Statistical Image Analysis for a Confocal Microscopy 2D Section of Cartilage Growth

Fahimah Al-Awadhi [*]   Christopher Jennison [†]   Merrilee Hurn[‡]

March 5, 2003

Abstract

Images are the source of information in many areas of scientific enquiry. A common objective in these applications is reconstruction of the true scene from a degraded image. When objects in the image can be described parametrically, reconstruction can proceed by fitting a high level image model. In this article we consider the analysis of *confocal fluorescence microscope* images of cells in an area of cartilage growth. Biological questions posed by the experimenters concern the nature of the cells in the image and changes in their properties with time.

Our model of the imaging process is based on a detailed analysis of the data. We treat the true scene as a realisation of a *marked point process*, incorporating this as the *high-level* prior model in a Bayesian analysis. Inference is by simulation using *reversible jump* versions of *Markov chain Monte Carlo* (MCMC) algorithms which can handle the varying dimension of the image description arising from an unknown number of cells, each with its own parameters.

Keywords: Image analysis; Markov chain Monte Carlo; Object recognition; Marked point process; Confocal fluorescence microscope; Reversible jump MCMC; Stochastic simulation.

1   Introduction

Confocal fluorescence microscopy allows visualisation of a 3D specimen without the need for physical sectioning. Shaw & Rawlins (1991) describe how a confocal microscope uses a system of lenses and pinholes to focus a laser beam at a point within a stained sample and records the fluorescence returning

[*]Department of Statistics and OR, Kuwait University, P.O.Box 5969 Safat, Kuwait, 13060

[†]Corresponding author: Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY

[‡]Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY

1
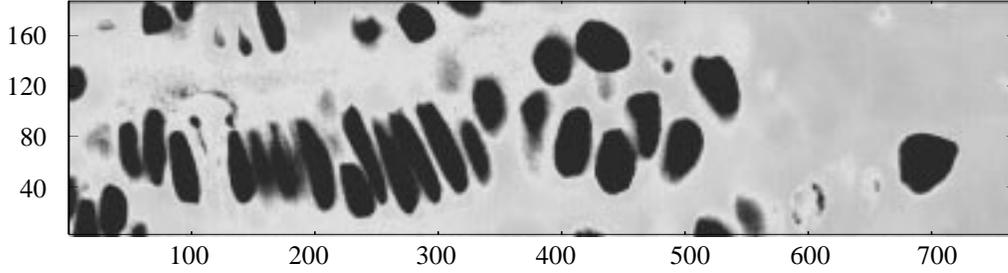
Figure 1: 2D optical section obtained by confocal fluorescence microscopy of cartilage growth.

from that point. Since the stain is taken up differentially by different tissue types, this measurement is an indication of the tissue type at that point. Scanning across a lattice of points creates a 2D section; repeating this process on a series of sections at different depths in the sample gives a 3D image composed of a stack of 2D optical sections. The *signal-to-noise* ratio is reduced by various types of degradation (see Pawley (1996)). Light is scattered as it travels to the point of focus and back again so fluorescence from what should be a point source is spread over a larger region. The image is also affected by diffraction and attenuation of the light as it travels through the specimen (see White et al. (1995)). Hence, each record Y_i is a blurred value of the true scene around the point of focus further degraded by sensor noise. In this article we consider the analysis of a single 2D section, but note that the methods we describe generalise to 3D analyses (see Al-Awadhi (2001)).

The particular example we focus on concerns the section of cartilage growth shown in Figure 1. In this case, the fluorescent stain is taken up by the cells while the background remains unstained; cells appear as elliptical objects on a roughly constant background. This image forms part of a time sequence showing changes in cell size and shape as growth occurs. The goal of the analysis is to generate interval estimates of certain size and shape attributes of the population of cells at a certain state of growth; as cells may move around between consecutive imaging times, changes in population parameters are considered rather than values for individual cells as this would involve a matching problem. We represent a possible true scene X as a list of ellipses and associated intensities. Our statistical model is a *high-level* image model (Hurn & Rue (1997)). We define a prior distribution π_X for X on a suitable sample space $\mathbb{E}$, to capture prior information about the true scene. The scene X determines values for the signal intensity at individual pixels which appear in the likelihood $\mathcal{L}(y|x)$ of the observed data.

We shall use Bayesian image analysis (Besag (1986), Besag et al. (1995)) to combine the likelihood of the recorded data Y with the prior model for the underlying true scene X. Conclusions about X are

2

based on the posterior distribution $\pi_{X|Y}$ of X given the observed data Y, given by

$$\pi_{X|Y}(x|y) \propto \mathcal{L}(y|x)\pi_X(x).$$

In Section 2 we present a model for the recorded data Y and in Section 3 we define a marked point process prior model for the true image X. In describing MCMC simulation in Section 4 we derive explicit formulae, in terms of sub-densities with respect to Lebesgue measure, for the acceptance probabilities of reversible jump transitions that change the number of cells in X. In Section 5 we discuss the performance of the MCMC sampler and note problems of poor mixing and difficulties in exploring the whole sample space: in the remainder of the paper we describe methods which overcome some of these problems. In Section 6 we modify the mechanism for introducing new cells into the scene, making better use of the data, Y, in proposing new cells. We also show how to employ a relaxed model in which constraints on the image X are weakened in order to move around the sample space more easily. Finally, in Section 7 we demonstrate the advantages of starting the MCMC algorithm in a carefully chosen state rather than from an empty image. The starting configuration is constructed using morphological operations to create an initial set of objects, then maximum variance concepts are applied to fit elliptical cells to these objects.

2   Distribution of the recorded data Y

Records Y_j, $j = 1, \ldots, S$, are observed as the microscope scans the 2D section. Our choice of model for the records is based on theoretical aspects of confocal microscopy and empirical findings from analysis of the data. Theoretically one might expect the number of photons counted at the receiver to follow a Poisson distribution if the point being imaged lies in one of the stained cells, and to be close to zero otherwise. However, counts are also linearly scaled by the "dark ground" technique to give values in the range τ_0 to 255, where $\tau_0 > 0$. This scaling adds a constant to all values to improve contrast in a grey-level image, hence the lowest observations are somewhat greater than zero, and at the other extreme, it is sometimes necessary to scale down high values to 255 to bring them into the permitted range.

We model Y_j, $j = 1, \ldots, S$, as independent given X and we approximate their distribution using the normal distribution. Due to scattering, the photon count when the microscope is focussed on pixel j is made up of contributions from a larger neighbourhood of pixels with labels $\{k; k \in \Lambda_j\}$. Each pixel has an associated intensity either at the background level, τ_0, which represents the "dark ground" scaling away from the theoretical value of zero for the background, or at the level for cell i if the pixel lies in cell i. We denote by λ_j the excess of pixel j's intensity over the background level, on the scale of the

recorded data (so $\lambda_j = 0$ if pixel j is a background pixel).

Shaw & Rawlins (1991) approximate the *point spread function* of a confocal microscope as a disc with different diameters in its three dimensions. Wilson (1990) reports confocal microscope experiments which revealed an approximately Gaussian point spread function remaining constant over the image field. Examination of data at the edges of cells led us to take weights from a Gaussian kernel with variance 0.6 in each axis direction. Defining the associated weights w_k, $k \in \Lambda_j$, summing to one, we set

$$E(Y_j|X) = \sum_{k \in \Lambda_j} w_k \lambda_k + \tau_0, \quad j = 1, \ldots S. \tag{1}$$

The two main sources of random variability in these data are Poisson variation in counts emitted by the stained cells, and image-wide instrumentation noise affecting background and cell regions equally. Exploratory analysis of the relation between the cell means and variances in our data set led us to the model

$$Var(Y_j|X) = \phi_0^2 + E(Y_j|X) - \tau_0, \quad j = 1, \ldots S. \tag{2}$$

Here, τ_0 and ϕ_0^2 represent the mean and variance for data values from areas of background (and correspond to the scaling away from zero and the instrumentation noise variance). The increase in the variance of one per unit increase in mean implies that the Poisson counts have been scaled by a factor close to unity, but it should be noted that this is an overall model accommodating other features of the data, including intensity variations within cells and across the background region. Estimated values of $\tau_0 = 38$ and $\phi_0^2 = 46.5$ were obtained for these data.

For a given true image $X = x$, the means (1) and variances (2) define the normal distributions of the records Y_j, $j = 1, \ldots, S$, which we take to be conditionally independent given X. Multiplying these normal densities gives the data likelihood $\mathcal{L}(y|x)$.

3   The Prior Image Model

3.1   Marked point process models

We require the prior distribution for the true image X to incorporate information about the likely number of cells and their location, size, shape, orientation and intensity. The fact that the total number of cells is itself a random variable means that the dimension of the vector defining these cells varies from image to image. Care is needed in expressing the prior model as a density with respect to a measure on the space of possible images. We follow the approach of Baddeley and Van Lieshout (1993) and Van Lieshout

(1995), first defining a "reference measure", Γ, that incorporates some features of the cell distribution but not their mutual interactions, then writing the full model as a density with respect to Γ. In related work, Grenander and Miller (1994) construct a pattern theoretic model in which they define templates regulated by various rules and hypotheses governing the variability in shape of objects under study.

In the marked point process model each object in the scene is represented as a pair (l, m) where l is the location and the "mark" m contains all further information required to identify the object. We model cells as ellipses, each centred on its location l and with semi-axes of length a and b, the semi-axis of length a making an angle θ with the x-axis. Adding the signal intensity level λ gives the cell's mark $m = (a, b, \theta, \lambda)$. Both a and b lie in the interval $M_{axis} = (\min_{axis}, \max_{axis})$, θ takes values in $(0, \pi)$ and λ has range $M_\lambda = (\min_\lambda, \max_\lambda)$. Thus, the mark m lies in $M = M_{axis}^2 \times (0, \pi) \times M_\lambda$. The location l is a pair of Cartesian co-ordinates, lying in the window L which has area $\mathcal{A}_L$. For cell i we write $X_i = (l_i, m_i) = (l_i, a_i, b_i, \theta_i, \lambda_i)$. We denote the class of all possible objects with their locations by $\mathbf{U} = L \times M$.

What is actually seen in the image is determined by the unordered set of objects $\{X_1, \ldots, X_N\}$, where N is the random variable denoting the number of objects present and each $X_i \in \mathbf{U}$, $i = 1, \ldots, N$. It is, however, convenient to define a model for the ordered list of objects $(X_1, \ldots, X_N)$, which takes values in $\mathbb{E} = \cup_n \mathbf{U}^n$, and then deduce the model for the unordered set of X_is from this. The following reference model concerns the ordered list of objects and we shall use this to define prior and posterior distributions for both the ordered and unordered X_is. The model for the recorded data presented in Section 2 associates a single cell or the background region with each pixel: this is determined by the value ascribed in the scene X to the centre of each pixel.

In the reference model, cell locations l follow a homogeneous Poisson point process with intensity one (see Diggle (1983)); letting ρ denote Lebesgue measure, the number of points within a region A has a Poisson distribution with mean $\rho(A)$. For each cell, we assume a, b, θ and λ to be independent with a and b following normal distributions with means μ_a and μ_b respectively and variance ξ^2, but restricted to the interval M_{axis}, θ having a density on $(0, \pi)$ equal to $\{|cos(\theta)| + \pi^{-1}\}/3$, i.e., favouring values near $\theta = 0$ and $\theta = \pi$, and λ following a uniform distribution over the range M_λ. Thus, the mark's density is

$$\frac{c}{2\pi\xi^2} \exp\{-\frac{1}{2\xi^2}[(a - \mu_a)^2 + (b - \mu_b)^2]\} \frac{|cos(\theta)| + \pi^{-1}}{3} \frac{1}{\max_\lambda - \min_\lambda}, \quad m \in M, \qquad (3)$$

where the constant c corrects for the restriction on the distributions of a and b.

Denote the probability measure for the mark distribution by ν. The reference measure can be viewed as a Poisson process on $\mathbf{U} = L \times M$ with intensity $\mu = \rho \oplus \nu$. Equivalently, the total number of

objects N in $\mathbf{U}$ has a Poisson distribution with mean $\mu(\mathbf{U}) = \rho(L) = \mathcal{A}_L$ and, conditional on a total of $N = n$ objects being present, $X_1, \ldots, X_n$ are independent and take values in $\mathbf{U}$ with probability measure $\mu(dx_i)/\mu(\mathbf{U})$. We define the reference measure $\Gamma(x, n)$ to be the measure of this process on $\mathbb{E} = \cup_n \mathbf{U}^n$. Under Γ, the probability that X lies in a given set $A \in \mathbf{U}^n$ is

$$
\begin{aligned}
P(A) &= \frac{\exp(-\mu(\mathbf{U})) \{\mu(\mathbf{U})\}^n}{n!} \int \cdots \int_A \frac{\mu(dx_1)}{\mu(\mathbf{U})} \cdots \frac{\mu(dx_n)}{\mu(\mathbf{U})} \\
&= \frac{\exp(-\mu(\mathbf{U}))}{n!} \int \cdots \int_A \mu(dx_1) \ldots \mu(dx_n).
\end{aligned}
$$

We now define the prior image model as a measure with density $p_X(x, n)$ with respect to the reference measure Γ on $\mathbb{E}$. The one feature to be incorporated concerning interaction between neighbouring cells is that they should not overlap. We specify this by setting

$$
p_X(x, n) = k\beta^n I[\text{No overlap between any pair of cells } x_i \text{ and } x_j], \tag{4}
$$

a *hard core interaction model*. Here k is a normalising constant and the parameter β modifies the overall density of cells. A consequence of the non-overlapping condition is that marks and locations of cells are no longer independent and the total number of cells will tend to be smaller than the Poisson distribution with mean $\beta\rho(L)$ that would arise otherwise. Under the density $p_X(x, n)$ the probability that X lies in the set $A \in \mathbf{U}^n$ is now

$$
P(A) = \frac{\exp(-\mu(\mathbf{U}))}{n!} \int \cdots \int_A p_X(x, n)\mu(dx_1) \ldots \mu(dx_n).
$$

In describing the distribution implied for the unordered set of objects $\{X_1, \ldots, X_n\}$, it is helpful to specify one representative ordered list of objects from each unordered set. We define $\mathcal{U}^n$ to be the subset of $\mathbf{U}^n$ in which $a_1 < a_2 < \ldots < a_n$ (ignoring the possibility of ties as this has probability zero). There is a one-to-one correspondence between unordered sets of objects and points in $\mathcal{U}^n$ and so it suffices to define a distribution on $\mathbb{E}^* = \cup_n \mathcal{U}^n$. Since $n!$ points in $\mathbf{U}^n$ map on to each point in $\mathcal{U}^n$ we define

$$
p_X^*(x, n) = \begin{cases} n!\, p_X(x, n) & x \in \mathcal{U}^n, \\ 0 & \text{otherwise} \end{cases}
$$

to give the density with respect to Γ for a representation of the unordered object lists. This simple relation between models for ordered and unordered lists facilitates switching between the two descriptions, a useful option since each can be the more convenient choice on different occasions.

3.2 Expressing the prior and posterior image models as sums of Lebesgue densities

The flexibility offered by marked point processes to describe very general distributions is reflected in their abstract definition. In using these models, we shall find it helpful to express them as sums of simpler components, each of which is of fixed dimension and can be written as a sub-density with respect to Lebesgue measure.

The location and mark of a single cell i is $x_i = (l_i, a_i, b_i, \theta_i, \lambda_i)$, where l_i comprises two co-ordinates, and the reference process involves the density

$$f_X(x_i) = \frac{1}{\mathcal{A}_L} \frac{c}{2\pi \xi^2} \exp\{-\frac{1}{2\xi^2}[(a_i - \mu_a)^2 + (b_i - \mu_b)^2]\} \frac{|cos(\theta_i)| + \pi^{-1}}{3} \frac{1}{\max_\lambda - \min_\lambda} \quad (5)$$

for $l_i \in L$ and $(a_i, b_i, \theta_i, \lambda_i) \in M$, with $f_X(x_i) = 0$ for all other values of $x_i \in \mathbb{R}^{6n}$. The reference measure $\Gamma(x, n)$ is a collection of sub-measures $\Gamma^n(x)$, $n = 0, 1, \ldots$, where $\Gamma^n(x)$ concerns images containing n cells defined, therefore, by $6n$ parameters. The sub-density of $\Gamma^n(x)$ with respect to Lebesgue measure on $\mathbb{R}^{6n}$ is

$$p_\Gamma^n(x) = \frac{e^{-\mathcal{A}_L} \mathcal{A}_L^n}{n!} \prod_{i=1}^n f_X(x_i) \quad \text{for } x_i \in \mathbf{U}, i = 1, \ldots, n,$$

(remembering that $\mu(\mathbf{U}) = \mathcal{A}_L$).

Since the prior model has density $p_X(x, n)$ given by (4) with respect to Γ, it can be written as a sum of sub-measures $\pi_X^n(x)$, $n = 0, 1, \ldots$, where $\pi_X^n(x)$ has sub-density with respect to Lebesgue measure on $\mathbb{R}^{6n}$ equal to

$$
\begin{aligned}
p_X^n(x) &= p_X(x, n) \, p_\Gamma^n(x) \\
&= k \, \beta^n \frac{e^{-\mathcal{A}_L} \mathcal{A}_L^n}{n!} \prod_{i=1}^n f_X(x_i) \, I[\text{No overlap between any pair of cells}]
\end{aligned}
$$

for $x_i \in \mathbf{U}, i = 1, \ldots, n$.

The density $p_{X|Y}(x, n|y)$ of the posterior image distribution with respect to Γ is proportional to the product of the prior density $p_X(x, n)$ and the likelihood of the observed data $\mathcal{L}(y|x)$. Hence, the posterior distribution is the sum of sub-measures $\pi_{X|Y}^n(x|y)$ with sub-densities

$$p_{X|Y}^n(x|y) \propto p_X(x, n) \, p_\Gamma^n(x) \, \mathcal{L}(y|x), \quad n = 0, 1, \ldots,$$

with respect to Lebesgue measure on $\mathbb{R}^{6n}$. The nth sub-density can be written as

$$p_{X|Y}^n(x|y) = K \, \beta^n \frac{e^{-\mathcal{A}_L} \mathcal{A}_L^n}{n!} \prod_{i=1}^n f_X(x_i) \, I[\text{No overlap between any pair of cells}] \, \mathcal{L}(y|x) \quad (6)$$

for $x_i \in \mathbf{U}, i = 1, \ldots, n$, where the normalising constant K does *not* depend on n, so ratios of sub-densities in different dimensional spaces can be calculated from this formula.

4 Simulation from the posterior image distribution

4.1 MCMC simulation

The posterior image distribution $\pi_{X|Y}$ is intractable analytically and we shall use Markov chain Monte Carlo (MCMC) simulation to provide samples from $\pi_{X|Y}$ as a means of statistical inference. Generating correlated samples from a distribution π by running a Markov chain with ergodic distribution π is now standard practice. For reviews of popular MCMC methods including the *Metropolis-Hastings* algorithm (Metropolis et al. (1953) and Hastings (1970)) and *Gibbs sampler* (Geman & Geman (1984)) see Smith & Roberts (1993), Tierney (1994) and Gilks et al. (1996).

In running a Markov chain on the image space $\mathbb{E}$, we use a variety of move types to update the current state. Some moves operate on the existing set of cells: translating a cell (the *shift* move), modifying the semi-axes of a cell (*resize*), changing a cell's orientation (*rotate*) and updating a cell's intensity level (*change intensity*). These transitions follow standard lines and we omit further details.

Other types of move change the number of cells in the image X: adding a cell (*birth*), deleting a cell (*death*), dividing a cell into two new cells (*split*) or combining two adjacent cells (*merge*). Special methods are needed to define such dimension-changing moves in a way that gives the desired ergodic distribution for the Markov chain. Geyer and Møller (1994) suggested a Metropolis-Hastings alternative to spatial birth and death processes for simulating spatial point processes. Their approach is a special case of reversible jump, or trans-dimensional, Markov chain Monte Carlo introduced by Green (1995) and more recently overviewed by Green (2003), which provides an elegant solution to the problem. Tierney (1998) describes the algorithm very generally in measure theoretic terms. We have seen that in our example $\pi_{X|Y}$ is made up of a collection of sub-distributions, each of which can be expressed as a sub-density with respect to Lebesgue measure in a fixed dimension. This represents an important special case for which a relatively simple treatment is possible. We give implementation details in this situation with the birth and death moves as a specific example.

4.2 Implementing the reversible jump algorithm

We wish to sample from the posterior image distribution $\pi_{X|Y}$ made up of sub-measures $\pi_{X|Y}^n$ with densities $p_{X|Y}^n(x|y)$ on $\mathbb{R}^{6n}$, $n = 0, 1, \ldots$, given by (6). Suppose ν is a type of move which adds a cell, i.e., a birth or split move, and denote the probability of selecting this move when in state x by $j_\nu(x)$. Let ν' be the reverse move, i.e., a death if ν is a birth or a merge if ν is a split, and denote the probability of

choosing this move when in state x' by $j_{\nu'}(x')$. Suppose $x \in \mathbb{R}^{6n}$, representing an image containing n cells. In carrying out a move of type ν from x, a random variable u is generated from a specified density $q(u)$ on $\mathbb{R}^6$. This determines a proposal state $x' = g(x, u) \in \mathbb{R}^{6(n+1)}$ which is accepted with probability $\alpha_\nu(x, x')$; otherwise the Markov chain remains in state x. Here, the deterministic function g is a bijection between the set of pairs (x, u) where $x \in \mathbb{R}^{6n} \cap \mathbb{E}$ and the set of possible states $x' \in \mathbb{R}^{6(n+1)} \cap \mathbb{E}$ containing $n + 1$ cells. Thus, each pair (x, u) leads to a unique x' and there is a single state $x = x(x')$ from which a given x' can be reached. When the Markov chain is in state x' and move ν' chosen, a move to the state $x = x(x')$ is considered and this proposal is accepted with probability $\alpha_{\nu'}(x', x)$. We now consider the derivation of $\alpha_{\nu'}(x', x)$ to ensure the correct ergodic properties of the Markov chain.

The Markov chain has the desired ergodic distribution $\pi_{X|Y}$ if each combination of move and reverse move maintains detailed balance with respect to $\pi_{X|Y}$. If ν and ν' are such a pair of moves, detailed balance requires that for all sets $A \in \mathbb{R}^{6n}$ and $B \in \mathbb{R}^{6(n+1)}$

$$P\{X \in A \text{ followed by move type } \nu \text{ to } X' \in B\} = P\{X \in B \text{ followed by move type } \nu' \text{ to } X' \in A\},$$

where in each case the initial events $X \in A$ and $X \in B$ are under sampling from $\pi_{X|Y}$. Let C_x denote the set of values u such that $g(x, u) \in B$ and $q(u) > 0$. It suffices to consider pairs of sets A and B such that C_x is non-empty for each $x \in A$ and $x(x') \in A$ for each $x' \in B$ (any other pairs A and B can be reduced to subsets with these properties). Then, the detailed balance condition is equivalent to

$$\int_A \int_{C_x} p_{X|Y}^n(x|y) \, j_\nu(x) \, q(u) \, \alpha_\nu(x, g(x, u)) \, du \, dx = \int_B p_{X|Y}^{n+1}(x'|y) \, j_{\nu'}(x') \, \alpha_{\nu'}(x', x(x')) \, dx'.$$

The left hand double integral can be converted to an integral over B by the change of variable to $x' = g(x, u)$ and, hence, the required condition is met if

$$j_\nu(x) \, p_{X|Y}^n(x|y) \, q(u) \, \alpha_\nu(x, x') \left| \frac{dx'}{d(x, u)} \right|^{-1} = j_{\nu'}(x') \, p_{X|Y}^{n+1}(x'|y) \, \alpha_{\nu'}(x', x)$$

for all x, u and $x' = g(x, u)$. Here $|dx'/d(x, u)|$ denotes the Jacobian of the transformation $x' = g(x, u)$. This equation is satisfied by setting

$$\alpha_\nu(x, x') = \min\{1, R_\nu\}$$

and

$$\alpha_{\nu'}(x', x) = \min\{1, R_\nu^{-1}\}$$

where

$$R_v = \frac{j_{\nu'}(x') \, p_{X|Y}^{n+1}(x'|y)}{j_\nu(x) \, p_{X|Y}^n(x|y) \, q(u)} \left| \frac{dx'}{d(u, x)} \right|. \tag{7}$$

9

4.3 Birth and death moves

We assign equal probabilities to all possible move types available from the current state. Thus, $j_\nu(x) = 1/8$ for each of the eight move types when $n \geq 2$, $j_\nu(x) = 1/7$ for all except the merge move when $n = 1$, and a birth move is selected with probability one when $n = 0$. If the current state x contains n cells and a birth move is chosen, the random variable u is defined as $u = (l, a, b, \theta, \lambda)$, the location and mark of a new cell, and a value is generated from the density $q(u) = f_X(u)$ defined in (5), i.e., the density for a cell's location and mark in the reference distribution Γ. The proposed new state is set as $x' = (x_1, \ldots, x_n, u)$, adding a new cell $x_{n+1} = u$ to the current state. Since $g(x, u)$ is basically the identity function, its Jacobian is equal to one.

If the death move is to provide the reverse of this step, it must always delete the last cell in the list. This may seem a peculiar requirement but for now we suppose this is the case. Following the derivation of Section 4.2, the birth move and death move proposals should be accepted with probabilities

$$\alpha_{\text{birth}}(x, x') = \min\{1, R_\nu\}$$

and

$$\alpha_{\text{death}}(x', x) = \min\{1, R_\nu^{-1}\}$$

where

$$R_\nu = \frac{j_{\text{death}}(x')\, p_{X|Y}^{n+1}(x'|y)}{j_{\text{birth}}(x)\, p_{X|Y}^n(x|y)\, q(u)} \times 1.$$

Substituting from (6) and cancelling terms, we obtain

$$R_\nu = \frac{j_{\text{death}}(x')}{j_{\text{birth}}(x)} \frac{\beta \mathcal{A}_L}{n+1} \frac{\mathcal{L}(y|x')}{\mathcal{L}(y|x)}\, I[\text{No overlap between any pair of cells in } x'].$$

Note that the ratio of likelihoods only involves the intensity λ_i at pixel i if this has a positive weight, w_i, in the mean $E(Y_j|X)$ of a record Y_j which also receives a contribution from pixels in the new cell x_{n+1}.

If we wish the death move to delete a randomly chosen cell, we must also re-define the birth move to provide the appropriate reverse transitions. To do this, the new cell should be inserted at a randomly chosen point in the list $(x_1, \ldots, x_n)$. This adds a factor $1/(n+1)$ to the proposal probability $j_{\text{birth}}q(u)$, cancelling the extra factor $1/(n+1)$ for the choice of cell to delete in the death move probability. So, the acceptance probabilities $\alpha_{\text{birth}}(x, x')$ and $\alpha_{\text{death}}(x', x)$ remain as above.

An alternative route to the same conclusion is to work with the set of cells created by disregarding the ordering in the list $(x_1, \ldots, x_n)$. As explained in Section 3.1, the appropriate prior density is now $p_X^*(x, n) = n!\, p_X(x, n)$ (on a smaller space) and the factor $n!$ also appears in the posterior density
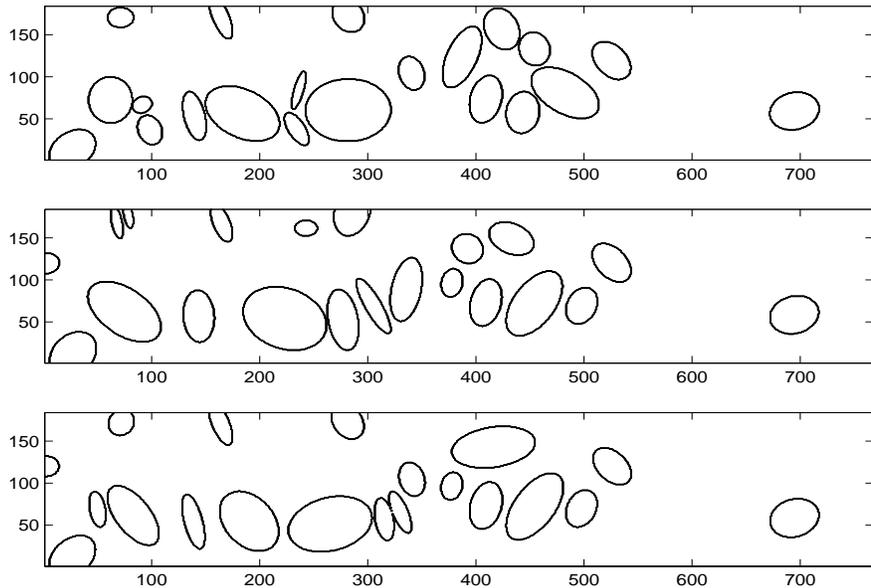
Figure 2: $25,000^{th}$ sample images starting from an empty scene, using different random number seeds.

function. The term R_ν contains an extra factor $(n+1)$ from the ratio $p_{X|Y}^{*,n+1}(x'|y)/p_{X|Y}^{*,n}(x|y)$ but this cancels the $1/(n+1)$ arising from the probability of choosing the correct cell in the death move, and $\alpha_{\text{birth}}(x,x')$ and $\alpha_{\text{death}}(x',x)$ remain unchanged.

Merge and split moves are implemented in the same general framework as birth and death moves. In defining two new cells in the split move, we use the random numbers in u to give equal and opposite perturbations about the location, size, orientation and intensity of the existing cell. Accordingly, the location and mark variables are averaged when combining two cells in the merge move. Further details of these moves are given in Al-Awadhi (2001); a similar mechanism is described by Rue & Hurn (1999). The merge and split moves are crucial in comparing possible true images and reaching suitable inferences when there is uncertainty about whether part of the scene contains a single large cell or two smaller ones.

5 Sampler performance

To assess how well the MCMC sampler performs, we ran 25,000 iterations using three different seeds for the random number generator starting each time from an empty scene ($n = 0$). The model constants were chosen by visual inspection of the size, intensity and shape of cells in the image. Figure 2 shows resulting sample images. A comparison of these images with the data in Figure 1 shows that some cells

11
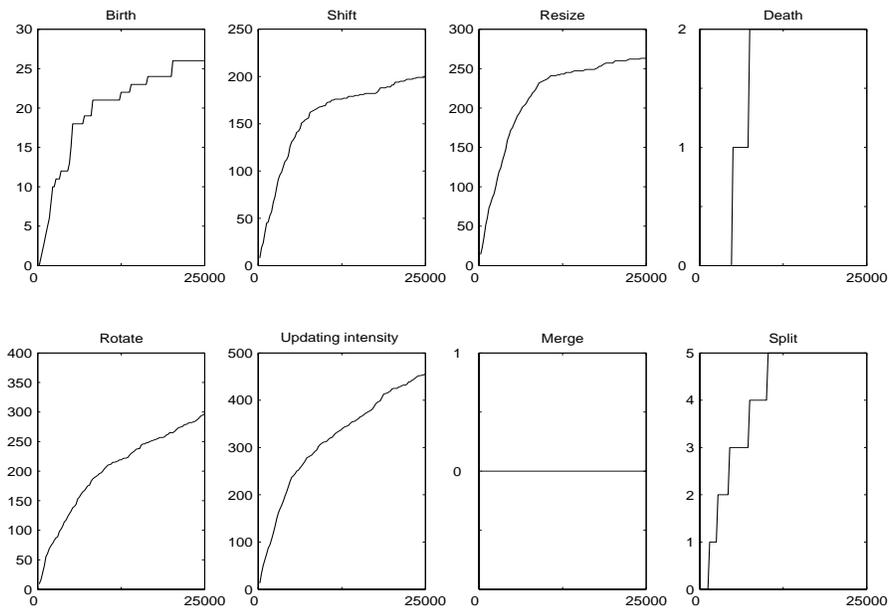
Figure 3: Traces of the cumulative total number of accepted moves of each type.

have been identified in roughly the correct places but more iterations are needed to match the data closely.

Figure 3 traces the cumulative number of accepted moves of each type for one of the three simulations above. Initially birth moves are accepted quite frequently, then they tail off as most of the required cells appear in the image. The acceptance rates for shift, resize and rotate moves remain fairly high throughout the simulation; they are most common just after a cell has been created because the proposal distributions for the location and marks of a new cell do not necessarily guarantee a very good fit to the data. Only two proposed death moves were accepted during this simulation and these occurred for cells created in split moves. Since birth moves are usually accepted because the proposed cell lies in an area of high record values, it is not surprising that the proposed death of such a cell should be rejected. No merge move was accepted at all in this run. Overall, the proportion of acceptances of all move types is very low.

The mixing of the algorithm cannot be described as anywhere near adequate. It seems quite likely that equally good, if not better, point estimates could be obtained using simpler deterministic algorithms. One of the significant benefits of using MCMC ought to be that the posterior distribution is well explored and parameter uncertainty can be expressed through the construction of credible intervals (particularly important in this application, where changes over time are of interest). However, if the algorithm is not mixing adequately, such credible intervals may tell a far from true story about the underlying process.

Paradoxically, the strong signal is a source of difficulty here. The posterior image distribution is

multi-modal and moving between different representations of the cell pattern can often involve passing through intermediate states of prohibitively low probability. A simple instance is when an area of the image may be reasonably interpreted as two thin cells lying side by side or as one wider cell. Transitions for individual cells will improve their fit to the data but the initial proposals in merge or split moves are unlikely to fit the data well enough to offer a realistic chance of acceptance. So, which ever pattern appears first is likely to be retained almost indefinitely. In the next section, we consider two modifications to the algorithm which can improve its performance given limited computing time.

6 Modifications to the standard algorithm

6.1 Improving the birth move

In the basic implementation, cells are proposed at random locations within the window L and many proposals are rejected since they do not fit the data well. Favouring points where the data value Y_j is high when selecting the cell centre improves the acceptance rate. In a modified algorithm we define probabilities p_i, $i = 1, \ldots, S$, as

$$
p_i \; \propto \; \begin{cases} 1 & \text{if } Y_i < 30, \\ Y_i/3 - 9 & \text{if } 30 \leq Y_i < 90, \\ 21 & \text{if } 90 \leq Y_i, \end{cases}
$$

then choose a pixel from this distribution on i and draw the cell centre uniformly from this pixel (of unit area). The probability p_i appears in place of $1/\mathcal{A}_L$ in the denominator of the acceptance probability for the proposed birth. The very high likelihood factor for a well-chosen birth proposal will often lead to an α value of 1 and automatic acceptance. Since these modified birth proposals now have a higher overall acceptance rate, the proportion of iterations proposing births can be reduced. Reallocating the total number of iterations, which in this case is 25,000, between the move types, we increased the probability of proposals other than births. Some samples from MCMC runs with this modification appear in Figure 4. Comparison of these samples with Figure 2 show the same numbers of cells are appearing despite the less frequent use of the birth move. The acceptance rates of the two algorithms are nearly the same. The extra time for other moves should improve the exploration of the posterior around these modes. However, there remains substantial variation between different runs of the sampler and there is no really clear evidence of an overall improvement.

13
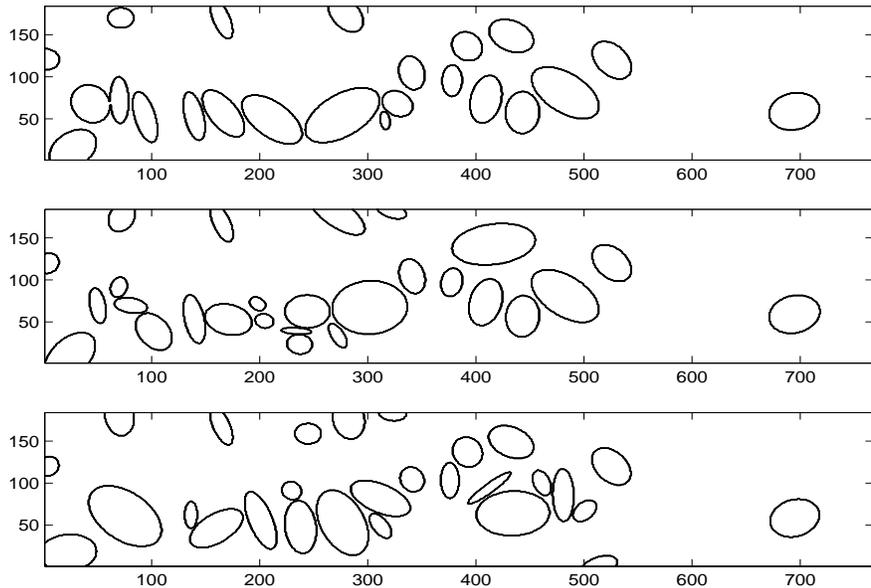
Figure 4: Sample images after modifying the births and changing the probabilities of choosing the moves.

6.2 A relaxed model sampler

For all the MCMC moves except changing a cell's intensity and the death move, there is a danger that a proposal will violate the condition that cells must not overlap, and hence will be automatically rejected. This is particularly a risk in areas of the image where cells are closely packed. Clearly it is wasting time generating such proposals, but it is not necessarily easy to see a more efficient direct proposal mechanism. Introducing intermediate steps into the MCMC in which cells are allowed to overlap could allow the sampler to bridge the gaps between non-overlapping configurations. Hurn et al. (1999) suggest such a relaxed method for sampling an Ising model constrained to have a fixed number of connected components. Because of the constraint, the standard MCMC sampler for the Ising model problem may be reducible: our difficulties where certain transitions are extremely unlikely to succeed have a similar cause and it is appropriate to look for a similar solution.

Let $\mathbb{E}_0$ denote the sub-space of $\mathbb{E}$ in which no cells overlap. We wish to sample from the posterior image distribution $\pi_{X|Y}$ which places all its probability on $\mathbb{E}_0$. In a relaxed model we define a distribution $\tilde{\pi}$ on the whole of $\mathbb{E}$, allowing images with overlapping cells and we specify a secondary Markov chain on $\mathbb{E}$ satisfying detailed balance with respect to $\tilde{\pi}$. We now define a modified MCMC algorithm which still generates a sequence of states $x \in \mathbb{E}_0$ with ergodic distribution $\pi_{X|Y}$ but uses excursions into $\mathbb{E}\backslash\mathbb{E}_0$

14

in generating proposals $x' \in \mathbb{E}_0$. Suppose the current state is $x \in \mathbb{E}_0$. In a separate process from the main Markov chain, we make transitions in $\mathbb{E}$ according to the secondary Markov chain starting at x and continuing until a state x' which is in $\mathbb{E}_0$ is reached. We then use x' as the proposal for the main chain. The arguments of Hurn et al. (1999) show the acceptance probability needed to ensure detailed balance of the main Markov chain with respect to $\pi_{X|Y}$ is

$$
\alpha(x, x') = \min \left\{ 1, \frac{p_{X|Y}^{n'}(x'|y) \, \tilde{p}^n(x)}{p_{X|Y}^n(x|y) \, \tilde{p}^{n'}(x')} \right\}, \tag{8}
$$

where n and n' are the numbers of cells in images x and x' respectively and $\tilde{p}^n(x)$ denotes the sub-density of $\tilde{\pi}$ with respect to Lebesgue measure on $\mathbb{R}^{6n}$. If the proposal of a move to x' is rejected, the chain remains at x. (Apart from generating x', the states in $\mathbb{E} \backslash \mathbb{E}_0$ visited en route do not contribute to the path of the main Markov chain.)

The relaxed distribution $\tilde{\pi}$ should promote transitions between regions of $\mathbb{E}_0$ which do not communicate well in the original Markov chain. It should give significant probability to the sub-space $\mathbb{E}_0$ in order not to take too long a time in reaching a state $x' \in \mathbb{E}_0$. This can be achieved by setting the density of $\tilde{\pi}$ proportional to that of $\pi_{X|Y}$ at states $x \in \mathbb{E}_0$ and penalising states in $\mathbb{E} \backslash \mathbb{E}_0$ by their degree of cell overlap. A simple modification to the posterior sub-density (6) gives our choice for the sub-density of $\tilde{\pi}$ with respect to Lebesgue measure on $\mathbb{R}^{6n}$:

$$
\tilde{p}^n(x) = \tilde{K} \, \beta^n \, \frac{e^{-\mathcal{A}_L} \mathcal{A}_L^n}{n!} \left\{ \prod_{i=1}^n f_X(x_i) \right\} \, \exp\{-\gamma \, v(x)\} \, \tilde{\mathcal{L}}(y|x),
$$

where $v(x)$ denotes the number of pixels in x covered by more than one cell. Here $\tilde{\mathcal{L}}(y|x)$ is a modified version of the likelihood function defined in Section 2 in which the intensity for a pixel covered by more than one cell is taken to be the mean intensity of the cells covering the pixel. The constant γ governs the degree to which instances of overlapping cells are penalised. Higher values of γ lead to greater probability on the set $\mathbb{E}_0$ of non-overlapping states whereas small values increase the chance of moving away from $\mathbb{E}_0$ and exploring the full state space $\mathbb{E}$ more widely. In defining the secondary Markov chain on $\mathbb{E}$, we use the same move types as in the original Markov chain on $\mathbb{E}_0$ but with an appropriately modified acceptance probability for each proposed move.

One can also guard against a run of the secondary chain spending long periods in states far from $\mathbb{E}_0$ by restricting excursions in $\mathbb{E}$ to a maximum length, m say. If $\mathbb{E}_0$ is not visited within m steps, x' is formally set to be the state after step m and when this is not a legitimate state under $\pi_{X|Y}$, it is rejected and the main Markov chain remains at x. Use of the relaxed model can be focussed on a part of the

γ	Number of successful excursions	Mean length of successful excursions
0	7	160
50	7	154
150	12	247
250	10	360
350	12	449
450	7	26
550	7	21

Table 1: Numbers of successful excursions and their average lengths using the relaxed model method.

image by only allowing updates to X which involve cells with centres in a sub-window B of the image domain L: it is easily checked that the acceptance probability (8) still applies. The sub-window itself can be randomly relocated within L for each new excursion.

We have experimented with the relaxed model in a rectangular window with bottom left corner (30,210) and top right corner (110,290), a region containing several tightly packed cells and thus presenting particular difficulty. We set the maximum excursion length to be $m = 2000$ and tried a range of values for the parameter γ. For each value of γ, the algorithm was run for 4,000,000 iterations starting from an empty image (taking an iteration to comprise one proposal whether in the main chain or within an excursion). Table 1 shows the number of successful excursions, defined as excursions which leave $\mathbb{E}_0$ and then return to $\mathbb{E}_0$ providing a legal proposal for the main chain. The average length of these successful excursions is also shown. With low values of γ, excursions are able to move far from $\mathbb{E}_0$ and many fail to return within the allotted time. The extent of overlapping is kept under closer control as γ increases; the highest values of γ allow little opportunity to move very far at all from the legal images in $\mathbb{E}_0$ and those excursions which are "successful" do not generate the large qualitative changes in the image which motivated this method. We therefore selected $\gamma = 150$ as an intermediate value for which excursions explore a significantly large section of image space but still return to provide a legal proposal.

We applied the full method to the whole image using $\gamma = 150$ with each set of relaxed updates confined to a randomly placed sub-window. Realisations after 5,000, 15,000 and 30,000 iterations, starting from an empty image, are shown in Figure 5. The results show that the sampler was able to make some difficult transitions involving merging or splitting cells in the bottom left hand corner of the
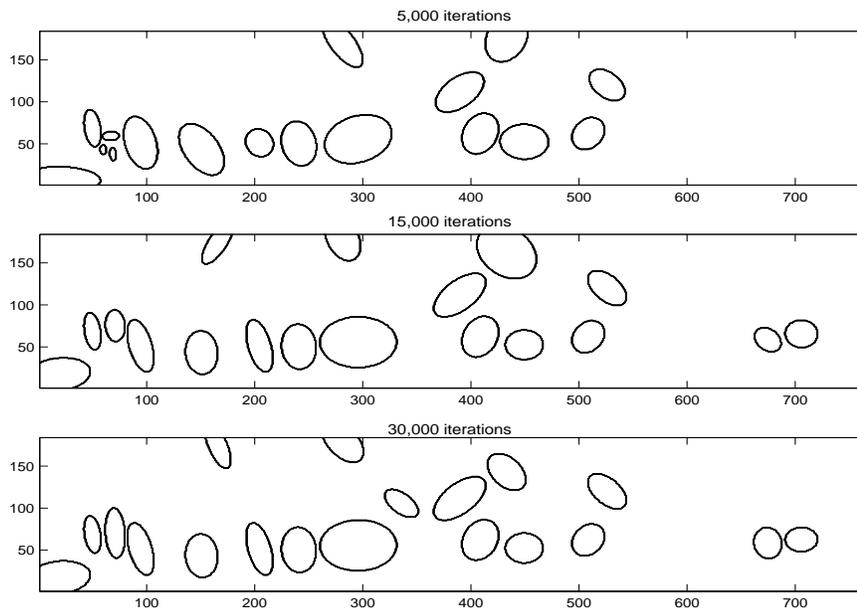
Figure 5: Relaxed model samples taken at iterations (from top to bottom) 5000, 15000, and 30000.

scene. Nevertheless, the final image still does not match the data well in many places. Although this new technique is able to improve a reasonably good image estimate, it does not solve the more basic problem of creating an initial estimate with almost all the cells present ready for fine tuning.

7   Constructing good starting states

7.1   A pragmatic approach

The previous sections have demonstrated the difficulties in creating an MCMC algorithm to sample from our posterior image distribution. It is important to understand the main cause of this problem. The methods we have developed look to be helpful in refining images in or near to the main areas of support of the posterior distribution but they still take very long times to reach such areas in the first place.

In principle, an MCMC sampler visits all parts of the sample space repeatedly, returning more frequently to those parts assigned the greatest probability. In practice, a high dimensional space usually contains vast areas with such low probability that they are unlikely to be visited in any practicable run time. The danger in our example is that all the available computation time is eaten up while the Markov chain works its way through extremely low probability regions towards a plausible section of the image space. A pragmatic solution to this problem is to start the Markov chain at a point which is already close

17

to the part of the image space we really wish to sample. In doing this, we acknowledge that the MCMC sampler does not actually visit large, low probability regions of the sample space; although it may not always be recognised so explicitly, we would argue that the same is true of MCMC samplers in many other applications to high dimensional problems.

Our construction of a good starting point proceeds in two stages. In the first stage we process data at the pixel level, using thresholding and morphological operations to produce an initial approximation to the true image. In the second stage we convert the thresholded image into a set of elliptical objects forming a legitimate scene in our image space $\mathbb{E}$.

7.2 Mathematical morphology

In thresholding, pixels with a record Y_j greater than a certain value are categorised as "object" and the remaining pixels as "background". Because of signal noise the thresholded image is very rough, containing irregularly shaped objects and many isolated object pixels in the background region or background pixels within objects.

Mathematical morphology (Serra (1988), Glasbey & Horgan (1995)) provides tools to tidy up such a binary image. A variety of morphological operators are available for different purposes. We applied the *opening* operator which consists of two basic operations, *erosion* and *dilation*. In the erosion step, object pixels with only a few neighbours classified as "object" are re-labelled as "background". This operation eats away at the edges of connected groups of pixels, removing small objects and narrow necks, the intention being to remove noise and leave a small number of distinct cells. In the dilation step, pixels in a small region around each object pixel are re-labelled as "object", regardless of their previous categorisation. This expands the existing objects in a smooth way but avoids reconnecting previously disconnected regions. Thresholding our image data and then applying the opening operator yielded the upper image in Figure 6.

7.3 Creating elliptical objects

In converting each set of connected pixels into an ellipse, we treat the set of pixel centres as a sample from a bivariate normal distribution, estimating the parameters of this distribution by standard methods and taking a contour of the bivariate normal density as the ellipse. If (X, Y) has a bivariate normal distribution with mean vector (μ_x, μ_y), $Var(X) = \sigma_x^2$, $Var(Y) = \sigma_y^2$ and $Cov(X, Y) = \rho\, \sigma_x \sigma_y$,
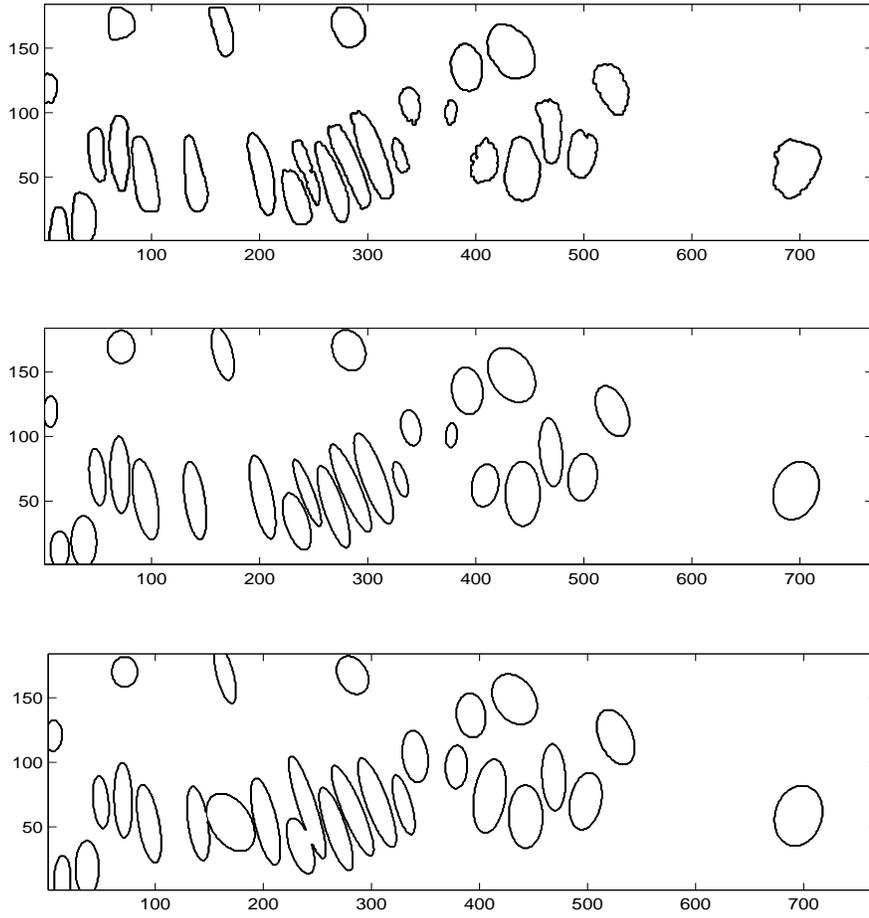
Figure 6: Images obtained by applying thresholding and morphologic operations (top) and converting the connected regions into ellipses (middle); a sample after 20,000 iterations (bottom).

density contours are ellipses of the form

$$\frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} - \frac{2\rho(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y} = C. \tag{9}$$

Re-parameterising (see Tatsuoka (1971)) in terms of $x' = (x - \mu_x)\cos\theta + (y - \mu_y)\sin\theta$ and $y' = -(x - \mu_x)\sin\theta + (y - \mu_y)\cos\theta$, where

$$\theta = \begin{cases} \frac{1}{2}\arctan\{2\rho\sigma_x\sigma_y/(\sigma_x^2 - \sigma_y^2)\} & \sigma_x \neq \sigma_y, \\ \\ 45^\circ & \sigma_x = \sigma_y, \end{cases}$$

equation (9) becomes

$$\frac{x'^2}{\sigma_{x'}^2} + \frac{y'^2}{\sigma_{y'}^2} = C,$$

where

$$\frac{1}{\sigma_{x'}^2} = \frac{\cos^2\theta}{\sigma_x^2} + \frac{\sin^2\theta}{\sigma_y^2} - \frac{2\rho\cos\theta\sin\theta}{\sigma_x\sigma_y}$$

19

and

$$\frac{1}{\sigma_{y'}^2} = \frac{\sin^2\theta}{\sigma_x^2} + \frac{\cos^2\theta}{\sigma_y^2} + \frac{2\rho\cos\theta\,\sin\theta}{\sigma_x\sigma_y}.$$

Since this choice of θ has removed the term in $x'y'$, we see the contour is an ellipse with centre (μ_x, μ_y) and semi-axes of length $\sigma_{x'}\sqrt{C}$ and $\sigma_Y'\sqrt{C}$ inclined at an angle θ to the original x and y axes.

The above formulae with estimates $\hat{\mu}_x$, $\hat{\mu}_y$, $\hat{\sigma}_x$, $\hat{\sigma}_y$ and $\hat{\rho}$ provide a fitted ellipse for each object in the image produced by morphological operations. A suitable value of C needs to be used, balancing the aim that a fitted cell should be large enough to cover the object it is matching with the need to avoid any overlapping cells. Here we set C to be 1.5.

Fitting ellipses to the objects in the upper panel of Figure 6 produced the image in the middle panel. We used this as the initial state for the MCMC algorithm incorporating the modifications described in Section 6, and taking the intensity λ_i for cell i to be the mean of recorded data values for pixels contained in cell i. The bottom panel of Figure 6 shows the $20,000^{th}$ iterations of one MCMC run from this initial configuration. Comparisons with Figure 1 show we now have a very good fit to the data, far superior to the sample images in Figures 2, 4 and 5. Most of the cells apparent to the eye in Figure 1 are present in Figure 6 and the MCMC iterations improve their fit to the data. A notable exception is the pair of cells around (170, 50) which have low intensity and fail to survive the thresholding process: they appear as one large cell in Figure 6 as a result of a birth move with no subsequent split.

8   Discussion

A major goal of the biologists who collected these data is an automatic method for identifying cells and fitting templates, the parameters of which describe cells' size and shape. The data are too noisy and cells too tightly packed in places for standard image processing software. However, the human eye is able to recognise cells and previous analyses were carried out by human interaction, points on the perimeter of each cell being specified by mouse clicks on a computer image. The algorithm of Section 7 provides an *automated* method for cell identification.

One further improvement is possible to correct the few errors that may be present in the starting point produced from a morphological analysis (e.g., the middle panel of Figure 6). These could be overcome by a small amount of human interaction in which instructions are entered through a graphical user interface to provide information about cells that need to be added. Only a rough version of each cell needs to be provided at this point as the fit will be improved by processing in the MCMC stage. Although a fully
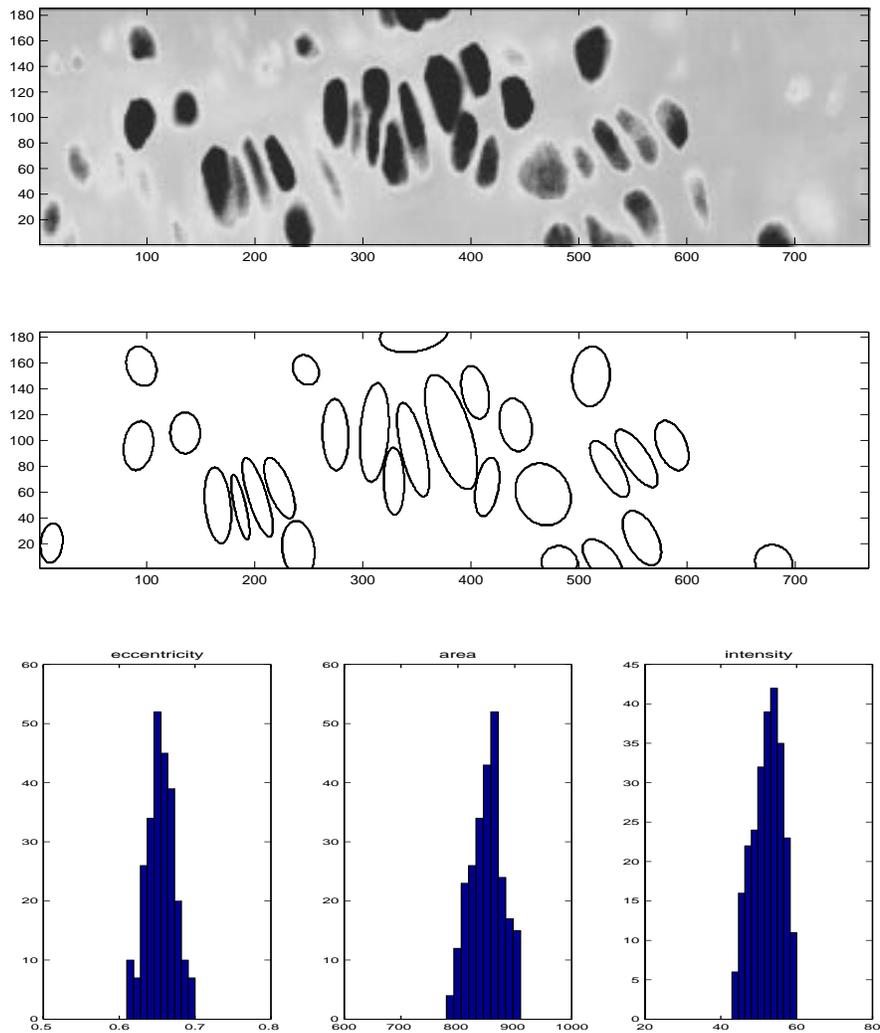
Figure 7: A second data slice (top); a sample from the posterior after 20,000 iterations (middle); histogram of the cell attributes eccentricity, size and intensity (bottom).

automatic method would have been ideal, this is a small amount of interaction compared to the previous approach of outlining the whole perimeter of each cell.

Given a sound initialisation, MCMC sampling provides a straightforward framework for estimating the parameters of fitted cells and relating these to the phases in cell development. For example, we can generate interval estimates of typical cell size or other population attributes. Figure 7 shows another example of a data image, together with a sample image from the posterior distribution and histograms of three cell attributes: the average eccentricity (ratio of minor to major axis length), average size and average intensity in the cell population. Here, attribute values were recorded at regular intervals during the MCMC run of 20,000 iterations.

The statistical model can be elaborated to answer additional questions. As one example, in classifying cells by their shape at different stages of development, we added a mixture distribution for the type of a cell based on its eccentricity. It is of interest to find the posterior probability that a particular cell belongs to either class. However, this is only meaningful if the cell retains an individual identity as the Markov chain progresses, whereas ordinarily a cell can come and go during sampling. Such questions can be addressed in a *conditional* analysis in which the cell population is fixed after a good starting state has been reached. Then, birth, death, split and merge moves are eliminated from the sampler and an identifiability constraint is imposed that each cell continues to cover a particular point, e.g., its centre in the starting configuration. The posterior class probabilities in our example indicate a transition from long narrow cells to more rounded shapes as one moves from left to right across the image.

An important feature of confocal microscopy is the ability to collect three dimensional data in a non-invasive manner. The methods we have described generalise readily to 3D: the signal process is complicated by attenuation of the laser beam to a degree which increases with depth in the sample; the marked point process prior model has a natural extension but with four more parameters per object. The increased dimensionality makes it even more difficult to generate successful proposals for birth and split moves and use of a separate method to provide a good starting state is crucial. Examples of 3D analysis of a stack of 12 two dimensional slices of the cartilage data are given by Al-Awadhi (2001).

9 Acknowledgements

References

[1] Al-Awadhi, F., (2001), Statistical Image Analysis and Confocal Microscopy, *PhD. Thesis*, Department of Mathematical Sciences, University of Bath, UK.

[2] Baddeley, A. and Van Lieshout, M. (1993), Stochastic Geometry Models in High-level Vision, In: *Statistics and Images*, Volume 20, 235–256, (editors K.V. Mardia and G.K. Kanji). Abingdon: Carfax Publishing.

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

[4] Besag, J., Green, P.J., Higdon, D. and Mengersen, K. (1995), Bayesian Computation and Stochastic Systems, *Statistical Science*, **10**, 3–66.

[5] Diggle, P.J. (1983), Statistical Analysis of Spatial Point patterns. London: Academic press.

[6] 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.

[7] Geyer, C. and Møller, J. (1994), Simulation Procedures and Likelihood Inference for Spatial Point Processes, *Scandinavian Journal of Statistics*, **21**, 359–373.

[8] Gilks, W., Richardson, S. and Spiegelhalter D. (1996), Markov Chain Monte Carlo in Practice. London: Chapman and Hall.

[9] Glasbey, C. and Horgan, G. (1995), Image Analysis for the Biological Sciences. Chichester: John Wiley.

[10] Green, P.J. (1995), Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination, *Biometrika*, **82**, 711–732.

[11] Green, P.J. (2003), Trans-dimensional Markov Chain Monte Carlo, In: *Highly Structured Stochastic Systems*, to appear, (editors P.J. Green, N. Hjort and S. Richardson). Oxford: Oxford University Press.

[12] Grenander, U. and Miller, M.I. (1994), Representations of Knowledge in Complex Systems, *Journal of the Royal Statistical Society*, Series B, **56**, 549–603.

[13] Hastings, W. (1970), Monte Carlo Sampling Methods using Markov Chains and their Applications, *Biometrika*, **57**, 97–109.

[14] Hurn, M. and Rue, H. (1997), High level Image Priors in Confocal Microscopy Applications, In: *The Art and Science of Bayesian Image Analysis*, 36–43, (editors K.V. Mardia, C.A. Gill and R.G. Aykroyd). Leeds: Leeds University Press.

[15] Hurn, M., Rue, H. and Sheehan., N.A. (1999), Block Updating in Constrained Markov Chain Monte Carlo Sampling, *Statistics and Probability Letters*, **41**, 353–361.

[16] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953), Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, **21**, 1087–1092.

[17] Pawley, J.B. (1996), Handbook of Biological Confocal Microscopy. London: Plenum Press.

[18] Rue, H. and Hurn, M. (1999), Bayesian Object Identification, *Biometrika*, **86**, 649–660.

[19] Serra, J. (1988), Image Analysis and Mathematical Morphology. Volume 2. London: Academic Press.

[20] Shaw, P.J. and Rawlins, D.J. (1991), Three Dimensional Fluorescent Microscopy, *Progress in Biophysics and Molecular Biology*, **56**, 187–213.

[21] Shaw, P.J. and Rawlins, D.J. (1991), The Point Spread Function of a Confocal Microscope: its Measurement and Use in Deconvolution of the 3-D Data, *Journal of Microscopy*, **163**, 151–165.

[22] Smith, A.F.M. and Roberts, G.O. (1993), Bayesian Computation via the Gibbs Sampler and Related MCMC Methods, *Journal of the Royal Statistical Society*, Series B, **55**, 3–23.

[23] Tatsuoka, M.M. (1971), Multivariate Analysis: Techniques for Educational and Psychological Research. New York: John Wiley.

[24] Tierney, L. (1994), Markov Chain for Exploring Posterior Distributions, *Annals of Statistics*, **22**, 1701–1728.

[25] Tierney, L. (1998), A Note on Metropolis-Hastings Kernels for General State Spaces, *Annals of Applied Probability*, **8**, 1–19.

[26] Van Lieshout, M.N. (1995), Markov Point Processes and their Applications in High-level Imaging, *Bulletin of the International Institute*, Book 2, 559–576.

[27] Wilson, T. (1990), Confocal Microscopy. London: Academic press.

[28] White, N.S., Errington, R.J., Fricker M.D. and Wood, J.L. (1995), Aberration Control in Quantitative Imaging of Botanical Specimens by Multidimensional Fluorescence Microscopy, *Journal of Microscopy*, **181**, 99–116.