Inference for Stereological Extremes

P. Bortot[1], S. G. Coles[2] and S. A. Sisson[3]

[1] Dipartimento di Statistica, Università di Bologna, Bologna, Italy
[2] Dipartimento di Statistica, Università di Padova, Padova, Italy
[3] Department of Statistics, University of New South Wales, Sydney, Australia.

Abstract: In the production of clean steels the occurrence of imperfections — so-called inclusions — is unavoidable. Furthermore, the strength of a clean steel block is largely dependent on the size of the largest imperfection it contains, so inference on extreme inclusion size forms an important part of quality control. Sampling is generally done by measuring imperfections on planar slices, leading to an extreme value version of a standard stereological problem: how to make inference on large inclusions using only the sliced observations. Under the assumption that inclusions are spherical, this problem has previously been tackled using a combination of extreme value models, stereological calculations, a Bayesian hierarchical model and standard Markov chain Monte Carlo (MCMC) techniques. Our objectives in this article are two-fold: to assess the robustness of such inferences with respect to the assumption of spherical inclusions, and to develop an inference procedure that is valid for non-spherical inclusions. We investigate both of these aspects by extending the spherical family for inclusion shapes to a family of ellipsoids. The issue of robustness is then addressed by assessing the performance of the spherical model when fitted to measurements obtained from a simulation of ellipsoidal inclusions. The issue of inference is more difficult, since likelihood calculation is not feasible for the ellipsoidal model. To handle this aspect we propose a modification to a recently developed likelihood-free MCMC algorithm. After verifying the viability and accuracy of the proposed algorithm through a simulation study, we analyze a real inclusion dataset, comparing the inference obtained under the ellipsoidal inclusion model with that previously obtained assuming spherical inclusions.

Keywords: Approximate Bayesian computation; Extreme value theory; Markov chain Monte Carlo; Simulated tempering; Steel inclusion; Stereology.

1 Introduction

A canonical problem in statistical stereology is the inference on a population of objects in three-dimensional space on the basis of a sample observed on a planar intersection. Applications are typically medical or biological – in which the objects might be cells observed on a slice of tissue – or industrial, where the objects might be imperfections in a volume of product. In the inferential process, stereological considerations are needed to allow for the biasing effect of the sampling procedure: for example, the larger objects are more likely to be sampled in this procedure than the smaller ones.

Our study here focuses on an industrial application, the production of clean steels, though the issues and techniques should carry over to a broader range of stereological applications. Ideally, clean steels should be free of imperfections. In practice, the occurrence of microscopically small particles — so-called inclusions — is unavoidable in the production process. Metallurgical considerations suggest that the strength of a block of clean steel is strongly affected by the size of the largest inclusion contained within the block, so inference on the largest inclusion size is important. Typically, sampling of inclusions is carried out by planar slicing and microscopic determination of the cross-sectional size of each observed inclusion on the planar slice. This leads then to a stereological variant of an extreme value problem: deducing the distribution of the largest inclusion in the block given the cross-sectional sample information. Similar considerations also arise in medical applications for which the largest size of a certain cell-type may be indicative of a particular infection or disease.

The problem of drawing inferences on stereological observations was first tackled by Wicksell (1925). The more specific task of making inference on the extremes of objects that are observed only stereologically has only been considered much more recently. Direct extreme value analysis without explicit reference to stereological aspects of the problem has been considered by Shi *et al.* (1999a, 1999b) and Anderson *et al.* (2000). Taking account of the stereology that generates the data, and assuming spherical inclusions, Drees and Reiss (1992) derived from first principles asymptotic families for the distributions of observed diameters. This work was recently generalized by Hlubinka (2003) to the case of spheroidal inclusions. Takahashi and Sibuya (1996, 1998, 2002) consider inference under the assumption of a generalized gamma distribution for the sizes of spherical inclusions. Similar considerations from a more conventional extreme value viewpoint were made by Murakami (1994) and Beretta and Murakami (1998), who assumed a Gumbel distribution for the inclusion diameters. Most recently, Anderson and Coles (2002) proposed a fully Bayesian analysis of the problem, enabling a quantification of estimation precision through the posterior distribution. In their approach the diameter of spherical inclusions is treated as a latent variable whose tail distribution is modelled via a standard family of extreme value models, and the classical calculations of Wicksell (1925) are exploited directly in the formulation of an MCMC algorithm to perform the inference.

Our objectives in this paper are two-fold: first, to assess the sensitivity of inferences made under the modelling procedure of Anderson and Coles (2002) to the assumption of spherical inclusions; and second, to develop an inference procedure that is valid for ellipsoidal inclusions. In Section 2 we detail the spherical inclusion model, and propose an alternative model based on a broader class of ellipsoidal inclusions. We also undertake a simulation study to assess the robustness of the spherical model analysis to non-spherical inclusions. In Section 3 we discuss inference for the ellipsoidal inclusion model, developing a new algorithm based on a likelihood-free version of the standard MCMC algorithm. In Section 4 we compare analyses of real inclusion data based on the assumptions of spherical and ellipsoidal inclusions respectively. We conclude with some comments in Section 5.

2 Stereological Models for Inclusions

2.1 A Model for Spherical Inclusions

Combining earlier work in the metallurgical literature with some reasonable assumptions that derive from extreme value theory, Anderson and Coles (2002) proposed the following model for inclusions:

1. Inclusions are spherical;
2. Inclusion centres follow a homogeneous Poisson process in the space corresponding to a volume of steel;
3. Inclusion diameters are mutually independent and independent of inclusion location;
4. The distribution of inclusion diameters, V, conditional on exceeding a threshold v_0, falls within the parametric family

$$G(v) = \Pr(V \leq v \mid V > v_0) = 1 - \left\{1 + \frac{\xi(v - v_0)}{\sigma}\right\}_+^{-1/\xi}, \quad v > v_0, \tag{1}$$

   where $\sigma > 0$, $\xi \in \mathcal{R}$ and $a_+ = \max(a, 0)$.

Assumptions 2 and 3 are thought to be plausible from a metallurgical point of view. Assumption 1 is bound to be wrong, but it is believed to be a reasonable approximation and hoped not to lead to misleading results. The model in assumption 4 is the generalized Pareto distribution. This distribution is derived from a standard argument in extreme value theory based on exceedances of an asymptotically rising threshold. The special case $\xi = 0$ is interpreted by taking the limit $\xi \to 0$, leading to a translated exponential distribution (Coles, 2001, for example).

In standard extreme value applications, model (1) is fitted to observed threshold exceedances, typically by maximum likelihood. In the stereological setting this approach is not immediately available, as the diameters are unobserved. What is observed is a set of 2-dimensional cross-sectional diameters, $S_1, \ldots, S_n$ say, corresponding to all inclusions that intersect the sampling plane with a cross-sectional diameter greater than some measurement threshold u_s. Associated with each S_i is an unobserved random variable V_i, which is the spherical diameter of the inclusion. The V_i do not have the same distribution as a randomly selected inclusion diameter, however, as a consequence of the size-biased sampling procedure. The problem then is making inference on the assumed tail model of spherical diameters on the basis of the observed variables $S_1, \ldots, S_n$, in which n is also a random variable.

Without the extreme value twist to the application, this problem is the classical corpuscle problem first considered by Wicksell (1925), who derived the distribution of the observed two-dimensional diameters in terms of the distribution of the unobserved three-dimensional diameters. Expressed in slightly non-standard terms to account for the thresholding aspect of the extreme value problem, this result can be stated as

$$\Pr(S \leq s \mid S > v_0) = 1 - \frac{\int_s^\infty (v^2 - s^2)^{1/2} g(v)}{\int_{v_0}^\infty (v^2 - v_0^2)^{1/2} g(v)}, \quad s \geq v_0, \tag{2}$$

where $g(.)$ is the generalised Pareto density function obtained from (1). This result follows from the Poisson assumptions combined with the spherical geometry of the model. Assumptions 2 and 3 also guarantee that the process of inclusions with diameter greater than v_0 is Poisson with a rate that we denote by λ. This observation, together with equation (2), enables a likelihood for the parameters λ, σ and ξ to be developed on the basis of observed planar diameters $s_1, \ldots, s_n$, but the complexity of (2) means that inferences beyond numerical evaluation of the maximum likelihood estimates are not really feasible (Anderson and Coles, 2002).

To resolve this computational difficulty, Anderson and Coles (2002) propose a Bayesian hierarchical formulation of the model, in which the unobserved V_i are treated as latent variables. With this structure and the use of equation (2) the authors develop a simple MCMC scheme for inference on model parameters. They also present an example, that we discuss ourselves in Section 4, which demonstrates the efficacy of the procedure.

2.2 A Model for Non-Spherical Inclusions

Since inclusions are microscopically small, and they are measured only stereologically, it is impossible to know how reasonable the assumption of spherical inclusions is. This raises two questions. Firstly, how sensitive are the inferences made under the assumption of spherical inclusions if, in fact, the inclusions are not spherical? Secondly, if a non-spherical class of inclusions is assumed, how can inferences be performed?

We investigate both of these aspects by substituting the family of spheres with a family of ellipsoids for the inclusion shapes. This seems the most natural generalization, but compared with the spherical class, a more detailed specification is necessary to determine both shape and orientation of an ellipsoidal object. There are limitless possibilities. Again, we opt for the simplest version and assume that the two smaller principal diameters are random uniform multiples of the largest principal diameter, that the inclusions are randomly orientated in space and that the surface measurement is the largest principal diameter of the ellipse generated by the planar intersection. More precisely,

1. Inclusions are ellipsoidal, randomly oriented in space, with principal diameters (V_1, V_2, V_3). Without loss of generality assume $V_3 = \max(V_1, V_2, V_3)$;

2. Inclusion centres follow a homogeneous Poisson process in the space corresponding to a volume of steel;

3. The vectors of diameters are mutually independent between inclusions and independent of inclusion location;

4. The conditional threshold exceedance distribution of V_3 falls within the generalized Pareto family (1) for sufficiently large threshold v_0.

5. $V_j = U_j V_3$, $j = 1, 2$, where U_1 and U_2 are independent uniform $U[0,1]$ variables.

6. The planar measurement S_i is assumed to be the largest principal diameter of the ellipse generated by the planar section of an inclusion.

Note that point 6 is now necessary to avoid ambiguity as the planar intersection generates ellipses rather than circles as in the spherical inclusion case.

2.3 Robustness of Spherical Inclusion Model

Simulation of the ellipsoidal inclusion model is trivial, and it requires only elementary geometry to calculate the planar measurements generated by a configuration of simulated ellipsoids. This provides us with a method to assess the robustness of the spherical inclusion model to shape mis-specification: we simulate data from the ellipsoidal model, fit the spherical model via the MCMC algorithm of Anderson and Coles (2002) using posterior means to estimate parameters, and repeat many times to get empirical estimates of bias. To obtain results that are essentially objective, in all subsequent analyses we use prior distributions that have large marginal variances and that are independent across variables. Checks were carried out throughout to ensure that results are not sensitive to changes within this vague prior specification.

Our simulations across a range of parameter configurations suggest that the biases for σ and ξ are relatively small, and negative and positive in sign respectively, while λ is substantially underestimated. The large bias in λ is not surprising: the overall dimensions of an ellipsoid are smaller than a sphere whose diameter is the same as the largest principal diameter of the ellipsoid. Hence, given the observed planar intersections, a smaller rate of inclusion is predicted under the spherical inclusion model than under the ellipsoidal inclusion model. The biases in σ and ξ are less easily understood. The precise magnitudes of the biases depend in

True value	Bias		
	λ	σ	ξ
$\xi = -0.2$	-71.93	-0.31	0.076
$\xi = 0$	-69.94	-0.236	0.035
$\xi = 0.2$	-66.96	-0.063	0.0042

TABLE 1. Bias in the posterior means of λ, σ and ξ when the spherical inclusion model is fitted to data simulated from the ellipsoidal inclusion model with $\lambda = 100$, $\sigma = 1.5$ and $\xi = -0.2,\ 0$ and 0.2 respectively.

a rather complex way on the true parameter configuration; for illustration, Table 1 gives the biases for fixed values of λ and σ across a range of values for ξ. The general pattern here of diminishing bias for both σ and ξ with increasing values of ξ was found to hold also for other values of λ and σ.

In applications, inference on quantities that involve combinations of parameter values is likely to be more relevant. For example, a standard unit of measurement in the quality control of metals is the characteristic size, v_C, defined so that the expected number of inclusions in a block of volume C with diameter greater than v_C is exactly one. By the various model assumptions

$$v_C = v_0 - \frac{\sigma}{\xi}\left\{1 - (\lambda C)^{\xi}\right\}.$$

When inclusions are ellipsoidal and the spherical model is fitted, the posterior mean of v_C, denoted by $\hat{v}_C$, underestimates the true value for small C, due to the negative bias on λ, but overestimates for large C, due to the positive bias on ξ. The magnitude of such effects depends on all parameter values, but the pattern of variation is most strongly dependent on the value of ξ. To illustrate, Figure 1 shows the graph of v_C against C for various choices of ξ. Also shown on these graphs are the empirical means, and 2.5% and 97.5% quantiles of $\hat{v}_C$, again as a function of C. When $\xi = -0.2$ the sampling distribution of $\hat{v}_C$ is substantially biased, both for small and large values of C. For $\xi = 0$ the same conclusions hold, but the effects are less exaggerated. For $\xi = 0.2$ the spherical model appears to make near perfect inferences on v_C.

In summary, the poor estimation of λ is influential on $\hat{v}_C$ only for small values of C. For larger values of C, discrepancies between $\hat{v}_C$ and the true value are due to the errors in estimating σ and ξ. These errors tend to be compensatory, except for very large values of C, when ξ becomes the dominant term. When ξ is moderately positive, corresponding to a reasonably heavy-tailed diameter distribution, even these errors are negligible in practice.

We stress, moreover, that the robustness argument cuts both ways. Fitting the ellipsoidal model to data generated from spherical inclusions - using the technique discussed in the following section - results in errors of a similar order of magnitude, but in the reverse direction. Again, the dominant error is in the inclusion rates, which are substantially over-estimated with the false ellipsoidal assumption. Furthermore, changing assumption 5 on the relative magnitudes of ellipsoidal diameters can lead to models that are even less similar to the spherical model, so the magnitude of errors arising from model misspecification could be even greater. We stress that our aim is not to choose between shape models - as information on shape is unavailable within the stereological observations - but to assess the robustness of the spherical model and to develop a procedure for inference when a non-spherical shape form for inclusions is proposed.

3 Likelihood-Free MCMC

3.1 MCMC Inference on the Inclusion Model

As observed in the previous section, the mis-specification of inclusion shapes can lead to inferences that are biased, possibly severely so. Consequently, if inclusions were known to be non-spherical, more accurate inference could be obtained, at least in principle, by working with the true family of inclusion shapes. But there is a difficulty: Wicksell's formula and the related equations linking the three-dimensional inclusion process and the two-dimensional observations process, which enabled the likelihood-based Bayesian analysis of Anderson and Coles (2002), are valid only in the spherical case. Some generalizations have been developed, most notably by Wicksell himself in Wicksell (1926) to classes of ellipsoids with fixed shape and by Hlubinka (2003) to the class of spheroids. The difficulty in extending results to more general families, including our ellipsoid family, is discussed by Baddeley and Jensen (2004). Hence there appears to be no simple way to perform the likelihood calculations that would be an integral part of an MCMC algorithm or any other standard likelihood-based inference procedure. Consequently, we turn to a family of stochastic simulation algorithms, termed 'approximate Bayesian computation' techniques (Beaumont *et al.* , 2002). These have been recently developed in the field of statistical genetics to enable inference on models that have a simple parametric structure, but an analytically or computationally intractable likelihood. Broadly, such techniques substitute likelihood evaluation with model simulation, and are therefore useful only if model simulation is both feasible and inexpensive. Fortunately, this is the case for the ellipsoidal inclusion model.

3.2 A Likelihood-Free MCMC Algorithm

Our starting point is the approximate MCMC algorithm developed by Marjoram *et al.* (2003), but see also Plagnol and Tavaré (2003). Let $f(y \mid \theta)$ denote the probability (density) function of a random (vector) variable Y parameterized by $\theta \in \Theta$, and suppose y_0 is the observed value of Y. Denote the prior distribution on θ by $\pi(\theta)$. Departing temporarily from the case of the inclusion model, we deal first with the situation where Y is discrete. In this case, Marjoram *et al.* (2003) demonstrate that, like a standard MCMC algorithm, the following algorithm generates a Markov chain with stationary distribution equal to the posterior distribution $f(\theta \mid y_0)$:

Algorithm LF (Likelihood Free MCMC)

LF1 Initialise θ_0; $i = 0$.

LF2 Propose θ^* according to a transition kernel $q(\theta_i \rightarrow \theta^*)$.

LF3 Generate $y^* \sim f(y \mid \theta^*)$.

LF4 With probability

$$\alpha = \min\left\{1, \frac{\pi(\theta^*)q(\theta^* \rightarrow \theta_i)}{\pi(\theta_i)q(\theta_i \rightarrow \theta^*)}\mathbf{1}(y^* = y_0)\right\}$$

set $\theta_{i+1} = \theta^*$; otherwise $\theta_{i+1} = \theta_i$.

LF5 Set $i = i + 1$ and go to LF2.

In step LF4, $\mathbf{1}(\cdot)$ is the indicator function. In effect, having made a proposal θ^* via a standard MCMC procedure, a value y^* is simulated from the model with parameter θ^*. If y^* differs from the true data value y_0, θ^* is immediately rejected (in favour of the current value θ_i). If $y^* = y_0$, then θ^* is accepted or rejected according to a probability that is the standard MCMC acceptance probability without the likelihood ratio term. Consequently, simulation replaces likelihood evaluation. The proof that $f(\theta \mid y_0)$ is the stationary distribution relies on reversibility arguments like those used for the standard Metropolis-Hastings algorithm.

Despite the guaranteed convergence of Algorithm LF, its speed of convergence may be very slow. Like the standard Metropolis-Hastings algorithm, this speed is affected by the choice of q. However, Algorithm LF is likely to have much slower mixing than the standard Metropolis-Hastings algorithm as $\alpha = 0$ unless $y^* = y_0$, so updates of θ are static unless a random simulation from the model $f(y \mid \theta^*)$ coincides exactly with the data y_0. Except in artificially simple cases, $f(y_0 \mid \theta^*)$ is likely to be very small – especially in problems that are high dimensional, highly structured or have many data components – leading to a small acceptance rate and a mixing of the chain that is therefore unacceptably slow. To address this difficulty Marjoram *et al.* (2003) propose two modifications to the basic algorithm. First, in step LF4, the term $\mathbf{1}(y^* = y_0)$ is substituted with $\mathbf{1}(S(y^*) = S(y_0))$, where $S(.)$ is a function that maps y to a vector of summary statistics. In other words, summary statistics of simulated data are required to match those of the original data. The gain in efficiency then derives from the fact that $\Pr(S(Y) = S(y_0) \mid \theta^*)$ could be very much greater than $\Pr(Y = y_0 \mid \theta^*)$. When $S(y)$ is exactly sufficient for θ in $f(y \mid \theta)$, the algorithm is still exact, in the sense of having stationary distribution $f(\theta \mid y_0)$. As an example, for both the spherical and ellipsoidal inclusion models, the elements of the vector of surface diameters $y = (s_1, \ldots, s_n)$ are exchangeable, so that $S(y) = (s_{(1)}, \ldots, s_{(n)})$, the vector of order statistics, is sufficient. However, slow mixing is still likely if n is large, and in this case, or for models of greater complexity, it is necessary to seek mappings $S(.)$ to lower dimensional spaces which, although not exactly sufficient, contain most of the data information on θ. In this case, the algorithm no longer provides an exact posterior inference, and its accuracy will depend on precisely how much information is lost in the mapping S.

The second modification suggested by Marjoram *et al.* (2003) is further replacement of the term $\mathbf{1}(S(y^*) = S(y_0))$ with $\mathbf{1}(\rho(S(y^*), S(y_0)) < \epsilon)$ for some metric ρ and $\epsilon > 0$. In other words, exact matching of summary statistics for a random draw from $f(\cdot \mid \theta^*)$ and original data y_0 is replaced with near matching as a pre-requisite for an update in the chain. This modification also permits the transition from discrete to continuous Y. The induced chain then converges to the stationary distribution $f(\theta \mid \rho(S(Y), S(y_0)) < \epsilon)$. Care is needed however, since this distribution may be quite different from the the target $f(\theta \mid y_0)$ when ϵ is not sufficiently small (Tanaka *et al.* , 2006, for example). On the other hand, imposing a value of ϵ that is too small may leave the acceptance rate of the algorithm unworkably low.

3.3 A Likelihood-Free MCMC Algorithm with State Space Augmentation

For inferences on the model $f(y \mid \theta)$ when the sample space of Y is continuous, as with the inclusion model, the basic algorithm LF is invalid, not least because $\Pr(Y = y_0 \mid \theta) = 0$. However, the modified version that accept points within an ϵ-neighbourhood of y_0 may provide a viable approximation, with the choice of ϵ again providing a compromise between accuracy and precision of the consequent inference.

To address this issue, we propose a modification to Algorithm LF based on an augmentation technique. The idea is analogous, though not equivalent, to the procedure of simulated tempering that is sometimes used to overcome mixing difficulties in standard MCMC algorithms (Geyer and Thompson, 1995; Stolovitzky and Berne, 2000; Brooks *et al.* , 2003). Typically, simulated tempering comprises simulation via standard stochastic methods from a model $f(\theta,\tau|y_0) \propto f(\theta|y_0)^{1/\tau}$, defined over an augmented state space $\Theta \times \mathcal{N}$, leading to a generated series (θ_i,τ_i). The target distribution is $f(\theta|y_0) = f(\theta|y_0,\tau=1)$. Subsetting the output to obtain $\{\theta_i : \tau_i = 1\}$ results in a series whose stationary distribution is the target distribution, but with potentially better mixing properties. This is because $f(\theta|y_0)^{1/\tau}$ is flat relative to $f(\theta|y_0)$ when τ is large, so that proposals made when τ is large have a greater acceptance probability. By analogy, in Algorithm LF our proposal is to augment the parameter space of $f(y \mid \theta)$ with ϵ, which is now treated as a model parameter. We then apply Algorithm LF to the enlarged space, updating both ϵ and the components of θ. This results in a Markov chain on the pairs $(\theta,\epsilon) \in \Theta \times \mathcal{R}^+$. In loose terms, values of θ that have been generated with small values of ϵ are reliable in the sense of having conditional distribution close to the target $f(\theta|y_0)$. Simulated θ's corresponding to large values of ϵ are less reliable, but the transition to such values enables a quality of mixing of the θ component that is unattainable with ϵ fixed at a small value.

In more detail, we assume that a suitable mapping $S(.)$ has been identified that exploits exact or near sufficiency of the model structure, together with a metric ρ in the space of $S(y)$. We also assume that a pseudo-prior for ϵ, $\pi(\epsilon)$ on $\mathcal{R}^+$, has been specified. Then, the new algorithm is

Algorithm LFA (Likelihood Free with Augmentation MCMC)

LFA1 Initialise (θ_0,ϵ_0); $i=0$.

LFA2 Propose (θ^*,ϵ^*) according to a transition kernel $q\left((\theta_i,\epsilon_i) \to (\theta^*,\epsilon^*)\right)$.

LFA3 Generate $y^* \sim f(y \mid \theta^*)$.

LFA4 With probability

$$\alpha = \min\left\{1, \frac{\pi(\theta^*)\,\pi(\epsilon^*)\,q\left((\theta^*,\epsilon^*) \to (\theta_i,\epsilon_i)\right)}{\pi(\theta_i)\,\pi(\epsilon_i)\,q\left((\theta_i,\epsilon_i) \to (\theta^*,\epsilon^*)\right)} \mathbf{1}(\rho(S(y^*),S(y_0)) < \epsilon^*)\right\}$$

set $(\theta_{i+1},\epsilon_{i+1}) = (\theta^*,\epsilon^*)$; otherwise $(\theta_{i+1},\epsilon_{i+1}) = (\theta_i,\epsilon_i)$.

LFA5 Set $i=i+1$ and go to LFA2.

In essence, this is Algorithm LF applied to the augmented (θ,ϵ) vector. It follows that Algorithm LFA produces a Markov chain on the state space $\Theta \times \mathcal{R}^+$ having stationary distribution

$$f(\theta,\epsilon \mid \rho(S(Y),S(y_0)) < \epsilon) \propto \pi(\theta)\pi(\epsilon)\text{Pr}(\rho(S(Y),S(y_0)) < \epsilon \mid \theta,\epsilon).$$

Recall that our real target is $f(\theta \mid y_0)$ and that this can be approximated by Algorithm LF to any degree of accuracy through choice of a sufficiently small ϵ. This suggests running Algorithm LFA with a pseudo-prior $\pi(\epsilon)$ (c.f. Geyer and Thompson, 1995) that favours small values. The occasional generation of large values of ϵ enables the problems of poor mixing that would be encountered with a small fixed ϵ to be avoided. The potential for bias induced by the simulation of large values of ϵ can be limited by filtering the series $\{(\theta_i, \epsilon_i)\}$ to obtain $\{\theta_i : \ \epsilon_i < \epsilon_T\}$ for some threshold value ϵ_T. The fact that the value of ϵ_T can be chosen retrospectively, in light of the generated chain, is an important feature of our approach, which we will discuss further in Section 3.4. The stationary distribution of the filtered series is proportional to

$$\int_0^{\epsilon_T} \pi(\theta)\pi(\epsilon)\Pr(\rho(S(Y), S(y_0)) < \epsilon \mid \theta, \epsilon)d\epsilon. \qquad (3)$$

For applications where Y is discrete, and the prior for ϵ puts mass on 0, the chain obtained with $\epsilon_T = 0$ has stationary distribution equal to $f(\theta \mid y_0)$. In this case, the algorithm is an exact analogue of simulated tempering. In the continuous case, expression (3) shows that $f(\theta \mid y_0)$ is approximated by a weighted average of $f(\theta \mid \rho(S(Y), S(y_0)) < \epsilon)$ over the range $0 < \epsilon < \epsilon_T$, with weights given by $\pi(\epsilon)$.

3.4 Application to Simulated Data

Since an exact MCMC inference is available, the spherical inclusion model provides us with a test-bed to judge the accuracy and viability of Algorithm LFA. We set parameters similar to those inferred from the data analysis of Anderson and Coles (2002): $v_0 = 5$, $\sigma = 1.5, \xi = -0.05, \lambda = 30$. A single realisation of this model led to $n_0 = 113$ inclusions intersecting the plane with a diameter greater than v_0. In both analyses we set the priors on σ and ξ to be proper but non-informative: $\log\sigma \sim \text{Ga}(0.001, 0.001)$, $\xi \sim \text{N}(0, 100^2)$. For Algorithm LFA we specified additionally $\log\lambda \sim \text{N}(0, 100^2)$, while the standard MCMC algorithm benefits from a reparameterization of λ to exploit conjugacy (see Anderson and Coles, 2002). For both algorithms our transition proposals for all parameters are based on simple componentwise random walks.

Algorithm LFA also requires additional specifications. First, it is necessary to specify a prior on ϵ. Recall that a compromise is necessary between precision, which is maximized by having a prior on ϵ that places all mass close to zero; and quality of mixing, which requires relatively large values of ϵ to have non-negligible probability. Our analyses here are based on $\epsilon \sim \text{Exp}(\tau)$ with $\tau = 1/10$. The choice of an exponential model is made to satisfy the requirement of favouring small values of ϵ, supporting the precision requirement, while also generating occasional large values to assist mixing. The choice of τ is arbitrary, but was selected after informal observation of several short runs with different candidate values, again with the balance of precision and mixing in mind. Further *post hoc* support for the choice is discussed in the context of Table 2 below.

As noted previously, we can also exploit sufficiency in the order statistics of the observed data by setting $S(y) = (s_{(1)}, \ldots, s_{(n)})$, the vector of order statistics. A suitable metric on this space is complicated by the fact that the data consists of both discrete and continuous components, the number of inclusions on the surface and their diameters respectively. We therefore considered a general class of metrics, allowing for vectors of different length, of the form

$$\rho(S(y), S(y_0)) = \sum_{i=1}^{n_0} (s_{(i)} - s_{0(i)})^2 + \kappa(n_y - n_0)^2, \tag{4}$$

where n_y is the length of y, $s_{0(i)}$ is the i-th smallest of the observed cross-sectional diameters, $s_{(i)}$ is the $100i/n_0$-th quantile of y and $\kappa > 0$ is a parameter to be specified. When $n_y = n_0$, ρ is simply the squared Euclidean distance, but otherwise it includes a second term that measures distance in cardinality. Recall that a proposal will be rejected in the algorithm whenever a simulated data realisation y^* and the true data are not close, or more precisely if $\rho(S(y^*), S(y_0)) > \epsilon$. It could be argued that if a simulated dataset does not provide exactly n_0 surface intersections, it is not much like the original data. In this case, we should set $\kappa \to \infty$ in (4). However, this degree of stringency is likely to impose unhelpfully slow mixing on the chain, particularly during burn-in. Hence, after some experimentation, we settled on $\kappa = 20$. With this value, after burn-in, we found 99.9% of accepted data realisations to have cardinality within one of n_0.

Both the standard and the new algorithms were run long enough to ensure convergence and a reasonable coverage of the posterior distribution. Specifically, following Anderson and Coles (2002), a chain of length 25000 was simulated with the standard MCMC, discarding the first 5000 iterations as burn-in. For the new algorithm, to add convergence assessment, five independent chains were generated, each of length 10 million. For each chain, the first 3 million iterations were discarded as burn-in, and the remaining process was thinned at every 100th iteration, leaving a total of 350000 iterations. Little is lost in this filtering, which helps to resolve storage issues raised by the length of the runs, due to the very strong dependence in the sample chains The necessity for such long simulation runs is the price to be paid in any algorithm that avoids likelihood evaluation by simulation. The additional requirement to further filter series to derive sequences $\{\theta_i : \epsilon_i < \epsilon_T\}$ makes additional demands on length of runs. Some gain in computational efficiency could be obtained by using cruder summary statistics S of the data, such as sample moments, but only at the additional cost of a less accurate inference.

Based on Algorithm LFA, Figure 2 illustrates marginal means of the approximate posterior distribution (3), together with plausible ranges based on means plus or minus two standard deviations, for a range of $\epsilon_T \in [0, 25]$. At each value of ϵ_T, the calculated values provide an approximation to the corresponding true posterior values. For small ϵ_T we anticipate high sampling variability, for large ϵ_T we anticipate bias. The figure appears to confirm this. The plots seem reasonably smooth down to $\epsilon_T \approx 3.3$, which we therefore believe provides the most reliable inference in this analysis. The variability for smaller values of ϵ_T suggests the corresponding estimates are unreliable. Also shown on Figure 2 are the corresponding inferences obtained from a standard MCMC analysis. At the value $\epsilon_T = 3.3$ the agreement between the two analyses is near perfect.

In Figure 3 we examine the marginal bias of the posteriors of σ, ξ and λ based on $\epsilon_T = 100, 50, 10$ and 3.3 respectively. Specifically, we show quantile-quantile plots of the standard MCMC output against the likelihood-free version. As expected, larger ϵ_T makes for larger biases, which are still evident for $\epsilon_T = 10$. The bias for $\epsilon_T = 3.3$ seems minimal, confirming that Algorithm LFA can provide accurate inference at the expense of a heavier computational burden than standard MCMC.

An extended simulation study suggests further that Algorithm LFA is superior to algorithm LF in this analysis. Accepting $\epsilon_T = 3.3$ as an optimal choice, samples with $\epsilon_T < 3.3$ can be generated as above with the LFA algorithm, or by application of the LF algorithm with any fixed value of $\epsilon >= 3.3$, followed by sub-selection of iterations satisfying $\epsilon < 3.3$. Given that in both the LF and LFA algorithms, sample chains are filtered to leave only those iterations with $\epsilon < 3.3$, the issue of mixing quality reduces essentially to that of acceptance rates. A comparison of relative rates is given in Table 2. The final column, in particular, gives the rate of iterations accepted with $\epsilon < 3.3$ in the LF algorithm for various choices of ϵ and in the LFA algorithm. The optimal choice for the LF algorithm is with $\epsilon = 3.3$, suggesting that adopting a higher value of ϵ and then filtering results in a reduced efficiency. Of course, in practice, it would not be clear a priori that $\epsilon = 3.3$ was the best choice for ϵ, so that running with a larger value would be unavoidable, leading to a sub-optimal acceptance rate. Notwithstanding this argument, an overall improvement in the rate of samples with $\epsilon < 3.3$ is obtained by direct application of the LFA algorithm, which requires only much weaker specification via a prior distribution on ϵ. This supports the view that the LFA algorithm can avoid mixing difficulties encountered by the LF algorithm. Note also that the strong preference for setting $\epsilon \leq 10$ in the LF algorithm provides additional support for the prior choice on ϵ made in the LFA algorithm.

4 Data Analysis

We conclude with an analysis of the steel inclusion diameters considered by Anderson and Coles (2002). The data comprise cross-sectional diameters of inclusions from the planar slice of a steel block. There are 112 such diameters recorded above a measurement threshold of 5μm, which is chosen as the model threshold v_0. We refer to the original paper for details of the standard MCMC analysis based on the spherical inclusion model. Our analysis here uses Algorithm LFA to fit the ellipsoidal model described in Section 2.2. The prior distributions and MCMC chain specifications are identical to those used in the simulation study of Section 3.4.

The results are summarized in Figure 4, which has a similar format to Figure 2. In this case there is an apparent smoothness in the Figures down to around $\epsilon_T = 3.3$. With this value, the corresponding posterior means are given in Table 3, which also includes the estimates for

ϵ	Pr(Accepted)	Pr($\rho < 3.3$|Accepted)	Efficiency Score
100	0.0282	< 0.0000	0.0007
50	0.0231	0.0029	0.0607
25	0.0101	0.0080	0.0808
10	0.0053	0.0509	0.2698
5	0.0015	0.1803	0.2705
3.3	0.0006	1.0000	0.6000
LFA	0.0147	0.0643	0.9452

TABLE 2. Acceptance rates in LF and (final row) LFA algorithm for different choices of ϵ. Final column is Pr(Acceptance) × Pr($\rho < 3.3$|Acceptance) × 1000

the spherical model fitted by standard MCMC. Also included in the table are approximate Monte Carlo standard errors, obtained by a naive batching of sample chains. Though crude, they amply demonstrate that differences between the inferences are model instrinsic rather than an artefact of the sample-based methodology.

A comparison of inferences on the characteristic size v_C as a function of C, made under the contrasting models, is shown in Figure 5. Results are similar to those obtained in the simulation study: for small C, v_C is underestimated by the spherical model relative to the ellipsoidal model, whereas for large C, the order is reversed. Taking sampling variability into account, the differences are not so large, and on this basis it might be argued that the spherical analysis shows some robustness to potential shape mis-specification.

By contrast, a clear distinction between the two models is apparent when quantities other than v_C are considered. For example, Figure 6 compares posterior means and credibility intervals of the inclusion rate for inclusions above a specified diameter. Even at extreme thresholds, when rates are low under both models, there is a substantial difference between the two inferences. There is an increasing body of literature that recognises that the impact of extreme inclusions on clean steels is much too complex to be summarized simply by the size of the largest inclusion, so that the presence of many inclusions of reasonably large size may be more important than the presence of an individual inclusion of exceptional size. Our analysis here - which is consistent with our simulation studies - suggest that measures of inclusion impact that are strongly dependent on the rate of extreme inclusions are likely to be strongly sensitive to assumptions made on inclusion shape.

5 Discussion

Our conclusions fall in two categories. First, with regard to general methodology, we have demonstrated that Algorithm LFA, which combines earlier ideas on likelihood-free MCMC and concepts from simulated tempering, is feasible and accurate, at least for problems of the scale we have considered here. The computational cost of likelihood-free algorithms is heavy however, and in many applications it may be acceptable to use simpler summary statistics to improve the acceptance rate of the algorithm at the cost of a reduction in accuracy.

Second, in terms of the stereological extreme analysis, we have seen via a simulation study that the mis-specification of inclusion shape family leads to biased estimates, especially of rate parameters. The importance of such mis-specification depends, to some extent, on the objective of the study. If inference is required specifically on the characteristic size, then the spherical inclusion model may provide accurate enough inference, especially if the diameter

Model	λ	σ	ξ
Spherical	30.7 (1.0)	1.47 (0.13)	-0.022 (0.048)
Ellipsoidal	95.7 (3.2)	1.90 (0.06)	-0.090 (0.014)

TABLE 3. Posterior means and associated Monte Carlo standard errors (in parentheses) for the spherical and ellipsoidal inclusion models fitted to steel inclusion data.

distributions turn out to be fairly heavy tailed. In contrast, for aspects where the rate of large inclusions is important, inferences are likely to depend much more critically on the accuracy of the specified shape family.

Though our main objective was to assess the robustness of the stereological analysis to the assumption of spherical inclusions, the fact that certain aspects of the inference do indeed seem to be sensitive to this choice raises questions about the viability of shape identification in the measurement process. This will obviously vary from application to application, but in the clean steel context that we have considered here it seems that the nature of typical inclusion shapes, and how they vary according to material, inclusion composition and production type, is reasonably well understood. For example, oxide inclusions are typically spherical, titanium inclusions are often cubical or of joined-pyramid form, while rolled steels often contain inclusions that are 'torpedo-shaped', and for which our ellipsoidal model may provide a reasonable approximation. Despite this level of understanding, it remains routine practice to report only a single measurement of surface inclusion size, which is interpreted as the radius of the planar circle induced by slicing a spherical inclusion. Minimally, our analysis has shown that for certain types of inference substantially better results might be obtained by using a shape family that better represents the believed form of inclusion shapes for the particular metal process under study.

A limitation of our analysis is that our investigations have been restricted to shape departures from the spherical inclusion model in the form of ellipsoidal inclusions. Consequently, our comments about robustness of the spherical model are limited to model mis-specification of this type. The advantage, however, of likelihood-free algorithms is that they can be applied to any model for inclusions that admits easy simulation. This would enable, for example, the analysis of non-Poisson models or models with broader families of shapes, for example the flexible parametric model for rotation invariant spatial particles described by Hobolth (2003). Furthermore, even when an assumed family for typical inclusions is accurate, the presence of occasional irregularly shaped inclusions is known to be commonplace. If a reasonable model could be specified for this irregular contamination process, the LFA algorithm should again provide a mecahnism for inference.

Acknowledgments

We are grateful for the financial support of the University of Padova and the University of New South Wales. SAS is supported by the Australian Research Council through the Discovery Project scheme (DP0664970). We are greatly indebted to Peter Green for introducing us to the area of likelihood-free MCMC inference and for sharing many ideas with us, and also to Clive Anderson for providing us with technical details of the inclusion shape and measurement process that we have included in the discussion. Finally, we are extremely grateful to two anonymous referees and an Editor who provided us with very insightful comments on an earlier version of the article.

Bibliography

Anderson, C. W. and Coles, S. G. (2002). The largest inclusions in a piece of steel. *Extremes*, **5**, 237 – 252.

Anderson, C. W., Shi, G., Atkinson, H. V., and Sellars, C. M. (2000). The precision of methods using the statistics of extremes for the estimation of the maximum size of inclusions in clean steels. *Acta Amter.*, **48**, 4235 – 4246.

Baddeley, A. and Jensen, E. B. V. (2004). *Stereology for Statisticians.* Chapman and Hall/CRC.

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025 – 2035.

Beretta, S. and Murakami, Y. (1998). Statistical analysis of defects for fatigue strength prediction and quality control of materials. *Fatigue Fract. Eng. Mater. Struct.*, **21**, 1049 – 1065.

Brooks, S. P., Friel, N., and King, R. (2003). Classical model selection via simulated annealing. *Journal of the Royal Statistical Society B*, **65**, 503–520.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* London: Springer.

Drees, H. and Reiss, R.-D. (1992). *Tail behavior in Wicksell's Corpuscle problem.* Dordrecht: Kluwer.

Geyer, C. J. and Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association*, **90**, 909 – 920.

Hlubinka, D. (2003). Stereology of extremes: shape factor of spheroids. *Extremes*, **6**, 5 – 24.

Hobolth, A. (2003). The spherical deformation model. *Biostatistics*, **4**, 583–595.

Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *PNAS*, **100**, 15324 – 15328.

Murakami, Y. (1994). Inclusion rating by statistics of extreme values and its application to fatigue strength prediction and quality control of materials. *J. Res. Natl. Inst. Stand. Technol.*, **99**, 345 – 351.

Plagnol, V. and Tavaré, S. (2003). Approximate Bayesian computation and MCMC. In H. Niederreiter (Ed.), *Proceedings of MCQMC2002*, Heidelberg. Springer Verlag.

Shi, G., Atkinson, H. V., Sellars, C. M., and Anderson, C. W. (1999a). Application of the Generalized Pareto distribution to the estimation of the size of the maximum inclusion in steels. *Acta Mater.*, **47**, 1455 – 1468.

Shi, G., Atkinson, H. V., Sellars, C. M., and Anderson, C. W. (1999b). Comparison of extreme value statistics methods for predicting maximum inclusion size in clean steels. *Ironmaking and Steelmaking*, **26**, 239 – 246.

Stolovitzky, G. and Berne, B. (2000). Catalytic tempering: A method for sampling rough energy landscapes by Monte Carlo. *Proceedings of the National Academy of Sciences*, **97**, 11164–11169.

Takahashi, R. and Sibuya, M. (1996). The maximum size of the planar sections of random spheres and its application to metallurgy. *Ann. Inst. Statist. Math.*, **48**, 127 – 144.

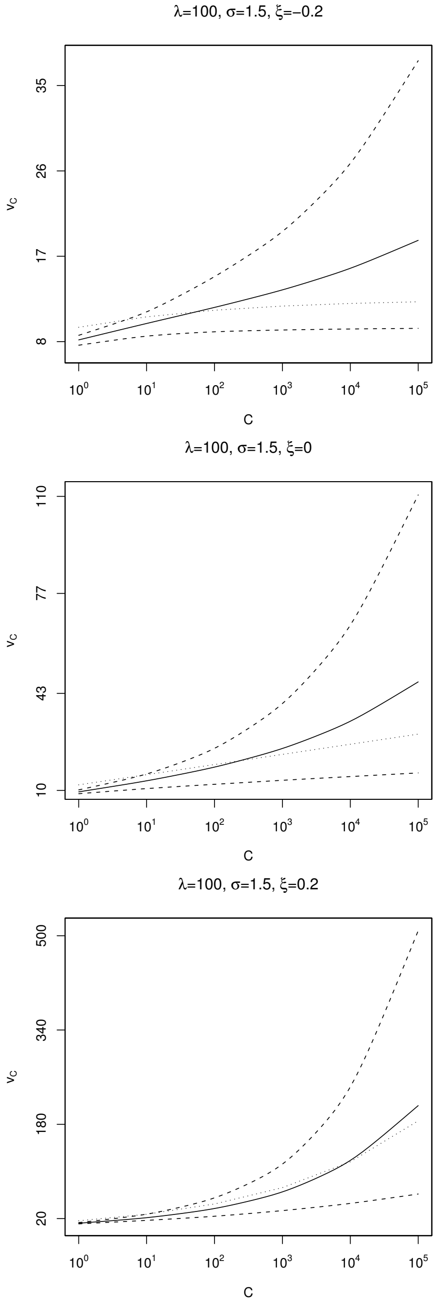
Takahashi, R. and Sibuya, M. (1998). Prediction of the maximum size in Wicksell's corpuscle problem. *Ann. Inst. Statist. Math.*, **50**, 361 – 377.

Takahashi, R. and Sibuya, M. (2002). Metal fatigue, Wicksell transform and extreme values. *Appl. Stochastic Models Bus. Ind.*, **18**, 301 – 312.

Tanaka, M. M., Francis, A. R., Luciani, F., and Sisson, S. A. (2006). Using approximate bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics (in press).*

Wicksell, S. D. (1925). The corpuscle problem: a mathematical study of a biometric problem. *Biometrika*, **17**, 84 – 99.

Wicksell, S. D. (1926). The corpuscle problem. Second memoir. Case of ellipsoidal corpuscles. *Biometrika*, **18**, 152–172.



λ=100, σ=1.5, ξ=−0.2

λ=100, σ=1.5, ξ=0

λ=100, σ=1.5, ξ=0.2

FIGURE 1. Characteristic size v_C as function of block size C. Dotted curve corresponds to correct values under ellipsoidal inclusion model. Solid curve corresponds to sample mean over repeated simulations of posterior mean based on spherical inclusion model. Outer dashed curves correspond to 2.5% and 97.5% quantiles of posterior means under repeated simulations and spherical model.

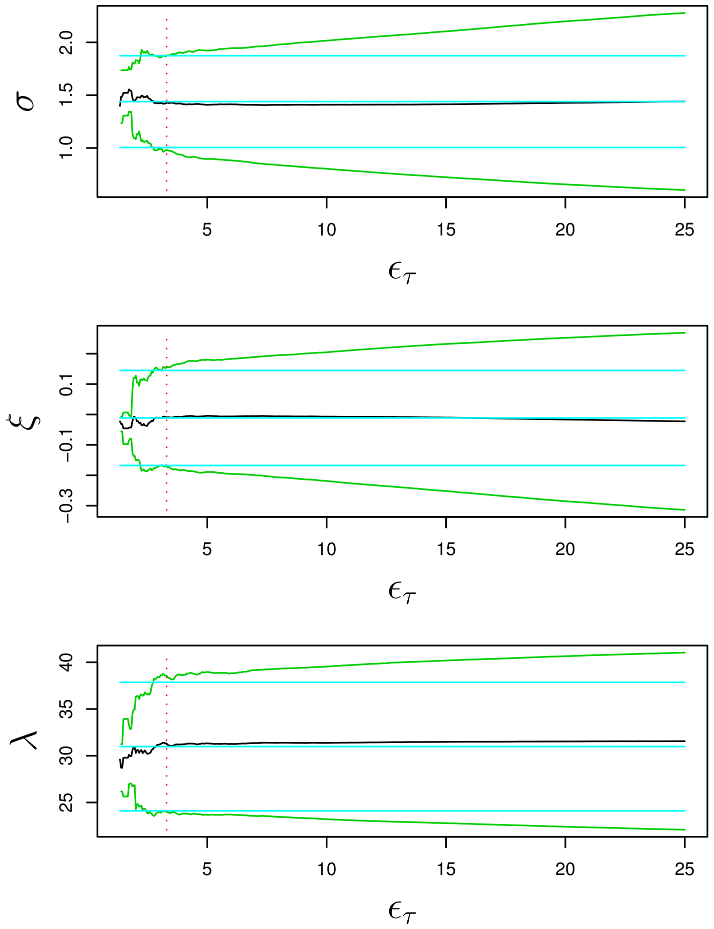


FIGURE 2. Posterior marginal means, and means plus or minus two standard deviations (solid lines), for each parameter conditional upon $\epsilon < \epsilon_T$. Horizontal lines correspond to mean and mean ± 2 standard deviations based on standard MCMC analysis. Vertical line corresponds to $\epsilon_T = 3.3$.

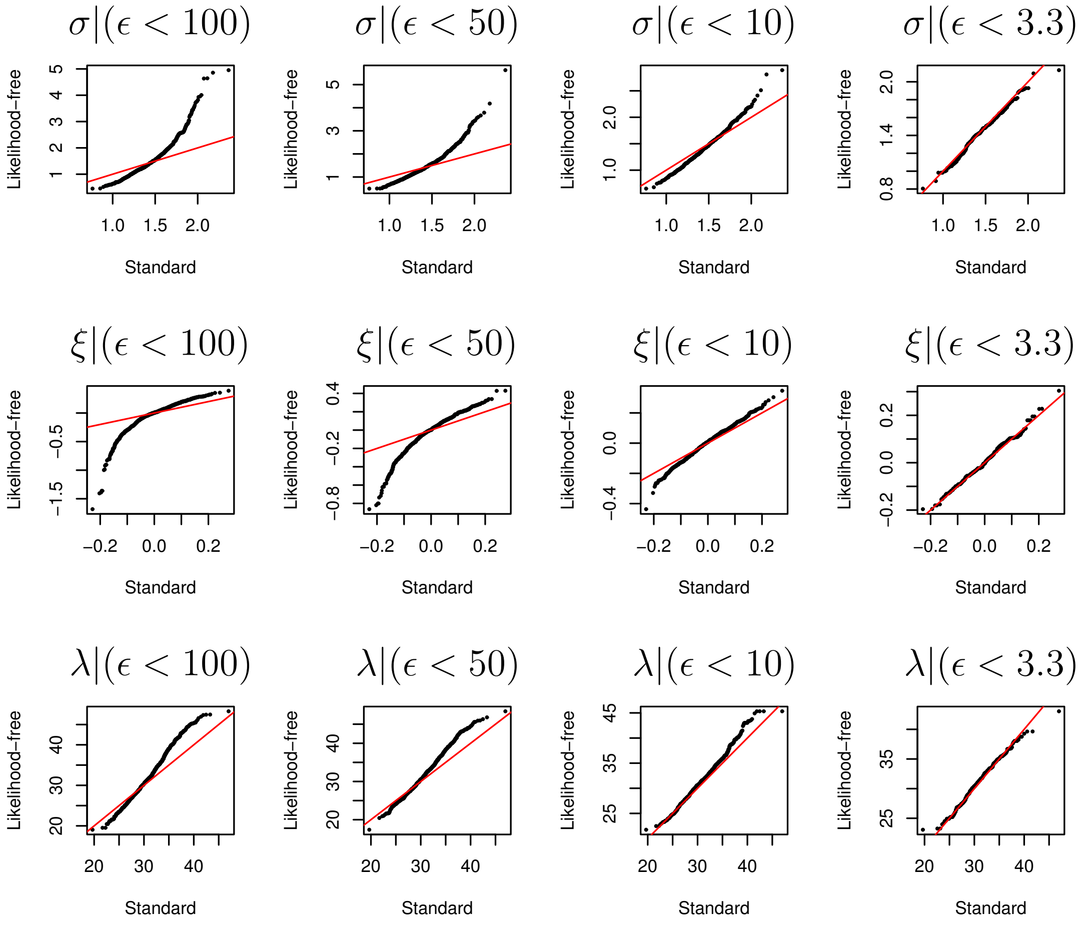


FIGURE 3. Posterior marginal quantile-quantile plots for each parameter against standard MCMC output, conditional upon $\epsilon < \epsilon_T = 100, 50, 10, 3.3$.

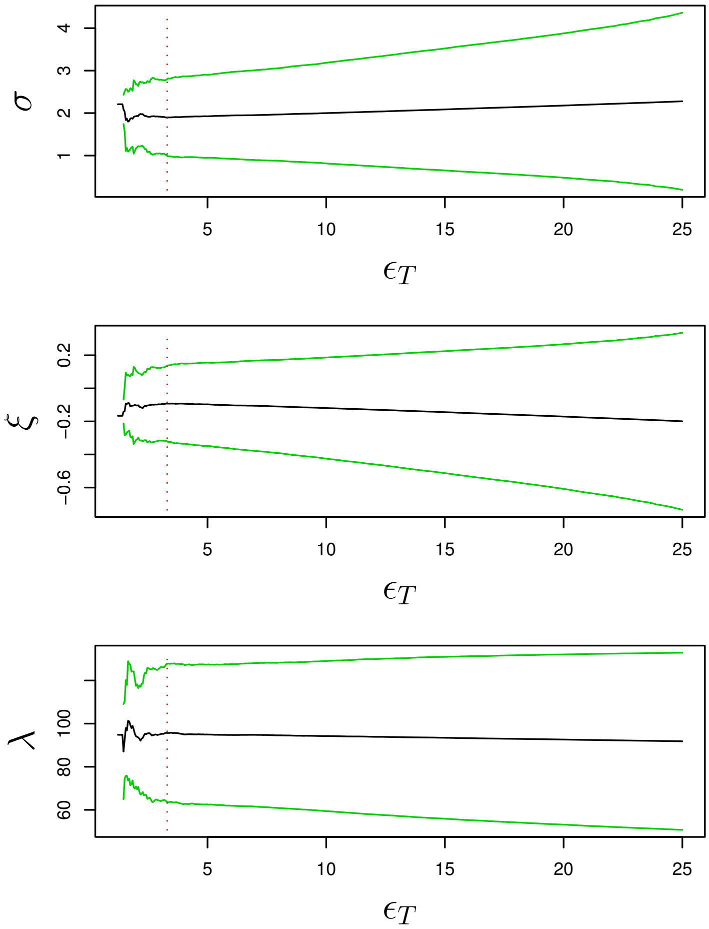


FIGURE 4. Posterior marginal means, two standard deviations (solid lines) for each parameter conditional upon $\epsilon < \epsilon_T$. Vertical line corresponds to $\epsilon_T = 3.3$.

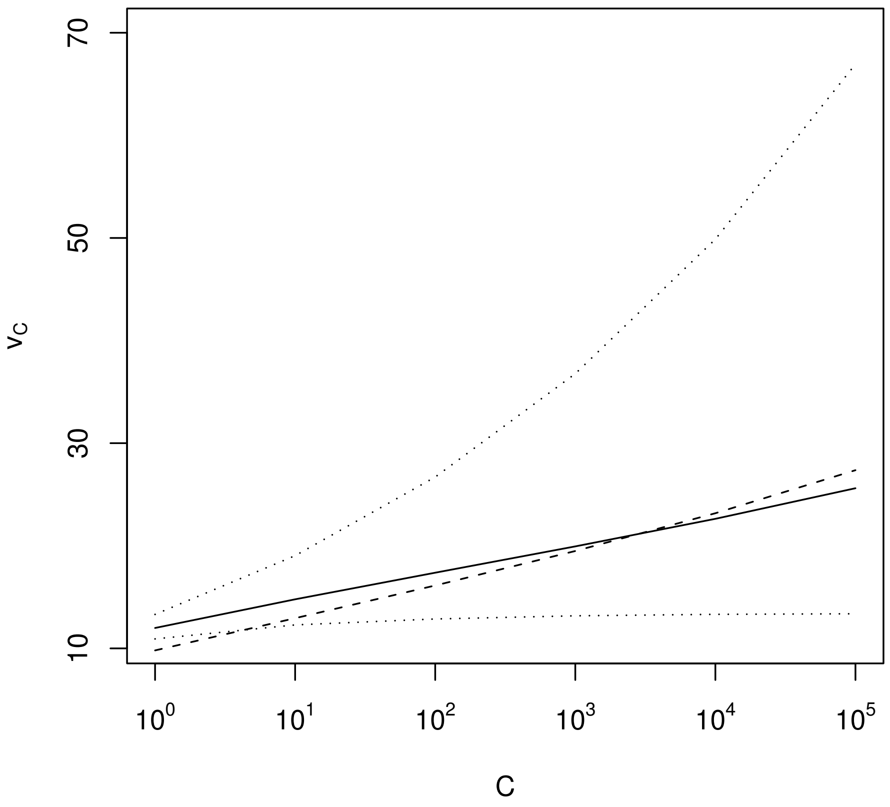


FIGURE 5. Characteristic size v_C as function of block size C in analysis of real data. Dashed curve corresponds to estimates under spherical model. Solid curve corresponds to posterior means under ellipsoidal model, with limits of 95% credibility intervals shown as dotted curves.

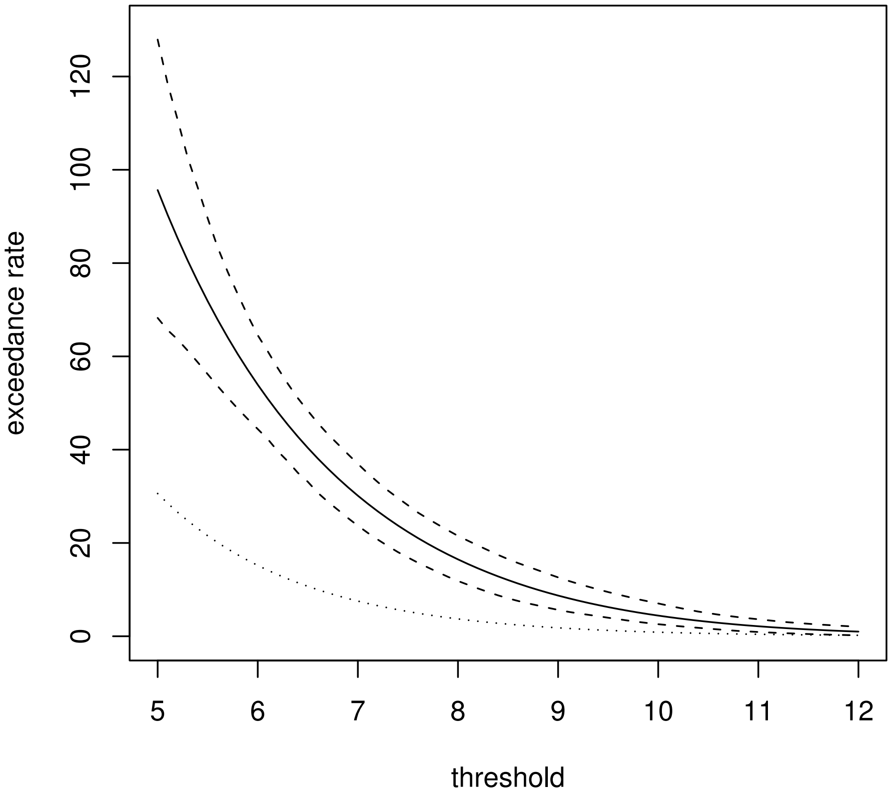


FIGURE 6. Solid curve shows posterior mean of rate of inclusions exceeding a threshold u as a function of u under the ellipsoidal model assumption. The broken curves are pointwise 95% credibility interval limits of the same quantity. The dotted curve shows the marginal mean of the same quantity under the spherical inclusion model