THREE DIMENSIONAL STATISTICAL IMAGE ANALYSIS AND CONFOCAL MICROSCOPY

FAHIMAH AL-AWADHI * MERRILEE HURN†
CHRISTOPHER JENNISON ‡

ABSTRACT

An important area of interest in microscopy may be the shape and spatial arrangement of the objects under study, aspects such size, interaction and orientation. Confocal fluorescence microscopy is a useful device for imaging three dimensional specimens because of its non-destructive and non-invasive nature which preserves such spatial information. However, imaging problems still arise; low signal-to-noise ratio, convolution by unknown point spread functions and attenuation of the signal by the object under study lead to degraded blurred images. The goal may be not only to enhance the visual appearance of the 3D images but also to give quantitative summaries of some of shape parameters. This article attempts to tackle the complicated 3D deconvolution problem in a statistical framework using a Bayesian approach. Simulation techniques are required to make inference about the true scene given the 3D data. Interval estimates can be made of size and shape attributes from the posterior distribution.
KEYWORDS: Bayesian statistics; Confocal microscopy; Image analysis; Markov chain Monte Carlo methods; Object recognition; Stochastic simulation; Three-dimensional deconvolution.

*Corresponding author: Department of Statistics and OR, Kuwait University, P.O.Box 5969 Safat, Kuwait, 13060
†Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK
‡Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK

1 INTRODUCTION

The use of mathematical and statistical techniques in biological imaging is now quite advanced (Glasbey and Horgan (1995) [7]). This article aims to describe some of the additional benefits resulting from some of the more recent advances in Bayesian modelling and computational approaches. The paper will be illustrated by considering cartilage data obtained by confocal fluorescence microcopy. Cartilage degradation is a hallmark of some major disease groups such as osteoarthritis (Woessner *et al* (1993) [15]), and aspects such as the types of cells, their numbers and their condition are important in determining the progress of disease.

The structure of this paper is as follows: In Section 2, the data set is described along with problems associated with imaging using confocal microscopy. We then define a statistical model for how the data arise from a particular object configuration (the likelihood), and describe how we may incorporate prior object information into the final model using Bayes theorem. Simulation techniques to handle the statistical model are discussed in Section 3; Sections 3.1 and 3.2 then introduce two approaches to optimise the computational resources by constructing informative initial configurations (the first is fully automated, the second allows user intervention). The output of these algorithms is a sequence of realisations from the statistical model, and it is via these that we make inference about the objects under study, the number of cells or their shape characteristics. The benefit over some existing techniques is that we can construct interval estimates of these attributes, attempting to acknowledge some of the segmentation uncertainty in the problem, and allowing more methodological image-to-image comparisons. Examples of some extracted properties of the cartilage cells are given in Section 4. We finish with some brief conclusions.

2 A BAYESIAN MODEL FOR CONFOCAL MICROSCOPY

2.1 THE LIKELIHOOD FUNCTION

The experimental data consist of a stack of two-dimensional images, which we denote collectively by Y, generated by confocal fluorescence microscopy for an area of cartilage growth. Despite ongoing developments in the technology, the images do

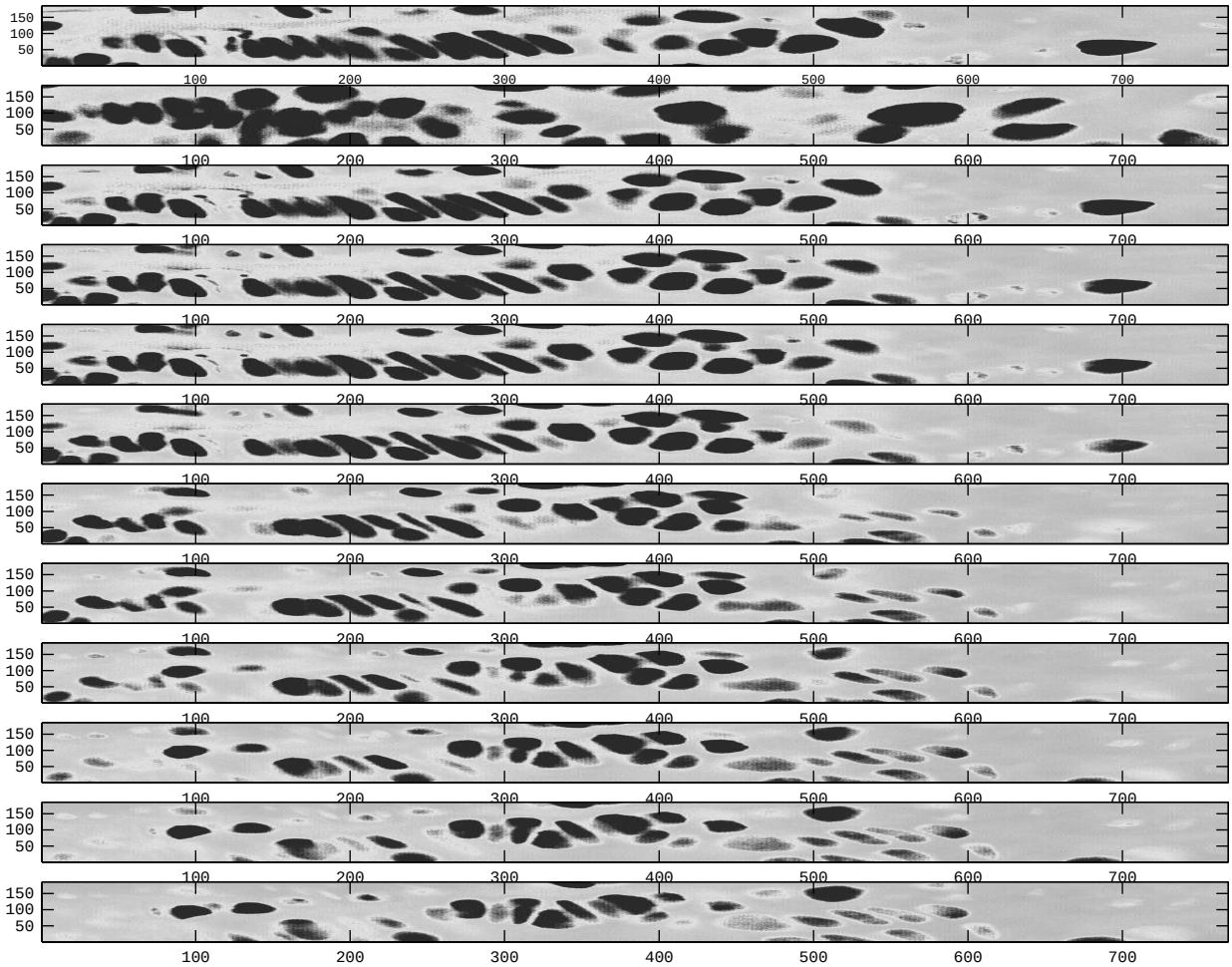

FIGURE 1: The 3D data represented as consecutive 2D optical sections.

suffer from some particular forms of degradation (Wilson (1990) [14], (1996) [11]). Firstly, the images appear blurred and out-of-focus as a result of scattering of the light (incoming and outgoing). Shaw and Rawlins (1991) [12] assessed the point spread function (psf) for different types of microscopes, and concluded that in most planes, it could be described by a central disc surrounded by rings of subsidiary maxima. Attenuation is another important distortion effect in the three dimensional image; due to diffraction and absorption of photons as they pass through the specimen, the signals gets weaker as focal points moves toward deeper optical sections (see Figure 1).

In order to approach the estimation problem in a Bayesian setting, we need to

formulate models for how the data arise from any particular object configuration (the likelihood model) and for any prior beliefs we have about the object (the prior model). We will begin with the likelihood, $L(y|x)$. The microscope measures low numbers of fluorescence photon counts which in theory come from the focal point, but in practice are affected by specimen-dependent spatially varying scatter and attenuation effects. There will then also be some instrument measurement noise. Additionally, the counts may be operator adjusted to improve the contrast (so that a zero count maps to the black-level, and the highest reading is below the maximum reading level).

Let i, j, k index the voxels, where k represents depth into the specimen, and i and j index the pixel grid at any particular depth. The records Y_{ijk} can be considered as scaled convolutions of sensor noise with the blurred, attenuated values of the underlying scene. We assume that the background medium has typical fluorescence level τ_0, and each (cartilage) cell has a typical increase in fluorescence level over and above τ_0 of λ_{ijk} for voxels ijk lying in the cell (with each cell assumed to have a constant increase, and $\lambda_{ijk} = 0$ for voxels in the background). We assume that emissions from each voxel follow a Poisson distribution with mean $\tau_0 + \lambda_{ijk}$. The voxel lattice of expected counts is then formed by assuming a geometric form of attenuation (with parameter Υ), followed by blurring these attenuated values by a three-dimensional discretised Gaussian point spread function (with neighbourhood Λ_{ijk} for voxel (i, j, k) and weights $\{w\}$ summing to 1). Instrument noise is assumed to be additive white noise. The user-adjustment is taken to be a linear scaling (such that a zero count would map to black-level d).

After some approximation, we model the record at voxel ijk as a Gaussian random variable with mean μ_{ijk} equal to the attenuated black-level plus the expected Poisson count for ijk, and with variance σ^2_{ijk} equal to the scaled sum of the variance of the instrument noise plus the expected Poisson variation.

$$\mu_{ijk} = \Upsilon^{k-1} \left\{ \sum_{i'j'k' \in \Lambda_{ijk}} w_{i'j'k'} (\tau_0 + \lambda_{i'j'k'}) \right\} + d \quad (1)$$

$$\sigma^2_{ijk} = \phi_0^2 + (\mu_{ijk} - d) \quad (2)$$

where ϕ_0^2 is the variance of the background area in the top slice. Exploratory analyses of the data suggested taking $d = 20$, $\Upsilon = 0.945$, $\tau_0 = 22$ and $\phi_0^2 = 35.5$. More details

of the two-dimensional version of the model are given in Al-Awadhi *et al* (2004) [2]. A more complicated model for the mean can be used to represent the heterogeneity of the signals inside the cells (see Al-Awadhi (2001) [1]).

Assuming the records Y_{ijk} are conditionally independent given the object configuration $X = x$ (and we will discuss in the next section how we will represent the specimen mathematically), it follows that the likelihood function of the attenuated blurred records Y is

$$L(y|x) = \prod_k \prod_j \prod_i \frac{1}{\sqrt{2\pi\sigma_{ijk}^2}} \exp\{-\frac{(y_{ijk} - \mu_{ijk})^2}{2\sigma_{ijk}^2}\}. \tag{3}$$

2.2 PRIOR IMAGE MODEL

A crucial question is how we intend to represent the objects under study as a mathematical entity in order to formulate prior beliefs about the system using a probability model. Since our objective for this study is inference about the number of cells as well as their characteristics, we require the prior model to handle both the dimensionality of the problem and the geometric features of the list of objects (such as shapes, locations orientations).

Baddeley and Van Lieshout (1993) [3] suggested using marked point processes as a suitable model for an unknown length list of geometric objects. Here the object configuration X is written as a list of unknown length, $X = \{X_1, \ldots, X_N\}$ where N is also a random variable. Each X_i represents a cell, carrying information about fluorescence, orientation etc. The model for X is a marked point process, where the "points" can be thought of as the cells, while the "marks" are the characteristics needed to represent each cell. Interactions between cells can be included by specifying an interaction term in the model which roughly speaking says how likely a configuration is in comparison to a Poisson point process (one in which, conditional on their number, the points are uniformly and independently distributed in the observation window). In this case, we want to prohibit overlap of any of the cells (knowing that the images are optical sections), and so we use the so-called hard-core model. Details of this process for the 2D case are given by Al-Awadhi *et al* (2004) [2].

How to parameterise each cell? The cells have no completely regular shape,

however a good approximation is as ellipsoids with centres c_x, c_y, c_z (these are the points of the point process) and with semi-axes a, b, c representing the width, length and height of the ellipsoids. To allow the cells to lie in different orientations, we consider rotations of the ellipsoids in 3D through the Euler angles $\theta_1, \theta_2, \theta_3$ (Spiegel (1967) [13]). These latter six parameters, along with the intensity level associated with the cell, $\lambda_{cell\ i}$, are used as the marks of the marked point process and so we also need to specify prior distributions for them. We shall assume that the lengths of the axes are bounded by some pre-determined range $(\min_{axis}, \max_{axis})$, and are then independently distributed as truncated Normals on this range with means μ_a, μ_b and μ_c and common variance γ^2. These prior parameters are set to $\mu_a = 25$, $\mu_b = 12.5$, $\mu_c = 3$ and $\gamma = 5$. Each $\theta_i, i = 1, 2, 3$ is assumed uniformly distributed over $[0, \pi]$. For the intensity rate $\lambda_{cell\ i}$, we assume it is uniformly distributed over a specified range $(\min_\lambda, \max_\lambda)$.

Finally we can write the prior distribution, at least up to proportionality, as

$$\pi(x) \propto \beta^n \prod_{i=1}^{n} I[\text{no overlap of cell } i]\pi(a_i, b_i, c_i)\pi(\lambda_i)\pi(\theta_{1i}, \theta_{2i}, \theta_{3i}) \qquad (4)$$

where β is a positive parameter reflecting the intensity of the process, ie related to how many cells we might expect.

2.3 THE POSTERIOR IMAGE DISTRIBUTION

Our posterior distribution $\pi()$ is constructed using Bayes theorem,

$$\pi(x|y) = \frac{L(y|x)\pi(x)}{\pi(y)}.$$

Notice however that we have only specified the likelihood component and the prior component, and so we can only write the posterior distribution up to proportionality as

$$\pi(x|y) \propto L(y|x)\pi(x) \qquad (5)$$

The constant of proportionality could be found by integrating out over all possible configurations x. Or at least it could be found this way if the computations were not intractable as a result of the dimensionality and complexity of the problem! It is clear

that in order to draw inferences from this posterior model, we will need to resort to a computational approach. We describe the approach used in the next section.

3 MARKOV CHAIN MONTE CARLO METHODS

Markov chain Monte Carlo methods are now widely used in statistical applications where the object of interest is too complex to allow for direct evaluation or calculation of estimates. The basic idea is to generate a sequence of (dependent) realisations from the distribution under study by simulating from a Markov chain which has this distribution as its equilibrium distribution. The fact that a large number of samples are drawn from this distribution enables interval estimates to be constructed as well as point estimates. There are now many good references available for these methods, for example Besag *et al* (1995) [4], Gilks *et al.* (1996) [6] and Green (1995) [8], which deals with the important variable-dimension case.

In implementing MCMC methods for this application, we considered so-called single-site updating, that is just one element of X is potentially changed at each step. This suggests using eight different move types: adding a new cell or deleting an existing cell, merging two close cells or splitting one cell into two, shifting a randomly chosen cell's location, resizing it, or changing its orientation or intensity level. In each case, a change is proposed, but the decision whether or not to accept the change is made based on a ratio of terms which include the target distribution $\pi(x|y)$ at the new configuration and at the old configuration (notice that this removes the necessity of knowing the normalising constant of this distribution). More details of this approach can be found for the two-dimensional case in Al-Awadhi *et al* (2004) [2].

MCMC samplers are generally computationally expensive, and they can take many iterations to converge from a poor starting configuration to higher probability regions of the state space. To illustrate this point, runs of 50000 iterations were carried out starting from an empty configuration (ie no cells). Shape and location aspects of a sample from the algorithm are displayed in Figure 2. The sampler has managed to locate most of the cells in the correct places, however, some still need adjusting to fit the data better, ie more runs still are required. Intuitively, it would make more

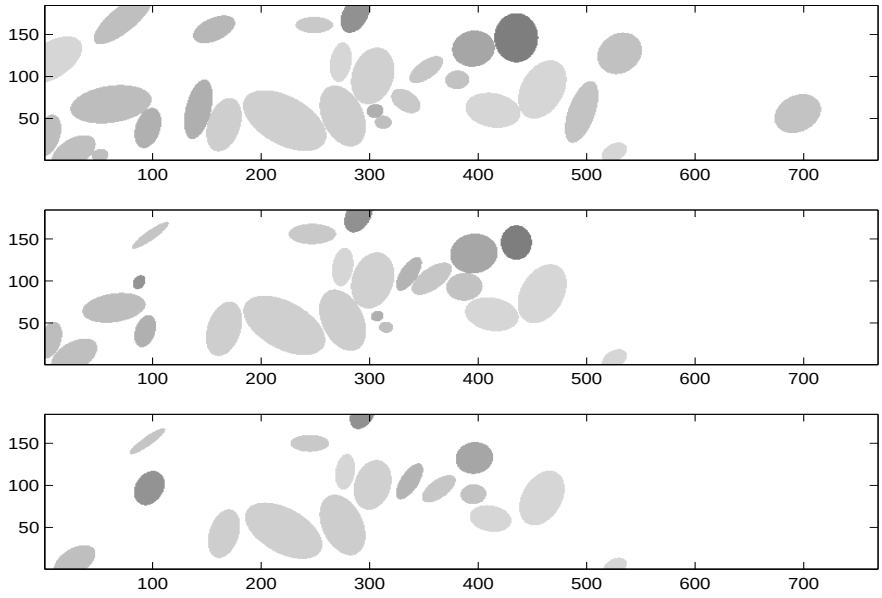

FIGURE 2: 2D sections orthogonal to the $z-$axis for $z = 3$ (top), $z = 7$ (middle) and $z = 12$ (bottom).

sense to begin with the best possible deterministic starting configuration, and to use the stochastic algorithm to assess the uncertainty around this solution. In the next subsections we describe two approaches to doing just this, one requiring no user input, and the other using expert knowledge to refine the method.

3.1 MCMC ALGORITHM USING A GOOD STARTING STATE

Al-Awadhi *et al* (2004) [2] developed an algorithm for automatically constructing good starting states in 2D, which we here generalise to 3D:

- Step 1: threshold the data for each optical section using its median record to obtain a stack of binary images (as a result of attenuation, the median will be higher in high sections than in deeper sections). Convert the binary images to sets of connected voxels using a mathematical morphology opening operator (Glasbey and Horgan (1995) [7]) with a cube of $3 \times 3 \times 3$ voxels used as the structuring element, B. The resulting 3D output after applying the morphological opening operator is shown in Figure 3.

- Step 2: convert the connected voxels into ellipsoids. Each set of connected

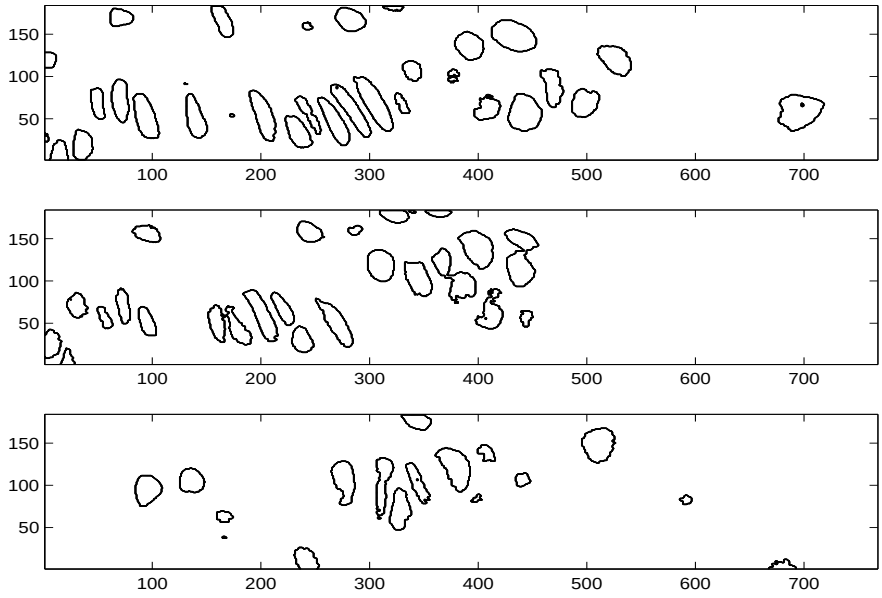

FIGURE 3: 2D sections at depths z=3, 7 and 12 after applying morphological operators.

voxels is viewed as a level set of a 3-D normal density (using the voxel labels as variables). The centre of the ellipsoid (ie the point of the point process) is taken as the mean of the normal distribution, and the values of the semi-axes and the angles of rotation are extracted using maximum variance concepts and the eigenvalues. The cell intensity is computed from the average record inside each ellipsoid. The resulting configuration is given by Figure 4.

Using this hopefully improved 3D starting configuration, the algorithm is again run for 50000 iterations. An output sample corresponding to the final iteration of one run is shown in Figure 5. It is clear that new cells have been created and that the positions of some cells have changed for a better fit to the data during the run. Given that the initial configuration was closer to a reasonable fit, more of the samples could be used for the inference stage. However it is clear that the algorithm must still work hard to reach the modal region of the posterior distribution, particularly where is uncertainty about some cells because voxels lying around them have high record values so the cells have no sharply differentiated edges and the background areas cannot be distinguished easily.

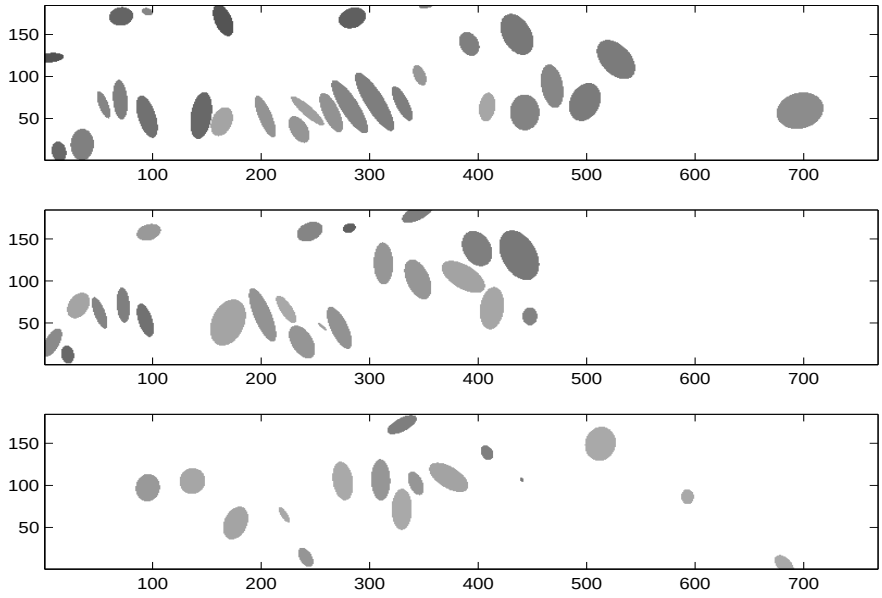

FIGURE 4: 3D configurations after converting the sets of connected voxels into ellipsoids (displayed using three optical sections at z=3, 7 and 12).

3.2 USER INTERVENTION TO CONSTRUCT A GOOD STARTING STATE

An alternative to the unconstrained model used so far would be to utilise the expertise of the user, but in a less time consuming way than for complete manual segmentation. The suggestion is to construct a starting point for the chain as above (ie morphological operators, conversion to ellipses) followed by some iterations of MCMC. At this stage, with now a reasonable configuration, the user is prompted to identify a point inside each cell which they would like to identify in the image. This point does not represent the point of the station process, it is used as a device for identifying cells which the user believes really do exist in the specimen (whether or not they are currently represented as part of X). Adjustments are then carried out to the configuration, removing spurious cells and adding newly identified ones. This can be achieved automatically, maintaining the condition that cells should not overlap. The MCMC is then restarted, but under two conditions. First, the number of cells is now held constant, that is we are restricting ourselves to a fixed value of n in the modified posterior (which rules out the birth/death and split/merge move types). Secondly, the cells are constrained to continue to contain the user-identified points in their interior. This latter constraint is

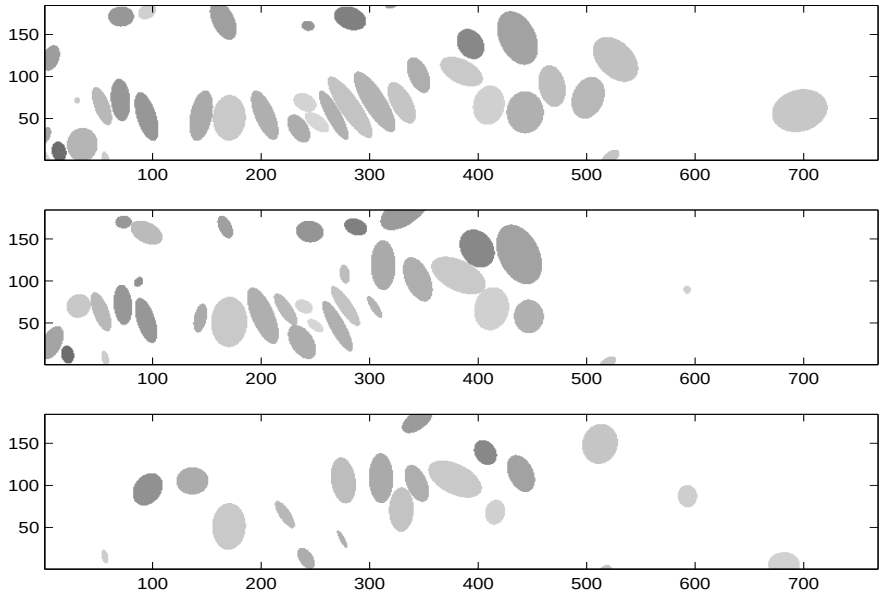

FIGURE 5: A sample after 50000 iterations of the MCMC sampler using the morphological initial state.

imposed by rejecting any proposed configuration which violates the condition.

The unconditional MCMC algorithm output given in Figure 5 is used as input for the user-intervention. The resulting configuration is seen in Figure 6, and a sample after a run with the constrained MCMC is given in Figure 7. The idea is that all of these constrained samples may be taken to contain useful information about the objects under study. The goal now is to extract suitable representations of the image content and devise decision making schemes which interpret the content in terms of their representations.

4 IMAGE INTERPRETATION

4.1 TYPES OF CARTILAGE CELLS

Cartilage tissue can be categorised into four different zones (Woessner *et al* (1993) [15]), the tangential layer, middle layer, deep layer and the calcified cartilage layer. The middle and the lower layers occupy around $80\% - 90\%$ of the tissue, and can be distinguished by the cells they contain. The middle layer contains more rounded cells

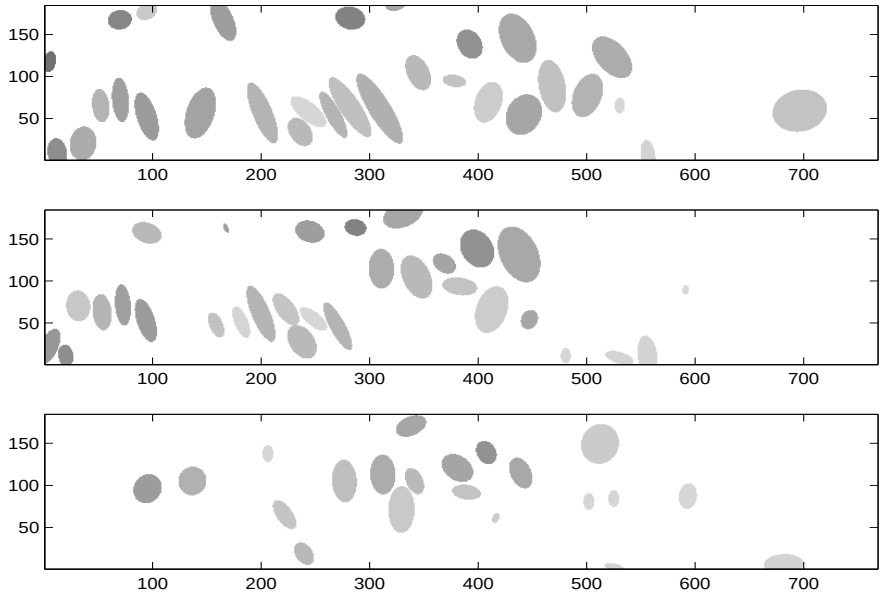

FIGURE 6: The initial configuration after running the MCMC sampler and allowing user intervention.

known as chondroblats, while the deep layer contains more elliptically shaped cells, chondrocytes, arranged in columns. Properties related to the types of the cells such as the ratio of the numbers of the two types, the ratios of shape parameters, their intensity and their size are informative criteria for the biologists to understand the processes in the cartilage tissue, providing information on the proliferation and differentiation of various cellular components of cartilage tissues and elsewhere (Noda (1993) [10]).

4.2 ESTIMATING THE PROPERTIES OF THE CELLS BY MCMC

In order to distinguish cell types, we need to add one further mark to the marked point process, namely a cell label; the chondroblasts cells will be labelled type I, and the others type II. A-priori, the only criterion we have for distinguishing the two types is shape; although both types can be approximated by ellipsoids, type II cells are more spherical than type I cells. This is handled in the prior model through the distributions on their semi-axes. Given the type of the cell, the semi-axes are assumed independent and given a multivariate normal distribution. The means and variance of semi-axes of type I cells are chosen to be $(30, 12.5, 4)$ with equal variances of 4.5^2. For type II cells

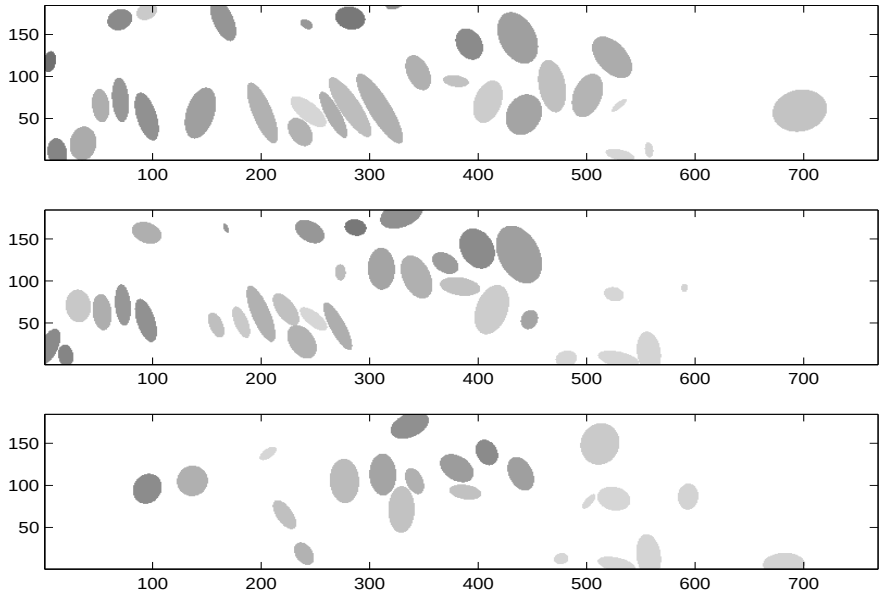

FIGURE 7: 2D optical sections at depths z =3, 7 and 12 after running the constrained MCMC algorithm from the user-refined starting point.

the mean vector is $(20.5, 17.5, 4)$, again with common variance 4.5^2. The two cell types are taken, a-priori, to be equally likely, and the prior distributions for orientation and fluorescence are taken to be common for the two types.

The Markov chain Monte Carlo algorithm also needs extending to deal with cell type; a proposed relabelling move is added which considers changing a cell's label while holding other values fixed. Using this additional move, the conditioned sampler is applied, and after deleting the first 10% of the iterations to allow for the sampler to converge, measurements are made on the samples. The information of interest to be extracted from the samples relates to the size and shape attributes of the two different cell types. Notice that it would also be possible to study attributes of a particular cell, eg the probability it is of Type I, using the conditioned MCMC (such inference would not be possible under the unconditioned algorithm where a cell may not exist for the entirity of the run). Point estimates of the quantities of interest can then be found by taking ergodic averages of the same quantities averaged over the observed samples. Perhaps more importantly, interval estimates of these quantities can also be found (particularly useful for comparative purposes). Since the MCMC samples are

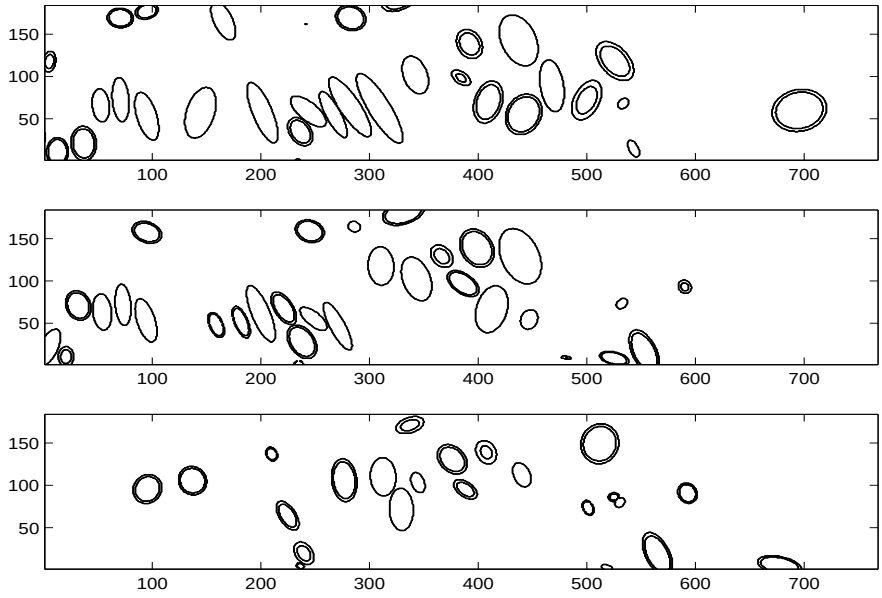

FIGURE 8: A sample image labelling the cells into two types (type I are labelled with a single-line border, and type II with double-line borders).

correlated, the variance of the averages is inflated by a factor called the integrated autocorrelation time, τ, ie, the standard error is $\sqrt{\tau\sigma^2/n}$ rather than $\sqrt{\sigma^2/n}$ as would be the case in the uncorrelated case (here σ^2 is the unknown variance of the attribute, and n is the sample size). The quantities σ^2 and τ can be estimated from the samples, see Green and Han (1992) [9] and Geyer (1991) [5] for details.

Figure 8 shows a sample after 100000 iterations, marking the labelling of each cell. Figure 9 shows histograms of various attributes for the two types of cell based on subsampling one sample in every 100 iterations after neglecting the first 10000 iterations. The estimated proportion of type I cells is 0.442. The estimated attribute means, together with estimated standard errors and τ are given in Table 1.

5 CONCLUSIONS

In this paper, we have presented a Bayesian approach to image interpretation for images of cartilage growth. We have described a modelling approach and associated algorithm for drawing inferences from this model. Although MCMC simulation is computationally intensive, the output allows us to explore more fully the estimation

Attribute	Type	Mean	S.E.	τ
a/c	I	7.845	0.474	4.436
	II	5.337	0.688	19.688
a/b	I	2.738	0.298	14.635
	II	1.658	0.141	11.959
b/c	I	3.235	0.299	4.802
	II	3.602	0.366	12.405
Volume	I	6785.499	529.284	5.818
	II	3767.395	286.122	8.806
Intensity	I	59.168	11.338	13.239
	II	64.711	15.730	43.979

TABLE 1: The estimated attributes, with standard errors and integrated autocorrelation times for the two types of cell.

variability than would a deterministic approach.

REFERENCES

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

[2] Al-Awadhi, F., Jennison, C. and Hurn, M. (2004), Statistical Image Analysis for a Confocal Microscopy Two-dimensional Section of Cartilage Growth, *Applied Statistics, Journal of the Royal Statistical Society, Series C*, 53, 31-49.

[3] Baddeley, A. and van Lieshout, M. (1993), Stochastic Geometry Models in High-level Vision, in K.V.Mardia and G.K. Kanji, eds. Statistics and Images. Carfax Publishing: Abingdon, 231-256.

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

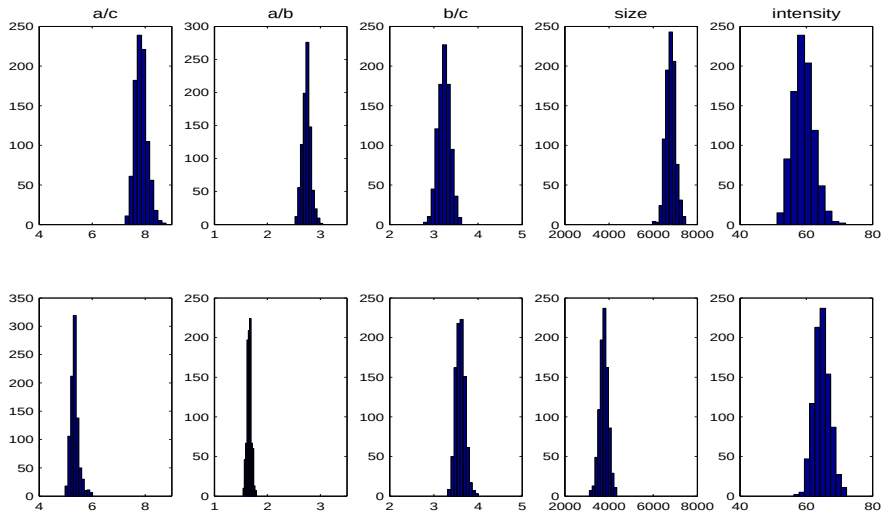

FIGURE 9: Histogram of convexity, volume and intensity for type I cells (top) and type II cells (bottom).

[5] Geyer, C.J. (1991), Markov Chain Monte Carlo Maximum Likelihood, *Computer Science and Statistics*, 23, 156-163.

[6] Gilks, W., Richardson, S. and Spiegelhalter, D. (1995), *Markov Chain Monte Carlo in Practice*. Chapman and Hall: London.

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

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

[9] Green, P.J. and Han, X. (1992), Metropolis Methods, Gaussian Proposals and Antithetic Variables. In: Stochastic Models, Statistical Methods and Algorithms in Image Analysis *Lecture notes in statistics*, 74, 142-164. Berlin: Springer-Verlag.

[10] Noda, M. (1993), Cellular and Molecular Biology of Bone. USA: Academic Press.

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

[12] 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.

[13] Speigel, M.R. (1967), Schaum's Outline of Theory and Theoretical Mechanics with an Introduction to Lagrange's Equations and Hamiltonian Theory. New York: MacGraw-Hill.

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

[15] Woessner, F., David, J. and Howell, S. (1993), Joint Cartilage Degradation, Basic and Clinical Aspects. USA: Marcell Dekker Inc.