

R Package bayesImageS: Bayesian Methods for Image Segmentation using a Hidden Potts Model

Matthew T. Moores
University of Warwick

Anthony N. Pettitt

Kerrie Mengersen
Queensland University of Technology

Abstract

The inverse temperature parameter of the Potts model governs the strength of spatial cohesion and therefore has a substantial influence over the resulting model fit. A difficulty arises from the dependence of an intractable normalising constant on the value of this parameter and thus there is no closed form solution for sampling from the posterior distribution directly. There are a variety of computational approaches for sampling from the posterior without evaluating the normalising constant. These algorithms differ in their levels of accuracy and their scalability for datasets of realistic size.

This R package provides implementations of Markov chain Monte Carlo algorithms for sampling from the intractable posterior distribution of the Potts model. The algorithms include pseudolikelihood, the exchange algorithm, path sampling, and approximate Bayesian computation. The following vignette explains these algorithms, providing the necessary theoretical background as well as implementation details specific to our R package. We address important questions such as how the computational cost increases with the size of the images, as well as how much accuracy is lost by using faster, more approximate methods. This document is intended to provide guidance on selecting a suitable algorithm for Bayesian image analysis. For nontrivial images, this necessarily involves some degree of approximation to produce an acceptable compromise between accuracy and computational cost.

Keywords: approximate Bayesian computation, composite likelihood, hidden Markov random field, image analysis, pseudo-marginal method, thermodynamic integration.

1. Introduction

Markov random field (MRF) models have seen widespread use in image analysis since their introduction by [Besag \(1974\)](#), as surveyed by [Winkler \(2003\)](#) and [Li \(2009\)](#). A MRF is a generalisation of the Markovian dependence structure to more than one dimension: satellite imagery has two spatial dimensions, while medical images such as computed tomography (CT) are three-dimensional. The hidden [Potts \(1952\)](#) model employs a latent MRF on discrete states to describe spatial dependence between adjacent neighbours. The degree of dependence in the model is governed by a parameter known as the inverse temperature due to its origin in statistical physics. It is difficult to set this parameter by trial and error, particularly for noisy images. Rather than using a fixed value, it would be preferable to estimate the inverse temperature as part of the model. However, the intractable normalising constant of the Potts model depends on the value of the inverse temperature, which means that there is no closed form solution for estimating its posterior distribution.

Rydén and Titterington (1998) derived a pseudolikelihood (PL) approximation (Besag 1975) to the intractable posterior density. Gelman and Meng (1998) instead approximated the ratio of normalising constants using thermodynamic integration (TI), also known as path sampling. Møller *et al.* (2006) introduced an auxiliary variable method that gives an exact MCMC algorithm for the special case of a 2 component Potts model, also known as an Ising model. Murray *et al.* (2006) proposed a variant of the exact method known as the exchange algorithm or multiple auxiliary variable method. By replacing the expensive perfect sampling step (Propp and Wilson 1996) with an approximation such as Gibbs sampling, Cucala *et al.* (2009) developed an approximate exchange algorithm (AEA) that can be applied for Potts models with $k > 2$. Friel *et al.* (2009) introduced the reduced dependence approximation (RDA) which uses recursion to calculate the normalising constant on small sub-lattices. McGroarty *et al.* (2012) generalised RDA to an irregular lattice. Grelaud *et al.* (2009) used the sufficient statistic of the Potts model to estimate the inverse temperature using approximate Bayesian computation (ABC). Everitt (2012) combined the approximate exchange algorithm with particle Markov chain Monte Carlo (PMCMC) and also implemented ABC with sequential Monte Carlo (SMC-ABC) for the Ising model. Although all of these methods work very well in theory, the largest image used in any of these papers to demonstrate an algorithm is less than ten thousand pixels. This does not give a reliable indication of how these algorithms might perform when applied to images of a more substantial size.

2. Hidden Potts model

Image segmentation can be viewed as the task of labelling the observed pixels \mathbf{y} according to a finite set of discrete states $\mathbf{z} \in \{1, \dots, k\}$. The hidden Potts model allows for spatial correlation between neighbouring labels in the form of a Markov random field. The latent labels follow a Gibbs distribution, which is specified in terms of its conditional probabilities:

$$p(z_i | z_{\setminus i}, \beta) = \frac{\exp \{ \beta \sum_{i \sim \ell} \delta(z_i, z_\ell) \}}{\sum_{j=1}^k \exp \{ \beta \sum_{i \sim \ell} \delta(j, z_\ell) \}} \quad (1)$$

where β is the inverse temperature, $z_{\setminus i}$ represents all of the labels except z_i , $i \sim \ell$ are the neighbouring pixels of i , and $\delta(u, v)$ is the Kronecker delta function. Thus, $\sum_{i \sim \ell} \delta(z_i, z_\ell)$ is a count of the neighbours that share the same label.

If the labels z_i are indexed row-wise, the nearest (first-order) neighbours $i \sim \ell$ in a regular 2D lattice with c columns are $\{z_{i-1}, z_{i-c}, z_{i+c}, z_{i+1}\}$. Pixels situated at the boundary of the image domain have less than four neighbours. Likewise, voxels in a regular 3D lattice have a maximum of 6 first-order neighbours. These neighbourhood relationships are reciprocal, so $h \in i \sim \ell$ implies $i \in h \sim \ell$. If \mathcal{E} is the set of all unique neighbour pairs, or edges in the image lattice, then $|\mathcal{E}| = 2(n - \sqrt{n})$ for a square lattice and $3(n - n^{2/3})$ for a cube.

The observation equation links the latent labels to the corresponding pixel values:

$$p(\mathbf{y} | \mathbf{z}, \boldsymbol{\theta}) = \prod_{i=1}^n p(y_i | z_i, \theta_{z_i}) \quad (2)$$

where θ_j are the parameters that govern the distribution of the pixel values with label j . The hidden Potts model can thus be viewed as a spatially-correlated generalisation of the finite mixture model (Rydén and Titterington 1998). Green and Richardson (2002) used a Poisson

likelihood for (2), with intensity λ_j . Instead we follow Geman and Geman (1984); Alston *et al.* (2007), and many others in assuming that the pixels with label j share a common mean μ_j corrupted by additive Gaussian noise with variance σ_j^2 :

$$y_i | z_i = j, \mu_j, \sigma_j^2 \sim \mathcal{N}(\mu_j, \sigma_j^2) \quad (3)$$

The Gibbs distribution is a member of the exponential family and so there is a sufficient statistic for this model, as noted by Grelaud *et al.* (2009):

$$S(\mathbf{z}) = \sum_{i \sim \ell \in \mathcal{E}} \delta(z_i, z_\ell) \quad (4)$$

This statistic represents the total number of like neighbour pairs in the image. The likelihood $p(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}, \beta)$ can therefore be factorised into $p(\mathbf{y} | \mathbf{z}, \boldsymbol{\theta})p(S(\mathbf{z}) | \beta)$, where the second factor does not depend on the observed data, but only on the sufficient statistic. The joint posterior is then:

$$p(\boldsymbol{\theta}, \beta, \mathbf{z} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{z}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})p(S(\mathbf{z}) | \beta)\pi(\beta) \quad (5)$$

The conditional distributions $p(\boldsymbol{\theta} | \mathbf{z}, \mathbf{y})$ and $p(z_i | z_{\setminus i}, \beta, y_i, \boldsymbol{\theta}_{z_i})$ can be simulated using Gibbs sampling, but $p(\beta | \mathbf{y}, \mathbf{z}, \boldsymbol{\theta})$ involves an intractable normalising constant $\mathcal{C}(\beta)$:

$$p(\beta | \mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) \propto p(S(\mathbf{z}) | \beta)\pi(\beta) \quad (6)$$

$$\propto \frac{\exp\{\beta S(\mathbf{z})\}}{\mathcal{C}(\beta)}\pi(\beta) \quad (7)$$

The normalising constant is also known as a partition function in statistical physics. It has computational complexity of $\mathcal{O}(nk^n)$, since it involves a sum over all possible combinations of the labels $\mathbf{z} \in \mathcal{Z}$:

$$\mathcal{C}(\beta) = \sum_{\mathbf{z} \in \mathcal{Z}} \exp\{\beta S(\mathbf{z})\} \quad (8)$$

It is infeasible to calculate this value exactly for nontrivial images, thus computational approximations are required.

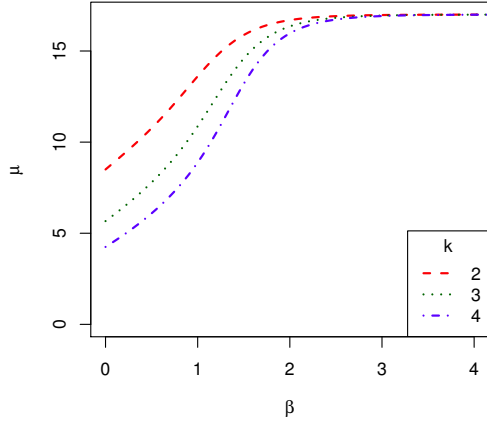
The conditional expectation of $S(\mathbf{z})$ given β can be expressed in terms of the normalising constant:

$$\mathbb{E}_{\mathbf{z}|\beta}[S(\mathbf{z})] = \frac{d}{d\beta} \log\{\mathcal{C}(\beta)\} \quad (9)$$

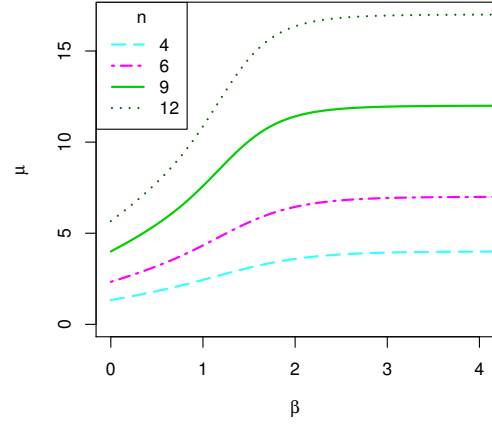
As β approaches infinity, all of the pixels in the image are almost surely assigned the same label, thus the expectation of $S(\mathbf{z})$ approaches the total number of edges $|\mathcal{E}|$ asymptotically, while the variance approaches zero. When $\beta = 0$, (1) simplifies to $\left(\sum_j \exp\{0\}\right)^{-1}$, hence the probability of any pair of neighbours being assigned the same label follows an independent Bernoulli distribution with $p = k^{-1}$. In this case, $S(\mathbf{z})$ follows a Binomial distribution with expectation $|\mathcal{E}|/k$ and variance $|\mathcal{E}|k^{-1}(1 - k^{-1})$. In a finite image lattice the distribution of $S(\mathbf{z})$ changes smoothly between these two extremes, as illustrated by Figure 1, but its computation is intractable for nontrivial images.

The Potts model undergoes a phase transition at the critical value of β , switching from a disordered to an ordered state. Potts (1952) showed that the critical value for a 2D lattice on a cylinder can be calculated exactly according to:

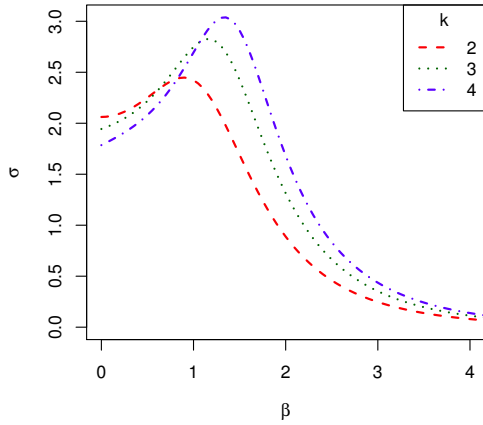
$$\beta_{crit} = \log \left\{ 1 + \sqrt{k} \right\} \quad (10)$$



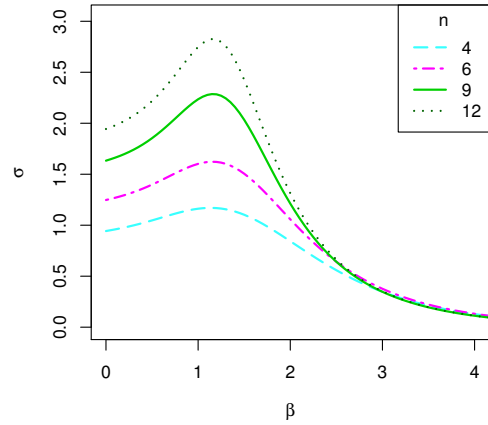
(a) Expectation for $n = 12$ and increasing values of $k \in \{2, 3, 4\}$.



(b) Expectation for $k = 3$ and increasing values of $n \in \{4, 6, 9, 12\}$.



(c) Standard deviation for $n = 12$ and increasing values of $k \in \{2, 3, 4\}$.



(d) Standard deviation for $k = 3$ and increasing values of $n \in \{4, 6, 9, 12\}$.

Figure 1: Distribution of the sufficient statistic of the Potts model for increasing values of the inverse temperature β , the number of pixels n , and the number of unique labels k . The expectation and standard deviation of $S(\mathbf{z})$ were calculated exactly, using a brute force method.

The periodic boundary condition does not apply to the lattices considered in this paper, so for example the critical value for the images in Figure 1 is different to (10). However, the error introduced by the finite boundary diminishes as n increases. Figure 2 shows that (10) is very accurate in predicting the behaviour of $S(\mathbf{z})$ for a 2D image with a maximum value of $S(\mathbf{z})$ for first-order neighbours of 1,954,672, with $k=6$ mixture components and $\beta_{crit} \approx 1.24$. $S(\mathbf{z})$ is approximated by simulation using the algorithm of Swendsen and Wang (1987). Figure 2a shows that the gradient of the expectation becomes very steep near the critical region, which is reflected in the super-exponential increase of the standard deviation illustrated by Figure 2c. As $n \rightarrow \infty$ the derivative of the likelihood at the critical point, and hence the variance, is unbounded (Pickard 1987). This heteroskedasticity has important implications for many of the methods discussed in Section 3.

There is no exact formula for β_{crit} in 3D images, although Hajduković (1983) developed an empirical approximation that works reasonably well:

$$\beta_{crit} \approx \frac{2}{3} \log \left\{ \frac{1}{2} \left(\sqrt{2} + \sqrt{4k - 2} \right) \right\} \quad (11)$$

The behaviour of $\mathbb{E}_{\mathbf{z}|\beta}[S(\mathbf{z})]$ for a 3D image with $|\mathcal{E}|$ approximately equal to three million, $k = 9$ and $\beta_{crit} \approx 0.86$ is illustrated by Figure 2b. The standard deviation is shown in Figure 2d. It is evident that (11) has underestimated β_{crit} since the maximum value of the standard deviation occurs at $\beta = 0.9$.

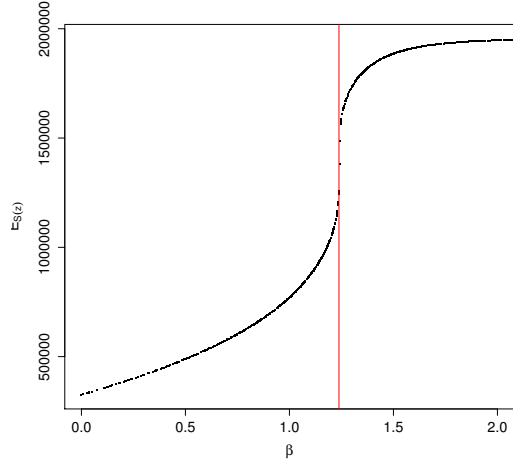
3. Computational methods

In this section we describe the four major algorithms for intractable likelihoods in Bayesian image analysis. These methods provide alternative means to simulate parameter values from (7) without computing the normalising constant. We describe the algorithms in terms of Markov chain Monte Carlo (MCMC) to enable direct comparison, although we also mention other approaches where applicable, such as particle-based (SMC and PMCMC) methods. Reference implementations of all of these methods are available from various sources described below, but for the purpose of comparison we have reimplemented the algorithms using **RcppArmadillo** (Eddelbuettel and Sanderson 2014).

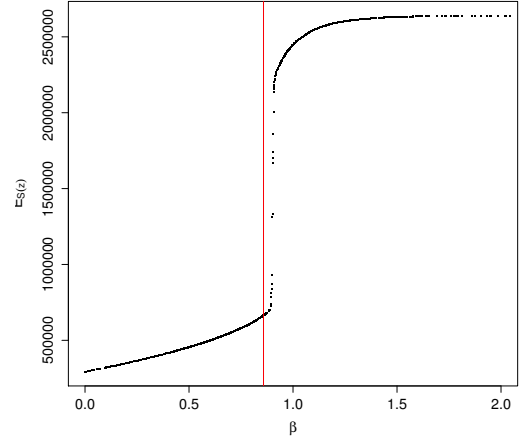
3.1. Pseudolikelihood and composite likelihood

Pseudolikelihood is the simplest of the methods that we have considered and also the fastest. Rydén and Titterton (1998) showed that the intractable distribution (7) could be approximated using the product of the conditional densities given by (1). This enables updates for the inverse temperature at iteration t to be simulated using a Metropolis-Hastings (M-H) step, as shown in Algorithm 1.

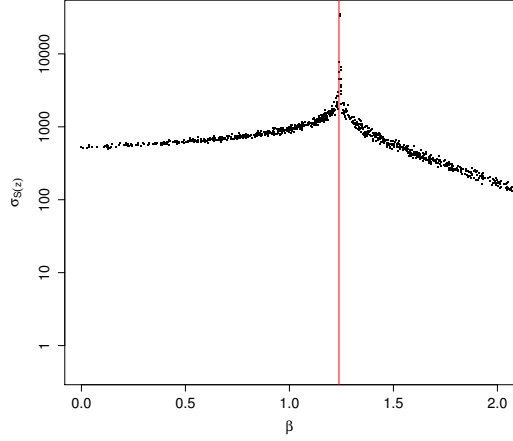
The M-H proposal density $q(\beta'|\beta_{t-1})$ can be any distribution such that $\int q(\beta'|\beta_{t-1}) d\beta' = 1$. However, there is a tradeoff between exploring the full state space and ensuring that the probability of acceptance (at line 10) is sufficiently high. We use the adaptive random walk (RWMH) algorithm of Garthwaite *et al.* (2016), which automatically tunes the bandwidth of the proposal density to target a given M-H acceptance rate. When a symmetric proposal density is used, $q(\beta'|\beta_{t-1}) = q(\beta_{t-1}|\beta')$ and so this term cancels out in the M-H ratio on



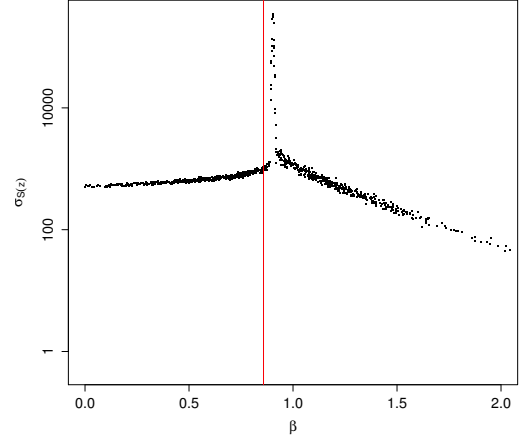
(a) Expectation for a 2D Potts model with $k = 6$.



(b) Expectation for a 3D Potts model with $k = 9$.



(c) Standard deviation for a 2D Potts model with $k = 6$.



(d) Standard deviation for a 3D Potts model with $k = 9$.

Figure 2: Approximate distribution of $S(\mathbf{z})$ using Swendsen-Wang for a 2D image with $k = 6$ and $|\mathcal{E}| \approx 2 \times 10^6$ (left) and a 3D image with $k = 9$ and $|\mathcal{E}| \approx 3 \times 10^6$ (right). The vertical line is the critical value of β .

Algorithm 1: MCMC with pseudolikelihood (PL)

```

1 Initialise  $\beta_0, \mu_0, \sigma_0^2, \mathbf{z}_0$ 
2 for  $t \in 1 \dots T$  do
3   Update the labels  $z_i \sim p(y_i|z_i, \mu_{z_i}, \sigma_{z_i}^2) p(z_i|z_{\setminus i}, \beta) \forall i \in \{1, \dots, n\}$ 
4   Calculate sufficient statistics  $\bar{y}_j, s_j^2 \forall z_i = j, \forall j \in \{1, \dots, k\}$ 
5   Update the noise parameters  $\mu_j, \sigma_j^2 \sim p(\bar{y}_j, s_j^2 | \mu_j, \sigma_j^2) \pi(\mu_j | \sigma_j^2) \pi(\sigma_j^2)$ 
6   Draw proposed parameter value  $\beta' \sim q(\beta' | \beta_{t-1})$ 
7   Approximate  $p(\mathbf{z} | \beta')$  and  $p(\mathbf{z} | \beta_{t-1})$  using:

```

$$p(\mathbf{z} | \beta) \approx \prod_{i=1}^n p(z_i | z_{\setminus i}, \beta) \quad (12)$$

```

8   Calculate the M-H ratio  $\rho = \min\left(1, \frac{p(\mathbf{z} | \beta') \pi(\beta') q(\beta_{t-1} | \beta')}{p(\mathbf{z} | \beta_{t-1}) \pi(\beta_{t-1}) q(\beta' | \beta_{t-1})}\right)$ 
9   Draw  $u \sim \text{Uniform}[0, 1]$ 
10  if  $u < \rho$  then
11     $\beta_t \leftarrow \beta'$ 
12  else
13     $\beta_t \leftarrow \beta_{t-1}$ 
14  end if
15 end for

```

line 8. The natural logarithm of ρ is used in practice to improve the numerical stability of Algorithm 1.

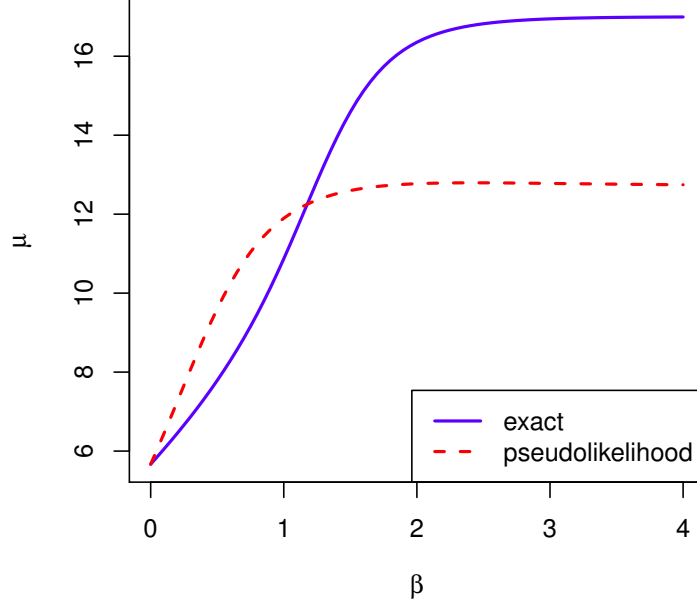
Pseudolikelihood is exact when $\beta = 0$ and provides a reasonable approximation for small values of the inverse temperature. However, the approximation error increases rapidly for $\beta \geq \beta_{crit}$, as illustrated by Figure 3. This is due to long-range dependence between the labels, which is inadequately modelled by the local approximation.

Rydén and Titterton (1998) referred to Equation (12) in Algorithm 1 as point pseudolikelihood, since the conditional distributions are computed for each pixel individually. They suggested that the accuracy could be improved using block pseudolikelihood. This is where the likelihood is calculated exactly for small blocks of pixels, then (12) is modified to be the product of the blocks:

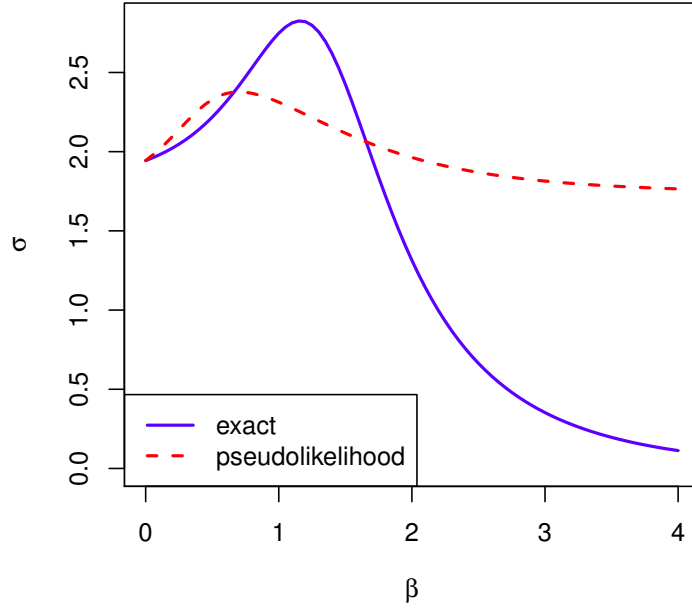
$$p(\mathbf{z} | \beta) \approx \prod_{i=1}^{N_B} p(\mathbf{z}_{B_i} | \mathbf{z}_{\setminus B_i}, \beta) \quad (13)$$

where N_B is the number of blocks, \mathbf{z}_{B_i} are the labels of the pixels in block B_i , and $\mathbf{z}_{\setminus B_i}$ are all of the labels except for \mathbf{z}_{B_i} . This is a form of composite likelihood, where the likelihood function is approximated as a product of simplified factors (Varin *et al.* 2011). Friel (2012) compared point pseudolikelihood to composite likelihood with blocks of 3×3 , 4×4 , 5×5 , and 6×6 pixels. Friel showed that (13) outperformed (12) for the Ising ($k = 2$) model with $\beta < \beta_{crit}$. Okabayashi *et al.* (2011) discuss composite likelihood for the Potts model with $k > 2$ and have provided an open source implementation in the R package **potts** (Geyer and Johnson 2014).

Evaluating the conditional likelihood in (13) involves the normalising constant for \mathbf{z}_{B_i} , which



(a) Expectation.



(b) Standard deviation.

Figure 3: Approximation error of pseudolikelihood for $n = 12$, $k = 3$ in comparison to the exact likelihood calculated using a brute force method: (a) $\sum_{\mathbf{z} \in \mathcal{Z}} S(\mathbf{z})p(S(\mathbf{z})|\beta)$ using either Equation (7) or (12); (b) $\sqrt{\sum_{\mathbf{z} \in \mathcal{Z}} (S(\mathbf{z}) - \mathbb{E}_{\mathbf{z}|\beta}[S(\mathbf{z})])^2 p(S(\mathbf{z})|\beta)}$

is a sum over all of the possible configurations \mathcal{Z}_{B_i} . This is a limiting factor on the size of blocks that can be used. The brute force method that was used to compute Figure 1 and 3 is too computationally intensive for this purpose. Pettitt *et al.* (2003) showed that the normalising constant can be calculated exactly for a cylindrical lattice by computing eigenvalues of a $k^r \times k^r$ matrix, where r is the smaller of the number of rows or columns. The value of (8) for a free boundary lattice can then be approximated using path sampling. Friel and Pettitt (2004) extended this method to larger lattices using a composite likelihood approach.

The reduced dependence approximation (RDA) is another form of composite likelihood. Reeves and Pettitt (2004) introduced a recursive algorithm to calculate the normalising constant using a lag- r representation. Friel *et al.* (2009) divided the image lattice into sub-lattices of size $r_1 < r$, then approximated the normalising constant of the full lattice using RDA:

$$\mathcal{C}(\beta) \approx \frac{\mathcal{C}_{r_1 \times n}(\beta)^{r-r_1+1}}{\mathcal{C}_{r_1-1 \times n}(\beta)^{r-r_1}} \quad (14)$$

McGrory *et al.* (2009) compared RDA to pseudolikelihood and the exact method of Møller *et al.* (2006), reporting similar computational cost to pseudolikelihood but with improved accuracy in estimating β . Source code for RDA is available in the online supplementary material for McGrory *et al.* (2012).

3.2. Thermodynamic integration

Gelman and Meng (1998) derived an approximation to the log ratio of normalising constants using the path sampling identity:

$$\log \left\{ \frac{\mathcal{C}(\beta_{t-1})}{\mathcal{C}(\beta')} \right\} = \int_{\beta'}^{\beta_{t-1}} \mathbb{E}_{\mathbf{z}|\beta}[\mathbf{S}(\mathbf{z})] d\beta \quad (15)$$

which follows from the definition of $\mathbb{E}_{\mathbf{z}|\beta}[\mathbf{S}(\mathbf{z})]$ in (9). The value of this definite integral can be approximated by simulating from the Gibbs distribution for fixed values of β and then interpolating between them. We used the Swendsen-Wang algorithm for simulating from $\mathbf{z}|\beta$, as mentioned in section 2. Figure 4a illustrates linear interpolation of $\mathbf{S}(\mathbf{z})$ on a 2D lattice for $k = 6$ and β ranging from 0 to 2 in increments of 0.05. This approximation can be precomputed using the same method that was used to produce Figure 2. Figure 4b illustrates interpolation of $\mathbf{S}(\mathbf{z})$ on a 3D lattice for $k = 9$.

TI is explained in further detail by Chen *et al.* (2000, chap. 5). A reference implementation in R is available from the website accompanying Marin and Robert (2007). This algorithm has an advantage over auxiliary variable methods such as AEA or ABC because the additional simulations are performed prior to fitting the model, rather than at each iteration. This is particularly the case when analysing multiple images that all have approximately the same dimensions. However, the computational cost is still slightly higher than PL, which does not require a pre-computation step.

3.3. Pseudo-marginal methods

Møller *et al.* (2006) demonstrated that it is possible to simulate from the posterior distribution of β by introducing an auxiliary variable, so that $\mathcal{C}(\beta)$ cancels out in the M-H ratio. The

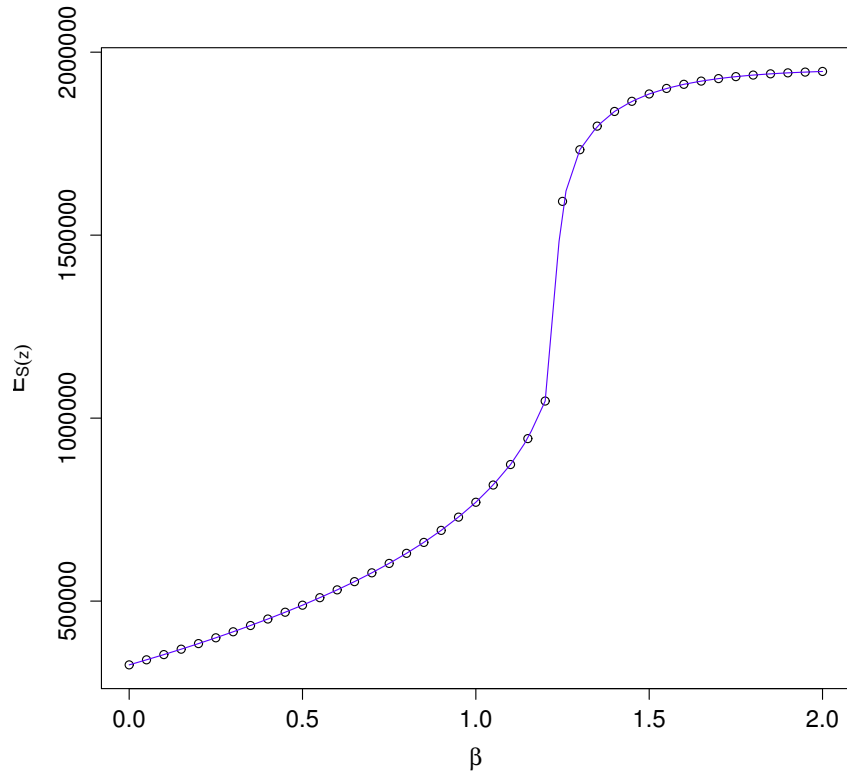
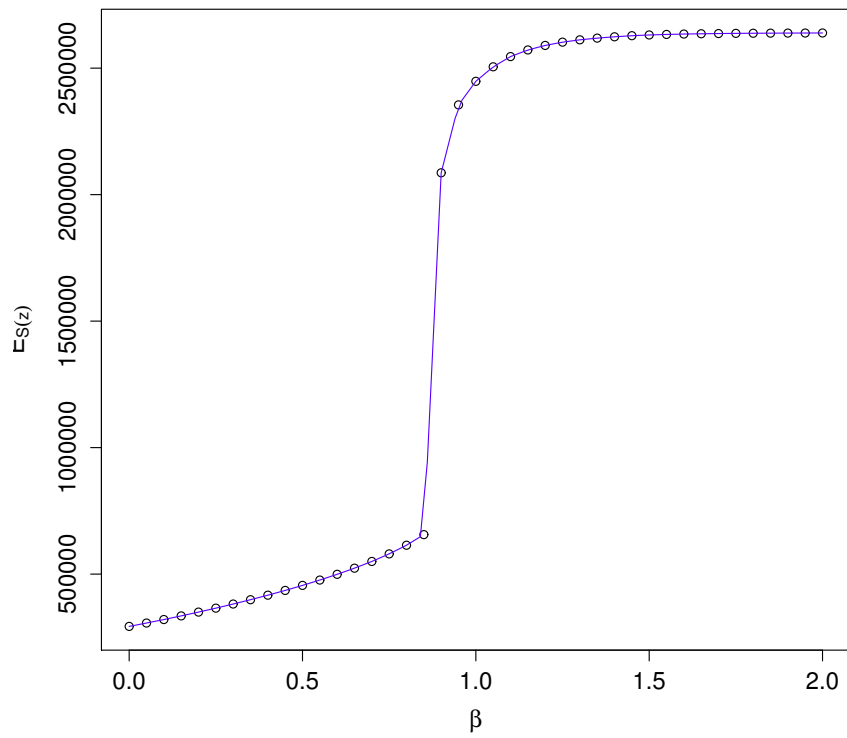
(a) 2D Potts model with $k = 6$.(b) 3D Potts model with $k = 9$.

Figure 4: Approximation of $\mathbb{E}_{\mathbf{z}|\beta}[\mathbf{S}(\mathbf{z})]$ by simulation for fixed values of β , with linear interpolation.

Algorithm 2: Path sampling (TI)

```

1  Initialise  $\beta_0, \mu_0, \sigma_0^2, \mathbf{z}_0$ 
2  for  $t \in 1 \dots T$  do
3    Update the labels  $z_i \sim p(y_i | z_i, \mu_{z_i}, \sigma_{z_i}^2) p(z_i | z_{\setminus i}, \beta) \forall i \in \{1, \dots, n\}$ 
4    Calculate sufficient statistics  $\bar{y}_j, s_j^2 \forall z_i = j, \forall j \in \{1, \dots, k\}$ 
5    Update the noise parameters  $\mu_j, \sigma_j^2 \sim p(\bar{y}_j, s_j^2 | \mu_j, \sigma_j^2) \pi(\mu_j | \sigma_j^2) \pi(\sigma_j^2)$ 
6    Draw proposed parameter value  $\beta' \sim q(\beta' | \beta_{t-1})$ 
7    Estimate  $\mathbb{E}_{\mathbf{z}|\beta}[\mathbf{S}(\mathbf{z})]$  for  $\beta \in \{\beta', \beta_{t-1}\}$  by interpolation
8    Evaluate the definite integral in Equation (15)
9    Calculate the log M-H acceptance ratio:


$$\log\{\rho\} = \min \left( 0, \log \left\{ \frac{\mathcal{C}(\beta_{t-1})}{\mathcal{C}(\beta')} \right\} + (\beta' - \beta_{t-1})\mathbf{S}(\mathbf{z}) + \log \left\{ \frac{\pi(\beta')q(\beta_{t-1}|\beta')}{\pi(\beta_{t-1})q(\beta'|\beta_{t-1})} \right\} \right) \quad (16)$$


10   Draw  $u \sim \text{Uniform}[0, 1]$ 
11   if  $u < \rho$  then
12      $\beta_t \leftarrow \beta'$ 
13   else
14      $\beta_t \leftarrow \beta_{t-1}$ 
15   end if
16 end for

```

exchange algorithm of [Murray et al. \(2006\)](#) improves the performance of this method and avoids the need for a fixed estimate of β . The drawback of this approach is that it requires perfect sampling from the stationary distribution of the Potts model. This is possible using coupling from the past ([Propp and Wilson 1996](#); [Huber 2016](#)) but can be computationally prohibitive, particularly for large images. [McGrory et al. \(2009\)](#) reported that the time required for perfect sampling increased sharply for larger values of β . For this reason, [Cucala et al. \(2009\)](#) substitute 500 iterations of Gibbs sampling on the auxiliary variable to produce an approximate sample from its stationary distribution. The details of this method are shown in Algorithm 3.

The M-H ratio given by Equation (17) can be considerably simplified if a uniform prior is used for β and β' is drawn from a symmetric M-H proposal density. In this case, $\log\{\rho\}$ simplifies to $(\beta' - \beta_{t-1})\mathbf{S}(\mathbf{z}) + (\beta_{t-1} - \beta')\mathbf{S}(\mathbf{w})$. The similarity with Equation (16) shows that the exchange algorithm is closely related to path sampling, since it is a form of importance sampling. An implementation of the exchange algorithm is available in the online supplementary material accompanying [Friel and Pettitt \(2011\)](#). [Everitt \(2012\)](#) provides source code for the approximate exchange algorithm with particle MCMC.

3.4. Approximate Bayesian computation

Like the exchange algorithm, ABC uses an auxiliary variable \mathbf{w} to decide whether to accept or reject the proposed value of β' . Instead of a M-H ratio such as (17), the summary statistics of the auxiliary variable and the observed data are directly compared. The proposal is accepted if the distance between these summary statistics is within the ABC tolerance, ϵ . This produces

Algorithm 3: Approximate exchange algorithm (AEA)

```

1 Initialise  $\beta_0, \mu_0, \sigma_0^2, \mathbf{z}_0$ 
2 for  $t \in 1 \dots T$  do
3   Update the labels  $z_i \sim p(y_i | z_i, \mu_{z_i}, \sigma_{z_i}^2) p(z_i | z_{\setminus i}, \beta) \forall i \in \{1, \dots, n\}$ 
4   Calculate sufficient statistics  $\bar{y}_j, s_j^2 \forall z_i = j, \forall j \in \{1, \dots, k\}$ 
5   Update the noise parameters  $\mu_j, \sigma_j^2 \sim p(\bar{y}_j, s_j^2 | \mu_j, \sigma_j^2) \pi(\mu_j | \sigma_j^2) \pi(\sigma_j^2)$ 
6   Draw proposed parameter value  $\beta' \sim q(\beta' | \beta_{t-1})$ 
7   Generate  $\mathbf{w} | \beta'$  by sampling from Equation (1)
8   Calculate the M-H acceptance ratio according to Equation (7):


$$\rho = \min \left( 1, \frac{q(\beta_{t-1} | \beta') \pi(\beta') \exp \{ \beta' S(\mathbf{z}) \} \mathcal{C}(\beta_{t-1})}{q(\beta' | \beta_{t-1}) \pi(\beta_{t-1}) \exp \{ \beta_{t-1} S(\mathbf{z}) \} \mathcal{C}(\beta')} \frac{\exp \{ \beta_{t-1} S(\mathbf{w}) \} \mathcal{C}(\beta')}{\exp \{ \beta' S(\mathbf{w}) \} \mathcal{C}(\beta_{t-1})} \right) \quad (17)$$


9   Draw  $u \sim \text{Uniform}[0, 1]$ 
10  if  $u < \rho$  then
11     $\beta_t \leftarrow \beta'$ 
12  else
13     $\beta_t \leftarrow \beta_{t-1}$ 
14  end if
15 end for

```

the following approximation when the labels \mathbf{z} are observed without error:

$$p(\beta | \mathbf{z}) \approx \pi_\epsilon(\beta | \|\mathbf{S}(\mathbf{w}) - \mathbf{S}(\mathbf{z})\| < \epsilon) \quad (18)$$

where $\|\cdot\|$ is a suitable norm. In this paper we simply use the absolute difference between $\mathbf{S}(\mathbf{w})$ and $\mathbf{S}(\mathbf{z})$, since the summary statistic (4) is univariate. $\mathbf{S}(\mathbf{z})$ is sufficient for β , as noted by Grelaud *et al.* (2009), therefore the ABC approximation (18) approaches the true posterior as $n \rightarrow \infty$ and $\epsilon \rightarrow 0$. In practice there is a tradeoff between the number of parameter values that are accepted and the size of the ABC tolerance.

Everitt (2012) introduced ABC for noisy data, such as the Ising model, but only when the observations are discrete and so $\mathbf{S}(\mathbf{y})$ is defined. Moores *et al.* (2015) showed that ABC can also be applied when the domain of the observed data is continuous, such as with the additive Gaussian noise of Equation (3). In this type of model the posterior distribution of β does not depend directly on the observed data \mathbf{y} , as shown by Equation (6). The ABC approximation (18) for the hidden Potts model becomes:

$$p(\beta | \mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) \approx \pi_\epsilon(\beta | \|\mathbf{S}(\mathbf{w}) - \mathbf{S}(\mathbf{z})\| < \epsilon) \quad (19)$$

The other parameters can be updated using the current value of β and then the summary statistic $\mathbf{S}(\mathbf{z})$ can be computed from the current values of the labels, using an ABC within Gibbs approach as shown in Algorithm 4. Equation (19) can be considered as a moving target, since $\mathbf{S}(\mathbf{z})$ could change whenever the labels are updated. As a consequence, ABC with noisy data can be more prone to getting stuck in low probability regions of the parameter space.

Grelaud *et al.* (2009) used a rejection sampler, where the proposals are drawn independently from the prior and thus $q(\beta' | \beta_{t-1}) = \pi(\beta)$. Under a sparse or uninformative prior, such as

Algorithm 4: ABC-MCMC with noisy data

```

1 Initialise  $\beta_0, \mu_0, \sigma_0^2, \mathbf{z}_0$ 
2 for  $t \in 1 \dots T$  do
3   Update the labels  $z_i \sim p(y_i | z_i, \mu_{z_i}, \sigma_{z_i}^2) p(z_i | z_{\setminus i}, \beta) \forall i \in \{1, \dots, n\}$ 
4   Calculate sufficient statistics  $\bar{y}_j, s_j^2 \forall z_i = j, \forall j \in \{1, \dots, k\}$ 
5   Update the noise parameters  $\mu_j, \sigma_j^2 \sim p(\bar{y}_j, s_j^2 | \mu_j, \sigma_j^2) \pi(\mu_j | \sigma_j^2) \pi(\sigma_j^2)$ 
6   Draw proposed parameter value  $\beta' \sim q(\beta' | \beta_{t-1})$ 
7   Generate  $\mathbf{w} | \beta'$  by sampling from Equation (1)
8   Draw  $u \sim \text{Uniform}[0, 1]$ 
9   if  $u < \frac{\pi(\beta') q(\beta_{t-1} | \beta')}{\pi(\beta_{t-1}) q(\beta' | \beta_{t-1})}$  and  $\|\mathbf{S}(\mathbf{w}) - \mathbf{S}(\mathbf{z})\| < \epsilon$  then
10     $\beta_t \leftarrow \beta'$ 
11  else
12     $\beta_t \leftarrow \beta_{t-1}$ 
13  end if
14 end for

```

the uniform prior used in this study, too many proposals are rejected for this approach to be viable. Instead we have based Algorithm 4 on the ABC-MCMC algorithm of [Marjoram *et al.* \(2003\)](#). This form of ABC algorithm is best suited for direct comparison with the other intractable likelihood methods in this article.

There have been many extensions to ABC, as reviewed by [Marin *et al.* \(2012\)](#). Of particular interest are adaptive ABC algorithms that automatically adjust the tolerance ϵ and the proposal bandwidth σ_{MH}^2 . ABC with sequential Monte Carlo (SMC-ABC) algorithms use a sequence of target distributions $\pi_{\epsilon_t}(\beta \mid \|\mathbf{S}(\mathbf{w}) - \mathbf{S}(\mathbf{z})\| < \epsilon_t)$ such that $\epsilon_1 > \epsilon_2 > \dots > \epsilon_T$, where the number of SMC iterations T can be determined dynamically using a stopping rule. The SMC-ABC algorithm of [Drovandi and Pettitt \(2011\)](#) uses multiple MCMC steps for each SMC iteration, while the algorithm of [Del Moral *et al.* \(2012\)](#) uses multiple replicates of the summary statistics for each particle. [Everitt \(2012\)](#) has provided a MATLAB implementation of SMC-ABC with the online supplementary material accompanying his paper.

The computational efficiency of ABC is dominated by the cost of drawing updates to the auxiliary variable, as reported by [Everitt \(2012\)](#). Thus, we would expect that the execution time for ABC would be similar to the approximate exchange algorithm. Various approaches to improving this runtime have recently been proposed, such as the Gaussian process emulation of [Wilkinson \(2014\)](#), the “lazy ABC” of [Prangle \(2016\)](#), and methods involving auxiliary models ([Cabras *et al.* 2014](#); [Moores *et al.* 2015](#); [Buzbas and Rosenberg 2015](#)).

References

- Alston CL, Mengersen KL, Robert CP, Thompson JM, Littlefield PJ, Perry D, Ball AJ (2007). “Bayesian mixture models in a longitudinal setting for analysing sheep CAT scan images.” *Comput. Stat. Data Anal.*, **51**, 4282–4296. doi:10.1016/j.csda.2006.05.013.

- Besag J (1974). “Spatial interaction and the statistical analysis of lattice systems.” *J. R. Stat. Soc. Ser. B*, **36**(2), 192–236.
- Besag J (1975). “Statistical Analysis of Non-Lattice Data.” *J. R. Stat. Soc. Ser. D*, **24**(3), 179–195.
- Buzbas EO, Rosenberg NA (2015). “AABC: Approximate approximate Bayesian computation for inference in population-genetic models.” *Theor. Popul. Biol.*, **99**, 31–42. doi:[10.1016/j.tpb.2014.09.002](https://doi.org/10.1016/j.tpb.2014.09.002).
- Cabras S, Castellanos M, Ruli E (2014). “A Quasi likelihood approximation of posterior distributions for likelihood-intractable complex models.” *METRON*, **72**(2), 153–167. doi:[10.1007/s40300-014-0040-5](https://doi.org/10.1007/s40300-014-0040-5).
- Chen MH, Shao QM, Ibrahim JG (2000). *Monte Carlo Methods in Bayesian Computation*. Springer Series in Statistics. Springer-Verlag, New York.
- Cucala L, Marin JM, Robert CP, Titterton DM (2009). “A Bayesian Reassessment of Nearest-Neighbor Classification.” *J. Am. Stat. Assoc.*, **104**(485), 263–273. doi:[10.1198/jasa.2009.0125](https://doi.org/10.1198/jasa.2009.0125).
- Del Moral P, Doucet A, Jasra A (2012). “An adaptive sequential Monte Carlo method for approximate Bayesian computation.” *Stat. Comput.*, **22**(5), 1009–20. doi:[10.1007/s11222-011-9271-y](https://doi.org/10.1007/s11222-011-9271-y).
- Drovandi CC, Pettitt AN (2011). “Estimation of Parameters for Macroparasite Population Evolution Using Approximate Bayesian Computation.” *Biometrics*, **67**(1), 225–233. doi:[10.1111/j.1541-0420.2010.01410.x](https://doi.org/10.1111/j.1541-0420.2010.01410.x).
- Eddelbuettel D, Sanderson C (2014). “RcppArmadillo: Accelerating R with high-performance C++ linear algebra.” *Comput. Stat. Data Anal.*, **71**, 1054–63. doi:[10.1016/j.csda.2013.02.005](https://doi.org/10.1016/j.csda.2013.02.005).
- Everitt RG (2012). “Bayesian Parameter Estimation for Latent Markov Random Fields and Social Networks.” *J. Comput. Graph. Stat.*, **21**(4), 940–960. doi:[10.1080/10618600.2012.687493](https://doi.org/10.1080/10618600.2012.687493).
- Friel N (2012). “Bayesian inference for Gibbs random fields using composite likelihoods.” In C Laroque, J Himmelsbach, R Pasupathy, O Rose, AM Uhrmacher (eds.), *Proc. Winter Simulation Conference*, pp. 1–8. doi:[10.1109/WSC.2012.6465236](https://doi.org/10.1109/WSC.2012.6465236).
- Friel N, Pettitt AN (2004). “Likelihood Estimation and Inference for the Autologistic Model.” *J. Comp. Graph. Stat.*, **13**(1), 232–246. doi:[10.1198/1061860043029](https://doi.org/10.1198/1061860043029).
- Friel N, Pettitt AN (2011). “Classification using distance nearest neighbours.” *Stat. Comput.*, **21**(3), 431–437. doi:[10.1007/s11222-010-9179-y](https://doi.org/10.1007/s11222-010-9179-y).
- Friel N, Pettitt AN, Reeves R, Wit E (2009). “Bayesian Inference in Hidden Markov Random Fields for Binary Data Defined on Large Lattices.” *J. Comp. Graph. Stat.*, **18**(2), 243–261. doi:[10.1198/jcgs.2009.06148](https://doi.org/10.1198/jcgs.2009.06148).

- Garthwaite PH, Fan Y, Sisson SA (2016). “Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro process.” *Commun. Stat. Theory*. doi:10.1080/03610926.2014.936562.
- Gelman A, Meng XL (1998). “Simulating normalizing constants: from importance sampling to bridge sampling to path sampling.” *Statist. Sci.*, **13**(2), 163–185. doi:10.1214/ss/1028905934.
- Geman S, Geman D (1984). “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images.” *IEEE Trans. PAMI*, **6**, 721–41.
- Geyer CJ, Johnson L (2014). *potts: Markov Chain Monte Carlo for Potts Models*. R package version 0.5-2, URL <http://CRAN.R-project.org/package=potts>.
- Green PJ, Richardson S (2002). “Hidden Markov models and disease mapping.” *J. Am. Stat. Assoc.*, **97**, 1055–1070.
- Grelaud A, Robert CP, Marin JM, Rodolphe F, Taly JF (2009). “ABC likelihood-free methods for model choice in Gibbs random fields.” *Bayesian Analysis*, **4**(2), 317–336. doi:10.1214/09-BA412.
- Hajduković D (1983). “On the location of the critical point of the q-state Potts model on the hypercubic lattice.” *J. Phys. A: Math. Gen.*, **16**(6), L193. doi:10.1088/0305-4470/16/6/004.
- Huber ML (2016). *Perfect Simulation*. Monographs on Statistics & Applied Probability. Chapman & Hall/CRC Press.
- Li SZ (2009). *Markov Random Field Modeling in Image Analysis*. 3rd edition. Springer, Dordrecht.
- Marin JM, Pudlo P, Robert CP, Ryder RJ (2012). “Approximate Bayesian computational methods.” *Stat. Comput.*, **22**(6), 1167–1180. doi:10.1007/s11222-011-9288-2.
- Marin JM, Robert CP (2007). *Bayesian core: a practical approach to computational Bayesian statistics*. Springer Texts in Statistics. Springer, New York.
- Marjoram P, Molitor J, Plagnol V, Tavaré S (2003). “Markov chain Monte Carlo without likelihoods.” *Proc. Natl Acad. Sci.*, **100**(26), 15324–15328. doi:10.1073/pnas.0306899100.
- McGrory CA, Pettitt AN, Reeves R, Griffin M, Dwyer M (2012). “Variational Bayes and the Reduced Dependence Approximation for the Autologistic Model on an Irregular Grid With Applications.” *J. Comput. Graph. Stat.*, **21**(3), 781–796. doi:10.1080/10618600.2012.632232.
- McGrory CA, Titterton D, Reeves R, Pettitt AN (2009). “Variational Bayes for estimating the parameters of a hidden Potts model.” *Stat. Comput.*, **19**(3), 329–340. doi:10.1007/s11222-008-9095-6.
- Møller J, Pettitt AN, Reeves R, Berthelsen KK (2006). “An Efficient Markov Chain Monte Carlo Method for Distributions with Intractable Normalising Constants.” *Biometrika*, **93**(2), 451–458. doi:10.1093/biomet/93.2.451.

- Moore MT, Drovandi CC, Mengersen K, Robert CP (2015). “Pre-processing for approximate Bayesian computation in image analysis.” *Stat. Comput.*, **25**(1), 23–33. doi:10.1007/s11222-014-9525-6.
- Murray I, Ghahramani Z, MacKay DJC (2006). “MCMC for Doubly-intractable Distributions.” In *Proc. 22nd Conf. UAI*, pp. 359–366. AUAI Press, Arlington, VA.
- Okabayashi S, Johnson L, Geyer CJ (2011). “Extending pseudo-likelihood for Potts models.” *Statistica Sinica*, **21**, 331–347.
- Pettitt AN, Friel N, Reeves R (2003). “Efficient calculation of the normalizing constant of the autologistic and related models on the cylinder and lattice.” *J. R. Stat. Soc. Ser. B*, **65**(1), 235–246. doi:10.1111/1467-9868.00383.
- Pickard DK (1987). “Inference for Discrete Markov Fields: The Simplest Nontrivial Case.” *J. Am. Stat. Assoc.*, **82**(397), 90–96. doi:10.1080/01621459.1987.10478394.
- Potts RB (1952). “Some generalized order-disorder transformations.” *Proc. Camb. Philos. Soc.*, **48**, 106–9. doi:10.1017/S0305004100027419.
- Prangle D (2016). “Lazy ABC.” *Stat. Comput.*, **26**(1), 171–185. doi:10.1007/s11222-014-9544-3.
- Propp JG, Wilson DB (1996). “Exact sampling with coupled Markov chains and applications to statistical mechanics.” *Random Struct. Algor.*, **9**(1–2), 223–252. doi:10.1002/(SICI)1098-2418(199608/09)9:1/2<223::AID-RSA14>3.0.CO;2-0.
- Reeves R, Pettitt AN (2004). “Efficient Recursions for General Factorisable Models.” *Biometrika*, **91**(3), 751–757. doi:10.1093/biomet/91.3.751.
- Rydén T, Titterton DM (1998). “Computational Bayesian Analysis of Hidden Markov Models.” *J. Comput. Graph. Stat.*, **7**(2), 194–211. doi:10.1080/10618600.1998.10474770.
- Swendsen RH, Wang JS (1987). “Nonuniversal critical dynamics in Monte Carlo simulations.” *Phys. Rev. Lett.*, **58**, 86–88. doi:10.1103/PhysRevLett.58.86.
- Varin C, Reid N, Firth D (2011). “An overview of composite likelihood methods.” *Statistica Sinica*, **21**, 5–42.
- Wilkinson RD (2014). “Accelerating ABC methods using Gaussian processes.” In S Kaski, J Corander (eds.), *Proc. 17th Int. Conf. AISTATS*, volume 33 of *JMLR W&CP*, pp. 1015–1023.
- Winkler G (2003). *Image Analysis, Random Fields and Markov chain Monte Carlo Methods: A Mathematical Introduction*. 2nd edition. Springer-Verlag, Berlin Heidelberg.

Affiliation:

Matthew T. Moores
Department of Statistics

University of Warwick
Coventry CV4 7AL, United Kingdom
E-mail: M.T.Moores@warwick.ac.uk
URL: <http://warwick.ac.uk/mmoores>