Markov chain Monte Carlo in image analysis [*]

Peter J. Green,
University of Bristol [†]

14 September 1994

1 Introduction

Digital images now routinely convey information in most branches of science and technology. The Bayesian approach to the analysis of such information was pioneered by Grenander (1983), Geman and Geman (1984), and Besag (1986), and both fundamental research and practical implementations have been pursued energetically ever since. In broad terms, the Bayesian approach treats the recorded raw image as numerical data, generated by a statistical model, involving both a stochastic component (to accommodate the effects of noise due to the environment and imperfect sensing) and a systematic component (to describe the true scene under view). Using Bayes' theorem, the corresponding likelihood is combined with a prior distribution on the true scene description to allow inference about the scene on the basis of the recorded image.

This informal and simplistic summary in fact embraces extraordinary variety, both in the style of modelling and in the type of application – from say, removal of noise and blur in electron microscopy to the recognition of objects in robotic vision. There is little common ground in this wide spectrum about which to generalise, but one feature that does stand out in any general view of image analysis literature is the prominent role played by Markov chain Monte Carlo (MCMC) as a means of calculation with image models.

The purpose of this article is to discuss general reasons for this prominence of MCMC, to give an overview of a variety of image models and the use made of MCMC methods in dealing with them, to describe two applications in more detail,

[*] To appear as a chapter in the book *Practical Markov chain Monte Carlo*, edited by W. Gilks, S. Richardson and D. Spiegelhalter, published by Chapman and Hall.

[†] Department of Mathematics, University of Bristol, Bristol, BS8 1TW, UK. Email: `P.J.Green@bristol.ac.uk`

and to review some of the methodological innovations in MCMC stimulated by the needs of image analysis, that may prove important in other types of application.

2 The relevance of MCMC to image analysis

The most obvious feature of digital images, considered as numerical data, is their very large size. Images very seldom consist of as few as 32×32 pixels (picture elements): sizes of between 256×256 and 1024×1024 are much more common. And if a time dimension is added, through consideration of 'movies' or image sequences, the volume of data available becomes very large indeed. This size factor immediately favours the use of MCMC methods in a statistical approach to image analysis since although other, more conventional, numerical methods sometimes provide practical routes to the calculation of point estimates of a true image (e.g. the *maximum a posteriori* estimate), MCMC is usually the only approach for assessing the variability of such estimates.

A second key feature of image data is its spatial structure. The interplay between stochastic and spatial variation is one of the most interesting aspects of image modelling from a mathematical perspective, and to construct such models that are tractable algorithmically can be a considerable challenge. The usual solution, as in the applications in other articles in this volume, is to adopt models with a parsimonious conditional independence structure. This is the key to efficient application of MCMC methods, which are all driven by conditional probabilities.

In the case of image models, neighbours in the conditional independence graph are usually spatially contiguous pixels or picture attributes. The random field models that are used are similar, or even identical, to those studied in statistical physics: this field was the original home ground for MCMC methods (Metropolis *et al.*, 1953; Binder, 1988) and physical analogies are drawn strongly in some work in statistical image analysis (e.g. Geman and Geman, 1984). From a historical perspective, developments in the specification of image models, and in the construction of MCMC algorithms to deal with them, have proceeded in parallel.

For extra parsimony, image models are usually specified to be spatially stationary. This makes implementation of MCMC methods particularly simple: effectively the same calculations are performed repeatedly at each location in the image, so the computer code required can be very compact. This in turn explains the apparent lack of interest among researchers in statistical image analysis in using general purpose MCMC packages. It is also evident that this is fertile ground for the implementation of MCMC on parallel architecture: impressive performance is obtained in the applications described by Grenander and Miller (1994).

We have thus identified a number of reasons why MCMC is such a natural approach to computing in image analysis, and it is not surprising that these methods were first encountered in statistics in this setting, before being adapted to other problems. In some cases, in fact, not only has the style of computation been borrowed, but image models themselves have been used with little modification: this is true of applications to geographical epidemiology (Besag, York and Mollié, 1991; Clayton and Bernardinelli, 1992; Gilks *et al.*, 1993), spatial archaeology (Besag, York and Mollié, 1991; Litton and Buck, 1993), biogeography (Högmander and Møller, 1993; (Heikkinen and Högmander, 1994) and agricultural field trials (Besag and Higdon, 1993). Some of these connections are further explored in Besag, Green, Higdon and Mengersen (1995).

3 Image models at different levels

In computer vision, a distinction is drawn between low-level analysis – for example, noise and blur removal – and high level analysis – for example, object recognition. In between there is a spectrum embracing such problems as tomographic reconstruction, segmentation and classification, texture modelling, etc. This variety is reflected in the range of different levels at which statistical modelling of images has been attempted. In this section, we will discuss the implications for the use of MCMC methods in three different styles of modelling. We focus on the usual problem of interest – calculation of probabilities and expectations under posterior distributions. But it is equally true that MCMC has an important role in computing with prior distributions for images.

3.1 Pixel-level models

Low-level image processing can be conducted with the aid of models for images that operate at the pixel level. If y denotes the recorded image, and x the description of the true scene, then x and y are large arrays, usually representing discretisations of the images onto a two-dimensional rectangular lattice, and usually of the same size. The original work of Geman and Geman (1984) and Besag (1986) had this character, and concentrated on the spatial classification problem, in which the value x_i of x at pixel i is drawn from a finite set of labels representing, say, different crop types in a remotely sensed agricultural scene. The corresponding recorded value y_i is the quantity of radiation emitted by the scene, that is recorded in the sensor as originating from pixel i on the ground. A typical form for the degradation model $p(y|x)$ to take would allow blur and independent noise, so that

$$p(y|x) = \prod_{k \in S} p(y_k|x_{\nu k}),$$

where S denotes the set of all pixels, νk the set of ground pixels influencing sensor pixel k, and for any set of pixels A, $x_A = \{x_k : k \in A\}$. In the product on the right-hand side here would be specified functions, quantifying the blur and noise. Note that throughout this article, the symbol $p(\ )$ will be used generically to denote probability functions or densities. Prior information in this type of application would usually be generic, reflecting belief that adjacent pixels are likely to have the same label, for example using the Potts model

$$p(x) \propto \exp(-\beta \sum_{i \sim j} I[x_i \neq x_j]). \tag{1}$$

Here $i \sim j$ means that i and j are neighbours in some prescribed undirected graph with the pixels as vertices, the sum is over all such pairs of neighbours, and $I[\,]$ is the indicator function, taking the value 1 if x_i and x_j are different; thus the sum gives the number of neighbour pairs that have differing labels. This is a simple example of a Markov random field or Gibbs distribution. (Throughout this article the symbol $\propto$ is used to denote proportionality in the first argument of the function on its left-hand side). Combining these two model terms yields the posterior distribution

$$p(x|y) \propto p(x)p(y|x),$$

and since

$$p(x_i|x_{-i}, y) \propto p(x|y),$$

where $x_{-i} = \{x_j : j \neq i\}$, we have

$$p(x_i|x_{-i}, y) \propto \exp(-\beta \sum_{j \in \partial i} I[x_j \neq x_i]) \prod_{k:\nu k \ni i} p(y_k|x_{\nu k}). \tag{2}$$

The notation ∂i means the set of neighbours of i, that is $\{j : j \sim i\}$; thus the sum above gives the number of neighbours of i labelled differently from i. The posterior full conditionals (or local characteristics) are seen to take a simple form, the product of a small number of easily computed factors. The normalisation to a proper distribution is easy since this is a discrete distribution over a small set of labels, and the way is clear for MCMC simulation using the Gibbs sampler. This was proposed by Geman and Geman (1984), and indeed they coined this name for the sampler already known to the statistical physics community as the *heat bath* (e.g. Creutz, 1979).

Geman and Geman discuss other samplers, and simulated annealing as well as simulating at equilibrium; they also go on to deal with a number of ramifications to the basic formulation above, including the use of edge variables that are not directly observed.

It is straightforward to amend the model to deal with the case where the true value x_i is a real number ('grey level') representing, say, the emitted radiance rather than a label. For example, we might use a pairwise difference prior

$$p(x) \propto \exp(-\beta \sum_{i \sim j} \phi(x_i - x_j)) \tag{3}$$

for some symmetric U-shaped function ϕ (see Besag 1986; Geman and McClure, 1987; Besag *et al.*, 1995). The full conditional retains its simple structure, but except in very special circumstances is no longer amenable to the Gibbs sampler, as it is a density function of rather arbitrary form with an unknown normalising constraint. The most convenient MCMC sampling method then is the Metropolis-Hastings method, which is in our experience capable of quite adequate computational efficiency.

Concrete applications of such models have included gamma camera imaging (Aykroyd and Green, 1991), magnetic resonance imaging (Godtliebsen, 1989; Glad and Sebastiani, 1993), film restoration (Geman, McClure and Geman, 1992) and remote sensing (Wilson and Green, 1993).

Variants of these simple models have been used to deal with other low-level vision problems. Geman (1991) provides a good general account of the theory. Texture modelling is important in segmentation (the subdivision of an image into homogeneous regions) and classification (the assignment of pixels of an image to pre-assigned classes). It can be accomplished using Markov random fields of higher order than that in (1); modelling and associated MCMC algorithms are discussed by Cohen and Cooper (1987). Boundary detection is discussed by Geman, Geman, Graffigne and Dong (1990), using low-level random field models with varying spatial resolutions. A general framework for the Bayesian restoration of image sequences via state space modelling was proposed by Green and Titterington (1988), and this is amenable to approximate MCMC calculations, at least for recursive estimation.

3.2 Pixel based modelling in SPECT

A particularly interesting variant applies to reconstruction from emission tomography data (PET and SPECT), based on the Poisson regression model introduced by Shepp and Vardi (1982). The novelty in tomography is the geometrical transformation between the 'body-space' in which the true scene x is located and the 'projection-space' where the recorded data y are obtained. The model (2) is still applicable, only the set of pixels νk is no longer a compact region surrounding k, but rather the set of all pixels in body space from which an emitted photon that is accumulated in the detected datum y_k can originate. This set can be large, which means there are many factors in (2) and hence relatively slow operation of MCMC methods.

In single photon emission computed tomography (SPECT), the generally accepted model states that the pixel values in the recorded image $y = \{y_t : t \in T\}$, which represent counts of detected gamma ray photons, satisfy

$$y_t \sim \text{Poisson}(\sum_{s \in S} a_{ts} x_s), \text{independently} \tag{4}$$

where $x = \{x_s : s \in S\}$ denotes the true concentration of a radioisotope as a function of location in the patient, the object of inference, and $\{a_{ts}\}$ are coefficients determined by the geometry and other physical properties of the instrumentation.

Bayesian reconstruction has been discussed by Geman and McClure (1987) and Geman, Manbeck and McClure (1993), using MCMC and other methods including ICM (Besag, 1986), by Green (1990) and Weir and Green (1994), using a modified EM method, and by Amit and Manbeck (1993), using deformable templates and wavelets, among others.

Here we briefly describe how to set up a Metropolis-Hastings sampler for the problem: more detail for a somewhat similar problem is given in Besag *et al.* (1995).

The prior model we shall use is related to that of (3), but on a log scale (since $x_s \geq 0$) and with a scale factor :

$$p(x) \propto \exp(-\sum_{s \sim r} \phi(\gamma(\log x_s - \log x_r)))/\prod x_s$$

where $\phi(u) = \delta(1+\delta)\log\cosh(u/\delta)$. The product in the denominator is a Jacobian arising from the logarithmic transformation. This expression gives absolute value and gaussian priors as $\delta \to 0$ and ∞ respectively (resulting in $\phi(u) = |u|$ or $\frac{1}{2}u^2$). Combined with (4) and factoring out terms not involving x_s gives the posterior local characteristics

$$\begin{aligned} p(x_s|x_{-s}, y) &\propto \exp\left(-\sum_{r \in \partial s} \phi(\gamma \log(x_s/x_r))\right) \\ &+ \sum_{t: a_{ts} > 0} (y_t \log \sum_r a_{tr} x_r - \sum_r a_{tr} x_r) - \log x_s. \end{aligned} \tag{5}$$

On the scale of problem presented by Green (1990), there would be around 300 terms in each of the sums over t and r above. There is therefore a heavy burden in computing even one value of the probability density above, although this can be partially offset by storing also $\{\mu_t = \sum_s a_{ts} x_s\}$ and keeping these up-to-date as x changes. The density is also of awkward, non standard form (and is neither convex in x_s or $\log x_s$) so the Gibbs sampler is not at all practical here.

Instead, we use a Metropolis-Hastings sampler, using a proposal in which the new value of $\log x_s$ is uniformly distributed over an interval centred at the old value. Thus the proposal density is

$$q_s(x_s \to x_s') \propto (x_s')^{-1}$$

on an interval $|\log(x_s'/x_s)| < c$. The corresponding acceptance function is calculated from (5):

$$\alpha(x, x') = \min\left[1, \exp\left\{-\sum_{r \in \partial s} (\phi(\gamma \log x_r/x_s') - \phi(\gamma \log x_r/x_s))\right.\right.$$

$$+ \sum_{t:a_{ts}>0} \left(y_t \log \left\{ 1 + a_{ts}(x'_s - x_s) / \sum_r a_{tr} x_r \right\} \right.$$
$$\left. \left. \left. -a_{ts}(x'_s - x_s) \right) \right\} \right] .$$

Note that the Jacobian terms have cancelled. The sampler is applied to the pixels one-by-one in a systematic raster scan. (In some of the recent Bayesian literature, such a sampler has been called "Metropolis-within-Gibbs"; however, this term is incorrect: the method was originally applied component-wise (Metropolis *et al.* (1953).)

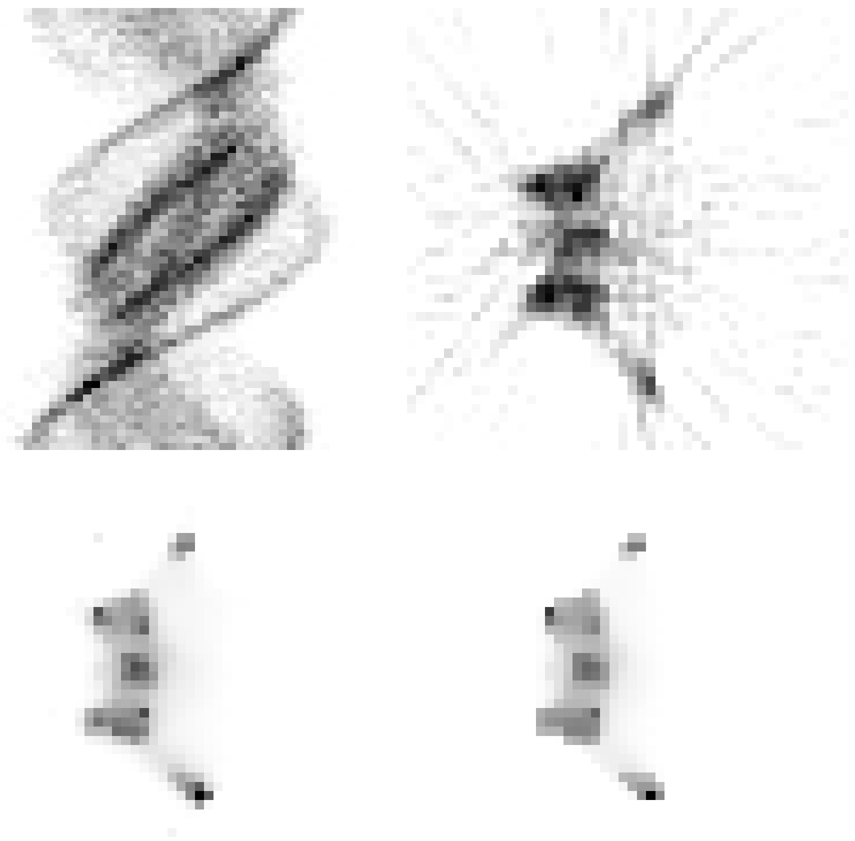

Figure 1: SPECT sinogram from a pelvis scan, and three reconstructions. (a) top left: sinogram; (b) top right: filtered back-projection; (c) bottom left: posterior mean, and (d) bottom right: posterior mode.

In a fully Bayesian treatment of the SPECT problem, the hyperparameters γ and δ in the prior should themselves be considered drawn from appropriate hyperpriors, but here such an approach would be problematical, since the normalisation constant in the joint distribution of x and y is an intractable function

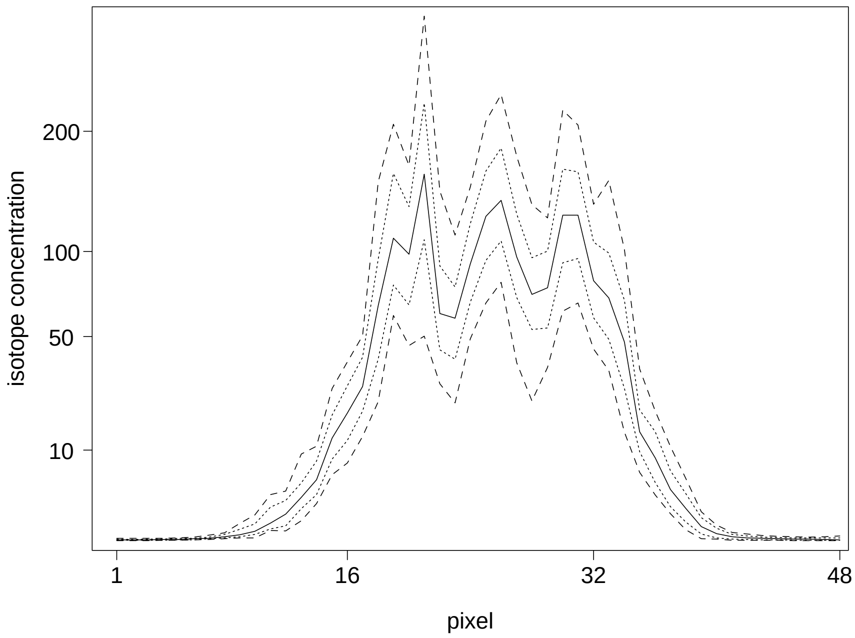

Figure 2: A one-dimensional transect through a Bayesian reconstruction from a pelvis scan. Solid line: pixelwise posterior median; dotted lines: pixelwise 80% credibility bounds; broken lines: simultaneous 80% credibility bounds.

of γ and δ. This constant might be estimated, off-line, by simulations from the prior. Alternatively, it may be appropriate to regard γ and δ as fixed, perhaps in practice chosen according to the organ being studied. For present purposes, we take γ and δ as fixed, with values chosen by trial and error.

The method just described was applied to SPECT data relating to a pelvis scan conducted at the Bristol Royal Infirmary. Data relevant to a particular horizontal slice are displayed in sinogram form in Figure 1, panel (a), and a filtered back-projection reconstruction of the slice in panel (b). Note the noisy character of this reconstruction, and the presence of radial artifacts. For alternative Bayesian reconstructions, the coefficients $\{a_{ts}\}$ were calculated using the model described in Green (1990), and the hyperparameters were set at $\gamma = 1.1$ and $\delta = 0.18$. Short pilot runs suggested that a value for the proposal spread parameter c of around 0.35 was most efficient from the point of view of minimising the integrated autocorrelation time for estimating $E(x_s|y)$; see Sokal (1989) and Green and Han (1992). The reconstruction resulting from averaging over 2000 sweeps of the Hastings sampler is shown as panel (c); for comparison, panel (d) shows an approximate MAP reconstruction, under the same prior, obtained using the one-step-late algorithm of Green (1990). The posterior mean and mode differ slightly, but the two Bayesian methods have provided reconstructions of comparable quality.

The reconstruction obtained using MCMC was some 20 times more expensive to compute; however, the advantage of the MCMC approach is seen when we wish to assess variability. Figure 2 displays information relevant to a one-dimensional transect through the body-space that corresponds to a horizontal left-right line in the patient and to a vertical line in the reconstructions shown in Figure 1, just to the left of the centre of the diagram. All of the curves are based on the final 1500 sweeps of the 2000-sweep Monte Carlo run giving Figure 1, and are plotted on a square-root vertical scale. The solid curve is the pixelwise median from the realisation, estimating the posterior median isotope concentration along the transect, and the dotted curves are the 10% and 90% quantiles. The broken lines display order-statistics-based *simultaneous* symmetric credible bounds, covering 80% of the posterior distribution for the whole transect. These were computed in the manner described in Besag, Green *et al.* (1995).

3.3 Template models

The major limitation of pixel-level image models is that inference based on them can only refer to statements about the pixels of the true scene. With calculations performed by MCMC, this is not quite the restriction it might seem, since posterior probabilities of arbitrary events defined by the pixel values can be estimated. (And if a complete realisation of the Markov chain is stored, such events need not even be identified until the simulation is complete!)

However, it is hardly realistic in practical terms to contemplate specifying which pixel values define, say, a platoon of tanks, and which a herd of cows. To perform inference at this higher level of objects, the role of the scene x in the previous section must be replaced or augmented by a description of the objects that the scene contains.

Having done so, at the conceptual level, at least, nothing changes. Bayes' theorem and the principle of estimating posterior expectations and probabilities by MCMC still apply. However, the details of modelling and of sampling may be very different.

One very promising route to image model building for these purposes is via deformable templates, introduced by Grenander: the original study was eventually published as Grenander, Chow and Keenan (1991). This dealt with simple images consisting of polygonal silhouettes: the black/white true silhouette is not observed directly, but only after spatial discretisation and the addition of noise. Prior modelling then focuses on the polygonal boundary of the silhouette, or rather, in the deformable template paradigm, on stochastic deformations of this polygon from an ideal outline determined by the context. In early experiments these were outlines of a human hand and of leaves.

There remains a central role for Markov random field models in this situation, incidentally. The deformation would usually be specified by its effects on the lengths and orientations of individual boundary segments of the polygon, and these effects would be modelled dependently using a simple local dependence assumption.

Deformable templates have been extended to a number of more challenging situations, including grey-level scenes, in which the recorded datum at each pixel is continuously distributed (Amit, Grenander and Piccioni 1991), with progress towards practical applications (Grenander and Manbeck, 1992; Ripley and Sutherland, 1990). When the number of elements defining the template or its deformation is small, the resulting image models are close to parametric rather than nonparametric in nature, and inference is less sensitive to the specification of the prior. Phillips and Smith (1992) discuss an interesting application to segmentation of facial images, and Aykroyd and Green (1991) compare performance of image analysis procedures using pixel and template based prior models. In all of these papers, MCMC methods are used.

Template models have been extended to deal with scenes containing variable numbers of objects. The appropriate parameter spaces are now unions of subspaces of varying dimension; novel MCMC methods have been developed for this situation, notably using jump-diffusion simulation (Grenander and Miller, 1994; Srivastava, Miller and Grenander, 1991). Green (1994) described an explicit class of MCMC methods using reversible Metropolis-Hastings jumps between subspaces, that should be widely applicable.

3.4 An example of template modelling

We illustrate the main ideas of image analysis based on a deformable template model by describing a toy example, simplified from the silhouette problem considered by Grenander, Chow and Keenan (1991).

The template is a closed polygon, whose boundary does not intersect itself. The prior distribution forming the template model is obtained from separate local perturbations to the lengths of each boundary segment and the angles between them, the whole template then being subjected to an uniformly distributed random rotation and a translation drawn from a diffuse prior. For simplicity, we assume here that the joint density for the deformation is proportional to the product of terms for the local perturbations which have the form of normal densities. Of course, the polygon must remain closed and non-intersecting, which places a very complicated constraint on the individual perturbations. Therefore, in spite of the product form, the local perturbations are not *a priori* independent and the normalising constraint for the joint density is completely intractable. One of the beauties of MCMC calculation is that no attention need be paid to either of these points in constructing a Metropolis-Hastings sampler, apart from checking that proposed moves satisfy the constraints.

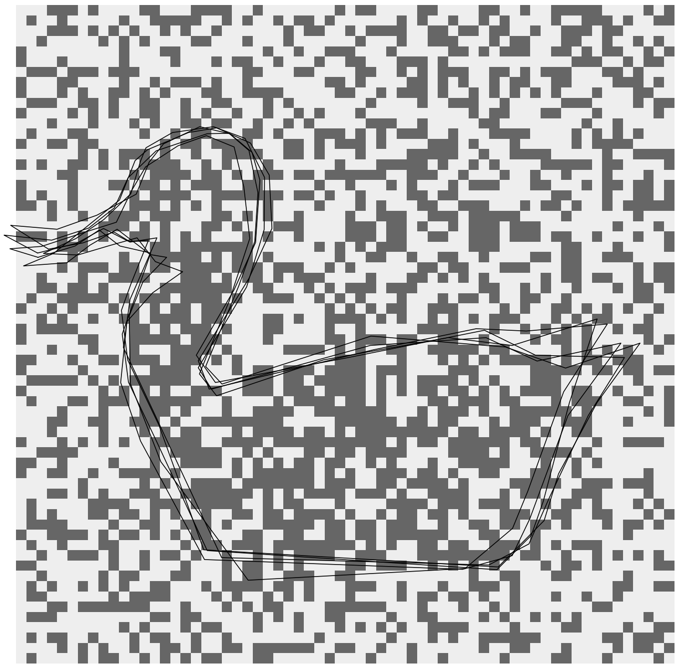

Figure 3: Template-based restoration of a polygonal silhouette: binary channel noise with 40% error rate.

Two simple alternative models for the likelihood $p(y|x)$ are used. In each case,

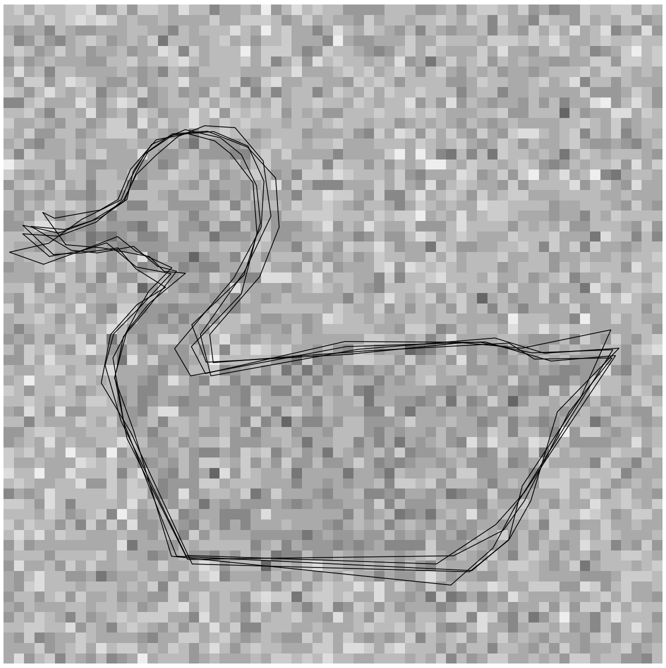

Figure 4: Template-based restoration of a polygonal silhouette: gaussian noise, with standard deviation 1.5 times the difference in means.

the data are pixel values independently generated, depending on the polygon only according to whether the pixel centre is inside or outside the polygon. The two models used are a symmetric binary channel, with 40% error rate, and additive gaussian noise, with standard deviation 1.5 times the difference in means.

A Metropolis-Hastings sampler was used to simulate from the resulting posterior. Each proposal consisted of a random displacement to a single vertex of the template, drawn uniformly from a small square centred at the current position. Such a move affects two boundary segment lengths and three inter-segment angles; it also changes the set of data pixel centres that are interior to the template, thus affecting a variable number of likelihood terms. The details are straightforward but notationally lengthy, so will not be further presented here.

The results from two experiments are displayed in Figures 3 and 4, showing cases with binary channel and gaussian noise respectively. In each case, five polygons are superimposed on an image of the data to give a rough impression of the variation in the posterior: these are the states of the simulation after 200, 400, 600, 800 and 1000 complete sweeps, each visiting every vertex once.

3.5 Stochastic geometry models

A third approach to the issue of image modelling is one that again aims at high level image attributes, but uses established models from stochastic geometry, rather than purpose-built processes. The cleanest general approach is one which objects are considered as realised points in a marked point process in an appropriate abstract object space. Models are specified by their densities with respect to the corresponding Poisson process. This general framework, some interesting classes of interacting object processes, and associated MCMC methods are described by Baddeley and van Lieshout (1994). Baddeley (1994) also notes that the object processes of Grenander and Miller (1994) can usefully be set in this framework.

A more specifically focused approach builds on the polygonal random field models of Arak and Surgailis (1989), and has potential for dealing with either pixellated images or point process data generated from an assumed underlying true image consisting of polygonal shapes. This is explored by Clifford and Middleton (1989), and extended by Clifford and Nicholls (1994),

All these approaches have yet to find concrete application, but they certainly seem to provide a rigorous and flexible basis for high-level imaging. Again, credible alternatives to MCMC methods for computing with these models have yet to be proposed.

3.6 Hierarchical modelling

All of the models described so far are simple in general structure: there are just two levels, x and y, modelling of their joint description is via specified $p(x)$ and $p(y|x)$, and y is observed. Within each level, dependencies are modelled non-directionally.

In the language of graphical modelling, this is a simple chain graph model (Whittaker, 1990). Richer chain graphs have been used; for example, Phillips and Smith (1992) use hierarchical models for features in images of human faces.

Wilson and Green (1993) use a four-level model for satellite remote sensing. The four levels are the ground classification, the emitted radiance, the detected radiance and the recorded data; only the latter is observed. The merit of introducing the additional layers between the object of inference and the data is that interactions can be kept more local, and each stage of modelling reflects a separate physical process.

In this case, all the levels are discretised on the same rectangular grid, but applications for other multi-level models can be envisaged, perhaps combining template and pixel based models at different levels.

4 Methodological innovations in MCMC stimulated by imaging

In very recent years, the search for efficient MCMC methods in image problems has continued to stimulate new methodology, that may find use in general Bayesian inference.

A typical difficulty with single-variable updating samplers is that they mix slowly in the presence of multimodality, which can occur when the data is relatively uninformative compared to the (spatially structured) prior. This is related to the phenomenon of phase transition in statistical physics, and similar solutions have been explored to combat it.

In auxiliary variable algorithms the variable x is augmented by additional variables, u, with an arbitrary specified conditional distribution $p(u|x)$. The joint distribution $p(x,u) = p(x)p(u|x)$ is then the target of a MCMC simulation; note that the marginal distribution of the original variables remains $p(x)$. With careful design, the resulting simulation can provide more efficient estimation than does the original process.

The first auxiliary variable method was the revolutionary algorithm designed to reduce critical slowing down in Potts models (1), proposed by Swendsen and Wang in 1987. If β is large, MCMC using single site updating algorithms mixes very slowly. The Swendsen-Wang algorithm introduces conditionally independent, binary auxiliary bond variables u_{ij} for each neighbour pair, with $p(u_{ij} = 1|x) = 1 - \exp(-\beta I[x_i = x_j])$; here $u_{ij} = 1$ indicates presence of a bond. Consideration of the form of the conditional density $p(x|u)$ shows that this is a uniform distribution over pixel labels, subject to the constraints that clusters of pixels formed by pairs bonded together have the same label. This is equivalent to x being partitioned into uniformly labelled clusters with labels drawn independently.

The Swendsen-Wang algorithm is easily extended to some more general models and to posterior distributions, but works poorly when symmetry is lost, as in some applications to Bayesian image analysis (Higdon, 1993; Gray, 1993).

The concept of auxiliary variables extends beyond the use of binary bonds (Edwards and Sokal, 1988; Besag and Green, 1993). In general, awkward interactions are reduced in strength, or eliminated altogether, in exchange for constraints on the variables.

There are modifications of the Swendsen-Wang algorithm that control the size of clusters by modifying the distribution of the bond variables, and these can be expressed in this framework. There include explicitly limiting cluster size stochastically (Kandel *et al.*, 1989), preventing clusters growing outside deterministic blocks of pixels (Han, 1993), or having bond probabilities depend on the data (Higdon, 1993). All these approaches can be used with multiple levels of auxiliary variables to capture a multigrid effect (Kandel, *et al.*, 1989; Besag and Green, 1993), with the aim that one or more of the bond levels will generate clusters with a good chance of changing label.

Auxiliary variable methods have proved extremely effective in combating the problems of critical slowing down in calculations with statistical physics models. Their potential in statistical applications with high levels of interaction seems great, but few examples have yet materialised. To work well, the auxiliary variables need to be constructed so that either $p(x|u)$ can be simulated directly without the need for rejection methods, or there is a rapidly-mixing Hastings algorithm for this conditional distribution.

5 Discussion

Both the general descriptions and the detailed examples above affirm the very valuable role of MCMC methods in computing with the models of Bayesian image analysis, in both low-level and high-level problems. As in other domains of application of modern Bayesian statistics, the availability of these methods encourages the "model-liberation movement"; in addition the sheer size of the data structures involved means that there are really no available alternative general computational methods if variability is to be assessed.

Among MCMC methods, the Gibbs sampler has very limited use in image analysis. Only in classification problems where marginal distributions take few values, or in purely gaussian models where simulation-based calculations are unnecessary, can the Gibbs sampler be competitive in computational efficiency. Metropolis-Hastings methods updating pixel values or picture attributes singly or in small groups are routine to code and efficient: the spatial stationarity usually assumed in image models cuts out any need for experimentation with proposal distributions tailored to individual variables.

Sensitivity analysis is an important aspect of any responsible statistical inference. In Bayesian image analysis, in contrast to some other applications covered by this volume, prior distributions remain highly influential even with a great volume of data (since great volume usually means a large field of view rather than replication or small variance). Thus sensitivity to the prior may be an important issue. MCMC is well attuned to sensitivity analysis, a point given particular emphasis in Besag *et al.* (1995).

On the negative side, MCMC is only available as a computational technique when a full probabilistic model is specified. This is sometimes awkward in image problems, as the field of view (especially in biological microscopy, for example) may contain very heterogeneous image structures.

Finally, there are concerns about convergence of MCMC methods. The very large dimensionality of parameter spaces in image problems makes monitoring convergence unusually difficult, while theoretical results, though limited at present, suggest that speed of convergence tends to decrease with dimensionality.

The possibility of phase-transition-like behaviour generated by stochastic/spatial interactions clearly increases such concerns. However, it remains unclear to what extent these issues are important in dealing with the posterior distribution in an image problem, although they certainly are in the prior.

Acknowledgements

This work was supported by the SERC Complex Stochastic Systems Initiative, and by the University of Washington Department of Statistics.

References

Amit, Y., Grenander, U. and Piccioni, M. (1991). Structural image restoration through deformable templates. *Journal of the American Statistical Association*, **86**, 376–387.

Amit, Y. and Manbeck, K. M. (1993). Deformable template models for emission tomography, *IEEE Transactions on Medical Imaging*, **12**, 260–269.

Arak, T. and Surgailis, D. (1989). Markov random fields with polygonal realizations. *Probability theory and related fields*, **80**, 543–579.

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, **337**, 323–342.

Baddeley, A. J. (1994). Contribution to the discussion of paper by Grenander and Miller (1994). *Journal of the Royal Statistical Society* B, **56**, 584–585.

Baddeley, A. J. and van Lieshout, M. N. M. (1994). Stochastic geometry models in high-level vision. To appear in *Statistics and Images* (K. V. Mardia and G. K. Kanji, eds.), Abingdon: Carfax Publishing Co.

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

Besag, J. E. (1989). Towards Bayesian image analysis. *Journal of Applied Statistics*, **16**, 395–407.

Besag, J. E., York, J. C. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with Discussion). *Annals of the Institute of Statistical Mathematics*, **43**, 1–59.

Besag, J. E. and Green, P. J. (1993). Spatial statistics and Bayesian computation (with Discussion). *Journal of the Royal Statistical Society* B, **55**, 25–37.

Besag, J. E. and Higdon, D. M. (1993). Bayesian inference for agricultural field experiments. *Bulletin of the International Statistical Institute*, **LV**, 121–136.

Besag, J., Green, P. J., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems, *University of Bristol Mathematics Research Report S-93-04*, to appear in *Statistical Science*.

Binder, K. (1988). *Monte Carlo Simulation in Statistical Physics: an Introduction*. Berlin: Springer.

Clayton, D. G. and Bernardinelli, L. (1992). Bayesian methods for mapping disease risk. In *Small Area Studies in Geographical and Environmental Epidemiology,* (ed. J. Cuzick and P. Elliott). Oxford: Clarendon.

Clifford, P. and Middleton, R. D. (1989). Reconstruction of polygonal images. *Journal of Applied Statistics*, **16**, 409–422.

Clifford, P. and Nicholls, G. (1994). *Bayesian inference for vector-based images.* To appear in proceedings of IMA conference on Complex Stochastic Systems and Engineering, Leeds, 1993.

Cohen, F. S. and Cooper, D. B. (1987). Simple parallel hierarchical and relaxation algorithms for segmenting non-causal Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**, 195–219.

Creutz, M. (1979). Confinement and the critical dimensionality of space-time. *Physical Review Letters*, **43**, 553–556.

Edwards, R. G. and Sokal, A. D. (1988). Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physical Review* D, **38**, 2009–2012.

Geman, D. (1991). Random fields and inverse problems in imaging. *Lect. Notes Math.*, **1427**. Berlin: Springer.

Geman, D., Geman, S., Graffigne, C. and Dong, P. (1990). Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 609–628.

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., Manbeck, K. M. and McClure, D. E. (1993). A comprehensive statistical model for single photon emission tomography. In: Chellappa, R. and Jain, A., eds. *Markov random fields: theory and application.* Harcourt Brace Jovanovich, Orlando.

Geman, S. and McClure, D. E. (1987). Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute*, **LII-4**, 5–21.

Geman, S., McClure, D. E., and Geman, D. (1992). A nonlinear filter for film restoration and other problems in image processing. *CVGIP: Graphical Models and Image Proc.*, **54**, 281–289.

Gilks, W. R., Clayton, D. G., Spiegelhalter, D. J., Best, N. G., McNeil, A. J., Sharples, L. D. and Kirby, A. J. (1993). Modelling complexity: Applications of Gibbs sampling in medicine (with Discussion). *Journal of the Royal Statistical Society* B, **55**, 39–52.

Glad, I. K. and Sebastiani, G. (1993). A Bayesian approach to synthetic magnetic resonance imaging. Dept. of Mathematical Sciences, Norwegian Institute of Technology, Trondheim, Norway.

Godtliebsen, F. (1989). *A study of image improvement techniques applied to NMR images.* Dr. Ing. thesis, Norwegian Institute of Technology.

Gray, A. J. (1993). Contribution to discussion meeting on the Gibbs sampler and other Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society* B, **55**, 58–61.

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

Green, P. J. (1994). Contribution to the discussion of paper by Grenander and Miller (1994). *Journal of the Royal Statistical Society* B, **56**, 589–590.

Green, P. J. and Titterington, D. M. (1988). Recursive methods in image processing *Bulletin of the International Statistical Institute*, **LII-4**, 51–67 (discussion pp. 95–102).

Green, P. J. and Han, X-L. (1992). Metropolis methods, Gaussian proposals, and antithetic variables. In *Stochastic models, statistical methods and algorithms in image analysis* (P. Barone, A. Frigessi and M. Piccioni, eds.), *Lecture Notes in Statistics*, **74**, 142–164. Berlin: Springer.

Grenander, U. (1983). *Tutorial in pattern theory.* Division of Applied Mathematics, Brown University, Providence.

Grenander, U., Chow, Y. and Keenan, D. M. (1991). Hands: a pattern theoretic study of biological shapes. *Research Notes on Neural Computing*, **2**. Berlin: Springer.

Grenander, U. and Manbeck, K. M. (1992). *Abnormality detection in potatoes by shape and color.* Division of Applied Mathematics, Brown University, Providence.

Grenander, U. and Miller, M. (1994). Representations of knowledge in complex systems (with Discussion). To appear in *Journal of the Royal Statistical Society*, B, **56**, 549–603.

Han, X-L. (1993) *Markov chain Monte Carlo and sampling efficiency.* Ph.D. thesis, University of Bristol.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Heikkinen, J. and Högmander, H. (1994). Fully Bayesian approach to image restoration with an application in biogeography. To appear in *Applied Statistics*, **43**.

Higdon, D. (1993). Contribution to discussion meeting on the Gibbs sampler and other Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society* B, **55**, 78.

Högmander, H. and Møller, J. (1993). *Estimating distribution maps from atlas data using methods of statistical image analysis.* Research Report No. 263, Department of Theoretical Statistics, University of Aarhus. To appear in *Biometrics.*

Kandel, D., Domany, E. and Brandt, A. (1989). Simulations without critical slowing down: Ising and three-state Potts models. *Physical Review* B, **40**, 330–344.

Litton, C. D. and Buck, C. E. (1993). Contribution to discussion meeting on the Gibbs sampler and other Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society* B, **55**, 82.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1091.

Phillips, D. and Smith, A. F. M. (1992). Bayesian faces. *Technical report.* Department of Mathematics, Imperial College, London.

Ripley, B. D. and Sutherland, A. I. (1990). Finding spiral structures in images of galaxies. *Philosophical Transactions of the Royal Society of London* A, **332**, 477–485.

Shepp, L. A. and Vardi, Y. (1982). Maximum likelihood reconstruction in positron emission tomography. *IEEE Transactions on Medical Imaging*, **1**, 113–122.

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.

Srivastava, A., Miller, M. I. and Grenander, U. (1991). Jump diffusion processes for object tracking and direction finding. In *Proc. 29th Annual Allerton Conference on Communication, Control and Computing*, pp. 563–570. University of Illinois.

Swendsen, R. H. and Wang, J. S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, **58**, 86–88.

Weir, I. S. and Green, P. J. (1994) Modelling data from single photon emission computed tomography, *University of Bristol Mathematics Research Report S-93-01.* To appear in *Statistics and Images* (K. V. Mardia and G. K. Kanji, eds.). Abingdon: Carfax Publishing Co.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics.* Chichester: Wiley.

Wilson, J. D. and Green, P. J. (1993) A Bayesian analysis for remotely sensed data, using a hierarchical model. *University of Bristol Mathematics Research Report S-93-02*,