Brazilian Journal of Probability and Statistics
2019, Vol. 33, No. 3, 498–519
https://doi.org/10.1214/18-BJPS399

Spatially adaptive Bayesian image reconstruction through locally-modulated Markov random field models

Salem M. Al-Gezeri[a] **and Robert G. Aykroyd**[b]

[a]*University of Benghazi*
[b]*University of Leeds*

Abstract. The use of Markov random field (MRF) models has proven to be a fruitful approach in a wide range of image processing applications. It allows local texture information to be incorporated in a systematic and unified way and allows statistical inference theory to be applied giving rise to novel output summaries and enhanced image interpretation. A great advantage of such low-level approaches is that they lead to flexible models, which can be applied to a wide range of imaging problems without the need for significant modification.

This paper proposes and explores the use of conditional MRF models for situations where multiple images are to be processed simultaneously, or where only a single image is to be reconstructed and a sequential approach is taken. Although the *coupling* of image intensity values is a special case of our approach, the main extension over previous proposals is to allow the direct coupling of other properties, such as smoothness or texture. This is achieved using a local modulating function which adjusts the influence of global smoothing without the need for a fully inhomogeneous prior model. Several modulating functions are considered and a detailed simulation study, motivated by remote sensing applications in archaeological geophysics, of conditional reconstruction is presented. The results demonstrate that a substantial improvement in the quality of the image reconstruction, in terms of errors and residuals, can be achieved using this approach, especially at locations with rapid changes in the underlying intensity.

1 Introduction

Usually, the first step in any image analysis application is image reconstruction. The use of Bayesian modelling with the incorporation of prior information into the reconstruction was first suggested more than 30 years ago by Besag (1986) and Geman and Geman (1984)—excellent collected works of the key papers appear in Mardia and Kanji (1993) and Mardia (1994). Although the use of homogeneous Gibbs prior models quickly became widely accepted they are often not sufficient. This is particularly true when the true scene contains regions with a wide variety of textures or when it contains features with distinct intensity compared to a background. A common consequence of the inappropriate use of homogeneous prior

Key words and phrases. Bayesian model, image reconstruction, inverse problems, magnetometry, Markov chain Monte Carlo, prior models.

Received June 2017; accepted April 2018.

models is that some interesting structures are removed during noise suppression. In such cases, it is preferable to use different levels of smoothing for different regions in the image, that is to use a locally adaptive prior model (see, for example, Aykroyd (1998) or Aykroyd and Zimeras (1999)).

There have been several approaches to adaptive smoothing based on the use of coupled MRF distributions (Melas and Wilson (1997)). In one version, a hidden MRF model is used to describe edges between regions (Besag (1986)). This allows local interaction between neighbours in the image MRF itself to be reduced or removed. In a similar way, region labels can be modelled with an MRF, in terms of categorical variables, allowing image segmentation and then separate MRF models used to allow different textures within each region. Alternatively, it is possible to describe the smoothing levels between each neighbouring pair in the main MRF using a second MRF model (Aykroyd (1998)). This allows for more complex texture patterns including continuously varying smoothing.

In all these cases the coupled MRF models refer to the same image. The main proposal in this paper is to couple two MRF models which are acting on different images. Hence, two images are being reconstructed simultaneously with the help of exogenous information from the other image. In particular, it is assumed that there is some known relationship between the level of some local characteristic in the two images. That is, for example, smooth regions in one image correspond to smooth regions in the other, and rough regions in one correspond to rough regions in the other. Such examples include multi-sensor satellite data (see, for example, Dousset and Gourmelon (2003)), polarimetric imaging (see, for example, Hu et al. (2018)). The location of these regions, however, is not known a priori but emerges as part of the reconstruction process. The modelling can also be applied to the situation where a map of the local characteristic is available from some other source, such as the case with PET/MR (Positron emission tomography combined with magnetic resonance) or PET/CT (PET combined with X-ray computed tomography) in medical imaging (see, for example, Vandenberghe and Marsden (2015)). In these cases, the noisy PET information about biological function can be enhanced by the low-noise anatomical information from MRI or CT (see, for example, Ehrhardt et al. (2016)). This reduces the joint prior modelling to a conditional prior model given this exogenous information.

This paper is structured as follows. Section 2 provides background to the general modelling approach based on a standard Markov random field prior distribution. Section 3 contains details of the proposed modulated Markov random field model including general properties of the modulating function with specific examples. Brief details of parameter estimation using the Markov chain Monte Carlo algorithm are given in Section 4 and results from a detailed simulation study, based on the archaeological prospecting technique of magnetometry, is presented in Section 5. The final summary and conclusions are presented in Section 6.

2 Image reconstruction using homogeneous MRF models

Suppose that we are interested in some physical quantity, X, which varies within some study region, $\mathcal{R} \subset R^3$. Let $\mathbf{X} = (X_i : i = 1, 2, \ldots, m)$ be the value at a finite number of locations $(\mathbf{s}_i : i = 1, 2, \ldots, m)$. Suppose also that data are measured at n locations $\mathbf{s}_j : j = 1, 2, \ldots, n$ from some region $\mathcal{S} \subset R^3$, and are denoted $\mathbf{Y} = (Y_j : j = 1, 2, \ldots, n)$. Note that there is no restriction on the relative values of n and m, and the estimation and data locations need not be the same. In particular, the locations may be from disjoint subsets of the study region. This makes the approach valid for remote sensing applications in environmental science or medicine as much as for cases where the two sets of locations are identical and the problem is one of image de-blurring or de-noising.

In this paper, we consider a class of imaging problems in which the data values $\mathbf{Y}$ are a degraded version of the true image $\mathbf{X}$ defined by the relationship

$$\mathbf{Y} = H\mathbf{X} + \boldsymbol{\varepsilon}, \tag{1}$$

where $E[\mathbf{Y}] = H\mathbf{X}$ is a corresponding blurred version but without noise. The blurring is due to the convolution with a known transfer matrix H whose elements are defined in terms of a point spread function, h, which depends only on the vector distance between the locations $H = [h_{ij}]_{m\times n} = h(\mathbf{s}_i - \mathbf{s}_j)$. The error term $\boldsymbol{\varepsilon}$ represents the noise component of the data which is appropriately represented by a Gaussian error model with zero mean and covariance matrix $\Sigma = [\sigma_{jj'}]_{n\times n}$. Often the errors are also assumed to be independent with constant variance and so $\Sigma = \sigma^2 I_n$. Thus, in general, $\mathbf{Y}$ is distributed normally with likelihood,

$$L(\mathbf{x}) = \frac{1}{\sqrt{|2\pi\,\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{y} - H\mathbf{x})^T \Sigma^{-1}(\mathbf{y} - H\mathbf{x})\right\}, \qquad \mathbf{x} \in R^m. \tag{2}$$

The usual maximum likelihood estimator of $\mathbf{x}$ from the data $\mathbf{y}$ is $\hat{\mathbf{x}} = (H^T \Sigma^{-1} H)^{-1} H^T \Sigma^{-1}\mathbf{y}$. In the case of independent and identically distributed data, that is with $\Sigma = \sigma^2 I_n$, this reduces to

$$L(\mathbf{x}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - H\mathbf{x})^T (\mathbf{y} - H\mathbf{x})\right\}, \qquad \mathbf{x} \in R^m. \tag{3}$$

In many image processing applications $m \gg n$ and so this linear inverse problem is ill-posed, and maximum likelihood estimation is not possible. If the number of unknown parameters, however, is fewer than the number of data values, that is $m < n$, but there is correlation between some rows, or between some columns, then the problem is ill-conditioned and maximum likelihood estimation is still not reliable—this problem is also known as multicollinearity in the regression literature. One approach is to introduce additional information and hence to regularise the problem. This can be done in a statistical framework where the regularisation is achieved using a prior distribution which describes the extra information using a probability model. The above likelihood is combined with the prior distribution,

using Bayes' theorem, to produce a posterior distribution—which is then used as the basis for estimation and inference.

In Bayesian image analysis, Markov random field models are widely used to model prior information in terms of local characteristics (Cheng and Huang (1993), Cross and Jain (1983), Derin and Elliott (1987), Kinderman and Snell (1992)). The underlying true scene represents the unknown distribution of some spatially varying process which is modelled by a prior distribution $\pi(\mathbf{x}|\beta)$. This prior information is commonly quantified in the form of a *Gibbs or Boltzmann distribution*, which reflects only the local behaviour of the true scene, with probability function of the form

$$\pi(\mathbf{x}|\beta) = \frac{1}{Z_\beta} \exp\{-\beta V(\mathbf{x})\}, \qquad \mathbf{x} \in R^m, \beta > 0. \tag{4}$$

The non-negative constant β is a smoothing parameter which controls the degree of correlation between neighbouring pixels and determines the level of influence the prior has in the posterior distribution and Z_β is a normalising constant such that $Z_\beta = \int_{\mathbf{x}} \exp\{-\beta V(\mathbf{x})\}\, d\mathbf{x}$. In general, it is not possible to evaluate Z_β as the high dimensional integral is intractable. It is possible, however, to use numerical approximations, see, for example, Aykroyd (2002), as defined below. Such distributions have their origins in statistical mechanics where they were first used to describe the thermodynamic energy of gaseous molecules.

In the statistical mechanics literature function $V(\mathbf{x})$ is called the *energy function*, which is designed to assign high probabilities to expected configurations which have *low energy*. A common form for the energy is as the sum of local energy contributions and further that the local energy be simply a spatially invariant function, v, applied to a local neighbourhood, and so

$$V(\mathbf{x}) = \sum_{i=1}^{m} v_i(\mathbf{x}) = \sum_{i=1}^{m} v\big(\mathbf{x}[\partial_i]\big), \tag{5}$$

where ∂_i is the set of pixels in the neighbourhood of i, $\mathbf{x}[\partial_i]$ is the set of pixel values in this neighbourhood and the sum is over all pixels. In the first-order neighbourhood the four nearest neighbours are used, that is the pixel to the left, right, above and below, whereas in the second-order neighbourhood the four diagonal nearest neighbours are also included. Further, it is common for the local energies only to depend on local differences, that is

$$v\big(\mathbf{x}[\partial_i]\big) = \sum_{j \in \partial_i} w_{ij} \phi(x_i - x_j), \tag{6}$$

where ϕ is a *potential* function and w_{ij} allows different comparisons to be given different weight. The only required properties of the potential function are that it is always non-negative, $\phi(u) > 0$ for all u, symmetric, $\phi(u) = \phi(-u)$, and monotonic increasing in $|u|$, that is $\phi(u_2) > \phi(u_1)$ if $|u_2| > |u_1|$. Taken together these ensure consistency of the corresponding MRF (Besag (1986)).

In particular, here a second order neighbourhood is used, with the four nearest neighbours receiving weight 1 and the four diagonal neighbours receiving weight $1/\sqrt{2}$. The most widely used potential function, which will be included in the later numerical study, is the quadratic

$$\phi(u) = u^2, \qquad -\infty < u < \infty,$$

leading to a Gaussian distribution. Also popular is the potential function $\phi(u) = |u|$ giving rise to the Laplace distribution. These functions, however, have been shown to lead to poor estimation for cases where the truth resembles a piece-wise constant function—see, for example, Geman and McClure (1987), Green (1990), Weir (1997) and Allum (1997). Here, the implicit-disontinuity prior proposed by Allum (1997) is proposed, which has potential defined as

$$\phi(u) = \begin{cases} |u| - \dfrac{\kappa - 1}{4\kappa} u^2, & |u| < 2, \\ \dfrac{a}{\kappa}|u| + \dfrac{\kappa - 1}{\kappa}, & |u| \geq 2, \end{cases}$$

where the parameter κ is arbitrarily chosen as a large positive number so that the function is similar to the median for small values of $|u|$ and a low gradient linear increase for values above the threshold. The function is strictly monotonic increasing with continuous derivative.

Although the strict monotonicity means that the corresponding density has a finite integral with respect to u, there is no closed form expression. However, using a numerical approximation (Aykroyd (2002)) for the normalising constant, $1/Z_\beta = \beta^m/\mathcal{K}$ where $\mathcal{K}$ is a constant, leads to the approximate prior distribution

$$\begin{aligned} \pi(\mathbf{x}|\beta) &= \frac{1}{Z_\beta} \exp\{-\beta V(\mathbf{x})\} \\ &= \frac{\beta^m}{\mathcal{K}} \exp\left\{-\beta \sum_{i=1}^{m} \sum_{j \in \partial_i} w_{ij} \phi(x_i - x_j)\right\}, \qquad \mathbf{x} \in R^m, \beta > 0. \end{aligned}$$

It is important to notes that this is not a proper distribution with respect to $\mathbf{x}$ as any constant added to $\mathbf{x}$ does not change the value of $\pi(\mathbf{x}|\beta)$. This has no impact on posterior estimation, but an alteration would be needed to allow simulation from the prior (see, for example, Aykroyd (2002)).

The unknown prior parameter, β, is in turn modelled though a hyper-prior distribution, $\pi(\beta)$. Here, the prior should favour large values of β, and hence the following density has been used

$$\pi(\beta) = \frac{1}{Z_\lambda} \exp\{-\lambda/\beta\}, \qquad \beta > 0, \lambda > 0.$$

The normalising constant is defined by $Z_\lambda = \int_\beta \exp\{-\lambda/\beta\}\, d\beta$. This has introduced an additional unknown parameter, λ and, although it is possible to model

this via an additional hyper-prior distribution, here the value will be fixed during preliminary analysis.

The posterior distribution, $\pi(\mathbf{x}, \beta|\mathbf{y})$ can be written in terms of the likelihood and the prior distributions using Bayes' theorem

$$\pi(\mathbf{x}, \beta|\mathbf{y}) = f(\mathbf{y}|\mathbf{x})\pi(\mathbf{x}|\beta)\pi(\beta)/f(\mathbf{y})$$
$$\propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - H\mathbf{x})^T(\mathbf{y} - H\mathbf{x}) - \beta V(\mathbf{x}) + m\log\beta - \lambda/\beta\right\},$$

after noting that $f(\mathbf{y}|\mathbf{x}) \equiv L(\mathbf{x})$, and dropping constant terms which do not involve the unknown parameters $\mathbf{x}$ and β. Inference now involves the simultaneous estimation of $\mathbf{x}$ and β from $\pi(\mathbf{x}, \beta|\mathbf{y})$, for example, using the marginal posterior means $\hat{\mathbf{x}} = E[\mathbf{x}|\mathbf{y}]$ and $\hat{\beta} = E[\beta|\mathbf{y}]$, or the joint maximum a posterior estimate $(\hat{\mathbf{x}}, \hat{\beta}) = \arg\max_{\mathbf{x},\beta}\pi(\mathbf{x}, \beta|\mathbf{y})$—the former approach will be discussed in later sections.

3 Image reconstruction using modulated MRF models

3.1 Joint estimation

In this section, we extend the previous description to consider two images $\mathbf{x}_1$ and $\mathbf{x}_2$ which are to be reconstructed simultaneously from two datasets $\mathbf{y}_1$ and $\mathbf{y}_2$. It is assumed that the local characteristics in one image provide information about the local characteristics in the other. In the simplest of examples, this information might say that the two images are replicates, but here we have in mind that the characteristic is some form of "local smoothness", following any reasonable definition of smoothness, or texture. In particular, if we expect that smooth areas in one image correspond to smooth areas in the other, and rough areas correspond to rough areas. Many other examples are, however, possible and some will be discussed later. It is assumed, however, that values of all prior parameters are not under estimation, but instead might be fixed at the estimated values from the homogeneous model discussed in the previous section.

In this case, the posterior distribution is then

$$\pi(\mathbf{x}_1, \mathbf{x}_2|\mathbf{y}_1, \mathbf{y}_2) = \frac{f(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}_1, \mathbf{x}_2)\pi(\mathbf{x}_1, \mathbf{x}_2)}{f(\mathbf{y}_1, \mathbf{y}_2)} = \frac{f(\mathbf{y}_1|\mathbf{x}_1)f(\mathbf{y}_2|\mathbf{x}_2)\pi(\mathbf{x}_1, \mathbf{x}_2)}{f(\mathbf{y}_1, \mathbf{y}_2)} \quad (7)$$

since we assume that $\mathbf{y}_1$ and $\mathbf{y}_2$ are conditionally independent given $\mathbf{x}_1$ and $\mathbf{x}_2$; and further that $\mathbf{y}_1$ given $\mathbf{x}_1$ does not depend on $\mathbf{x}_2$, and that $\mathbf{y}_2$ given $\mathbf{x}_2$ does not depend on $\mathbf{x}_1$. For estimation, the probability function $f(\mathbf{y}_1, \mathbf{y}_2)$ is not important as it does not depend on any of the quantities to be estimates. Hence it forms part of an overall normalising constant.

The remainder of this section considers the novel contribution of the work. In the prior modelling it is not assumed that $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent, but instead

that a joint prior in terms of a single Gibbs distribution can be defined, that is

$$\pi(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\mathcal{K}} \exp\{-U(\mathbf{x}_1, \mathbf{x}_2)\} \tag{8}$$

with joint energy of the form

$$U(\mathbf{x}_1, \mathbf{x}_2) = \beta_1 V(\mathbf{x}_1) + \beta_2 V(\mathbf{x}_2) + \gamma W(\mathbf{x}_1, \mathbf{x}_2), \tag{9}$$

where β_1, β_2, and $V(\mathbf{x}_1)$, $V(\mathbf{x}_2)$, with definitions as in Section 2, describe within-image properties, whereas the new term describes between-image properties. There is an *association* parameter $\gamma \geq 0$ and *association measure* $W(\mathbf{x}_1, \mathbf{x}_2)$. Note that, although implied here, there is no reason why the same form of energy function must be applied to both $\mathbf{x}_1$ and $\mathbf{x}_2$.

To exploit various simplifications, which will be useful later, consider each energy defined in terms of local energies, in a similar way to the homogeneous case, as

$$U(\mathbf{x}_1, \mathbf{x}_2) = \beta_1 \sum v_i(\mathbf{x}_1) + \beta_2 \sum v_i(\mathbf{x}_2) + \gamma \sum w_i(\mathbf{x}_1, \mathbf{x}_2) \tag{10}$$

with, for example,

$$v_i(\mathbf{x}) = \frac{1}{m_{\partial_i}} \sum_{j \in \partial_i} \phi(x_i - x_j) \tag{11}$$

and where ∂_i represents the set of pixel indexes of the neighbours of pixel i and m_{∂_i} is the number of neighbours, that is $m_{\partial_i} = |\partial_i|$. Now bringing together the last two terms in Equation (10) to give

$$U(\mathbf{x}_1, \mathbf{x}_2) = \beta_1 \sum v_i(\mathbf{x}_1) + \beta_2 \sum \left[1 + \frac{\gamma}{\beta_2} \frac{w_i(\mathbf{x}_1, \mathbf{x}_2)}{v_i(\mathbf{x}_2)}\right] v_i(\mathbf{x}_2) \tag{12}$$

and then define a *modulating function*, the form of which will be discussed later,

$$\psi_i(\mathbf{x}_1, \mathbf{x}_2) = 1 + \frac{\gamma}{\beta_2} \frac{w_i(\mathbf{x}_1, \mathbf{x}_2)}{v_i(\mathbf{x}_2)} \tag{13}$$

giving our definition of the joint prior distribution as

$$\pi(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\mathcal{K}} \exp\left\{-\beta_1 \sum v_i(\mathbf{x}_1) - \beta_2 \sum \psi_i(\mathbf{x}_1, \mathbf{x}_2) v_i(\mathbf{x}_2)\right\}. \tag{14}$$

Note that, clearly there is an alternative representation where the roles of $\mathbf{x}_1$ and $\mathbf{x}_2$ are exchanged, but there is no added value as the two images can always be re-labelled.

Now, motivated by early discussion, we will define the local modulating functions and local energies in terms of neighbouring pixel values

$$\psi_i(\mathbf{x}_1, \mathbf{x}_2) = \psi\big(\mathbf{x}_1[\partial_i], \mathbf{x}_2[\partial_i]\big) \quad \text{and} \quad v_i(\mathbf{x}) = v\big(\mathbf{x}[\partial_i]\big), \qquad i = 1, \ldots, m, \tag{15}$$

for example

$$\psi_i(\mathbf{x}_1, \mathbf{x}_2) = \psi\left(\frac{v_i(\mathbf{x}_1)/\bar{v}(\mathbf{x}_1)}{v_i(\mathbf{x}_2)/\bar{v}(\mathbf{x}_2)}\right) \quad \text{and} \quad \psi_i(\mathbf{x}_1) = \psi\big(v_i(\mathbf{x}_1)/\bar{v}(\mathbf{x}_1)\big), \tag{16}$$

where

$$\bar{v}(\mathbf{x}_1) = \frac{1}{m}\sum_{i=1}^{m} v_i(\mathbf{x}_1)$$

is the mean local energy across $\mathbf{x}_1$ and similarly for $\bar{v}(\mathbf{x}_2)$. The aim of these is to introduce normalisation of the input variables and to contrast the properties of the two inputs.

3.2 Conditional estimation

Suppose now that instead of simultaneously modelling two images with a joint prior distribution we have a situation where we wish to reconstruct a single image, but we have additional spatial information about the image in the form of an explanatory variable $\mathbf{z} = (z_1, z_2, \ldots, z_m)$ recorded at the same locations $(\mathbf{s}_i : i = 1, 2, \ldots, m)$ at which we wish to estimate the values of $\mathbf{x}$. A conditional Markov random field model (Divino, Frigessi and Green (2000)) can now be defined where the energy function is allowed to depend on the explanatory variable

$$\pi(\mathbf{x}|\mathbf{z}) = \mathcal{K}(\mathbf{z}) \exp\big\{-U(\mathbf{x}|\mathbf{z})\big\}. \tag{17}$$

The idea is to reduce, or increase, the energy as indicated by the explanatory variable. In particular, again we have in mind that the value of the explanatory variable will have some information regarding local properties of $\mathbf{x}$. Divino, Frigessi and Green (2000) only consider the situation where the explanatory variable contains information about the value of the image $\mathbf{x}$. Hence, again the modelling is extended to consider the situation where the explanatory information indicates the degree of smoothness expected in the image $\mathbf{x}$. This can be thought of as representing a kind of *target smoothness*. So, $\mathbf{z}$ may be another image from which smoothness can be measured, or it could directly be a *smoothness map*. Hence, we now redefine the energy in the conditional model, using our modulating function approach

$$\pi(\mathbf{x}|\mathbf{z}) = \mathcal{K}(\mathbf{z}) \exp\big\{-\beta \Psi(\mathbf{x}|\mathbf{z}) V(\mathbf{x})\big\}. \tag{18}$$

In general ψ depends on both $\mathbf{x}$ and the explanatory variable $\mathbf{z}$, but we may also consider cases where it only depends on the explanatory variable and so $\Psi(\mathbf{x}|\mathbf{z}) = \Psi(\mathbf{z})$, hence

$$\Psi(\mathbf{x}|\mathbf{z}) V(\mathbf{x}) = \sum_{i=1}^{m} \psi_i(\mathbf{z}) v_i(\mathbf{x}) \quad \text{with } \psi_i(\mathbf{z}) = \psi\big(\mathbf{z}[\partial_i]\big). \tag{19}$$

Now the whole choice of modulating function has reduced to one of specifying the univariate function, ψ. Recall that this function should be designed to adjust

the prior using knowledge of the relationship between corresponding local characteristics. We have chosen to define this function requiring that it is non-negative, monotonic and it takes value 1 when local characteristics are equivalent. If ψ is monotonic increasing, then the local characteristics will be called *positively associated*; if monotonic decreasing, then the characteristics will be *negatively associated*.

There are many alternatives satisfying the above conditions, but we choose to base our modulating functions on cumulative distribution functions. This choice will produce bounded functions which will avoid extreme adjustments—a case identified as producing unappealing results in early experiments.

Suppose we want the modulating function, ψ, to range from a minimum, α_0, to a maximum, α_M, where $\alpha_0 = 1/\alpha_M$, and let F be our chosen cumulative distribution function which in turn depends on a set of parameters. Then, for positive association, we write ψ as

$$\psi(u) = \alpha_0 + (\alpha_M - \alpha_0)F(u) \tag{20}$$

and for negative association consider

$$\psi(u) = \big(\alpha_0 + (\alpha_M - \alpha_0)F(u)\big)^{-1}. \tag{21}$$

Note that an apparently obvious alternative to this negative association scheme is to consider $\psi(u) = (\alpha_0 + (\alpha_M - \alpha_0)\{1 - F(u)\})$, that is using the survivor function $1 - F(u)$. However for positive skew distributions, as is likely to be the case, then it will only be possible to solve $\psi(u = 1) = 1$ for the case in Equation (20) but not when using the survivor function for negative association. Now the choice of α_0 and α_M is determined by the range of modulation required. To fix the other parameters we will impose the conditions that the mode of the underlying distribution is at $u = 1$, and that the function at $u = 1$ has value 1, that is $\psi(1) = 1$. Again there are many possible options for the cumulative distribution function F, but it appears, in experiments not presented here, that there will be little practical difference across a broad range of alternatives.

Consider the log-logistic function with scale parameter α and shape parameter β, and hence probability density function

$$f(x|\alpha, \beta) = \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{(1 + (x/\alpha)^{\beta})^2}, \qquad x \geq 0, \alpha, \beta > 0,$$

and cumulative distribution function $F(x|\alpha, \beta) = 1/(1 + (x/\alpha)^{-\beta})$. Setting the mode at 1, requires $\alpha = ((\beta + 1)/(\beta - 1))^{1/\beta}$, and then $\beta = ((\alpha_M - \alpha_0)/(1 - \alpha_0))/((\alpha_M - \alpha_0)/(1 - \alpha_0) - 2)$. This latter condition simplifies to $\beta = (\alpha_M + 1)/(\alpha_M - 1)$ when $\alpha_0 = 1/\alpha_M$. Figure 1 shows examples of the modulating function for a few values of α_M (each with $\alpha_0 = 1/\alpha_M$) and both positive and negative association. It is worth noting that if the log-logistic distribution is replaced by, for example, the Weibull or the gamma distribution then very similar functions are obtained reinforcing the comment that the exact form is unlikely to be important.

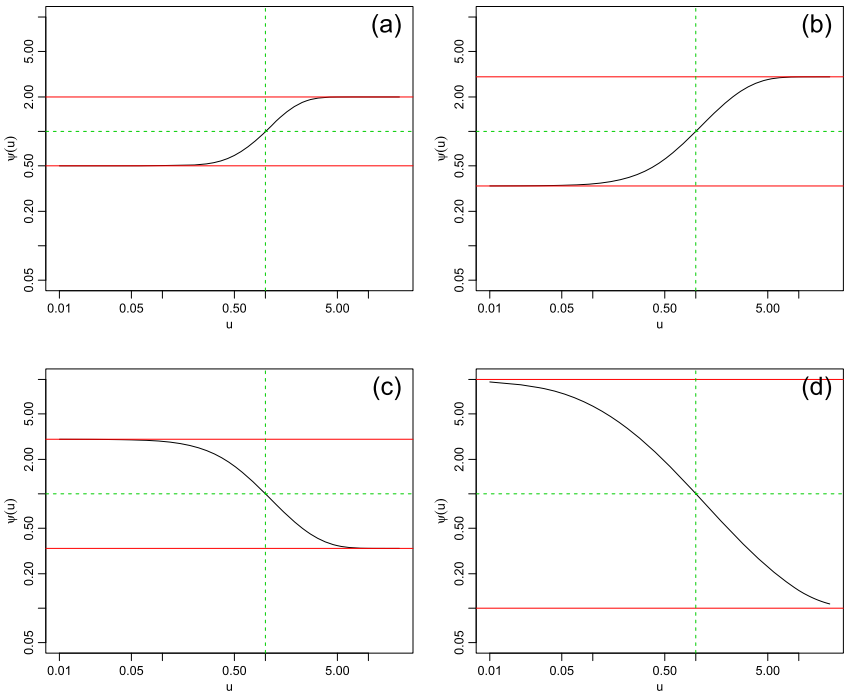

Figure 1 *Examples of cdf-based modulating functions with*: (a) $\alpha_M = 2$, (b) $\alpha_M = 3$, (c) $\alpha_M = 5$, (d) $\alpha_M = 10$ (*all have* $\alpha_0 = 1/\alpha_M$).

4 Parameter estimation and the MCMC algorithm

The Markov chain Monte Carlo (MCMC) approach is now widely used for many statistical estimation problems; for theoretical details see Gamerman and Lopes (2006), Geyer (2011) and Brooks et al. (2011), and for practical examples see Gilks, Richardson and Spiegelhalter (1995). The following description follows Aykroyd (2015) where more details can be found, also see, for example, Voss (2013). The MCMC method provides an approach when the model is complex or the number of parameters is large and deterministic gradient-based algorithms are infeasible. In brief, the transitions in the Markov chain are designed so that an equilibrium distribution exists and is equal to the target distribution, for example the posterior distribution in a Bayesian analysis. If the transitions are designed well, then after an initial transient period, referred to as burn-in, values are produced in the same proportions as would be obtained by direct simulation. Hence, the resulting sample will have the same statistical properties as a sample obtained directly from the posterior distribution. The only difference is that, by the very nature of a Markov chain, there will be serial correlation within the sample which must be taken into account when the algorithm output is summarised. If transitions are badly designed, however, then the initial transient period could be long

Algorithm 1 A simple random-walk Metropolis–Hastings algorithm

Set an initial value for $\boldsymbol{\theta}$, call this $\boldsymbol{\theta}^0 = (\theta_1 \cdots \theta_p)$
Repeat the following steps for $k = 1, \ldots, K$
  Repeat the following steps for $i = 1, \ldots, p$
  Generate ϵ from a Gaussian distribution $N(0, \tau^2)$
  Generate a propose new value $\theta_i^* = \theta_i + \epsilon$
  Evaluate

$$\alpha = \frac{\pi(\boldsymbol{\theta}^*|\mathbf{y})}{\pi(\boldsymbol{\theta}|\mathbf{y})}$$

  Generate u from a uniform distribution $U(0, 1)$
  If $\alpha > u$ then accept the proposal and set $\theta_i = \theta_i^*$, otherwise $\theta_i = \theta_i$
  End repeat
End repeat
Discard initial values and use remainder to make inference.

and the within sample correlation could be high. This means that the algorithm is inefficient and larger samples would be required to achieve acceptable accuracy and precision. There are no special aspects to the algorithm used here, and so the interested reader is directed to one of the above references for details. The specific MCMC numerical method is summarised in Algorithm 1 with computational code, written using the R software (R Core Team (2016)), available from the corresponding author. The only points needing to be mentioned are that here the parameter vector is $\boldsymbol{\theta} = (\mathbf{x}, \beta)$ for the homogeneous model, $\boldsymbol{\theta} = (\mathbf{x}_1, \mathbf{x}_2)$ for the joint modulated model and $\boldsymbol{\theta} = (\mathbf{x}_1)$ for the conditional modulated model. In all cases, only a single component of $\boldsymbol{\theta}$ is updated at each iteration, giving vector $\boldsymbol{\theta}^*$, and that a proposed value is generated as a Gaussian perturbation of the current value using a suitable variance.

Although the theoretical proof of convergence etc. is complicated, the statement and implementation of the algorithm is often straightforward. When considering complex estimation problems, however, great care may be needed; Lui (2001) gives an in depth catalogue of basic approaches that are likely to cover a range of situations.

Then, the only remaining practical issue is the choice of the proposal variance. It is important to realise that both low and high values lead to long transient periods and highly correlated samples and hence unreliable estimation (see, for example, Gelman, Gilks and Roberts (1997), Sokal (1989), Cowles and Carlin (1996), Raftery and Lewis (1995)). A reasonable proposal variance can be chosen adaptively during the early burn-in period. Once the sample has been generated from the posterior distribution, a number of possible estimators are available. Let $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^K$ be the MCMC sample collected after equilibrium of the Markov chain has been declared, then the pixel-wise posterior mean and variances can be

estimated by the sample mean and variance, and credible intervals using the appropriate sample percentiles. Similarly, marginal posterior distributions and other functions of the parameters can be estimated easily using the corresponding function of the MCMC sample.

5 A simulation study based on archaeological geophysics

5.1 Background

In archaeological geophysics various remote sensing methods, including magnetometry and seismology, use surface measurements to indirectly investigate the underground scene. The set of magnetometer readings across an archaeological site provides an image of the magnetic features beneath the ground surface. In particular, the readings depend on the nearby magnetic susceptibility values, but they are a blurred and noisy version of the true, but unknown, susceptibility distribution. In the following example, simulated data will be generated where $\mathbf{x}$ represents the true magnetic susceptibility distribution with $\mathbf{y}$ the corresponding surface magnetometry data. Details of the data modelling process, and the definition of blur matrix H in particular, can be found in Aykroyd, Haigh and Allum (2001) and Aykroyd and Al-Gezeri (2014).

The true underground scene, see Figure 2(a), is make-up of a large circular region of moderate susceptibility material in the centre, surrounded by four small circular high susceptibility regions on the perimeter of the large circle. The diameters of the small circles vary with the greatest at the top-right decreasing in an clockwise direction to the smallest at the top-left—these small circles are particularly challenging to reconstruct amongst the blurring and noise. This means that there is a substantial susceptibility contrast between adjacent regions, but precise uniformity within region. The blur matrix, see Figure 2(b), introduces a slight southerly shift in the peak intensity in the recorded magnetic field strength of a feature compared to the truth, and the random error, which in practice is caused by natural variation and measurement error, masks details—an example dataset is shown in Figure 2(c).

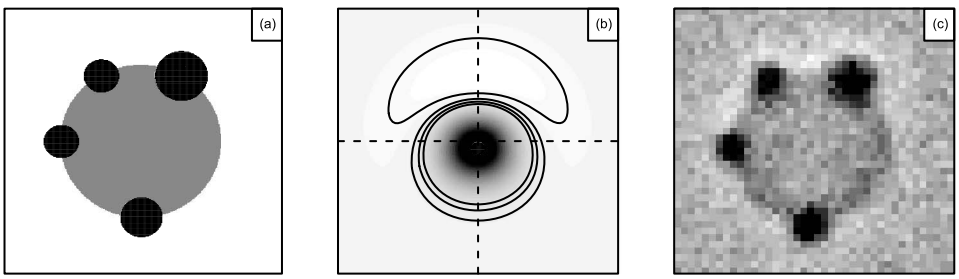

Figure 2 *Stages in the data modelling*: (a) *true intensity distribution with four small circles around a large circle*, (b) *point-spread matrix*, H, *and* (c) *resulting example dataset*.

5.2 MCMC convergence and sample size

To start the analysis, the standard prior based model is used, and the MCMC algorithm produces an estimate of the posterior mean which is calculated using the pixel-wise sample mean after a burn-in period. Here, a burn-in of 500 iterations is used with a main run of 10,000 using the hyper-parameter value of $\lambda = 10^5$. Separate proposal variances for each parameter were set during the burn-in period to achieve acceptance rates of 20–25%. To reduce autocorrelation, a 1-in-10 thinning is applied to give a working sample of 1000 for estimation. Figure 3 shows traces through the MCMC run for β and two selected pixels, one from the background and one from the small region towards the top-right, along with corresponding autocorrelation functions from the values after the burn-in. Together these show a rapid convergence to the equilibrium distribution with only moderate autocorrelation, and hence that estimation should be adequately reliable. Sample size calculations (Aykroyd and Green (1991)) indicate that a sample size of almost 1000 is needed to reduce the Monte Carlo variation to 1% of the sampling variance, but that about 200 is sufficient to achieve an acceptable 5% figure. This suggest that the existing run is more than enough for good estimation and that a smaller sample would be adequate. In the next section, all example reconstructions use these large samples, whereas in Section 5.4 shorter runs, with a main working sample of 200, are used in a simulation study involving replication.

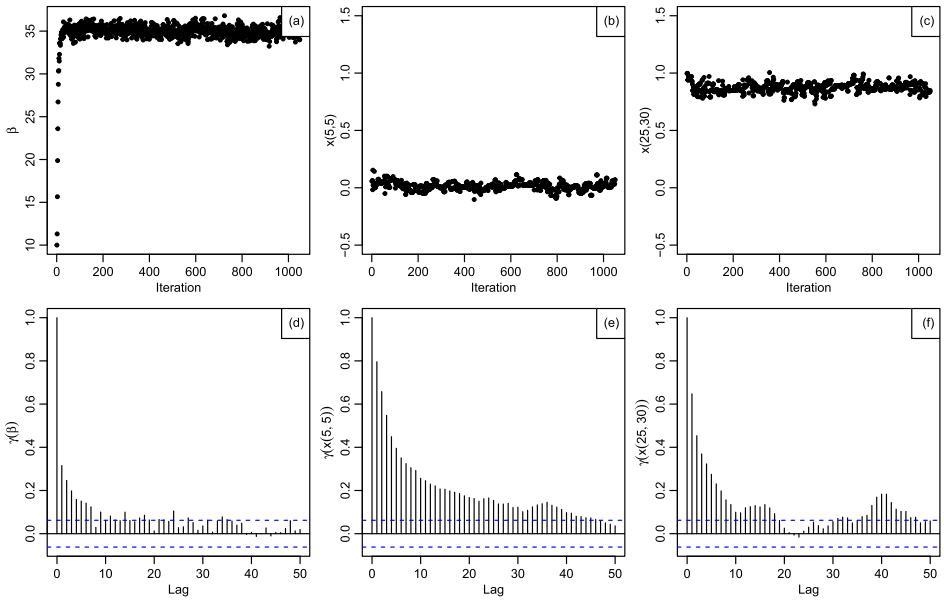

Figure 3 *Examples of MCMC monitoring output with* 500 *iteration for the burn-in and* 1000 *for the main run*: (a) *trace of* β, (b) *&* (c) *trace of two selected pixels*, (d) *ACF of* β, *and* (e) *&* (f) *ACF of selected pixels*.

5.3 Reconstruction results

The posterior mean estimate based on the Allum prior distribution is shown in Figure 4(a) and that based on the Gaussian prior distribution is in (d). In both it can be seen that a reasonably clear reconstruction is obtained with the southerly shift removed and the noise reduced substantially. The Allum prior reconstruction, however, is much clearer with better defined changes between regions and flatter within regions—these are key properties. However, the errors shown in (b) and (e), comparing the estimates with the true values, show clear patterns. Similar comments can be made about the residuals shown in (c) and (f), which compare the fitted values with the data. It is possible to reduce the patterns in the residuals by decreasing the value of λ in the hyper-prior distribution of β, however this is at the expense of introducing additional noise into the posterior mean estimate. Given that our primary aim is to produce a piecewise constant reconstruction, the larger value of λ is preferred. Hence, a suitable value of λ is one which produces a small, to moderate, pattern in the residuals and general smoothness in the reconstruction.

Figure 5 shows marginal posterior distributions for the same three parameters shown in Figure 4, that is the parameter β and two individual pixel values. The top row are from the estimation using the Allum prior and the bottom row using the Gaussian prior. In each, the histogram summarises the posterior sample and the continuous curve is the smoothed density estimate. The thick horizontal bar at

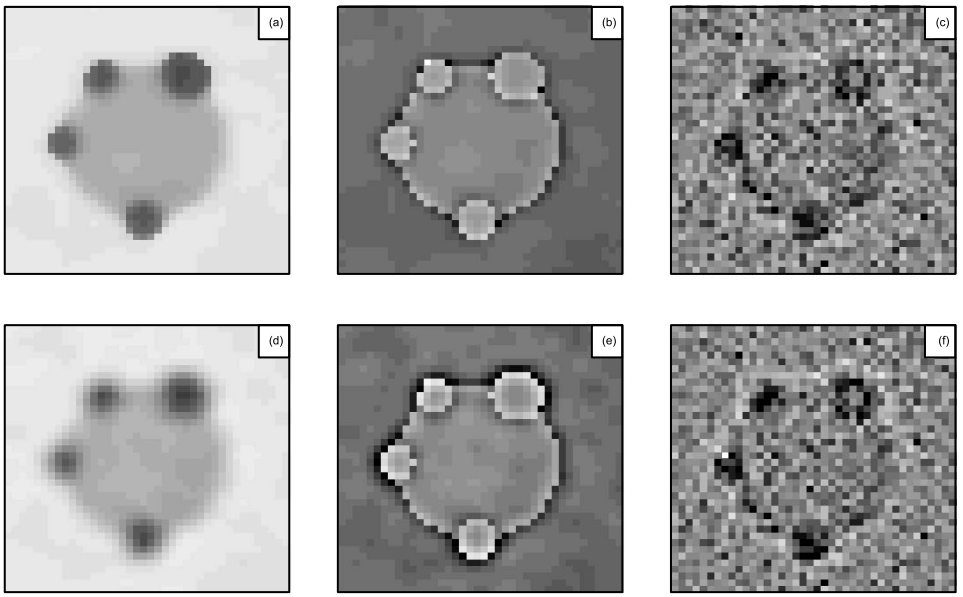

Figure 4 *Standard reconstruction*: *top row using the Allum prior and bottom row using a Gaussian prior*, *then* (a)/(d) *posterior mean*, (b)/(e) *error relative to true values and* (c)/(f) *residuals relative to data values*.

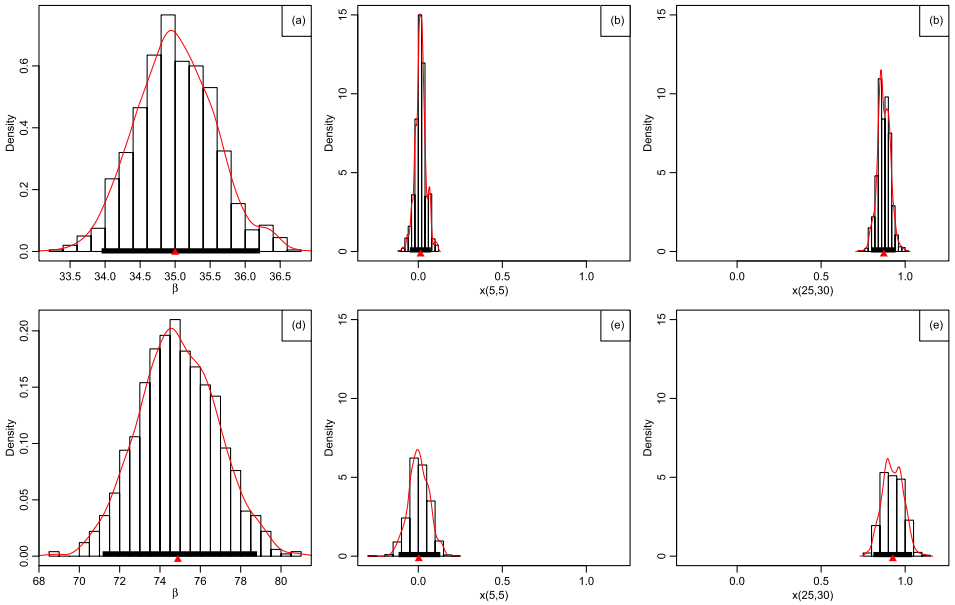

Figure 5 *Standard reconstruction*: *top row using the Allum prior and bottom row using a Gaussian prior, then* (a)/(d) *marginal posterior distribution of* β, (b)/(e) *marginal posterior distribution of single background pixel and* (c)/(f) *marginal posterior distribution of single small region pixel.*

the bottom shows the posterior 95% credible interval and the small triangle is the posterior mean. The posterior estimates for β, using (i) the Allum prior is $\hat{\beta} = 35.0$, with a credible interval of $(33.9, 36.2)$, and (ii) the Gaussian prior is $\hat{\beta} = 74.9$, with a credible interval of $(71.2, 78.8)$. Each has a reasonably symmetric distribution which passes a Shapiro–Wilk normality test ($p = 0.3$ and $p = 0.8$, respectively). For the individual pixel values it is very noticeable that the marginal distributions using the Allum prior are much narrower than those with the Gaussian prior. Three of these fail a Shapiro–Wilk normality test ($p < 0.001$, $p = 0.003$, $p = 0.096$ and $p = 0.004$, respectively) and hence it is best not to base any inference on a normality assumption.

Further investigation is shown in Figure 6, again with results for the Allum prior in the top row and the Gaussian prior in the bottom row. Panels (a) and (d) show profiles through the image reconstruction with the solid curve being the posterior mean surrounded by the 95% credible interval and with the true value shown as a step function. These show the superiority of the Allum prior especially where the true values change rapidly. Panels (b) and (e) show the width of the 95% credible interval with a clear pattern of larger values around the boundaries of the regions with the Allum prior, but in contrast there is no pattern for the Gaussian prior. The posterior means were used to calculate the local energy map in (c) and (f) which show the smooth regions in white and the locations of more abrupt change

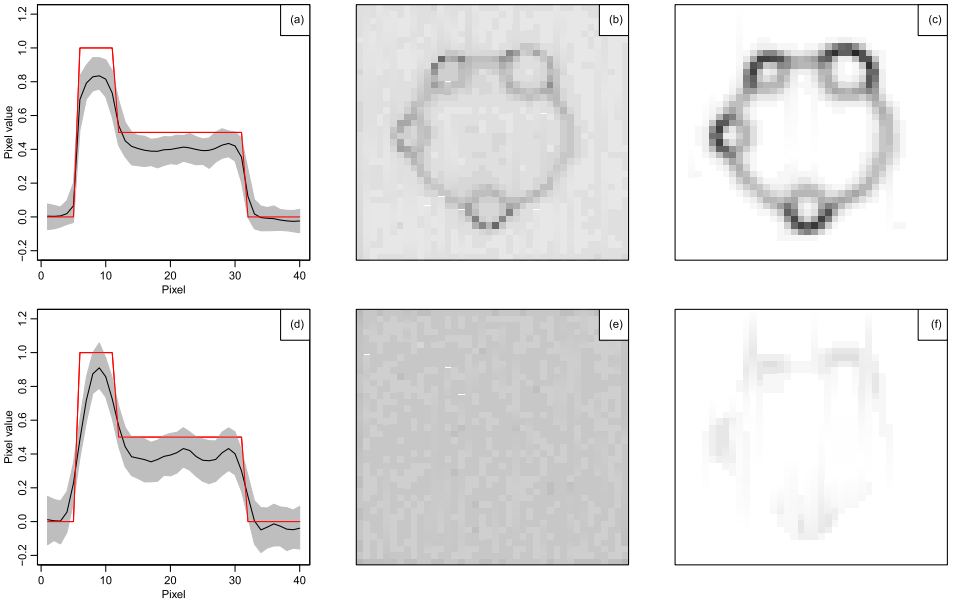

Figure 6 *Standard reconstruction*: *top row using the Allum prior and bottom row using a Gaussian prior, then* (a)/(d) *posterior mean along the middle row*, (b)/(e) *posterior credible interval width* (c)/(f) *posterior local energy.*

in grey. Again, it can be seen that the Allum prior allows sharp change at the regions boundaries.

Now moving to reconstruction using the modulated MRF model. Given that we require reduced smoothing in areas with high values of local energy, a negative association modulating function is selected. Figure 7 shows the posterior means, errors and residuals using $\alpha_M = 10$, with the value of β kept fixed at the posterior estimate from the standard model. Both sets of results show noticeable improvement over the corresponding standard reconstruction. The edges are clearer, with a slight increase in smoothness elsewhere, and the residuals show a reduced spatial pattern. Figure 8 shows further results. The patterns in the Allum prior posterior mean profile, panel (a), is particularly noteworthy showing very wide credible intervals at the region boundaries. This pattern is also shown in (b) by the dark shades. The sharpening of the reconstruction is also demonstrated by the increase in local energy at region boundaries, particularly as shown for the Allum prior shown in (c).

5.4 A simulation study

To compare the various image reconstruction methods quantitatively, the whole estimation process is repeated $K = 10$ times to produce reconstructions $(\hat{\boldsymbol{x}}^k : k =$

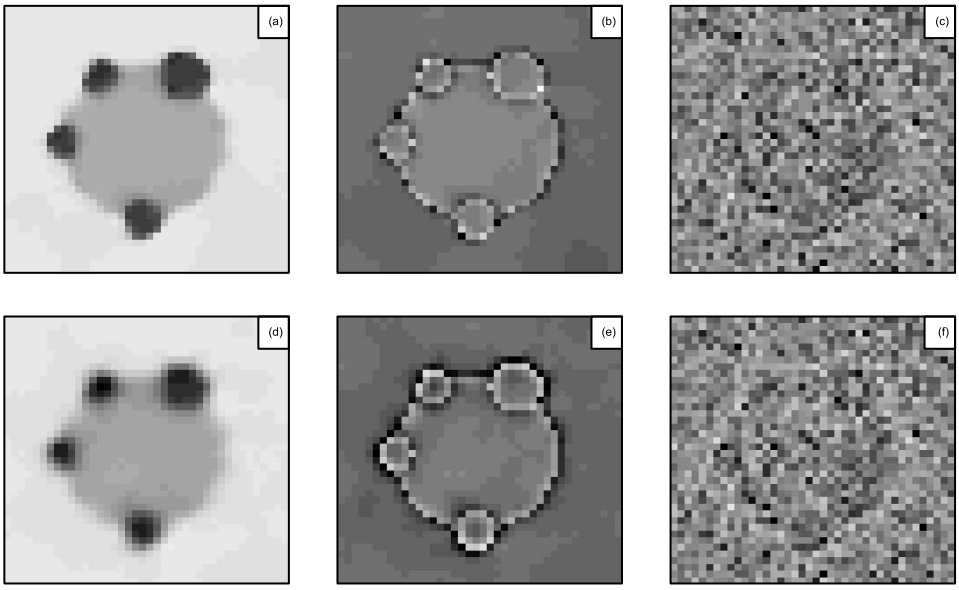

Figure 7 *Modulated reconstruction using the Allum* (*top row*) *and Gaussian priors* (*bottom row*), *then* (a)/(d) *posterior mean*, (b)/(e) *error relative to true values and* (c)/(f) *residuals relative to data*.

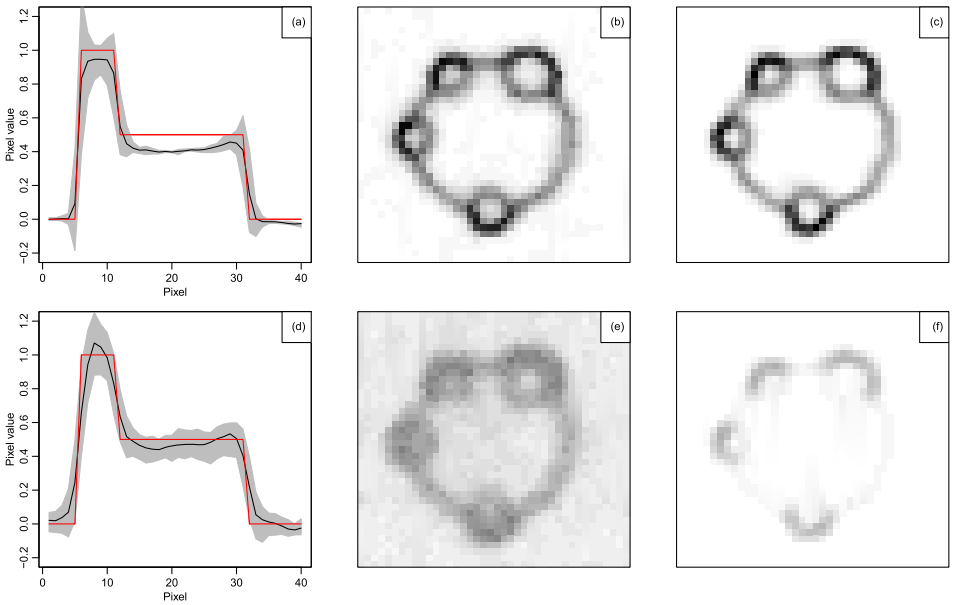

Figure 8 *Modulated reconstruction using the Allum* (*top row*) *and Gaussian priors* (*bottom row*), *then* (a)/(d) *posterior mean profile*, (b)/(e) *credible interval width* (c)/(f) *local energy.*

$1, \ldots, K$) where $\hat{\boldsymbol{x}}^k = (\hat{x}_1^k, \ldots, \hat{x}_n^k)$. Several goodness of fit measures will be considered, starting with the average sum of squared errors (SSE)

$$\text{SSE} = \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{n} (\hat{x}_i^k - x_i)^2 \tag{22}$$

which gives an overall measure of accuracy. We are, however, mainly interested in the accuracy around the various region boundaries and hence a *narrow-band* SSE is also defined which is only calculated over the pixels within a narrow band one pixel each side of a boundary,

$$\text{SSE}_B = \frac{1}{K} \sum_{k=1}^{K} \sum_{i \in B} (\hat{x}_i^k - x_i)^2, \tag{23}$$

where B is the set of pixel in the narrow-band set. We also consider the average residual sum of squares (RSS) given by

$$\text{RSS} = \frac{1}{K} \sum_{k=1}^{K} \sum_{j=1}^{n} (\hat{y}_j^k - y_j^k)^2, \tag{24}$$

where $\hat{y}_1^k, \ldots, \hat{y}_m^k$ are the fitted values corresponding to reconstruction using data $y_1^k, \ldots, y_m^k$, respectively. Finally, to focus on within-region variability, we consider a sum of squared deviations, $\text{SSD}_{\bar{B}}$, calculated ignoring the narrow-band pixels, that is over the set $\bar{B}$, defined as

$$\text{SSD}_{\bar{B}} = \frac{1}{K} \sum_{k=1}^{K} \sum_{i \in \bar{B}} (\hat{x}_i^k - \bar{\hat{x}}_i^k)^2, \tag{25}$$

where $(\bar{\hat{x}}_1^k, \ldots, \bar{\hat{x}}_m^k)$ are mean values calculated from pixels within corresponding regions—note that this requires knowledge of true region locations.

A summary of the results, in terms of SSE, SSE_B, RSS and $\text{SSD}_{\bar{B}}$, are shown as boxplots in Figure 9. In each panel, the left-hand pair use the Allum prior and the right-hand pair the Gaussian prior, then within each pair the left-hand is with the standard prior and the right-hand is with the modulated prior. The standard Gaussian prior produces the higher SSE followed by the standard Allum prior. There is then a substantial reduction produced by the modulated priors, with the modulated Allum prior having the lowest SSE. For the RSS, the highest value is for the standard Allum prior, but both modulated prior distributions have substantially lower values. When concentrating on the narrow-band SSE the picture is clearer than for the regular SSE with a substantial improvement due to the modulated approach for both prior distributions. Similarly, concentrating on the variability within region, it is clear that the Allum prior distributions lead to flatter regions and that the modulated functions improve the flatness. Taking together the visual appearance and the quantitative measures, it is clear that the modulated prior approach is worthwhile and that the Allum prior performs better than the Gaussian. Hence, the modulated Allum prior is the best approach.

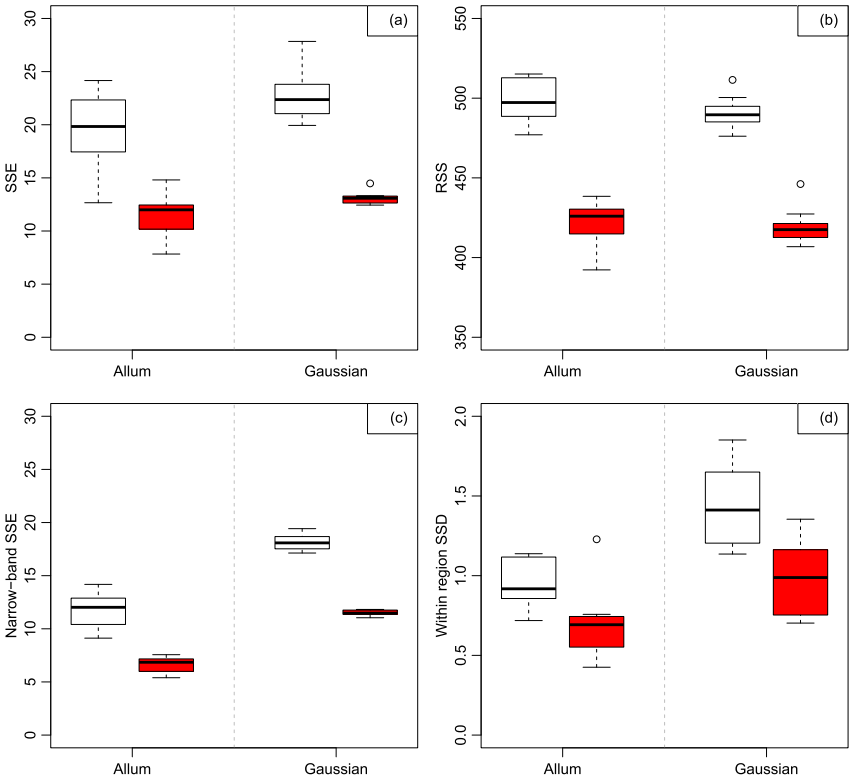

Figure 9 *Comparison over replication*: (a) *sum of squared error* (*SSE*), (b) *residual sum of squares* (*RSS*), (c) *narrow-band sum of squared error* (SSE_B) *and* (d) *within region sum of squared deviations* ($\mathrm{SSD}_{\bar{B}}$). *In each panel, the left-hand pair use the Allum prior and the right-hand pair the Gaussian prior, then within each pair the left-hand is with the standard prior and the right-hand is with the modulated prior.*

6 Discussion

One of the greatest challenges when performing any kind of Bayesian image reconstruction is the choice of prior parameter. One approach is to use trial-and-error, that is to try many values and see which results in the most "appealing" estimated image. This may produce a "nice" image but is ad-hoc, very subjective and open to question. Such lack of reproducibility naturally suggests the use of an automatic procedure which can also incorporate uncertainty. Fortunately, the Bayesian approach can accomplish this easily by the addition of a hyper–prior distribution into the hierarchical modelling framework, and the MCMC algorithm provides a straightforward method for full estimation. However, the standard homogeneous prior will not work for all cases. In particular, if the scene itself is not homogeneous, but instead contains regions of very different intensities with sharp changes—this is the usual case when the scene contains objects—then the

boundaries will be over-smoothed or the uniform regions will be under-smoothed. Any choice of a single prior parameter cannot balance the dual requirements of strong smoothing within uniform regions with the need for low, or no, smoothing across region boundaries. Although, one approach is to consider a full inhomogeneous prior model, the resulting increase in the number of parameters requires more modelling and much more computational time.

In this paper, a radical alternative to the fully inhomogeneous MRF model has been proposed. This prior distribution allows local adjustments to the model based on some pre-specified function of a second image. Although this is related to previous approaches it is not limited to comparing corresponding intensities, but instead proposes the use of a local statistic, such as local energy, to modulate the prior. Hence, the previously suggested approaches are special cases of our proposed method. There are many real data examples which have this structure and it is our intention to perform an investigation as future work. Obvious examples include archaeological geophysics, the motivation for the simulation study presented here, which can combine magnetometry, seismology and radar, or in medical imaging where high-resolution magnetic resonance imaging can to used to provide a contextual smoothing map to guide a PET reconstruction. The clear wide-spread applicability and encouraging numerical results suggest that the use of the locally-modulated Markov random field (lm-MRF) model has the potential for major impact and can become a standard tool for all those working in image processing and statistical image analysis.

Acknowledgments

The authors thank the anonymous referees, and the Editor, for their comments on an earlier version which have resulted in a much improved paper.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

Allum, G. T. (1997). A statistical approach to inverse data problems in archaeological geophysics. Ph.D. thesis, Department of Statistics, University of Leeds, UK.

Aykroyd, R. G. (1998). Bayesian estimation for homogeneous and inhomogeneous Gaussian random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**, 533–539.

Aykroyd, R. G. (2002). Approximations for Gibbs distribution normalising constants. *Statistics and Computing* **12**, 391–397. MR1951711

Aykroyd, R. G. (2015). Statistical image reconstruction. In *Industrial Tomography: Systems and Applications*. Cambridge: Woodhead Publications.

Aykroyd, R. G. and Al-Gezeri, S. M. (2014). 3D modelling and depth estimation in archaeological geophysics. *Chilean Journal of Statistics* **5**, 19–35. MR3200504

Aykroyd, R. G. and Green, P. J. (1991). Global and local priors, and the location of lesions using gamma-camera imagery. *Philosophical Transactions of the Royal Society of London A* **332**, 323–342.

Aykroyd, R. G., Haigh, J. G. B. and Allum, G. T. (2001). Bayesian methods applied to survey data from archaeological magnetometry. *Journal of the American Statistical Association* **96**, 64–76. MR1973783

Aykroyd, R. G. and Zimeras, S. (1999). Inhomogeneous prior models for image reconstruction. *Journal of the American Statistical Association* **94**, 934–946. MR1723299

Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, Series B* **48**, 259–302. MR0876840

Brooks, S., Gelman, A., Jones, G. and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. Boca Raton: Chapman & Hall/CRC. MR2742422

Cheng, C. C. and Huang, C. L. (1993). Markov random-fields for texture classification. *Pattern Recognition Letters* **14**, 907–914.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* **91**, 883–904. MR1395755

Cross, G. R. and Jain, A. K. (1983). Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **5**, 25–39.

Derin, H. and Elliott, H. (1987). Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **9**, 39–55.

Divino, F., Frigessi, A. and Green, P. J. (2000). Penalized pseudolikelihood inference in spatial interaction models with covariates. *Scandinavian Journal of Statistics* **27**, 445–458.

Dousset, B. and Gourmelon, F. (2003). Satellite multi-sensor data analysis of urban surface temperatures and landcover. *Journal of Photogrammetry and Remote Sensing* **58**, 43–54.

Ehrhardt, M. J., Markiewicz, P., Liljeroth, M., Barnes, A., Kolehmainen, V., Duncan, J. S., Pizarro, L., Atkinson, D., Hutton, B. F., Ourselin, S., Thielemans, K. and Arridge, S. R. (2016). PET reconstruction with an anatomical MRI prior using parallel level sets. *IEEE Transactions on Medical Imaging* **35**, 2189–2199.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed. *Chapman & Hall/CRC Texts in Statistical Science*. Boca Raton: Chapman & Hall/CRC. MR2260716

Gelman, A., Gilks, W. R. and Roberts, G. O. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* **7**, 110–120. MR1428751

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Geman, S. and McClure, D. E. (1987). Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute* **52**, 5–21. MR1027188

Geyer, C. J. (2011). Introduction to Markov chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng, eds.). Boca Raton: Chapman and Hall/CRC. MR2858443

Gilks, W., Richardson, S. and Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in Practice*. Boca Raton: Chapman & Hall/CRC. MR1397966

Green, P. J. (1990). Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging* **9**, 84–93.

Hu, H., Zhao, L., Li, X., Wang, H. and Liu, T. (2018). Underwater image recovery under thenonuniform optical field based on polarimetric imaging. *IEEE Photonics Journal* **10**, ASN: 6900309.

Kinderman, R. and Snell, J. L. (1992). *Markov Random Fields and Their Applications*. Providence: American Mathematical Society. MR0620955

Lui, J. (2001). *Monte Carlo Strategies in Scientific Computing*. Berlin: Springer. MR1842342

Mardia, K. V. (1994). *Advances in Applied Statistics: Statistics and Images, Vol. 2*. Abingdon: Carfax Publishing Company. MR1333678

Mardia, K. V. and Kanji, G. K. (1993). *Advances in Applied Statistics: Statistics and Images, Vol. 1*. Abingdon: Carfax Publishing Company. MR1333677

Melas, D. E. and Wilson, S. P. (1997). Texture based image segmentation using the double MRF model. In *Proceedings of the Art and Science of Bayesian Image Analysis* (K. V. Mardia, C. A. Gill and R. G. Aykroyd, eds.). 169–175. Leeds: Leeds University Press.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raftery, A. and Lewis, S. (1995). The number of iterations, convergence diagnostics and generic Metropolis algorithms. In *Practical Markov Chain Monte Carlo* (W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds.). London: Chapman and Hall.

Sokal, A. D. (1989). Monte Carlo methods in statistical mechanics: Foundations and new algorithms. *Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne*.

Vandenberghe, S. and Marsden, P. K. (2015). PET-MRI: A review of challenges and solutions in the development of integrated multimodality imaging. *Physics in Medicine and Biology* **60**, R115.

Voss, J. (2013). *An Introduction to Statistical Computins: A Simulation-Based Approach*. Chichester: Wiley. MR3237438

Weir, I. S. (1997). Fully Bayesian SPECT reconstructions. *Journal of the American Statistical Association* **92**, 49–60.

Department of Statistics
University of Benghazi
Benghazi
Libya

Department of Statistics
University of Leeds
Leeds, LS2 9JT
United Kingdom
E-mail: r.g.aykroyd@leeds.ac.uk