K is the $K \times K$ identity matrix, parameters β reduce to a single scalar interaction

parameter b, we get the Potts model traditionally used for image segmentation. In this case, parameter b can be interpreted as a strength of interaction between neighbors. The higher b, the more weight is given to the neighbors. If b is set to 0, only the individual features are taken into account, reducing our model to the independent non spatial case.

Note that the standard Potts model is most of the time appropriate for classification since it tends to favor neighbors that are in the same class (*i.e* have the same risk level). However this model penalizes pairs that have different risk levels with the same penalty whatever the values of these risk levels. In practice, it may be more appropriate from a disease mapping point of view to encode higher penalties when the risk levels are further apart so as to model abrupt changes in the risk level of neighboring regions are undesirable. It follows that cases where the $\mathbb{B}_{ij}$'s are far from $b \times I_K$ can be useful in situations where encoding finer interactions between regions is desirable. This is part of the flexibility and modelling capabilities of the model we describe.

In practice, these parameters can be tuned according to experts, *a priori* knowledge, or they can be estimated from the data. In the first case, our model can deal with the most general parametrization, namely $\beta = \{\alpha_i, \mathbb{B}_{ij}, i, j = 1, \ldots, N\}$. In the latter case, the part to be estimated is usually assumed independent of the region indices i and j, so that in what follows we will reduce α and $\mathbb{B}$ respectively to a single vector and a single matrix. Note that formulated as such the model is not identifiable in the sense that different values of the parameters, namely $(\alpha, \mathbb{B})$ and $(\alpha + \gamma_1 \mathbb{1}_K, \mathbb{B} + \gamma_2 \mathbb{1}_{K \times K})$ lead to the same probability distribution. Notations γ_1 and γ_2 are scalars and $\mathbb{1}_K$ and $\mathbb{1}_{K \times K}$ respectively denote the K component vector and $K \times K$ matrix with all components being 1. This issue is generally easily handled by imposing some additional constraint such as $\alpha(k) = 0$ and $\mathbb{B}(k, l) = 0$ for one of the pairs (k, l).

In the disease mapping context, we propose to use for $\mathbb{B}$ a matrix with three non zero diagonals defined for some positive real value b by:

$$\begin{aligned} \mathbb{B}(k,k) &= b \quad \text{for all } k = 1, \ldots, K \\ \mathbb{B}(k,l) &= b/2 \quad \text{for all } (k,l) \text{ such that } |k-l| = 1 \\ \mathbb{B}(k,l) &= 0 \quad \text{otherwise.} \end{aligned} \tag{2}$$

The idea is to favor first, neighbors in the same risk class and then neighbors in risk classes that are close. All other pairs of risk classes being equally weighted. This is the simplest non standard $\mathbb{B}$ structure that can encode smooth variations in the risk level.

The difficulty with MRF models is that the computation of $P(\mathbf{z}; \beta)$ in (1) is not possible due to the normalizing constant $W(\beta)$ which involves K^N terms and is intractable except in some trivial cases. By contrast the conditional probability $p(z_i \mid \mathbf{z}_{\mathcal{N}(i)}; \beta)$, where $\mathbf{z}_{\mathcal{N}(i)}$ denotes the values $\{z_j, j \in \mathcal{N}(i)\}$ associated to the set $\mathcal{N}(i)$ of neighbors of i, is easily computed using the following formula:

$$P(z_i \mid \mathbf{z}_{\mathcal{N}(i)}; \beta) = \frac{\exp(z_i^t(\alpha + \mathbb{B} \sum_{j \in \mathcal{N}(i)} z_j))}{\sum_{k=1}^{K} \exp(\alpha(k) + e_k^t \mathbb{B} \sum_{j \in \mathcal{N}(i)} z_j)} . \tag{3}$$

2.2 An observation model for count data

The previous lines define the hidden structure of the model. For the model to be fully defined, the observation model needs also to be specified. The class dependent distribution $P(y_i|z_i)$ is usually a standard distribution, typically in rare disease mapping, a Poisson distribution $\mathcal{P}(y_i;\ n_i\ \lambda_{z_i})$, where n_i is the population size in area i and the z_i subscript in λ_{z_i} indicates that the distribution parameter depends on the specific value of z_i which determines the level of the risk and thus the risk value among a finite number of them $\{\lambda_1, \ldots, \lambda_K\}$. With our vectorial notation, we will write $z_i^t\lambda$ for λ_{z_i} with $\lambda = [\lambda_1, \ldots, \lambda_K]^t$. In the one dimensional Poisson case, this corresponds to:

$$\begin{aligned} P(Y_i = y_i|Z_i = z_i; \lambda) &= \mathcal{P}(y_i;\ n_i\ z_i^t\lambda) \\ &= \exp(-n_i\ z_i^t\lambda)\ \frac{(n_i\ z_i^t\lambda)^{y_i}}{y_i!} \end{aligned} \tag{4}$$

Note that in practice, epidemiologists usually prefer to consider relative risks rather than absolute risks. Relative risk correspond to the ratio between a local and an overall risk. However, in the case of a unique population without any structure, the use of relative risk is equivalent to the use of absolute risk.

For the distribution of the observed variables $\mathbf{y}$ given the classification $\mathbf{z}$, the usual conditional independence assumption leads to:

$$P(\mathbf{Y} = \mathbf{y}|\mathbf{Z} = \mathbf{z}; \lambda) = \prod_{i \in S} \mathcal{P}(y_i; n_i\ z_i^t\lambda).$$

It follows that the energy of the hidden field $\mathbf{z}$ given the observed field $\mathbf{y}$ is:

$$H(\mathbf{z}|\mathbf{y}; \lambda, \beta) = H(\mathbf{z}; \beta) - \sum_{i \in S} \log \mathcal{P}(y_i; n_i\ z_i^t\lambda),$$

and its conditional probability is:

$$P(\mathbf{z}|\mathbf{y}; \lambda, \beta) = W(\beta)^{-1} \exp(-H(\mathbf{z}; \beta) + \sum_{i \in S} \log \mathcal{P}(y_i; n_i\ z_i^t\lambda)).$$

The unknown parameters (to be estimated) of this model are then denoted by $\mathbf{\Psi} = (\lambda, \alpha, \mathbb{B})$.

In labelling problems, most approaches fall into two categories. The first ones focus on finding the best $\mathbf{z}$ using a Bayesian decision principle such as Maximum A Posteriori (MAP) or Maximum Posterior Marginal (MPM) rules. This explicitly involves the use of $P(\mathbf{z}|\mathbf{y})$ and uses the fact that the conditional field denoted by $\mathbf{Z}|\mathbf{Y} = \mathbf{y}$ is a Markov field. This includes methods such as ICM (Besag, 1986) and Simulated Annealing (Geman & Geman, 1984) which differ in the way they deal with the intractable $P(\mathbf{z}|\mathbf{y})$ and use its Markovianity. A second type of approaches is related to a missing data point of view. Originally, the focus is on estimating parameters using a maximum likelihood principle when some of the data are missing (the z_i's here). The reference algorithm in such cases is the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). In addition to providing estimates of the parameters, the EM algorithm provides also a classification $\mathbf{z}$ by offering the possibility to restore the missing data. However, when applied to HMRFs, the algorithm is not tractable and requires

approximations. It follows variations such as the Gibbsian EM of Chalmond (1989), the MCEM algorithm and a generalization of it (Qian & Titterington, 1991), the PPL-EM algorithm of Qian & Titterington (1991) and various Mean Field like approximations of EM (Celeux et al., 2003). We focus on the latter in the next section for their computational efficiency and good results in practice.

3 Estimating disease maps using variational EM variants

In disease mapping, the question of interest is to recover the unknown map $\mathbf{z}$ interpreted as a classification into a finite number K of labels representing K risk classes. These labels are unobserved and have to be considered as missing data. Our aim is therefore to classify each region in one of the K risk classes. To do so, we consider a MPM principle consisting in assigning each region i to the class e_k that maximizes $P(Z_i = e_k|\mathbf{y};\mathbf{\Psi})$. Such maximizations depend on the model parameters $\mathbf{\Psi}$ which is usually unknown (or partly unknown when prior knowledge can be incorporated) and has to be estimated. In this paper, to deal with the spatial dependence structure, we use the EM algorithm with some of the approximations presented in (Celeux et al., 2003). These approximations are based on the mean field principle which consists in replacing the intractable Markov distributions by factorized ones for which the exact EM can be carried out. It corresponds to one of the simplest variational approximations and allows to take the Markovian structure into account while preserving the good features of EM. The work in (Celeux et al., 2003) generalizes the mean field principle and introduces different factorized models resulting in different procedures. Note that in practice, these algorithms have to be extended to incorporate the estimation of matrix $\mathbb{B}$ and to include irregular neighborhood structure coming from arbitrary graphs and not from regular pixel grids like in image analysis.

3.1 A tractable EM variant

Briefly, these algorithms can be presented as follow. They are based on the EM algorithm which is an iterative procedure aiming at maximizing the log-likelihood (for the observed variables $\mathbf{y}$) of the model by maximizing at each iteration the expectation of the complete log-likelihood (for the observed and hidden variables $\mathbf{y}$ and $\mathbf{z}$) knowing the data and a current estimate of the model parameters. When the model is an Hidden Markov Model with parameters $\mathbf{\Psi}$, there are two difficulties in evaluating this expectation. Both the normalizing constant $W(\beta)$ in (1) and the conditional probabilities $P(z_i \mid \mathbf{y};\mathbf{\Psi})$ and $P(z_i, z_j \mid \mathbf{y};\mathbf{\Psi})$ for j in the neighborhood $\mathcal{N}(i)$ of i, cannot be computed exactly.

Informally, the mean field approach consists in approximating the intractable probabilities by neglecting fluctuations from the mean in the neighborhood of each region i. This is obtained by assuming the neighboring z_j for j in $\mathcal{N}(i)$, fixed to their mean values. More generally, we talk about mean field-like approximations when the value z_i does not depend on the values z_j for $j \neq i$ which are all set to constants (not necessarily to the means) independently of the value of z_i. These constant values denoted by $\tilde{z} = \{\tilde{z}_1, \ldots, \tilde{z}_N\}$ are not arbitrary but satisfy some appropriate consistency conditions (see Celeux et al., 2003). It

follows that $P(z_i \mid \mathbf{y}; \mathbf{\Psi})$ is approximated by

$$\begin{aligned} P(z_i \mid \mathbf{y}, \tilde{z}_{\mathcal{N}(i)}; \mathbf{\Psi}) &\propto \mathcal{P}(y_i; n_i\ z_i^t \lambda)\ P(z_i | \tilde{z}_{\mathcal{N}(i)}; \beta) \\ &\propto \mathcal{P}(y_i; n_i\ z_i^t \lambda)\ \exp(z_i^t(\alpha + \mathbb{B} \sum_{j \in \mathcal{N}(i)} \tilde{z}_j)). \end{aligned} \tag{5}$$

The normalizing constant is not specified but its computation is not an issue as it involves only a sum over K terms. Then, for all $j \in \mathcal{N}(i)$, $P(z_i, z_j \mid \mathbf{y}; \mathbf{\Psi})$ is approximated by $P(z_i \mid \mathbf{y}, \tilde{z}_{\mathcal{N}(i)}; \mathbf{\Psi})\ P(z_j \mid \mathbf{y}, \tilde{z}_{\mathcal{N}(j)}; \mathbf{\Psi})$. Both approximations are easy to compute. Using such approximations leads to algorithms which in their general form consist in repeating the following two steps. At iteration q,

(1) Create from the data $\mathbf{y}$ and some current parameter estimates $\mathbf{\Psi}^{(q-1)}$ a configuration $\tilde{z}^{(q)} = \{\tilde{z}_1^{(q)}, \ldots, \tilde{z}_N^{(q)}\}$, *i.e.* values for the Z_i's that satisfy some consistency conditions (see Celeux et al., 2003). Replace the Markov distribution $P(\mathbf{z}; \beta)$ of (1) by the factorized distribution $\prod_{i \in S} P(z_i | \tilde{z}_{\mathcal{N}(i)}^{(q)}; \beta)$ using expression (3). It follows that the joint distribution $P(\mathbf{y}, \mathbf{z}; \mathbf{\Psi})$ can also be approximated by a factorized distribution:

$$\prod_{i \in S} \mathcal{P}(y_i;\ n_i\ z_i^t \lambda)\ P(z_i | \tilde{z}_{\mathcal{N}(i)}^{(q)}; \beta)$$

and the two problems encountered when considering the EM algorithm with the exact joint distribution disappear. The second step is therefore,

(2) Apply the EM algorithm for this factorized model with starting values $\mathbf{\Psi}^{(q-1)}$, to get updated estimates $\mathbf{\Psi}^{(q)}$ of the parameters.

In particular the so-called *mean field* and *simulated field* algorithms perform step **(1)** in two different ways. They correspond to two different sets of consistency conditions that can be satisfied by iterating some straightforward equations. The *mean field* algorithm consists of updating the $\tilde{z}_i^{(q)}$'s by setting, for all $i = 1, \ldots, N$, $\tilde{z}_i^{(q)}$ to the mean of distribution $P(z_i \mid \mathbf{y}, \tilde{z}_{\mathcal{N}(i)}^{(q)}; \mathbf{\Psi}^{(q-1)})$. Note that as z_i is an indicator vector, the mean value $\tilde{z}_i^{(q)}$ is a vector made of the respective probabilities to be in each of the K classes. In the *simulated field* algorithm, $\tilde{z}_i^{(q)}$ is simulated from $P(z_i \mid \mathbf{y}, \tilde{z}_{\mathcal{N}(i)}^{(q)}; \mathbf{\Psi}^{(q-1)})$. Note also that to save additional notation, the updating described above is synchronous while we actually implemented a sequential updating of the $\tilde{z}_i^{(q)}$'s: each site i is updated in turn using the new values of the other sites as soon as they become available rather than waiting until all sites have been updated. Step **(1)** can then be iterated a number of times to guarantee that the corresponding consistency equations are satisfied. In practice the fixed point is reached in few iterations. Similarly, in step **(2)**, performing a small number of EM iterations is usually enough.The approximate EM step decomposes to the following E and M steps:

The **E-step** consists in computing, using equation (5), the approximate posteriors denoted by $\tilde{t}_{ik}^{(q)}$:

for all i and k, $\quad \tilde{t}_{ik}^{(q)} = P(Z_i = e_k | \mathbf{y}, \tilde{z}_{\mathcal{N}(i)}^{(q)}; \mathbf{\Psi}^{(q-1)})$.

The **M-step** consists of updating the parameters. The risk updates are available in closed-form (equation (6) below) while the MRF prior parameters need to be computed numerically (equation (7)):

$$\text{for all } k, \quad \lambda_k^{(q)} = \frac{\sum_{i \in S} \tilde{t}_{ik}^{(q)} \, y_i}{\sum_{i \in S} n_i \, y_i}, \tag{6}$$

and

$$\beta^{(q)} = \arg\max_{\beta} \sum_{i \in S} \sum_{k=1}^{K} \tilde{t}_{ik}^{(q)} \log \tilde{\pi}_{ik}^{(q)}(\beta), \tag{7}$$

where $\tilde{\pi}_{ik}^{(q)}(\beta) = P(Z_i = e_k | \tilde{z}_{\mathcal{N}(i)}^{(q)}; \beta)$.

Computing the gradient in (7) with $\beta = (\alpha, \mathbb{B})$ and $\mathbb{B}$ of the form specified in equation (2), the solutions α and b of (7) satisfy:

$$\text{for } k = 1, \ldots, K, \quad \sum_{i \in S} \tilde{\pi}_{ik}^{(q)}(\alpha, b) = \sum_{i \in S} \tilde{t}_{ik}^{(q)},$$

and,

$$\sum_{i \in S} \sum_{k=1}^{K} \left(\tilde{t}_{ik}^{(q)} - \tilde{\pi}_{ik}^{(q)}(\alpha, b) \right) \left(\sum_{j \in \mathcal{N}(i)} \tilde{z}_j^{(q)}(k) + \tfrac{1}{2} \left(\tilde{z}_j^{(q)}(k-1) + \tilde{z}_j^{(q)}(k+1) \right) \right) = 0$$

The HMRF estimation provides us with estimations for the risk values λ_k, for $k = 1, \ldots, K$, but also for the hidden field parameters, *i.e* matrix $\mathbb{B}$ and vector α. Once the parameters are estimated, approximations of the $P(Z_i = e_k | \mathbf{y}; \mathbf{\Psi})$ values required to classify each region using the MPM principle can be calculated. The area i is assigned to the class k for which the later posterior probability is the highest. Note that when using the mean field approximation principle, MAP and MPM lead to the same solution due to the product form of the approximation in step **(1)** above.

3.2 The Search/Run/Select strategy

The likelihood function to be maximized generally possesses many stationary points of different natures (including local maxima and minima). Consequently, convergence to the global maximum with the EM algorithm may strongly depend on the starting parameters values. To anticipate this initialization issue, we cast the EM algorithm above into a more general procedure. As in (Biernacki et al., 2003), we adopt a three stage Search/Run/Select strategy whose goal is to identify in a reasonable amount of time the highest likelihood:

Search. Build a search method for generating M sets of initial parameters values. These sets can be either generated at random or using some initialization strategy (see Section 4).

Run. For each initial position from the search step, run the variational EM algorithm described in Section 3.1 until a stopping criterion (*e.g.* a fixed number of iterations or stabilization of the log-likelihood) is satisfied. As

specified in Section 3.1, two variantes of the algorithm are possible depending on the approximation chosen in step **(1)**.

Select. Select the set of estimated parameter values that provides the highest likelihood among the M trials. Build from this selected solution the risk map as specify at the end of Section 3.1.

In the next Section, we focus on the Search step and describe the initialization strategy we propose for a more efficient exploration of the parameter space.

4 A search procedure for initializing EM

The sensitivity of the EM algorithm to starting values is a well documented issue. It is particularly critical and very likely to strike when classes are not well separated. To overcome this sensitivity, different initialization strategies have been proposed and investigated in the context of independent Gaussian mixtures (see for instance McLachlan & Peel (2000) chap.2) or Biernacki et al. (2003)). For Gaussian mixtures, Biernacki et al. (2003) propose a strategy based on randomly drawing the mixture model parameter values a number of times to provide random initializations of either Classification EM (CEM), Stochastic EM (SEM) or EM itself stopped after a fixed number of iterations. The resulting set of parameter values leading to the highest likelihood is then selected and the EM algorithm ran until convergence starting from this selected position. Other initialization strategies have been investigated in (Karlis & Xekalaki, 2003) for both Gaussian and Poisson mixtures leading also to the conclusion that it was advisable to start from several different initial values to ensure more reliable results.

More generally, most strategies can be divided into two categories: the ones based on initial parameter values and the ones based on an initial partition of the data. As the EM goal is primary to estimate parameters, it produces a sequence of parameter estimates and it is therefore natural to initialize the algorithm with initial parameter values. However, as mentioned in (Biernacki, 2004) for the Gaussian case, at each iteration of the algorithm, the E and M steps produce estimated values that are not arbitrary but linked through some equations. The sequence of such estimates corresponds then to an EM trajectory in the parameter space. Then, parameter values randomly drawn do not necessarily belong to an EM trajectory and this can result in computationally inefficient strategies as the maximum likelihood solution necessarily belongs to one of the possible EM trajectories. In the case of standard Gaussian mixtures in a non spatial context, Biernacki (2004) shows that all EM trajectories are included in a space defined by two equations linking the parameter estimates. An initialization strategy is then derived and showed in (Biernacki, 2004) to be more efficient than exploring widely the parameter space without any account of the EM trajectories.

A second type of strategies for choosing starting values is to use random partitions of the data into K groups and then to compute for each group, the parameter estimates, here the risk level as the ratio of the observed counts in the group over the population size of the group. These values are then used as starting values. They are by construction in the EM trajectory space but they

tend to provide values close to each others and then not to explore the space efficiently (see Figure 1 (d) for an illustration).

In the disease mapping context, the initialization issue has not really been addressed. In this context, Alfo et al in (Alfo et al., 2009) use 500 runs of a short-length CEM algorithm before their approximate EM. They report satisfying results with this initialization procedure although it is mentioned in (Biernacki et al., 2003) that this choice is generally not a stable strategy because CEM is actually more sensitive to the starting value than EM itself. We suspect then that the example in (Alfo et al., 2009) is so that there is no real initialization problem and any strategy would provide a satisfying solution. In this paper, we address initialization following the EM trajectory-based idea developed in (Biernacki, 2004) for Gaussian mixtures in a non spatial case. We show that it is possible to derive, for our spatial Poisson mixture, constraints on the parameter estimates to ensure that they belong to an EM trajectory. We propose this way a new initializing strategy and define our *Search* procedure.

4.1 Searching the risk parameters space using the EM trajectories properties

Compared to the work in (Biernacki, 2004), our task is complicated by the addition of a spatial Markov prior whose parameters need also to be initialized. Our first approach is then to focus on the initialization of the risks or Poisson distributions parameters λ. It is interesting to note that whatever the model for the spatial prior an equation similar to that in (Biernacki, 2004) can be found that links the λ_k's values. Let $n = \sum_{i \in S} n_i$ be the total population size. At each iteration q, we denote by $n_k^{(q)}$ the quantity:

$$n_k^{(q)} = \frac{\sum_{i \in S} \tilde{t}_{ik}^{(q)} n_i}{n},$$

which can be interpreted as the proportion of the population in the kth risk level. It follows easily that $\sum_{k=1}^{K} n_k^{(q)} = 1$. Using then equation (6) for the current risk level estimations, it comes,

$$\sum_{k=1}^{K} n_k^{(q)} \lambda_k^{(q)} = \frac{\sum_{i \in S} y_i}{n} = \bar{\lambda}. \tag{8}$$

$\bar{\lambda}$ can be interpreted as an average risk and has the property to depend on the observed data only. At each iteration of the algorithm, the current parameter estimates $\lambda_k^{(q)}$ satisfy this equation. Consequently all EM trajectories are included in the space defined by this equation. The idea is then to produce values for the λ_k's by sampling in this space. A simple way to achieve this is to follow the simulation steps below:

Step 1. Values for the $n_k^{(0)}$'s are first drawn using a Dirichlet distribution $\mathcal{D}(\pi, \ldots, \pi)$ with $\pi = 1$ for a uniform sampling on the space defined by $\sum_{k=1}^{K} n_k^{(0)} = 1$.

Step 2. Then k is chosen at random in the set $\{1, \ldots, K\}$ and the $\lambda_l^{(0)}$'s for $l \neq k$ are drawn uniformly and without replication in the sample $\{\frac{y_1}{n_1}, \ldots, \frac{y_N}{n_N}\}$. The last $\lambda_k^{(0)}$ is set to verify:

$$\lambda_k^{(0)} = \frac{\bar{\lambda} - \sum_{l \neq k} n_k^{(0)} \lambda_l^{(0)}}{n_k^{(0)}}. \tag{9}$$

These steps generate a vector of random parameters. The number of initial values generated this way is set by the user. Note however, that as in (Biernacki, 2004) for Gaussian parameters, the later equation (9) in step 2 does not guarantee that $\lambda_k^{(0)}$ is strictly positive. If this is not the case, the simulated sample is discarded and the procedure restarted from step 1.

To illustrate the proposed strategy and the differences with other initializations using random parameter values or random partitions, we consider a simple non spatial two-class case where the true parameter values are $\lambda_1 = 0.1$, $\lambda_2 = 0.2$ and the proportions for the two classes are $\pi_1 = \pi_2 = 0.5$. We consider a hundred of sites $N = 100$ and create values for the population sizes n_i by sampling at random with replication among the integers between 10 and 109. We then simulate the y_i's from a two component Poisson mixture model using equation (4). The histogram of the y_i's is shown in Figure 1 (a). The corresponding $\bar{\lambda}$ value is around 0.141.

In the two-class case, it follows from equation (8) that the EM trajectories and therefore the maximum likelihood estimates of λ_1 and λ_2 are constrained to live in the greyed area of Figure 1 (b) delimited by horizontal lines $\lambda_1 = \bar{\lambda}$ and $\lambda_2 = \bar{\lambda}$. The latter Figure also shows 100 initial values of λ_1 and λ_2 drawn uniformly at random between 0 and 0.4 and ordered so that $\lambda_1 < \lambda_2$. Among the 100 points, only 42 lie within the EM trajectory area. The point corresponding to the true values $\lambda_1 = 0.1$ and $\lambda_2 = 0.2$ is marked with an "X". For comparison, we show respectively in Figures 1 (c) and (d), 100 values obtained by using our strategy and starting from 100 random partitions. As expected, random partition initializations tend to produce values close to $\bar{\lambda}$ and fail in exploring the space efficiently. By contrast, our strategy better explores the grey area and in particular around the true values (0.1, 0.2). The phenomenon is even more striking as we increase the number of simulated values for initialization.

In Figure 1 (e), (f) and (g), we then show the (λ_1, λ_2) values obtained after 1 iteration of EM starting from the initial values obtained with the three initializing strategies. To better visualize the differences on the plots, we show only 20 such values. They lie all within the grey area but only our proposed strategy tends to produce values near the true values. Random partition values are somewhat clustered at the wrong place and values from random parameter values are more dispersed. Among the 20 points, more points are likely not to lead close to the desired solution.

4.2 A full search procedure

To complete our *Search* procedure, we need in addition to initial λ values, starting positions for the Markov prior parameters $\beta = (\alpha, \mathbb{B})$. When $\mathbb{B}$ reduced to some value b (definition (2)), our full *Search* procedure decomposes then in two steps as follows:

Search 1. Generate M starting values $\lambda^{(0)}$ using the two steps above.

Search 2. For each initial value $\lambda^{(0)}$, set $\alpha^{(0)} = 0$, $b^{(0)} = 1$ and run our variational EM until the chosen stopping criterion is satisfied, with b kept fixed to its initial value. Only λ and α are updated. We propose to use a stopping criterion based on a relative change in log-likelihood. This criterion means that the two EM steps are repeated until the relative increase in log likelihood between two consecutive iterations is smaller than an accuracy threshold ϵ.

The idea in adding the second step is to prevent undesirable behavior of the algorithm in the case of complex or very noisy data. We observed in our real data sets (see Section 5) that imposing a certain amount of spatial structure first could help to avoid converging to meaningless solutions. This is typically done by fixing $b = 1$ for a number of iterations before letting all the parameters free. This strategy is a simple solution we found to deal with very low risk values and badly separated Poisson mixture components. In simpler better separated cases, the Search 2 step is usually not necessary. Note that in the disease mapping context, talking about Poisson mixtures is a shortcut due to the introduction of the population sizes n_i in equation (4). This modification makes the intuition on these models not as straightforward as for traditional mixtures.

A typical sequence of λ and b values obtained this way is shown in the Supplementary Materials section,"Extra Figure 6", for a realistic example closed to the real data to be considered in Section 5. In this Figure, we show also the parameters values obtained with the same initialization (resulting from **Search 1**) but skipping the **Search 2** stage and running directly our EM algorithm letting all parameters free. We observe clearly that without the **Search 2** stage, the b parameter is rapidly increasing to very high values which tends to trap the algorithm into some meaningless solution with a too high spatial interaction. Also the α values not shown here stay between -4 and 1 when **Search 2** is included while in the other case, the $\alpha(1)$ value is decreasing rapidly to a very low negative value as a compensation for the corresponding high value of b. Note that this pathological behavior is due to the very low risk values used in this example, which can be observed in animal epidemiology.

5 Illustrations

Our goal is to address the analysis of typical rare animal disease data for which the observed cases and the risk values may be very low typically less than 10 cases among a small population size of few hundreds. A real example of such a data set is given in Section 5.2 with observations corresponding to the Bovin Spongiform Encephalopathy (BSE) epidemic in France. To assess the performance of our procedure, we first carry out tests on simulated data with properties similar to the targeted real data. In our illustrations (both simulated and real data), the underlying structure is derived from the French territory. France is divided into 1264 hexagons each of width 23km ($450km^2$). The neighborhood structure is based on adjacent hexagons. For each hexagon, the population size n_i is set to the corresponding cattle population in France in years $2001 - 2005$: the n_i's vary from 1 to 32039. The cattle population map and the histogram of the population size are shown in Figure 2 (a) and (fb). We consider then

different count data $\mathbf{y}$ depending whether they are generated (Section 5.1) or observed (Section 5.2).

For comparison we consider three different strategies to provide an estimation of the unknown parameters and mapping into regions of homogeneous risk levels. Two of them, denoted respectively by S_{tra} and S_{rand} correspond to Search/Run/Select decompositions as introduced in Section 3.2. A third one denoted by S_{EMM} represent commonly used strategies when dealing with initialization issues. In particular, it is close to the strategy used in (Alfo et al., 2009). More specifically:

S_{tra} **strategy:** M initial values for all the parameters are generated using the EM trajectory properties and the full search procedure described in Section 4.2. Our variational EM is then run for each parameter set until convergence and the parameter values corresponding to the highest likelihood are selected.

S_{rand} **strategy:** This strategy differs from the previous one only in the way the M initial values are generated. M initial values for λ are generated uniformly at random (typically between 0 and 1.5 in our disease mapping context). The α and b values are respectively initialized to the null vector and to 1. Our variational EM is then run for each parameter set until convergence and the parameter values corresponding to the highest likelihood are selected.

S_{EMM} **strategy:** M initial values for λ are generated uniformly at random, α is initialized to the null vector (equal proportions) and b is fixed to 0 (non spatial case). The standard EM algorithm with no spatial interaction is run until convergence for each parameter set. The estimate parameter values with the highest likelihood are then selected and used as initial values for our variational EM with spatial interaction.

It is not completely fair to compare the last strategy with the first two due to the fact that only one run of the spatial variational EM algorithm is used for the former. However, we include it in our comparison because it is similar to the strategy use in (Alfo et al., 2009). In this latter paper, the only difference is that the non spatial EM is replaced by a non spatial CEM (Classification EM). We rather use EM since CEM is known to be even less stable than EM with respect to initialization (Biernacki et al., 2003).

Regarding variational EM, we investigated both the so-called *Mean Field* and *Simulated Field* variants. In contrast to some other studies (Celeux et al., 2003), we observed that for the type of data sets under consideration, the *Mean Field* algorithm was providing better and more stable results. This is probably due to the fact that this variant tends to smooth more the data, which is here an advantage to better recover the spatial structure. In the following sections, results are then reported only for the *Mean Field* algorithm.

5.1 Simulated data sets

Typical simulated examples with 3 and 5 classes. We consider two synthetic risk maps with respectively 3 and 5 risk classes (See Figure 2 (c) and (e)). In the 3-class case, risk levels are set to $\lambda_1 = 1e^{-5}$, $\lambda_2 = 1e^{-4}$ and $\lambda_3 = 1e^{-3}$,

which correspond in epidemiological terms to *low, medium* and *high* risk levels. In the 5-class case, the risks are set to $\lambda_1 = 1e^{-5}$, $\lambda_2 = 5e^{-5}$, $\lambda_3 = 1e^{-4}$, $\lambda_4 = 5e^{-4}$ and $\lambda_5 = 1e^{-3}$ corresponding to *very low, low, medium, high* and *very high* risk levels. From the population counts (n_i's), the *true* risk values above, the known classes (maps of Figure 2 (c) and(e)), we can easily simulate the counts y_i's from the Poisson distribution in (4). Examples of such counts are shown in Figure 2 (d) and (f). Figures 3 and 4((a):(c)) show the corresponding classifications obtained with the three strategies S_{tra}, S_{rand} and S_{EMM} with $M = 1000$ assuming $K = 3$ and $K = 5$ respectively. The classification obtained with the BYM method (Mollie & Richardson, 1991) is also reported (Figures 3 and 4(d) respectively). The performance is evaluated considering both classification performance and risk value estimation. For classification performance, we consider for each class the proportion of hexagons which are correctly assigned. More specifically, we report the ratio between the number of hexagons assigned to the true class and the number of hexagons in the estimated class. In the Supplementary Materials section, "Extra Tables 2 and 3" show these proportions and the estimated risk values.

For both the 3 and 5 class examples, the BYM model is clearly not providing satisfying mapping. In particular, the highest risk regions are found in regions with very few cattle population (*e.g.* South-East of France). In terms of risk estimation, BYM tends to overestimate risk levels, especially the lowest ones. For high risks the overestimation is not as large but the corresponding regions are not properly identified.

In the 3 class case ("Extra Table 2"), all strategies give reasonable results, for the high and medium risk regions, both in terms of estimation and classification. The main differences are observed for the low risk region. Our proposed strategy S_{tra} performs better than S_{EMM} at estimating the low risk value. It is also better although comparable to the S_{rand} strategy. In terms of mapping, S_{rand} and S_{tra} clearly outperform S_{EMM}. S_{rand} result looks visually better but in terms of classification rates (Table 1) this is the case only for the low risk region.

In the 5 class case ("Extra Table 3"), all strategies have trouble separating the low and very low risk regions and tend to loose a class. For the S_{EMM} strategy we can visualize the 5 classes but this is due to the division of the true high risk region into two classes which correspond to almost the same risk values (4.99 e^{-04} and $5.14e^{-04}$ for high risk). The low risk region is not better identified in this case and two classes are separated although they correspond to the same risk value. We will see in what follows (Table 1) that this seems to be a tendency of the S_{EMM} strategy. In terms of classification the S_{tra} strategy outperforms the other strategies for the high and very high risk regions, which correspond to the risk levels of importance in epidemiology.

To emphasize the difference between the S_{tra} and S_{rand} strategies, we consider the same 5-class data set but reduce the number of starting values to $M = 10$. This can typically be necessary if the time or computational resources need to be limited. As mentioned in Section 4 and illustrated in Figure 1 ((b):(g)), this should benefit to our S_{tra} strategy which is more efficient in exploring the parameter space and in finding good initializations. Indeed, we observe more satisfying mapping results (see Figure 4 ((e):(g)) with S_{tra} than with S_{rand} for similar estimations of the different risk levels. S_{tra} is clearly better at identifying the very high risk regions but also in this case the low risk ones. S_{EMM} is clearly providing less satisfying results in this case.

Intensive simulation study. The very low values of the risk levels induce some difficulty in interpreting the results and is responsible for some instability. To further investigate the algorithms, we repeat the simulations above a hundred times with the same true risk values. For the true values of K, the performance is then evaluated considering both average classification performance and risk value estimation. Table 1 shows for the 3 and 5 class cases, the mean and standard deviation of the proportions of well classified hexagons for the 100 simulated data sets. It also shows the mean and standard deviation of the risk values.

For the 3-class example, the average estimation of the risk values is in general close to the real parameter values for the three strategies. However, S_{EMM} tends to overestimate low risks. S_{tra} and S_{rand} give similar results. In terms of proportions of well classified hexagons, S_{rand} outperforms S_{tra} on average and shows smaller variances. However the boxplots of "Extra Figure 7 ((a),(b) and (c))" show that the median risk values are very close for both strategies.

In the 5-class case, S_{tra} provides better average risks for the high and very high risk values. In terms of proportions of well classified hexagons, the variance is generally lower for S_{tra}. For estimated risks, it is also the case for medium to very high risks. The S_{tra} strategy seems to provide better and more stable results for higher risk values, which is a desirable feature in epidemiology. The relative errors $RE = \frac{|\lambda - \hat{\lambda}|}{\lambda}$ plotted in "Extra Figure 7 ((d),(e),(f))", show that this is generally compensated by a worse estimation of the medium risk class compared to S_{rand}. For the 3-class case, we can notice that the estimator $\hat{\lambda}$ predicts observations of the parameter λ with a good accuracy. However, the relative errors show that the estimation of parameters associated to highest risk region is much more precise than for lowest risk. For the 5-class case, larger variations are observed for the class which disappears in general (λ_3). The estimation is also more precise for higher risks than for the lowest ones.

We then also consider the issue of selecting the right number of classes. In this case K is not fixed. For each simulations in the 3-class case, we run our algorithm with the S_{tra} initialization, for $K = 2$ and $K = 3$. We observed that for $K \geq 4$, the algorithm systematically looses a class and these values of K are never selected. We then used the mean field approximation of the Bayesian Information Criterion (BIC) as described in (Forbes & Peyrard, 2003) for hidden Markov models, to select K among $K = 2$ or 3. For the S_{tra} strategy, it follows that among the 100 simulations, $K = 3$ was selected 75 times and $K = 2$ 25 times. Similarly in the 5-class example, we used the approximated BIC to select a value of K from 2 to 7. $K = 5$ was selected 46 times, $K = 4$ was selected 42 times, $K = 3$ was selected 12 times and $K = 2, 6, 7$ were never selected. Similar results were observed for the other strategies. It confirms especially in the 5-class case that the data we have to deal with do not correspond to an easy well-separated case.

Overall, we observe that the three strategies are recovering more easily the high risk regions than the low risk regions. In general, when classes disappear, they correspond to the regions of lowest risks. The S_{tra} strategy performs satisfyingly compare to other strategies. In particular, the proportions of correctly allocated hexagons is improved. Also, with a limit amount of computational

resource, the S_{tra} is more likely to provide satisfying results with a better exploration of the parameter space.

5.2 The BSE data set

The Bovine Spongiform Encephalopathy is a non contagious neurodegenerative disease in cattle. This sudden and unexpected disease (Anderson et al., 1996; Ducrot et al., 2008) threatened the bovine production in Europe and has been intensively studied (Abrial et al., 2005; Allepuz et al., 2007; Paul et al., 2007) for spatial analysis. In our data set, the numbers of observed cases are available for each hexagonal geographical units in France. These cases occurred between July $1, 2001$ and December $31, 2005$. Figure 5(a) shows the corresponding observed map. We first apply our model initialized with the S_{tra} strategy described in Section 4. Regarding the number of classes, the approximated BIC of Forbes & Peyrard (2003) suggests to select $K = 3$. For comparison, we also consider the BYM model widely used in epidemiology. Since this model only provides continuous estimated values for the risk level in each hexagon, some additional post-processing is required to obtain the mapping into a prescribed number K of risk levels. Such a mapping can be obtained by applying some clustering procedure on the estimated continuous values. A commonly used method for that is the EM algorithm for Gaussian mixtures. Figure 5((b) and (c)) shows the mapping obtained with the BYM model and our model.

When comparing the two maps obtained with the expert knowledge related to the BSE disease in France, the result from our model appears to be very satisfying. Indeed, three regions are clearly delimited and correspond to the regions expected by the experts. Indeed in the BSE case it is known that high risks regions are located in Brittany, in the center, in the South-West of France and in the Alps. When studying diseases, this ability to recover accurately high risk regions is an essential feature as it is important to clearly identify the regions were important and quick decisions have to be taken.

In the BYM map, additional high risk regions are highlighted but with boundaries that are more doubtful, sometimes including too few hexagons or including regions known for low risk. Typically, the Alps region (known as high risk) is not clearly identified but merged with the South-East region (known as low risk). Moreover, the French Riviera appears as a higher risk region than the Alps and the South-West although it is known that on this very urban coast both the cattle population and the number of observed cases are low (Abrial et al., 2003, 2005). Our HMRF mapping is in that sense much more reliable with the French Riviera identified as a low risk region as it should. We suspect that this bad feature of BYM may come from the strength of its spatial prior in the absence of strong information on the spatial structure in the observed data. We suspect this is the case for the BSE data set so that the resulting map using BYM may also mainly reflect the prior rather than the observations.

6 Discussion

In this paper, we propose an unsupervised method for automatically classifying geographical units into risk classes. To do so, we recast the disease mapping issue into a clustering task using a discrete hidden Markov model and Poisson class-

dependent distributions. The designed hidden Markov prior is non standard and consists of a variation of the Potts model where the interaction parameter can depend on the risk classes. One advantage of our discrete HRMF modelling is that the classification step is part of the model instead of being a post-processing step as in most methods currently used by animal epidemiologists. The model parameters were then estimated using an EM algorithm and a mean field approximation principle. This provides a way to face the intractability of the standard EM in this spatial context, with a computationally efficient alternative to more intensive MCMC procedures.

We then focused on the issue of dealing with very low risk values and small numbers of observed cases and population sizes. In particular we addressed the problem of finding good initial parameter values which can be critical in this context. We developed a new initialization strategy appropriate for spatial Poisson mixtures in the case of not so well separated classes as encountered in animal disease risk analysis.

Our discrete HMRF based method provides risk maps more reliable than the traditional BYM method, with less classification errors and more clearly delimited at risk zones. Our experiments show that our model performs well in determining high risk regions, both in terms of accurate localisation of these regions and estimation of the associated risk level. This is an important point since these high risk regions are of primary interest in practice when the goal is to eventually impose safety procedures. The low risk regions are more difficult to delineate, especially when they are not in areas of large population size. Overall, our experiments suggest that the usual BYM method, in its simplest version, is not adapted to rare diseases in very inhomogeneous populations, as it tends to estimate high risks in regions with very low population.

The solution we propose instead is a flexible model whose parameters are easy to interpret and to adapt to other situations involving spatial count data. In particular, the interpretation of the pair-wise potential functions in terms of neighborhood interaction allows users to define their own spatial smoothing depending on the targeted task. The definition of the neighborhood, simply based on geographical proximity in this paper, can be adapted to the context and potentially include non spatial information through some measures of similarity between sites based on non geographical features. Typically, for the BSE example, sites could be set as neighbors if they share the same animal food provider. A second example, is the possibility to introduce dissymmetric interactions to account for an ecological gradient such as wind dissemination.

In addition, to better understand the mechanisms underlying the spread of a disease, it is possible to introduce covariates at various stages of the hierarchy without changing too much the structure of the model. The use of a mean field principle for inference generalizes easily in this case and has the advantage to maintain the model tractability.

Then, the model applies naturally to all kinds of graphical structures and can therefore adapt easily to integrate temporal information such as given for instance by observations corresponding to cases for the same area but at different periods of time. Further investigations for such a spatio-temporal analysis are planned.

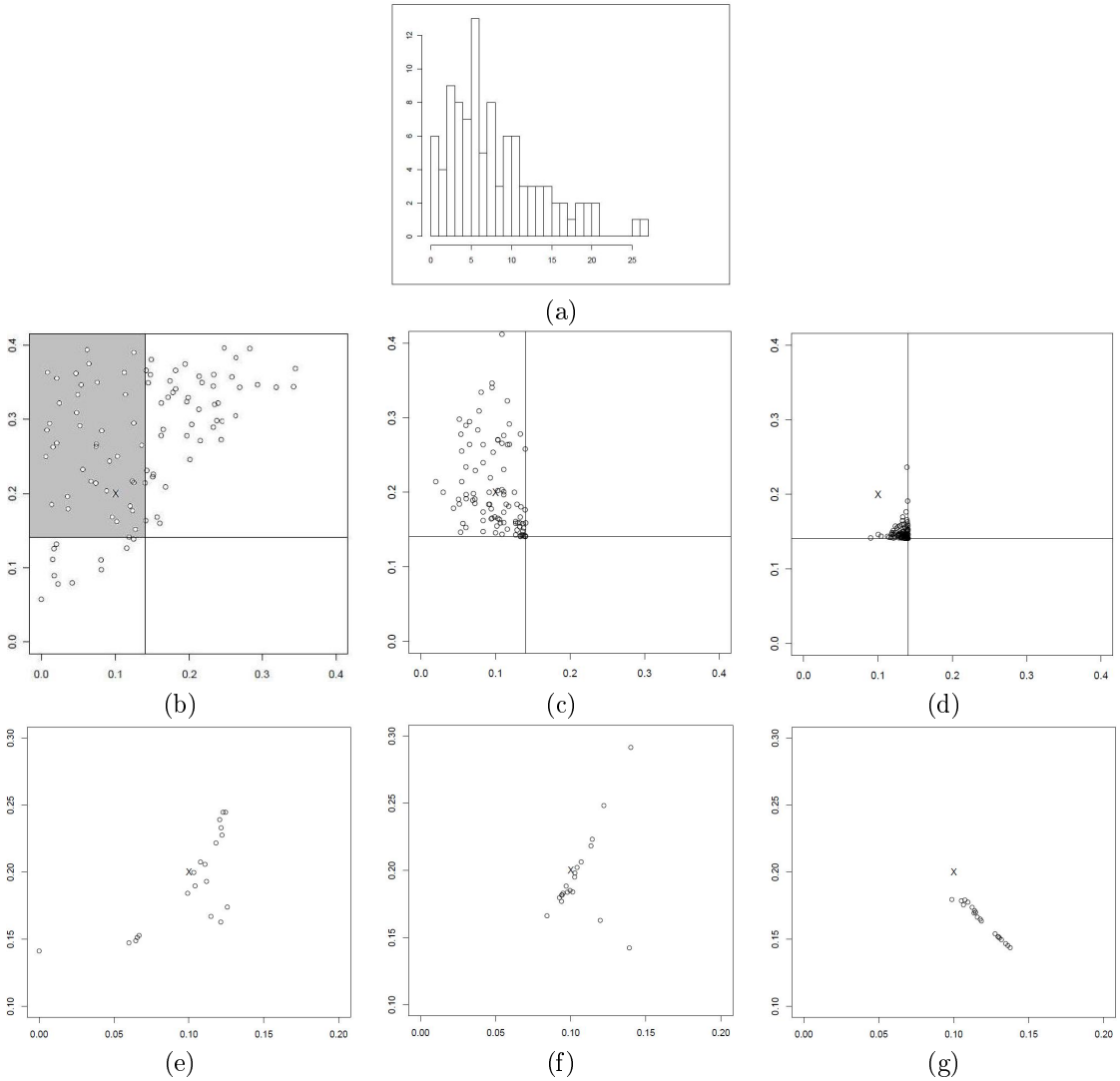


Figure 1: A non spatial two class example. Figure (a): histogram of 100 values from our model with $\lambda_1 = 0.1$, $\lambda_2 = 0.2$, $b = 0$ and α is so that the two classes are in equal proportion; (b): 100 values of (λ_1, λ_2) sampled at random between 0 and 0.4 and ordered so that $\lambda_1 < \lambda_2$; (c): 100 values of (λ_1, λ_2) generated from the strategy using the EM trajectories property and (d): 100 values of (λ_1, λ_2) generated from random partitions of the simulated data. The true values (0.1, 0.2) are indicated by a "X". (e): 20 values of (λ_1, λ_2) after 1 iteration of EM when the initial values are random, (f): when the initial values are generated using the proposed strategy and (g): when the initial values are obtained from random partitions.

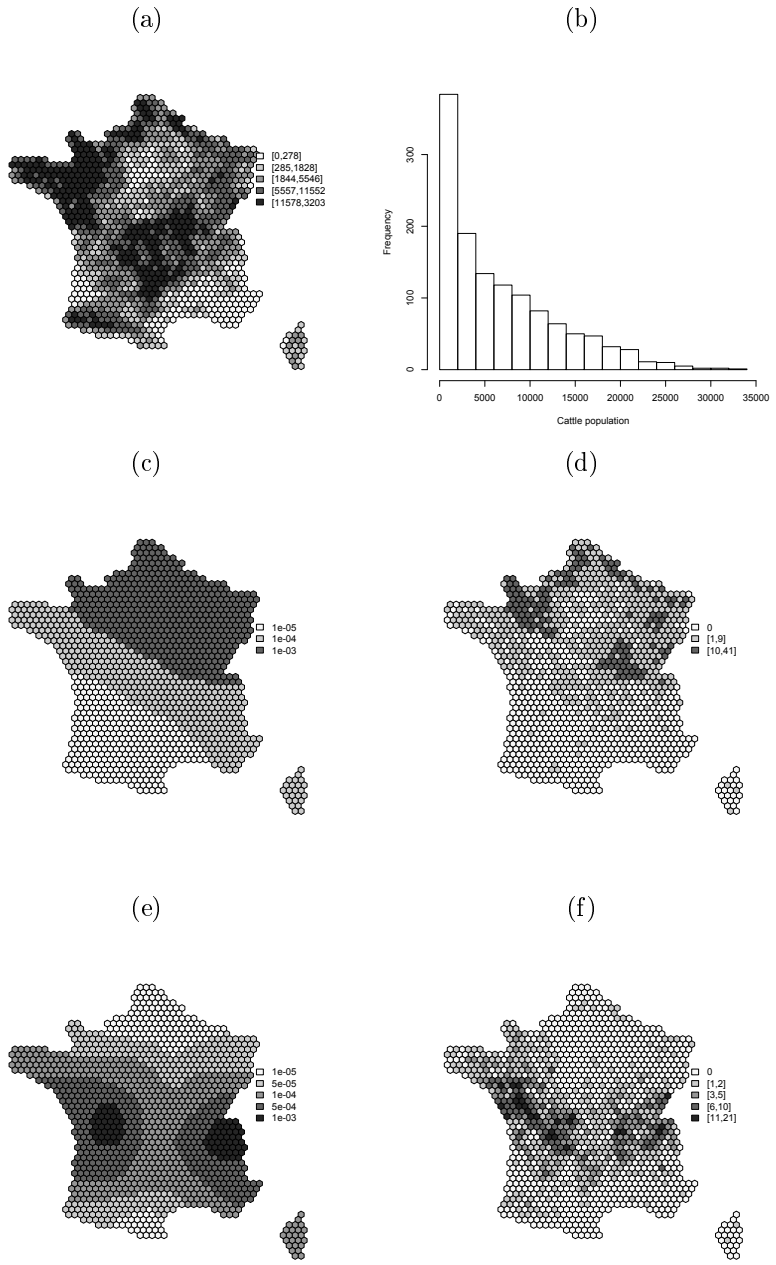


Figure 2: Cattle population data and simulated data sets. (a) and (b): Population map and histogram of population size. (c) and (e): synthetic underlying risk maps with 3 and 5 classes, (d) and (f): examples of simulated counts for the 3-class and 5-class cases.

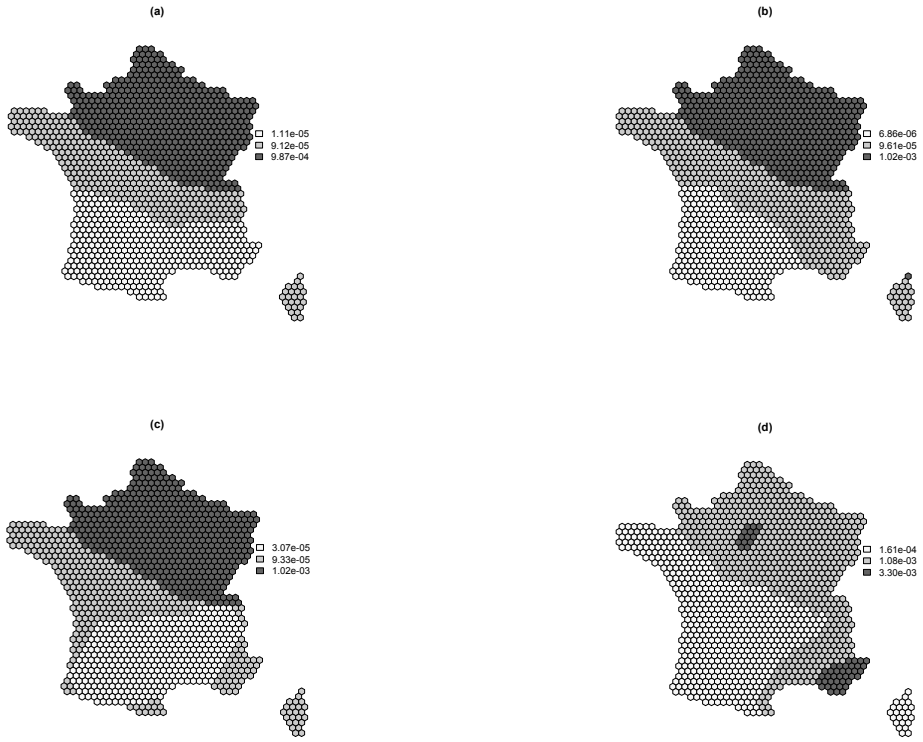


Figure 3: Classification results in the 3-class case. (a), (b) and (c): risk maps obtained respectively with the S_{tra}, S_{rand} and S_{EMM} strategy. (d): risk map obtained with the BYM model.

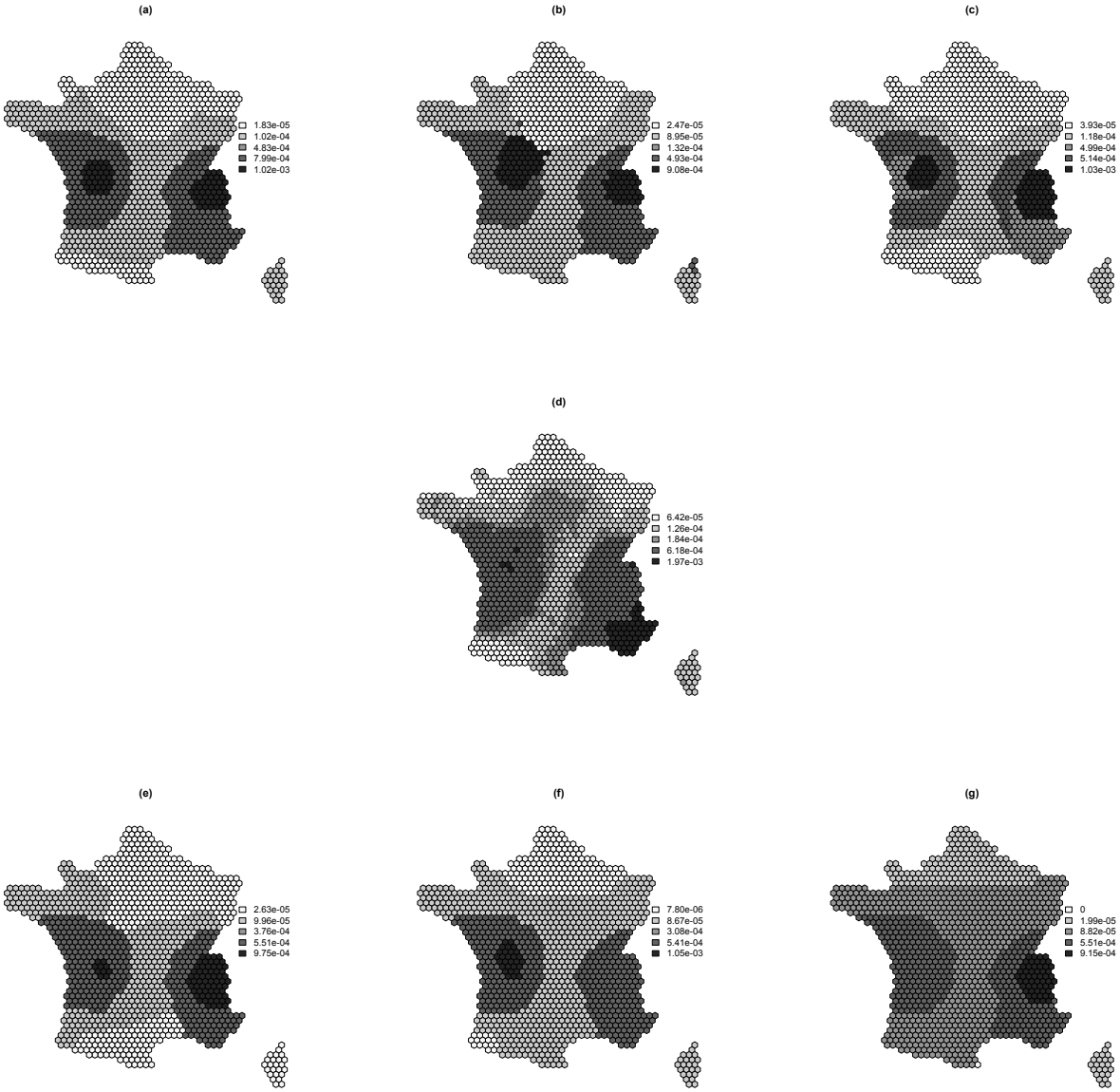


Figure 4: Classification results in the 5-class case. Risk maps obtained with (a) S_{tra}, (b) S_{rand}, (c) S_{EMM}, starting from 1000 initial positions and (d) with the BYM model. Risk maps obtained starting from 10 initial positions with the (e) S_{tra}, (f) S_{rand} and (g) S_{EMM} strategies.

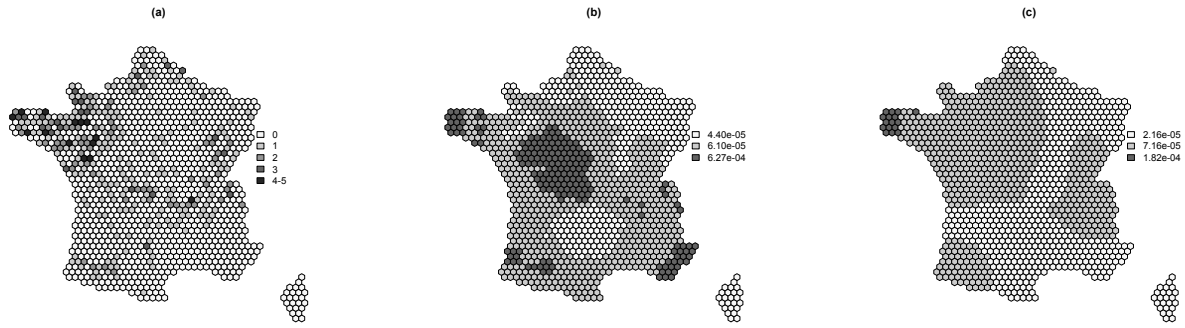


Figure 5: BSE data set. (a): BSE cases registered between the first of July 2001 and the 31st of December 2005. Estimated risk maps obtained using (b) the BYM model and (c) our hidden Markov model designed for risk mapping and inference via variational EM.

Table 1: 100 five-class and 100 three-class data sets. Average percentage and standard deviation of correct classifications, average risk value and standard deviation of the risk value for each class using different initialization strategies.

Results for the three-class data set			
True risk level	Strategy	Classification rate	Estimated risks
low	S_{rand}	85.84 (36.76)	$1.02e^{-05}(3.31e^{-06})$
$1e^{-05}$	S_{emm}	41.68 (38.48)	$4.12e^{-05}(3.11e^{-06})$
	S_{tra}	71.87 (25.23)	$1.49e^{-05}(1.48e^{-05})$
medium	S_{rand}	93.83 (23.16)	$9.82e^{-05}(6.06e^{-06})$
$1e^{-04}$	S_{emm}	66.42 (28.21)	$2.19e^{-04}(2.13e^{-04})$
	S_{tra}	86.19 (30.05)	$1.15e^{-04}(6.84e^{-05})$
high	S_{rand}	99.34 (11.25)	$9.94e^{-04}(1.71e^{-05})$
$1e^{-03}$	S_{emm}	99.12 (21.75)	$9.99e^{-04}(2.57e^{-05})$
	S_{tra}	95.89 (14.38)	$9.97e^{-04}(1.74e^{-05})$
Results for the five-class data set			
True risk level	Strategy	Classification rate	Estimated risks
very low	S_{rand}	47.15 (28.77)	$2.17e^{-05}(2.15e^{-05})$
$1e^{-05}$	S_{EMM}	26.08 (18.20)	$2.58e^{-05}(2.98e^{-06})$
	S_{tra}	46.30 (21.59)	$2.07e^{-05}(1.53e^{-05})$
low	S_{rand}	27.92 (16.16)	$7.99e^{-05}(7.53e^{-05})$
$5e^{-05}$	S_{EMM}	18.17 (14.43)	$5.43e^{-04}(3.49e^{-05})$
	S_{tra}	22.12 (12.48)	$9.62e^{-05}(4.39e^{-05})$
medium	S_{rand}	53.90 (31.62)	$1.74^{-04}(1.57e^{-04})$
$1e^{-04}$	S_{EMM}	36.82 (34.37)	$3.03e^{-04}(2.06e^{-04})$
	S_{tra}	15.47 (23.89)	$3.33e^{-04}(1.37e^{-04})$
high	S_{rand}	60.92 (33.22)	$4.58e^{-04}(1.97e^{-05})$
$5e^{-04}$	S_{EMM}	68.97 (32.21)	$5.74e^{-04}(5.86e^{-05})$
	S_{tra}	67.81 (34.34)	$5.57e^{-04}(1.05e^{-04})$
very high	S_{rand}	42.43 (29.65)	$8.71e^{-04}(4.27e^{-04})$
$1e^{-03}$	S_{EMM}	64.00 (35.48)	$9.78e^{-04}(1.76e^{-04})$
	S_{tra}	89.95 (16.51)	$1.05e^{-03}(7.66e^{-05})$

Supplementary materials

Additional Tables, and Figures referenced in Sections (4 and 5) are available below.

1. Extra Table 2

True risk level	Strategy	Classification rate	Estimated risks
low	S_{rand}	98.36	6.86 e^{-06}
$1e^{-05}$	S_{EMM}	64.82	3.07 e^{-05}
	S_{tra}	75.43	1.11 e^{-05}
medium	S_{rand}	96.91	9.61 e^{-05}
$1e^{-04}$	S_{EMM}	82.58	9.33 e^{-05}
	S_{tra}	97.47	9.12 e^{-05}
high	S_{rand}	99.80	1.02 e^{-03}
$1e^{-03}$	S_{EMM}	100	1.02 e^{-03}
	S_{tra}	100	9.87 e^{-04}

Table 2: Three-class data set. Percentage of correct classifications and estimated risk for each class using different initialization strategies.

2. Extra Table 3

True risk level	Strategy	Classification rate	Estimated risk
very low	S_{rand}	47.02	2.47 e^{-05}
$1e^{-05}$	S_{EMM}	37.30	3.93 e^{-05}
	S_{tra}	45.71	1.83 e^{-05}
low	S_{rand}	28.94	8.95 e^{-05}
$5e^{-05}$	S_{EMM}	4.30	1.18 e^{-04}
	S_{tra}	18.58	1.02 e^{-04}
medium	S_{rand}	0	1.32 e^{-04}
$1e^{-04}$	S_{EMM}	21.23	4.99 e^{-04}
	S_{tra}	0	4.83 e^{-04}
high	S_{rand}	81.60	4.93 e^{-04}
$5e^{-04}$	S_{EMM}	94.03	5.14 e^{-04}
	S_{tra}	85.55	7.99 e^{-04}
very high	S_{rand}	63.63	9.08 e^{-04}
$1e^{-03}$	S_{EMM}	82.79	1.03 e^{-03}
	S_{tra}	98.73	1.83 e^{-03}

Table 3: Five-class data set. Percentage of correct classifications and estimated risk for each class using different initialization strategies.

3. Extra Figure 6

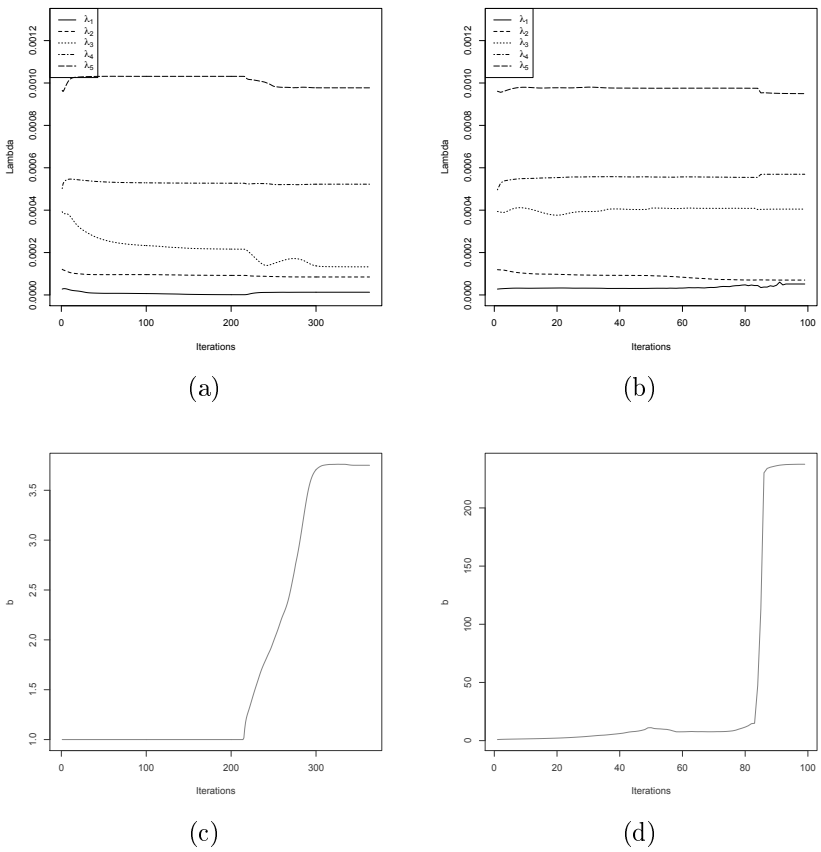
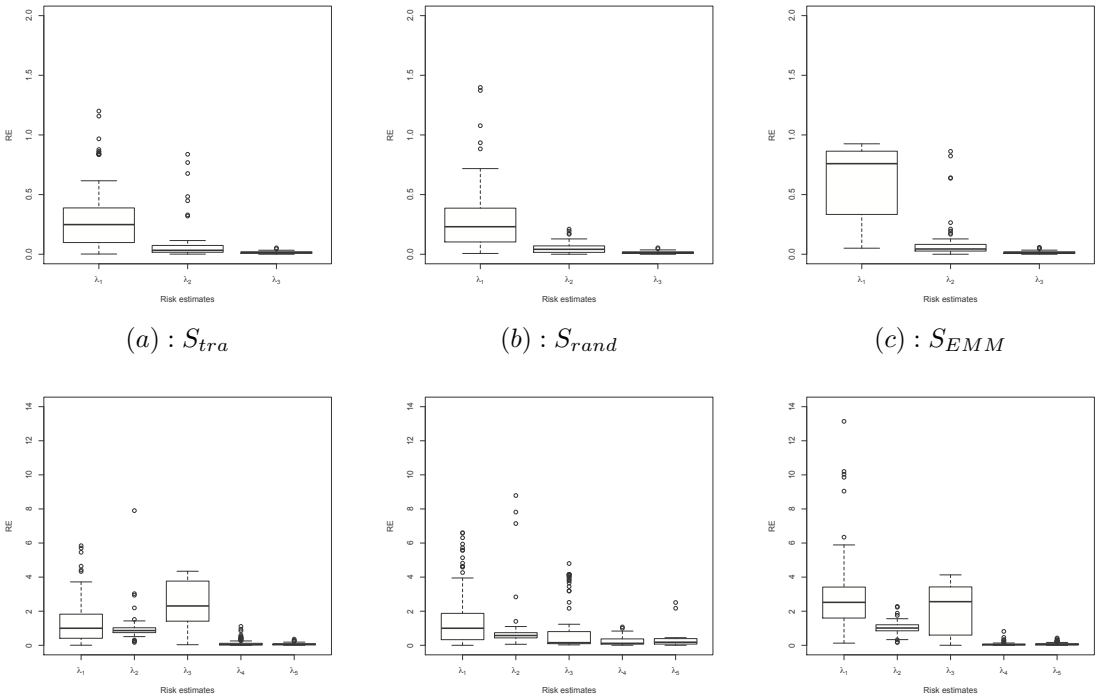


Figure 6: Parameters obtained for a simulated data example of a hidden Markov model with 5 mixture components. (a): λ values obtained using our full search procedure. (b): λ values obtained skipping the **Search 2** step. (c): spatial interaction parameter b value obtained using our full search procedure . (d): b value obtained skipping the **Search 2** step. Note the much larger scale in the latter plot.

4. Extra Figure 7



$(a) : S_{tra}$ $(b) : S_{rand}$ $(c) : S_{EMM}$

$(d) : S_{tra}$ $(e) : S_{rand}$ $(f) : S_{EMM}$

Figure 7: Relative errors for the risk parameters for the 3 different initialization strategies. (a) to (c): 3-class example . (d) to (f): 5-class example.

References

Abrial, D.; Calavas, D.; Jarrige, N. & Ducrot, C. (2005): Poultry, pig and the risk of BSE following the feed ban in France: A spatial analysis. *Veterinary Research* **36**:615–628.

Abrial, D.; Calavas, D.; Lauvergne, N.; Morignat, E. & Ducrot, C. (2003): Descriptive spatial analysis of BSE in western France. *Veterinary Research* **34**:1–11.

Alfo, M.; Nieddu, L. & Vicari, D. (2009): Finite Mixture Models for Mapping Spatially Dependent Disease Counts. *Biometrical Journal* **52**:84–97.

Allepuz, A.; Lopez-Quilez, A.; Forte, A.; Fernandez, G. & Casal, J. (2007): Spatial analysis of bovine spongiform encephalopathy in Galicia, Spain(2002-2005). *Preventive Veterinary Medicine.* **79**(2-4):174–185.

Anderson, R. M.; Donnelly, C. A.; Ferguson, N.; Woolhouse, M.; Watt, C. J.; UDY, H. J.; Mawhinney, S.; Dunstan, S.; Southwood, T.; Wilesmith, J.; Ryan, J.; Hoinville, L.; Hillerton, J. & Andstin, A. (1996): Transmission dynamics and epidemiology of BSE in British cattle. *Nature* .

Besag, J. (1986): On the statistical analysis of dirty pictures. *Jounal of Royal Statistical Society B* **48**(3):259–302.

Besag, J.; York, J. & Mollie, A. (1991): Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Statist Math* **43**(1):1–59.

Biernacki, C. (2004): Initializing EM Using the Properties of its Trajectories in Gaussian Mixtures. *Statistics and Computing* **14**(3):267–279.

Biernacki, C.; Celeux, G. & Govaert, G. (2003): Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics and Data Analysis* **41**:561–575.

Blanchet, J. & Forbes, F. (2008): Triplet Markov fields for the supervised classification of complex structure data. *IEEE Trans. on Pattern Analyis and Machine Intelligence* **30(6)**:1055–1067.

Celeux, G.; Forbes, F. & Peyrard, N. (2003): EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattern Recognition* **36**:131–144.

Chalmond, B. (1989): An iterative technique for reconstruction of m-ary images. *Pattern Recognition* **22**(6):747–761.

Clayton, D. & Bernadinelli, L. (1992): Bayesian Methods for Mapping Disease Risk. *Geographical and Environment Epidemiology: Methods for Small Area Studies, eds. P.Elliot, J.Cuzik, D.English, and R.Stern, Oxford, UK:Oxford University Press* pages 205–220.

Dempster, A.; Laird, N. & Rubin, D. (1977): Maximum likelihood from incomplete data via the EM algorithm. *Royal Statistical Society Series B* **39**(1):1–38.

Ducrot, C.; Arnold, M.; de Koejer, A.; Heim, D. & Calavas, D. (2008): Review on the epidemiology and dynamics of BSE epidemics. *Veterinary Research.* **39:15**.

Fernandez, C. & Green, P. (2002): Modelling spatially correlated data via mixtures: a Bayesian approach. *Jounal of Royal Statistical Society B* **64**(4):805–826.

Forbes, F.; Doyle, S.; Garcia-Lorenzo, D.; Barillot, C. & Dojat, M. (2010): A Weighted Multi-Sequence Markov Model For Brain Lesion Segmentation. *13th International Conference on Artificial Intelligence and Statistics (AISTATS10).* Sardinia, Italy.

Forbes, F. & Peyrard, N. (2003): Hidden Markov Model Selection based on Mean Field like approximations. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **25**(8).

Fraley, C. & Raftery, A. (2007): Bayesian regularization for Normal mixture estimation and model-based clustering. *Journal of Classification* **24**:155–181.

Francois, O.; Ancelet, S. & Guillot, G. (2006): Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics* **174**:885–816.

Geman, S. & Geman, D. (1984): Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Transaction on Pattern analysis and Machine Int* **6**(6):721–741.

Green, P. J. & Richardson, S. (2002): Hidden Markov models and disease mapping. *Journal of the American Statistical Association* **97**(460):1–16.

Karlis, D. & Xekalaki, E. (2003): Choosing initial values for the EM algorithm for finite mixtures. *Computational Statistics and Data Analysis* **41**:577–590.

Knorr-Held, L. & Rasser, G. (2000): Bayesian Detection of Clusters and Discontinuities in Disease Maps. *Biometrics* **56**:13–21.

Knorr-Held, L.; Rasser, G. & Becker, N. (2002): Disease mapping of stage-specific cancer incidence data. *Biometrics* **58**(3):492–501.

Knorr-Held, L. & Richardson, S. (2003): A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Journal of Royal Statistical Society* **52**(2):169–183.

Lawson, A.; Biggeri, A.; Boehning, D.; Lesaffre, E.; Viel, J.; Clark, A.; Schlattmann, P. & Divino, F. (2000): Disease mapping models: an empirical evaluation. *Statistics in Medicine* **19**:2217–2241.

Lawson, A. & Song, H. (2010): Bayesian hierarchical modeling of the dynamics of spatio-temporal influenza season outbreaks. *Spatial and Spatio-temporal Epidemiology* **1**:187–195.

MacNab, Y. (2010): On Gaussian Markov random fields and Bayesian disease mapping. *Statistical Methods in Medical Research* **0**:1–20.

McLachlan, G. & Peel, D. (2000): *Finite Mixture Models.* New York: John Wiley and Sons.

Mollie, A. (1996): Bayesian Mapping of Disease. *Markov Chain Monte Carlo in Practice, eds. W. Gilks, S. Richardson, and D. J. Spiegelhalter, London: Chapman and Hall* pages 359–379.

Mollie, A. (1999): Bayesian and Empirical Bayes approaches to disease mapping. A. Lawson; A. Biggeri & D. Bohning (editors), *Disease mapping and risk assessment for public health*, pages 15–29. Wiley.

Mollie, A. & Richardson, S. (1991): Empirical Bayes estimates of cancer mortality rates using spatial models. *Statistics in Medicine* **10**:95–112.

Pascutto, C.; Wakefield, J.; Best, N.; Richardson, S.; Bernardinelli, L.; Staines, A. & Elliott, P. (2000): Statistical issues in the analysis of disease mapping data. *Statistics in Medicine* **19**:2493–2519.

Paul, M.; Abrial, D.; Jarrige, N.; Rican, S.; Garrido, M.; Calavas, D. & Ducrot, C. (2007): Bovine spongiform encephalopathy and spatial analysis of the feed industry. *Emerging Infectious Diseases.* **13**(6):867–872.

Qian, W. & Titterington, D. (1991): Estimation of parameters in hidden Markov models. *Philosophical Transactions of the Royal Society of London* **Series A 337**:407–428.

Richardson, S.; Monfort, C.; , M.; Draper, G. & Muirhead, C. (1995): Spatial variation of natural radiation and childhood leukaemia incidence in Great Britain. *Statistics in Medicine* **14**:2487–2501.

Robertson, C.; Nelson, T.; MacNab, Y. & Lawson, A. (2010): Review of methods for space-time disease surveillance. *Spatial and Spatio-temporal Epidemiology* **1**:105–116.

Vignes, M. & Forbes, F. (2009): Gene clustering via integrated Markov models combining individual and pairwise features. *IEEE Trans. on Computational Biology and Bioinformatics* **6**(2):260–270.

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399