

Extreme Value Modelling with Complex Computations: an Application to Stereological Extremes

*Valori estremi e complessità computazionale: una applicazione agli
estremi stereologici¹*

Stuart Coles

Dipartimento di Statistica

Università di Padova

coles@stat.unipd.it

Riassunto: Nello studio di sistemi ambientali complessi, giocano un ruolo rilevante sia la statistica computazionale che la teoria dei valori estremi. In questo articolo verrà illustrata l'applicazione di un nuovo algoritmo MCMC *likelihood-free* ad un problema stereologico di valori estremi. Fare inferenza sulla dimensione di oggetti campionati in modo stereologico è un problema classico. In alcune applicazioni industriali e biologiche, l'obiettivo primario è la comprensione del comportamento estremo di tali oggetti. Ciò collega la stereologia classica alla teoria dei valori estremi. In questa presentazione verranno discussi due casi: il primo in cui gli oggetti possono essere assunti di forma sferica, il secondo in cui tale assunzione non può essere formulata. Nel primo caso, l'approccio stereologico standard combinato con la teoria dei valori estremi può portare alla formulazione di un modello gerarchico facilmente trattabile attraverso l'inferenza basata su MCMC. Nel secondo caso, dove non sono disponibili risultati stereologici standard, l'inferenza viene condotta mediante un nuovo algoritmo *likelihood-free*. Entrambe le versioni del problema sono illustrate utilizzando un'applicazione relativa alla produzione di acciaio, la cui purezza è compromessa dalla presenza di microscopiche impurità.

Keywords: Bayesian computation; Clean Steels; Extremes; Stereology; Markov chain Monte Carlo.

1. Introduction

Two important statistical themes under the broad umbrella of 'Complex Environmental Systems' are extreme value modelling and statistical computation. The importance of extreme value modelling to the study of almost any environmental system is self-evident: it is generally the extremes of an environmental process — be it sea level, pollution concentration or wind speed, for example — that cause greatest damage. Extreme value theory provides a class of models for the study of high levels of a stochastic process. The models are underpinned by an asymptotic development which is used as a proxy argument to support model extrapolation, a usual requirement when modelling extremes.

More generally, complex systems generally imply complex models, even with parsimony in mind, and the inevitable corollary of this is a heavy computational demand. In some situations this may just mean a long computational time, exploiting now standard techniques such as Markov chain Monte Carlo. In other situations, however, the model structure itself may require that some innovation is needed in order to carry out the computations.

¹This work was undertaken with the financial support of MIUR and the University of Padova

In part this article is review, setting out the basic themes in extreme value modelling in so much as they have been shown to be relevant for complex environmental systems. In greater part, the paper treats a problem of computation that arises in a specific application of extreme value theory. There's a caveat though: the application itself is not to an environmental process. As such, in this setting, the application is really a metaphor for the way standard computational algorithms might be modified and generalized to handle the difficulties that arise in complicated models. More specific applications of extreme value models to complex environmental systems for which the types of computational procedure described here may be of use are currently being considered, but are still some way from full development.

The paper is structured as follows. In Section 2 we set out some of the basics of extreme value theory. In Section 3 we set out the extreme value application of this paper, which is the estimation of the largest impurity in a steel block. The problem is stereological since the impurities are sampled by planar slicing of the block. Assuming the impurities are spherical, it turns out that standard stereological computations combined with standard extreme value models and a Markov chain Monte Carlo algorithm to carry out the computations can provide a plausible inference. However, for non-spherical impurities, the stereological calculations break down, and an innovative version of the Markov chain Monte Carlo algorithm that avoids likelihood evaluation is necessary. This is all set out in Section 4. The application to the steel inclusion problem is given in Section 5, and we conclude with some general remarks in Section 6.

2. Models for extremes

For environmental systems it is generally the extremes that do the most damage. Extreme value theory attempts to provide a class of models to describe such events. The logical and statistical difficulty is that probabilities are required for events that are so extreme they have occurred rarely, if at all. This prohibits the usual statistical cycle of data exploration, model development, inference, diagnostic, and back to model exploration. Extreme value theory substitutes this usual rationale for model development with asymptotic argument. The idea is to use classes of models that, in some sense, are obtained as limits as the level of the process grows. These models are then assumed to be reasonable approximations at finite levels and fitted to relevant data, usually the most extreme observations. A faith in the asymptotic argument underlying the models then enables an extrapolation to even higher levels.

The classical line in extreme value theory is as follows. The variables X_1, X_2, \dots are assumed independent with common unknown distribution function F . The variable $M_n = \max\{X_1, \dots, X_n\}$ is the n -block maximum, and an understanding of the extremes of the process of the X_i derives from a model for M_n . Suitably normalized, the limit family for M_n is the so-called generalized extreme value family, which can then be fitted to observed realisations of M_n for a suitably large choice of n . In environmental applications the notion of a univariate time-independent homogeneous process is too restrictive to be widely useful, but there are various extensions and generalizations. For example:

1. The limit family remains the same even if the X_i series is dependent, subject to some restrictions on long-range dependence;
2. When the X_i are vectors, there is an analogous theory, leading to the family of multivariate extreme value distributions;

3. When the X_i are stochastic fields, the analogy extends to a limit family of stochastic processes known as max-stable processes.

Put together, these families of models provide a comprehensive toolkit for modelling environmental extremes, whether of univariate, multivariate or spatial processes, though it should be admitted that high-dimensional or spatial modelling of extremes is still in its infancy, and the validity of asymptotic representations in these cases is yet to be really established.

Despite the various generalizations that are available, a problem with any block maximum approach to extremes is that for inference it tends to be wasteful of data. In any one block there may several extreme events, but it is only the largest that contributes to inference. There are alternative representations which avoid this difficulty, usually stemming from point process type arguments. These lead, in particular, to a threshold exceedance representation for extremes. Again assuming X_1, X_2, \dots to be a sequence of independent and identically distributed variables, it can be shown that under broad regularity conditions, the conditional exceedance distribution of a high threshold u can be approximated by a member of the generalized Pareto family of distributions given by

$$G(x) = \Pr(X \leq x \mid X > u) = 1 - \left\{ 1 + \frac{\xi(x - u)}{\sigma} \right\}_+^{-1/\xi}, \quad x > u, \quad (1)$$

where $\sigma > 0$, $\xi \in \mathcal{R}$ and $a_+ = \max(a, 0)$ (Davison and Smith, 1990). This characterisation enables inference to be based on all extreme observations, in the sense of exceeding the threshold u , rather than those that happen to be the largest in a block.

The threshold exceedance model or, more generally, the point process model from which the generalized Pareto model derives, has proved to be very effective in modelling extremes of environmental processes (Smith, 1989). For this representation too, there are equivalents in the multivariate and spatial settings, though again the utility in these cases is harder to verify. A complete development of the generalized Pareto model, and a description of its association with other models for extremes, including the point process and block maxima representations, can be found in Coles (2001).

3. An extreme value stereological problem

In the production of clean steels the introduction of impurities is an unavoidable side effect of the manufacturing process. The impurities, known as inclusions, are microscopically small, but their presence affects the strength of the material. In particular, the size of the largest inclusion is a determining factor in the strength of the steel block. Hence, for quality control, it is necessary to estimate the size of the largest inclusion. Sampling is usually carried out by planar slicing. The problem then is to infer the distribution of large inclusions within a steel block on the basis of the sample observed on the plane. Consequently, the statistical analysis is stereological.

Figure 1 shows a dataset with this structure. The data comprise the 112 planar diameters recorded above a measurement minimum of $5 \mu\text{m}$ in the planar slicing of an industrial block of clean steel. It should be remembered that the histogram so obtained does not provide an estimate of our target distribution, which remains the distribution of the unobserved three-dimensional diameters. Moreover, the distribution of both two- and

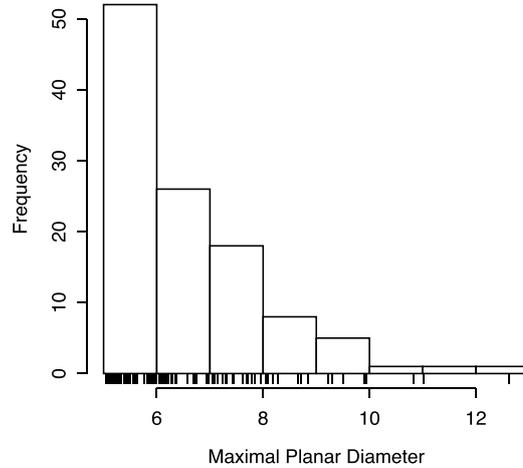


Figure 1: Histogram of 112 inclusion diameters, recorded in μm , exceeding a measurement threshold of $5 \mu\text{m}$, obtained by planar slicing of clean steel block.

three-dimensional diameters for a random inclusion will differ from that of a random inclusion in the sampled distribution. This is due to the sized-biased nature of the sampling: the planar slicing is more likely to intersect a large inclusion than a small one. Allowance for this size-biasing effect is an essential part of the inference of stereological data.

Procedures for the statistical analysis of stereological data have a long history, but the emphasis on extreme inclusion sizes introduces some novelty. This section summarizes the analyses in two recent papers, Anderson and Coles (2002) and Bortot *et al.* (2005). In each paper the basic model structure adopted is similar:

1. Inclusions fall within a regular shape family;
2. Inclusion centres follow a homogeneous Poisson process in the space corresponding to a volume of steel;
3. Inclusion (maximal) diameters are mutually independent and independent of inclusion location;
4. Inclusions have a random orientation within block;
5. The distribution function of inclusion diameters, V , conditional on exceeding a threshold v_0 , is given by

$$G(v) = \Pr(V \leq v \mid V > v_0) = 1 - \left\{ 1 + \frac{\xi(v - v_0)}{\sigma} \right\}_+^{-1/\xi}, \quad v > v_0, \quad (2)$$

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

Assumptions 2 to 4 are thought to be reasonable from a metallurgical point of view. Assumption 5 is natural from an extreme value perspective, as explained in Section 2. This leaves the question of inclusion shapes, and it is this aspect that distinguishes the two previous analyses. Anderson and Coles (2002) assume that inclusions are spherical, whereas Bortot *et al.* (2005) allow inclusions to be ellipsoidal in shape. In detail, both possibilities

are bound to be wrong, but it is likely that the inclusions are not too different from spherical. The analysis of the ellipsoidal case therefore provides some measure of sensitivity to shape assumption.

Denoting by S the (three-dimensional) maximal diameter of an inclusion, the objective is to make inference on the tail of the distribution of S given a set of (maximal) cross-sectional diameters of inclusions, V_1, \dots, V_n say, obtained by planar slicing. The analysis is helped in the spherical case by a classical result due to Wicksell (1925) that links the distributions of S and V under the specified sampling scheme. Modified slightly to account for the thresholding aspect, this states:

$$\Pr(S \leq s \mid S > v_0) = 1 - \frac{\int_s^\infty (v^2 - s^2)^{1/2} g(v)}{\int_{v_0}^\infty (v^2 - v_0^2)^{1/2} g(v)}, \quad s \geq v_0, \quad (3)$$

where $g(\cdot)$ is the generalised Pareto density function associated with (2). In the present context, there are two thresholds to take into account: first, the measurement threshold, and second, the threshold that arises from any application of the generalized Pareto model for tail approximation. We will assume that the measurement threshold of $v_0 = 5\mu\text{m}$ actually serves both purposes. For further detail on this point, see Anderson and Coles (2002), who show that under this assumption formula 3 can be exploited in a Markov chain Monte Carlo Metropolis Hastings algorithm in which the S_i corresponding to the V_i are treated as unobserved latent variables. In combination with conjugate choices for prior distributions, where available, and a simple reparameterization for the rate parameter λ , this led to a well-behaved and efficient stochastic method of computation on the model.

The more interesting, but more challenging, situation is when the inclusion shapes are not spherical, in which case, equation (3) is invalid and has no obvious analogue. Without any assumption about inclusion shapes it is impossible to make progress: the stochastic relationship between observed planar diameters and inclusion solid size must be expressed, at least implicitly, otherwise the stereological information is worthless. The next simplest class of solid shapes after spheres is the class of ellipsoids, resulting in elliptical planar intersections. This specification alone fails to define the shape distribution, but adding an assumption that minor axes of the ellipsoids have lengths that are distributed as independent uniform $U[0, 1]$ multiples of the principal axis does complete a valid definition. Although the ellipsoidal assumption is no more plausible than the spherical one, it at least provides a model by which the sensitivity of inferences to the choice of shape distribution can be assessed.

The difficulty arises in computation: the planar slicing sampling mechanism means that the model likelihood is not easily obtained. In the spherical case, formula (3) generates a conditional likelihood based on unobserved latent variables, whose distribution can be integrated out via MCMC, but there is no analogue in the ellipsoidal case. Hence, even generic computational methods like MCMC, which require likelihood evaluation, are ruled out. We turn, therefore, to consider stochastic algorithms that avoid explicit likelihood calculation. There has been some interest in such algorithms lately, primarily to solve inference problems in genetics models. Like the steel inclusion model, a common characteristic is that the parameter space is small, but the sample space is large and highly structured with an intractable likelihood.

4. Likelihood-free MCMC computation

Consider the general setting where $f(y | \theta)$ is the probability (density) function of a random (vector) variable Y parametrized by $\theta \in \Theta$, y_0 is the observed value of Y and the prior distribution on θ is $\pi(\theta)$. Assume further that explicit calculation of the likelihood $f(y | \theta)$ is either not feasible or is unduly expensive, whereas the simulation of realisations of Y from the model is cheap and straightforward. In this case Marjoram *et al.* (2003) propose the following algorithm:

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 $y^* \sim f(y | \theta^*)$.

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.

The sequence θ_i obtained this way is a Markov chain. When Y is discrete, it is easy to show that the stationary distribution of the process is the target distribution $f(\theta | y_0)$, so the output can be treated in the usual way with MCMC algorithms to make inference on θ . The trick in this particular algorithm is the replacement of likelihood evaluation in the acceptance probability calculation of standard MCMC algorithms with model-based simulation. However, both the speed of convergence and the mixing properties of the generated chain may be poor, since $\alpha = 0$ unless $y^* = y_0$. This means that updates of θ are static unless a random simulation from the model $f(y | \theta^*)$ coincides exactly with the data y_0 . Except in artificially simple cases, $f(y_0 | \theta^*)$ 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 $S(\cdot)$ 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 $\Pr(S(Y) = S(y_0) | \theta^*)$ could be very much greater than $\Pr(Y = y_0 | \theta^*)$. When $S(y)$ is exactly sufficient for θ in $f(y | \theta)$, the algorithm is still exact, in the sense of having stationary distribution $f(\theta | y_0)$. As an example, for both the spherical and ellipsoidal inclusion models, the elements of the vector of surface diameters $y = (s_1, \dots, s_n)$ are exchangeable, so that $S(y) = (s_{(1)}, \dots, s_{(n)})$, 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 $S(\cdot)$ 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 is the 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 $f(\cdot | \theta^*)$ and original data y_0 is replaced with near matching as a pre-requisite for an update in the chain. The induced chain then converges to the stationary distribution $f(\theta | \rho(S(Y), S(y_0)) < \epsilon)$. Care is needed however, since this distribution may be quite different from the target $f(\theta | y_0)$ when ϵ is not sufficiently small. On the other hand, imposing a value of ϵ that is too small may leave the acceptance rate of the algorithm unworkably low.

For inferences on the model $f(y | \theta)$ when the sample space of Y is continuous, as with the inclusion model, the basic algorithm LF is invalid, not least because $\Pr(Y = y_0 | \theta) = 0$. However, the modified version that accepts points within an ϵ -neighbourhood of y_0 may provide a viable approximation, though the choice of ϵ is likely to be even more critical. To try to overcome the tension in the choice of ϵ that arises from the trade-off between accuracy and mixing rate, Bortot *et al.* (2005) suggested an alternative algorithm in which ϵ itself is treated as a model parameter. More specifically, the parameter space of $f(y | \theta)$ is augmented with ϵ , which is now treated as a model parameter. Then, Algorithm LF is applied to the enlarged space, updating both ϵ and the components of θ . This results in a Markov chain on the pairs $(\theta, \epsilon) \in \Theta \times \mathcal{R}^+$. Loosely, values of θ that are 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.

Like algorithm LF, the modified procedure assumes that a suitable mapping $S(\cdot)$ has been identified that exploits exact or near sufficiency of the model structure, together with a metric ρ in the space of $S(y)$. Additionally, it is necessary to assume that a pseudo-prior for ϵ , $\pi(\epsilon)$ on \mathcal{R}^+ , has been specified. This leads to:

Algorithm LFA (Likelihood-free with Augmentation MCMC)

LFA1 Initialise (θ_0, ϵ_0) ; $i = 0$.

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

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

LFA4 With probability

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

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

$$f(\theta, \epsilon | \rho(S(Y), S(y_0)) < \epsilon) \propto \pi(\theta) \pi(\epsilon) \Pr(\rho(S(Y), S(y_0)) < \epsilon | \theta, \epsilon). \quad (4)$$

Recall that realizations from the target distribution $f(\theta | y_0)$ can be obtained to any degree of accuracy by running Algorithm LF with sufficiently small ϵ . This suggests running Algorithm LFA with a pseudo-prior $\pi(\epsilon)$ that favours small values. However, in contrast to algorithm LF, 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 stationary distribution of the filtered series is proportional to

$$\int_0^{\epsilon_T} \pi(\theta)\pi(\epsilon)\Pr(\rho(S(Y), S(y_0)) < \epsilon \mid \theta, \epsilon)d\epsilon. \quad (5)$$

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 analogue of the stochastic optimization procedure known as simulated tempering. In the continuous case, expression (5) 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 $\pi(\epsilon)$.

5. Analysis of the steel inclusion data

The stereological model for inclusions in clean steels can now be analysed under either the spherical or ellipsoidal models for inclusion shapes. Using the latent variable formulation described in Section 2, standard MCMC algorithms can be applied in the spherical case, though a one-dimensional integration is required at each iteration of the algorithm; see Anderson and Coles (2002) for details. In contrast, the LFA algorithm can be applied to the ellipsoidal inclusion model, resolving the difficulty of an intractable likelihood. Note that, although the simulation under the Poisson-ellipsoidal model is easy, it is still necessary to calculate the planar intersection diameter for each simulated ellipsoid, but this requires only elementary geometry calculations.

| Model | λ | σ | ξ |
|-------------|-------------|-------------|----------------|
| Spherical | 30.7 (3.54) | 1.47 (0.24) | -0.022 (0.089) |
| Ellipsoidal | 95.7 (16.0) | 1.90 (0.45) | -0.090 (0.110) |

Table 1: Posterior means and standard deviations (in parentheses) of spherical and ellipsoidal inclusion models fitted to steel inclusion data.

For illustration we consider the analysis of the 112 planar diameters shown in Figure 1. For each analysis we assume vague but proper priors on each of the parameters. In the MCMC specification for the spherical case, simple random walk update steps were used for each of the parameters; see Anderson and Coles (2002) for complete details. In the LFA algorithm analysis of the ellipsoidal case, various diagnostics were used to identify the most appropriate filtering threshold ϵ_T ; see Bortot *et al.* (2005). A comparison of the two analyses, each made with weak vague prior specifications, is given in Table 1. Clearly, the biggest difference in the two models relates to the estimated Poisson rate of inclusions. This makes sense: an ellipsoidal inclusion with the same solid diameter as a spherical inclusion is likely to generate smaller planar intersections. Hence, with an identical number of planar observations, the rate parameter will be greater under the ellipsoidal model than under the spherical model. The effects of the generalized Pareto diameter tail distribution parameters are smaller, but also evident.

To assess more easily the difference between the two models it is helpful to look at some quantity of interest that depends on all the various model parameters. For example, a

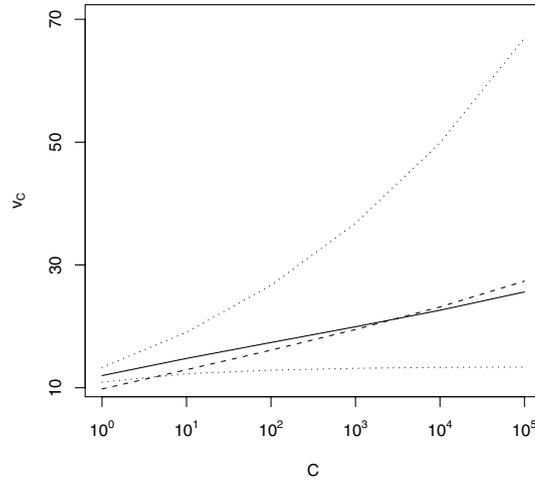


Figure 2: Characteristic size v_C as function of block size C in analysis of real data. Dashed curve corresponds to estimates under spherical model. Solid curve corresponds to posterior means under ellipsoidal model, with limits of 95% credibility intervals shown as dotted curves.

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 made,

$$v_C = v_0 - \frac{\sigma}{\xi} \{1 - (\lambda C)^\xi\}. \quad (6)$$

A comparison of inferences on the characteristic size v_C as a function of C , made under the contrasting models, is shown in Figure 2. 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 reasonably argued that the spherical analysis shows some robustness to potential shape mis-specification. However, in other applications, particularly if the distribution of the diameters is shorter tailed than for these data, the sensitivity of inferences to the form of shape inclusions may be stronger.

6. Discussion

Whilst not itself an application to complex environmental systems, the example presented here illustrates both extreme value techniques and computational techniques to infer difficult statistical models, each of which has undoubted relevance to environmental modelling. The extreme value material used is by now well known, but the likelihood-free computational methods presented are less well known. The option to replace likelihood evaluation with model simulation is attractive, though computation time can increase dramatically. The algorithm LFA presented here offers some improvements over the original LF algorithm, but the search for more efficient algorithms is likely to continue.

In the specific area of extreme value models for environmental processes, one possible application of the LFA algorithm is in the inference of max-stable processes which were

alluded to in Section 2. These are the infinite-dimensional analogue of extreme value distributions, and have been proposed for modelling the extremes of a process spatially: for example, to describe the annual maximum rainfall process over a region (Coles, 1993). For this class of models process simulation is an option, whereas likelihood evaluation is not, suggesting that the LFA algorithm might be a natural candidate. The difficulty is that the data themselves are high-dimensional, being a discrete realization of a spatial process, and it is not clear whether a suitable pseudo summary statistic $S(\cdot)$ can be found that is sufficiently low in dimensionality to enable the algorithm to proceed, but which is not so crude a representation of the original data that it blurs the resolution of the inference so much as to make it worthless. In time-honoured tradition: an area for future research.

Acknowledgements

This work draws heavily on previous collaborations with Clive Anderson, Paola Bortot and Scott Sisson.

References

- Anderson C.W. and Coles S.G. (2002) The largest inclusions in a piece of steel, *Extremes*, 5, 237–252.
- Bortot P., Coles S.G. and Sisson S. (2005) Inference for stereological extremes, *JASA* (submitted).
- Coles S.G. (1993) Regional modelling of storms via max-stable processes, *Journal of the Royal Statistical Society, series B*, 54, 797–816.
- Coles S.G. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer.
- Davison A.C. and Smith R.L. (1990) Models for exceedances over high thresholds (with discussion), *Journal of the Royal Statistical Society, series B*, 52, 393–442.
- Marjoram P., Molitor J., Plagnol V. and Tavaré S. (2003) Markov chain Monte Carlo without likelihoods, *PNAS*, 100, 15324–15328.
- Smith R.L. (1989) Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone (with discussion), *Statistical Science*, 4, 367–393.
- Wicksell S.D. (1925) The corpuscle problem: a mathematical study of a biometric problem, *Biometrika*, 17, 84–99.