Bayesian approaches to spatial inference: Modelling and computational challenges and solutions

Matthew T. Moores
Queensland University of Technology, mmoores@uow.edu.au

Kerrie Mengersen
Queensland University of Technology

Follow this and additional works at: https://ro.uow.edu.au/eispapers1
Bayesian approaches to spatial inference: Modelling and computational challenges and solutions

Abstract
We discuss a range of Bayesian modelling approaches for spatial data and investigate some of the associated computational challenges. This paper commences with a brief review of Bayesian mixture models and Markov random fields, with enabling computational algorithms including Markov chain Monte Carlo (MCMC) and integrated nested Laplace approximation (INLA). Following this, we focus on the Potts model as a canonical approach, and discuss the challenge of estimating the inverse temperature parameter that controls the degree of spatial smoothing. We compare three approaches to addressing the doubly intractable nature of the likelihood, namely pseudo-likelihood, path sampling and the exchange algorithm. These techniques are applied to satellite data used to analyse water quality in the Great Barrier Reef.

Publication Details

This journal article is available at Research Online: https://ro.uow.edu.au/eispapers1/1685
Bayesian approaches to spatial inference: Modelling and computational challenges and solutions
Matthew Moores, and Kerrie Mengersen

Citation: AIP Conference Proceedings 1636, 112 (2014); doi: 10.1063/1.4903719
View online: https://doi.org/10.1063/1.4903719
View Table of Contents: http://aip.scitation.org/toc/apc/1636/1
Published by the American Institute of Physics
Bayesian Approaches to Spatial Inference: Modelling and Computational Challenges and Solutions

Matthew Moores* and Kerrie Mengersen*,†

*Mathematical Sciences School, Queensland University of Technology, Brisbane, Queensland 4001, Australia.
†Email: k.mengersen@qut.edu.au

Abstract. We discuss a range of Bayesian modelling approaches for spatial data and investigate some of the associated computational challenges. This paper commences with a brief review of Bayesian mixture models and Markov random fields, with enabling computational algorithms including Markov chain Monte Carlo (MCMC) and integrated nested Laplace approximation (INLA). Following this, we focus on the Potts model as a canonical approach, and discuss the challenge of estimating the inverse temperature parameter that controls the degree of spatial smoothing. We compare three approaches to addressing the doubly intractable nature of the likelihood, namely pseudo-likelihood, path sampling and the exchange algorithm. These techniques are applied to satellite data used to analyse water quality in the Great Barrier Reef.

Keywords: Bayesian inference, Image processing, Markov random fields, Monte Carlo methods

1. INTRODUCTION

The link between MaxEnt and Bayesian estimation is well known, especially for spatial models. For example, any Markov random field (MRF) that is defined using the Gibbs measure and Hammersley-Clifford theorem has a maximum entropy interpretation [1, chap. 4]. In this paper, we review the autonormal model and hidden Potts model, two kinds of MRF that can be used for image analysis. These models are applied to remotely sensed imagery from the Moderate-resolution Imaging Spectroradiometer (MODIS-AQUA). We fit the autonormal model using integrated nested Laplace approximations (INLA) for smoothing and interpolation of the pixel values. We also compare three different Markov chain Monte Carlo (MCMC) algorithms for fitting the hidden Potts model. The Potts model is used to classify the image pixels into regions of similar values.

A MRF represents spatial correlation between neighbouring nodes $i \sim \ell$ in a lattice, as shown in equation (1):

$$p(x_i | x_{i \sim \ell}, \vartheta) = \frac{\exp \{- \vartheta \sum_{i \sim \ell} f(x_i, x_\ell)\}}{Z(\vartheta)}$$

where $\vartheta$ is a scale parameter, $\sum_{i \sim \ell} f(x_i, x_\ell)$ is the potential of the Gibbs field and $Z(\vartheta)$ is known as the normalising constant or partition function. A wide variety of different MRF models can be defined by choosing different forms of potential energy function.

Besag [2] introduced the autonormal model, also known as the intrinsic conditional autoregressive (CAR) prior. The intrinsic CAR is a Gaussian MRF where the potential function is defined in terms of pairwise differences between nodes, as shown in equation (2):

$$p(x_i | x_{i \sim \ell}, \sigma_{sp}^2) \propto \exp \left\{ -\frac{1}{2\sigma_{sp}^2} \sum_{i \sim \ell} (x_i - x_\ell)^2 \right\}$$

where $\sigma_{sp}^2$ is the variance of the spatially-correlated error. Each node is normally distributed with expectation equal to the average of the values of its neighbours, according to equation (3):

$$x_i | x_{i \sim \ell}, \sigma_{sp}^2 \sim \mathcal{N} \left( \frac{1}{n_i} \sum_{i \sim \ell} x_\ell, \frac{\sigma_{sp}^2}{n_i} \right)$$

where $n_i$ is the number of neighbours of node $i$. The differences between the latent Gaussian field and the observed data $y$ are conditionally independent, as shown in equation (4):

$$y_i | x_i, \sigma_y^2 \sim \mathcal{N}(x_i, \sigma_y^2)$$
The hidden Potts model [3] is a MRF with discrete states \(x \in \{1, \ldots, k\}\), defined according to equation (5):

\[
p(x_i|x_{i \sim \ell}, \beta) = \frac{\exp \{\beta \sum_{j \sim \ell} \delta(x_i, x_j)\}}{\sum_{j=1}^{k} \exp \{\beta \sum_{j \sim \ell} \delta(j, x_{\ell})\}}
\]

where \(\delta(u, v)\) is the Kronecker delta function. Thus, the potential function is a weighted count of the neighbours of node \(i\) that share the same label. The parameter \(\beta\), known as the inverse temperature, governs the strength of spatial association between labels. Under the assumption of additive white noise, the observed data \(y\) is normally distributed conditional on the latent states, as shown in equation (6):

\[
y_i|x_i = j \sim \mathcal{N}(\mu_j, \sigma_j^2)
\]

Thus the hidden Potts model can also be interpreted as a spatial mixture model [4].

2. INTEGRATED NESTED LAPLACE APPROXIMATION

Integrated nested Laplace approximation (INLA) is a computational method that was introduced by Rue et al. [5] for fast evaluation of latent Gaussian models. Observing that most posterior densities are unimodal, the Laplace approximation [6] replaces the posterior by a Gaussian density with mean situated at the posterior mode and variance calculated as the local curvature of the distribution, evaluated at the mode. This can be extended to more than one variable using a multivariate Normal distribution with precision-coprecision matrix equal to the negative of the Hessian matrix at the posterior mode. Often this matrix will be very sparse, with zero elements implying conditional independence between the corresponding variables. Rather than drawing samples from the posterior, as would be done in MCMC, INLA uses multivariate Newton solvers to find the parameter values that maximise the posterior density. This method is implemented in an R package [7] available from www.r-inla.org.

The joint distribution of the intrinsic CAR model with known variance is already multivariate Normal, but the variance can be treated as a free parameter in order to estimate its posterior distribution. The Laplace approximation of the log-precision is nested within the joint distribution of the nodes, \(x|\log \left\{ \frac{1}{\sigma^2_{sp}} \right\}\). The marginal posterior of the nodes can then be approximated by numerical integration.

The computational speed of INLA also makes it possible to extend the neighbourhood scheme beyond the usual first-order (nearest) neighbours. For example, the 2D random walk model approximates a thin plate spline smoother using a second-order polynomial, as shown in equation (7):

\[
E[x_i|x_{i \sim \ell}, \sigma^2_{sp}] = \frac{1}{20} \left( \begin{array}{cccc}
8 & \cdots & \cdots & \cdots \\
\cdots & \cdots & -2 & \cdots \\
\cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots
\end{array} \right)
\]

This principle can be extended even further to construct an MRF approximation to the continuous spatial models used in geostatistics. This has been demonstrated for the Gaussian process with Matérn covariance [8] and the log-Gaussian Cox process [9].

3. DOUBLY-INTRACTABLE LIKELIHOOD

The INLA approach is not applicable to the hidden Potts model because the latent state space is discrete, resulting in a multimodal posterior. If natural conjugate priors are used for the means and variances of the mixture components then they can be simulated using Gibbs sampling. Computational methods for simulating the latent labels are described by Winkler [10]. Simulating from the inverse temperature parameter requires specialised methods because the partition function \(Z(\beta)\) is doubly-intractable (both analytically and computationally). It involves a sum over all possible combinations of the labels \(x \in \mathcal{X}\), as shown in equation (8):

\[
Z(\beta) = \sum_{x \in \mathcal{X}} \exp\{\beta S(x)\}
\]

where \(S(x) = \sum_{i \sim \ell \in \mathcal{N}} \delta(x_i, x_{\ell})\) is the sufficient statistic of the Potts model. We review three methods for simulating from the inverse temperature without evaluating the partition function. These methods are pseudolikelihood [11], path sampling [12] and the exchange algorithm [13].
Pseudolikelihood approximates equation (8) using the product of the conditional densities in equation (5). Proposed values of $\beta'$ can be drawn using a random walk and evaluated using a pseudolikelihood approximation to the Metropolis-Hastings (MH) ratio. Pseudolikelihood is very fast, but the approximation error increases for large values of $\beta$.

Path sampling involves precomputation of the expectation of the sufficient statistic for fixed values of $\beta$. During model fitting, these precomputed values are interpolated to approximate $E[S(x)|\beta']$ for the proposed values of $\beta'$. The MH ratio can then be approximated using numerical integration, according to the path sampling identity in equation (9):

$$\log \left\{ \frac{Z(\beta^\circ)}{Z(\beta')} \right\} = \int_{\beta^\circ}^{\beta'} E[S(x)|\beta] \, d\beta$$

The precomputation can be costly for large datasets, but the output can be reused if there are multiple datasets that are approximately the same size and use the same number of mixture components $k$.

The exchange algorithm is an exact method for simulating from the inverse temperature using an auxiliary variable. The auxiliary variable $w$ is drawn from equation (5) for the proposed value of $\beta'$ so that the partition function cancels out in the MH ratio, as shown in equation (10):

$$\rho = \frac{q(\beta^\circ|\beta') \pi(\beta') \exp \{ \beta^\circ S(x) \} Z(\beta^\circ) \exp \{ \beta^\circ S(w) \} Z(\beta')}{q(\beta'|\beta^\circ) \pi(\beta^\circ) \exp \{ \beta^\circ S(x) \} Z(\beta^\circ) \exp \{ \beta^\circ S(w) \} Z(\beta')}$$

If $\beta'$ is drawn from a symmetric random walk then the ratio $\frac{q(\beta^\circ|\beta')}{q(\beta'|\beta^\circ)} = 1$. If a uniform prior is used for $\beta$ then the ratio $\frac{\pi(\beta')}{\pi(\beta^\circ)}$ also cancels, simplifying the logarithm of the MH ratio to $\log \rho = (\beta' - \beta^\circ) S(x) + (\beta^\circ - \beta') S(w)$. Since the auxiliary vector $w$ has the same dimension as the latent labels $x$, simulation of this variable can be expensive for large datasets. The exchange algorithm is only exact when perfect sampling [14] is used to simulate $w$. Cuaca et al. [15] use 500 iterations of Gibbs sampling for the auxiliary variable, which results in an approximate exchange algorithm with reduced computational cost.

Fast approximations are available for fitting the hidden Potts model, such as iterated conditional modes (ICM) [16] or variational Bayes (VB) [17]. However, the options for estimating the inverse temperature are limited. Path sampling and the exchange algorithm estimate the ratio of normalising constants, which is ideal for MCMC but makes these methods unsuitable for use in deterministic, iterative algorithms like ICM or VB. Pseudolikelihood produces a direct approximation of equation (8), so it is able to be used in combination with these faster algorithms.

### 4. SATELLITE REMOTE SENSING

We illustrate these methods using satellite imagery of the water off northern Queensland, Australia, including the Great Barrier Reef lagoon. The image was obtained from the Moderate-resolution Imaging Spectroradiometer (MODIS-AQUA) at a resolution of 1km per pixel. The pixel intensity values illustrated in figure 1 are an estimate of the concentration of chlorophyll-a (in milligrams per cubic metre) derived from the ratio of reflectance of blue and green light [18]. Chlorophyll levels are an indicator of the abundance of phytoplankton, which is an important measure of water quality. The image lattice has 267,867 nodes that are connected by 481,237 edges, not including the missing data. The main causes of missingness are pixels that fall on land, such as the Queensland coastline visible in figure 1, or pixels that were obscured by cloud cover.

We used INLA to fit two Gaussian MRF models to this data. We also compared pseudolikelihood, path sampling and the exchange algorithm for estimating the inverse temperature of a hidden Potts model. Computations were performed on a cluster of 8 core, 2.66GHz Intel Xeon computers with 48GB RAM, running 64bit GNU/Linux 3.0.74 and R 3.0.1. Some related experiments have been published in Falk [19].

#### 4.1. Gaussian MRF models

Using INLA, we fit both the intrinsic CAR model and the 2D random walk to illustrate the effect of the larger neighbourhood region. These models can be used to smooth the observed values, but more importantly they can also provide estimates for the missing data. Table 1 shows Bayesian credible intervals for the precision of the latent region.

---

114
FIGURE 1. Distribution of chlorophyll in the Great Barrier Reef, Queensland, Australia.

TABLE 1. Inference for the Gaussian MRF models.

<table>
<thead>
<tr>
<th>Model</th>
<th>95% CI for Intercept</th>
<th>95% CI for $\sigma_{y}^{-2}$</th>
<th>95% CI for $\sigma_{sp}^{-2}$</th>
<th>Elapsed Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic CAR</td>
<td>[0.1216; 0.1219]</td>
<td>[59586; 70500]</td>
<td>[1141; 1159]</td>
<td>14h 12min</td>
</tr>
<tr>
<td>2D Random Walk</td>
<td>[-0.2281; 0.9676]</td>
<td>[6127; 6280]</td>
<td>[1396; 1459]</td>
<td>43h 3min</td>
</tr>
</tbody>
</table>

Gaussian MRF $\sigma_{sp}^{-2}$ and the independent residuals $\sigma_{y}^{-2}$. We also report the elapsed (wall clock) time for fitting the models. The intrinsic CAR model provides a closer fit to the data, shown by the higher precision of the residuals. The fitted values for both models are illustrated in figure 2.

4.2. Hidden Potts model

We ran pseudolikelihood, path sampling and the exchange algorithm for 20,000 MCMC iterations each, discarding the first ten thousand as burn-in. The results are shown in table 2. The exchange algorithm took two orders of magnitude longer, which follows from using 500 iterations of Gibbs sampling to draw the auxiliary variable. Precomputation of $E[S(z)|\beta]$ for path sampling took 16 hours for 61 values of $\beta$. Pseudolikelihood was the fastest method but produced a much higher posterior estimate of $\beta$ than either path sampling or the exchange algorithm. Pixel classifications from all three algorithms are illustrated in figure 3. Some small differences are visible, but most pixels have been allocated to the same mixture component.

TABLE 2. Inference for the inverse temperature hyperparameter.

<table>
<thead>
<tr>
<th>Method</th>
<th>95% CI for $\beta$</th>
<th>Elapsed</th>
<th>CPU time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudolikelihood</td>
<td>[1.762; 1.787]</td>
<td>13min</td>
<td>1h 45min</td>
</tr>
<tr>
<td>Path sampling</td>
<td>[1.374; 1.381]</td>
<td>10min</td>
<td>1h 15min</td>
</tr>
<tr>
<td>Exchange algorithm</td>
<td>[1.419; 1.429]</td>
<td>41h 23min</td>
<td>323h 51min</td>
</tr>
</tbody>
</table>
FIGURE 2. Fitted values from (a) intrinsic CAR model, and (b) 2D random walk.

FIGURE 3. Classification of image pixels using different algorithms to estimate $\beta$. From left to right: pseudolikelihood, the exchange algorithm, and path sampling.

5. DISCUSSION

Both maximum entropy and Bayesian models are now firmly entrenched in a wide range of disciplines. For example, the description of MaxEnt for spatial modelling of disease by Stevens and Pfeiffer [20] is paralleled by the description of Bayesian spatial models for this purpose by Besag et al. [21], Cramb et al. [22] and many others. In the context of multispecies conservation and planning, Carroll et al. [23] showed that the fit of Bayesian spatial models based on a CAR prior was 35% to 55% better than the fit of analogous independent models, and that they outperformed analogous spatial models developed using MaxEnt methods proposed by Phillips and Dudík [24] and also described by Elith et al. [25].
We have reviewed methods for fitting Markov random fields with either continuous or discrete state spaces. We demonstrated that Gaussian MRF models can be fitted using INLA, which provides a faster alternative to MCMC algorithms. The intrinsic CAR and 2D random walk models provide a smoothly-varying latent field that can be used to extrapolate missing data in satellite imagery.

We also compared three algorithms for estimating the inverse temperature of a hidden Potts model, namely pseudolikelihood, path sampling and the exchange algorithm. Although pseudolikelihood overestimated $\beta$, it was much faster than the other algorithms and still produced reasonable labelling of the pixels. McGrory et al. [17] have reported similar findings with pseudolikelihood on other images. More research is needed into the tradeoff between accuracy and speed for image analysis, particularly with regard to the sensitivity of the other model parameters, the pixel labels as well as the means and variances of the mixture components, to errors in $\beta$.

**ACKNOWLEDGMENTS**

The authors would like to thank our research collaborators, particularly Clair Alston, Matthew Falk, Samuel Clifford, Finn Lindgren, Håvard Rue, and Daniel Simpson. MODIS-AQUA imagery courtesy of NASA Goddard Space Flight Center. Computational resources and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

**REFERENCES**