

EXTREME VALUE METHODS IN BODY BURDEN ANALYSIS: WITH APPLICATION
TO INFERENCE FROM LONG-TERM DATA SETS

by

MATTHEW J. ATKINSON

(Under the direction of Machelles Wilson)

ABSTRACT

The Generalized Extreme Value (GEV) model's relevance to the extremes of a distribution, and the Generalized Pareto (GP) model's relevance to the exceedences above a threshold in a distribution are equivalent to the Gaussian model's relevance to the center of a distribution. Limit theorems are presented which unify the extreme values of samples from sufficiently smooth distributions under the GEV model, and similarly unify exceedences under the GP model. These models are fit (via maximum likelihood estimation of model parameters) to radiocesium body-burden data in a population of deer at increased risk of exposure. Analysis suggests that a member of the Frechet EV family best quantifies maxima from this dataset. Return levels are estimated, and a formula for estimating tolerance limits is developed from the GEV functional form.

INDEX WORDS: extreme value, return period, tolerance limit, radiocesium,
non-human biota

EXTREME VALUE METHODS IN BODY BURDEN ANALYSIS: WITH APPLICATION
TO INFERENCE FROM LONG-TERM DATA SETS

by

MATTHEW J. ATKINSON

Bachelor of Mathematics, The University of Waterloo, 1999

A Thesis Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2004

© 2004

Matthew J. Atkinson

All Rights Reserved

EXTREME VALUE METHODS IN BODY BURDEN ANALYSIS: WITH APPLICATION
TO INFERENCE FROM LONG-TERM DATA SETS

by

MATTHEW J. ATKINSON

Approved:

Major Professor: Machelles Wilson

Committee: William P. McCormick
Lynne Seymour

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2004

ACKNOWLEDGMENTS

More so than is usually the case, this work would not have been possible without the patient support of my committee. I thank you each for sticking with me. Furthermore, this Research Cowboy would be nothing without his Rockin' Researcher: thank you for everything.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iv
LIST OF FIGURES	vi
 CHAPTER	
1 AN INTRODUCTION TO EXTREME VALUE THEORY	1
1.1 EXTREME VALUE THEORY TAKES FAMILIAR CONCEPT TO NEW LIMITS!	1
1.2 THE DEFINITIVE MODELS OF EXTREME BEHAVIOR	3
1.3 DEEPER THEORETICAL FOUNDATIONS OF EV MODELS	10
1.4 PARAMETER ESTIMATION IN THE EV DISTRIBUTIONS	14
1.5 RETURN LEVEL PLOTS AND DIAGNOSTICS FOR GEVD MODELS	16
1.6 TOLERANCE LIMIT ESTIMATION	19
1.7 FINDING THE DOMAIN OF ATTRACTION FOR NEARLY EVERY EXTREME	22
2 GETTING TO KNOW THE SRS DEER DATA	25
2.1 PRELIMINARY ANALYSIS LEADS TO INVESTIGATION OF EXTREME DEER	25
2.2 IMPLICATIONS OF TIME TAKEN SERIOUSLY, BUT RULED OUT	26
2.3 BLOCKING SCHEMES CAREFULLY CONSIDERED	29
3 RESULTS AND CONCLUSION	33

3.1	MAXIMUM RADIOCESIUM BODY-BURDENS IN THE SREL DEER DATA	33
3.2	TOLERANCE LIMITS FOR THE SREL DEER DATA	37
4	FUTURE CONSIDERATIONS	43
4.1	NEW THEORY ON THE BLOCK	43
4.2	SREL'S GOOD DATA	44
5	REFERENCE LIST	45

LIST OF FIGURES

2.1	Histogram of $\ln(\text{pCi/kg})$	31
2.2	QQ-plot of $\ln(\text{pCi/kg})$	31
2.3	Extreme value summary plot of raw body-burden dataset.	32
2.4	Extreme value summary plot of body-burden dataset arranged in contiguous blocks of 984 observations.	32
3.1	Comparison of GEVD densities for different block sizes in the SREL body-burden data. For $n = 984$ (left), parameter estimates are $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (36.74, 16.676, 0.0322)$. For $n = 1000$ (right), estimates are $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (35.90, 16.287, 0.0669)$	38
3.2	Profile likelihood for ξ in the $n = 1000$ Frechet model.	38
3.3	Diagnostic plots for GEVD fit to SREL body-burden data ($n = 1000$).	39
3.4	Diagnostic plots for Gumbel fit to SREL body-burden data ($n = 1000$).	39
3.5	Profile likelihood for 15-block return level: $\hat{z}_{p=1/15} = 83.588$ in the $n = 1000$ Frechet model.	40
3.6	Profile likelihood for 30-block return level: $\hat{z}_{p=1/30} = 97.764$ in the $n = 1000$ Frechet model.	40
3.7	Profile likelihood for 45-block return level: $\hat{z}_{p=1/45} = 106.280$ in the $n = 1000$ Frechet model.	41
3.8	Profile likelihood for 60-block return level: $\hat{z}_{p=1/60} = 112.440$ in the $n = 1000$ Frechet model.	41
3.9	Delta-method confidence intervals for $\gamma = 90\%$ (indicated by arrows) and $\gamma = 99.5\%$ β -content tolerance limits.	42

CHAPTER 1

AN INTRODUCTION TO EXTREME VALUE THEORY

1.1 EXTREME VALUE THEORY TAKES FAMILIAR CONCEPT TO NEW LIMITS!

The Extreme Value (EV) model's relevance to the maxima or minima of a distribution, and the Generalized Pareto (GP) model's relevance to the exceedences above a threshold in a distribution are similar to the Gaussian (Normal) model's relevance to the center of a distribution. That is, limit theorems exist which unify the maximum values of samples from sufficiently smooth distributions under the EV models, and unify exceedences under the GP model. These parametric models prove rich enough to characterize the extremes of a very wide range of stochastic distributions.

Many traditional statistical analyses are concerned with the central tendencies of a data sample collected from some physical process, and rely on a bevy of Central Limit Theorems to characterize the convergence of a distribution's center (upon multiple sampling) to the Gaussian distribution. In such analyses, observed values that deviate too far from the sample mean are often considered outliers, and efforts are made to justify the removal of these data points from the analysis. In an extreme value analysis, attention is instead focused on those values occurring far above or below the average (i.e., those events that constitute the tail of the underlying, or parent distribution). While the study of extreme events dates back to 1709, when Nicolas Bernoulli considered the mean longest distance from the origin when points are scattered at random on a line of fixed length, limit theorems attributed to Fisher

and Tippet (1928), and Balkema and de Haan (1974) are essential in characterizing the behavior of a distribution's tail. These limit theorems provide a framework to support statistical techniques and models for the estimation and prediction of unusual, or rare events in an observed physical process, based on historical data.

Inherent in the definition of extreme value is the concept of scarcity: estimates of processes at levels far above those typically observed are often of the most interest. For example, this paper examines extreme radiological body burden measurements in a population of deer that is at an increased risk of exposure to contamination. Comparing the frequency and intensity of maximum exposures to regulatory limits, *which exceed every observation in the dataset*, is of particular interest. Typically, one is cautioned against extrapolating in the manner described - from observed to unobserved (or scarcely populated) levels - in any statistical estimation procedure. With that in mind, one must be cautious when interpreting EV model predictions, and pay particular attention to accurately reflecting the inherent uncertainty of extrapolated predictions. That being said, there are no competitive models or theories for this increasingly important field of study - first applied in engineering fields to assess stress loads on building materials and structures in the 1940's, and more recently being applied in arenas as diverse as actuarial science, genetics and quantum mechanics.

Extreme value analysis is frequently of particular interest in the biological sciences. Specific ecological applications include modeling extreme natural events like floods, windstorms, earthquakes and drought conditions; the study of species innovation, disease outbreak or any other rare event of importance; and the analysis of high concentrations of harmful substances in biological systems. In any context where the frequency and intensity of events from the tail of a distribution are of interest, extreme value analysis comprises the necessary statistical tools.

1.2 THE DEFINITIVE MODELS OF EXTREME BEHAVIOR

To formalize the concept of extreme value analysis, we seek models to describe the statistical behavior of the sample maxima:

$$M_{n:n} = \max\{X_1 \dots X_n\}$$

where the X_i 's, $i = 1 \dots n$, form a sequence of independent random variables from a common distribution function F , which describes data obtained from some physical process. The development of models to characterize the minimum values of such a process is similar to that leading to the models for maximum values. For explicit development of the parametric models for minima, see Coles (2001), Reiss & Thomas (1997), or just about any other introductory text on extreme value analysis.

Alternatively, any analysis of minima can be transformed to an analysis of maxima. If one is concerned with the smallest values of a sequence of negative variables:

$$M_{n:1} = \min\{X_1, \dots, X_n : X_i \leq 0\}$$

then also:

$$M_{n:1} = -\max\{-X_1, \dots, -X_n : X_i \leq 0\}$$

so that by simply negating the original data, the problem is reduced to an analysis of maximum values. If instead one is concerned with the smallest values of a sequence $\{X_1, \dots, X_n : -\infty < X_i < +\infty\}$, first shifting the data so that all values are less than zero then negating the shifted data, once again reduces the problem to a maximum value analysis. Regardless of the transformation required, its inverse is applied to model predictions to obtain results in the context of the problem which originally motivated the minimum value analysis. Since this paper concerns the estimation of maximal values, little further will be said about the analysis of minimum values. Henceforth, the terms extreme and maximal are used interchangeably.

Of course, the theory of order statistics permits the derivation of the exact distribution of $M_{n:n}$ for any value of n , assuming knowledge of the underlying distribution function F :

$$\begin{aligned} P\{M_{n:n} \leq x\} &= P\{X_i \leq x \quad \forall \quad i = 1, \dots, n\} \\ &= \prod_{i=1}^n P\{X_i \leq x\} \\ &= [F(x)]^n, \quad X_i \sim F, \text{ i.i.d.} \end{aligned} \tag{1.1}$$

As a means of data analysis this line of pursuit is fraught with complications, foremost of which is the fact that in most any practical application, the underlying (or parent) distribution function F is frightfully complicated if it is known at all. Although standard techniques can be used to derive an estimate $\hat{F}$ of F , and hence of $[F(x)]^n$, the multiplicative form magnifies the tiniest error in the estimation of F , leading to unsatisfactory discrepancies for even moderate values of n . Furthermore, for any value of x less than x^+ , the upper endpoint of F , (the smallest value in the domain of F such that $P\{X \leq x\} = 1$ is denoted $x^+ = \inf\{x : F(x) = 1\}$), the theoretically exact distribution of $M_{n:n}$ degenerates to a point-mass. That is, for any fixed value of x , it is clear that in the limit of $[F(x)]^n$ as $n \rightarrow \infty$ one obtains a degenerate distribution:

$$\lim_{n \rightarrow \infty} [F(x)]^n = \begin{cases} 1, & F(x) = 1 \\ 0, & F(x) < 1 \end{cases}$$

The results of Fisher and Tippet (1928) overcome these problems associated with the limiting distribution of the variable $M_{n:n}$. They demonstrated that if there exist normalizing sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ that stabilize the location and scale of sample maxima as sample size increases, then the limiting distribution of the normalized variable $M_{n:n}^* = (M_{n:n} - b_n)/a_n$ will necessarily converge to a member of one of three parametric models. That is, if $M_{n:n}$ can be transformed as described,

so that the limiting distribution of the new random variable $M_{n:n}^*$ is non-degenerate, the functional form of that limiting distribution is given by the Extremal Types Theorem.

Extremal Types Theorem (THM1):

If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\lim_{n \rightarrow \infty} P \left\{ \frac{M_{n:n} - b_n}{a_n} \leq x \right\} = G(x)$$

where G is a non-degenerate distribution function, then G is necessarily a member of one of the three families:

$$\begin{aligned} G_I(x) &= \exp\{-e^{-(\frac{x-\mu}{\sigma})}\}, & -\infty < x < \infty \\ G_{II}(x) &= \begin{cases} 0, & x \leq \mu \\ \exp\{-(\frac{x-\mu}{\sigma})^{-\alpha}\}, & x > \mu \end{cases} \\ G_{III}(x) &= \begin{cases} \exp\{-[-(\frac{x-\mu}{\sigma})]^\alpha\}, & x < \mu \\ 1, & x \geq \mu \end{cases} \end{aligned}$$

with scale parameter $\sigma > 0$, location parameter μ , and in the case of G_{II} and G_{III} , shape parameter $\alpha > 0$.

While distributions G_I , G_{II} and G_{III} are collectively called the extreme value distributions, respectively they are known as the Gumbel, Frechet and Weibull distributions. Their functional similarity is striking, and together they comprise a wide range of heavy-tailed, bell shaped curves. The simple reparameterization $\xi = 1/\alpha$ unifies these distributions in a single three-parameter model jointly attributed to von-Mises and Jenkinson, and hereafter referred to as the Generalized Extreme Value distribution (GEVD). The reparameterization also permits the interpretation of the Gumbel (G_I) model as the limit of the Frechet (G_{II}) and Weibull (G_{III}) distributions. For convenience, the parametric form of the GEVD is presented in the following restatement of the Extremal Types Theorem.

Extremal Types Theorem Redux (THM2):

If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\lim_{n \rightarrow \infty} P \left\{ \frac{M_{n:n} - b_n}{a_n} \leq x \right\} = G(x)$$

where G is a non-degenerate distribution function, then G is necessarily a member of the GEVD family

$$G(x) = \exp \left\{ -[1 + \xi(\frac{x - \mu}{\sigma})]^{-1/\xi} \right\}$$

and G is defined on $\{x : 1 + \xi(\frac{x - \mu}{\sigma}) > 0\}$ with scale parameter $\sigma > 0$, location parameter μ and shape parameter $-\infty < \xi < \infty$.

Now, a little algebra employing the limiting definition of e :

$$\lim_{\xi \rightarrow 0} (1 + \xi x)^{1/\xi} = e^x$$

demonstrates that as $\xi \rightarrow 0$ the functional form of $G(x)$ in THM2 attains that of $G_I(x)$ in THM1. Furthermore, $G_{II}(x)$ and $G_{III}(x)$ correspond respectively to values of $\xi > 0$ and $\xi < 0$ in the GEVD parameterization. Hence the right-skewed Gumbel distribution ($\xi = 0$) serves as the central model in the sense that as ξ approaches zero the Frechet and Weibull models approach the Gumbel. As $\xi > 0$ increases, the Frechet model becomes increasingly right-skewed, and as $\xi < 0$ decreases, the Weibull model changes from left to right-skewed, with a symmetric shape nearly indistinguishable from a normal distribution when $\xi = -0.28$ (corresponding to $\alpha = -3.6$ in THM1). While both Gumbel tails are infinite, the Frechet model has an infinite right tail and the Weibull has an infinite left tail. The location parameter gives the left endpoint of the Frechet model, and the right endpoint of the Weibull model. The primary difference between the three extreme value models is the rate of decay of the upper distributional tail: while the Weibull model always has a finite upper endpoint, the Gumbel model tails decrease exponentially and Frechet's upper tail decreases polynomially.

When the underlying distribution F , leads to extremes that are described well by a particular member of the extreme value model family G , then F is said to be in the domain of attraction of the limiting distribution G , denoted $F \in D(G)$. Extremes extracted from the Uniform distribution, together with the minima extracted from a very wide range of parent distributions all converge to the Weibull distribution. The Pareto and Cauchy distributions reside in the Frechet maxima domain of attraction. The Gumbel model proves truly attractive: the Gaussian, log-normal, exponential and gamma distributions all reside in the Gumbel domain of attraction. In general, the rate of decay of the parent distribution's tail determines the extreme domain of attraction.

Only the most concocted distributions give rise to extremes that can not be made to converge to one of the three non-degenerate distributions in THM1. For example, let $E_1, E_2, E_3, \dots$ be i.i.d standard exponential random variables with density function F_E . Then, $F_E \in D(G_I)$; however the random variables defined by:

$$X_i := e^{\lfloor E_i \rfloor}, \quad i = 1, 2, 3, \dots$$

where $\lfloor E_i \rfloor$ is the integral part of E_i prove insufficiently smooth in distribution to yield extremes that converge to any member of the GEVD family (Dietrich et al, 2002). Fortunately, such examples remain theoretical curiosities, as overwhelming empirical evidence has demonstrated the applicability of the extreme value models to data extracted from physical and biological processes.

One can thus replace $[F(x)]^n$ in Equation 1.1 with the appropriate $G(x)$ from THM1 to achieve an estimate for sufficiently large n , essentially considering a string of maximum values extracted from multiple samples of equal size as independent realizations of a random variable distributed according to $G(x)$. Extremes from a very wide range of parent distributions can now be treated as realizations of a random variable having an EV distribution whose parameters must be estimated.

Coles' (2001) S-Plus routines permit the practitioner to calculate maximum likelihood estimates of GEVD and GPD parameters from a dataset and make inference into the extreme behavior of the physical process being modeled.

1.2.1 THE PARETO MODEL FOR EXCEEDENCES

Another approach to modeling extreme behavior in a physical process is to consider some threshold of relevance, and consider measurements that exceed that threshold. When confronted with data from some parent distribution F , the idea is to model the conditional distribution of:

$$\begin{aligned} P\{X > u + y | X > u\} &= \frac{P\{X > u + y\}}{P\{X > u\}} \\ &= \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0 \end{aligned} \quad (1.2)$$

where u is some value suggested by the data, above which the random variable $X|X > u$ converges to a non-degenerate distribution as the sample size increases. This approach is a more efficient use of data, since several of the upper-most values are used, instead of just the most extreme occurrence in each sample. The value u ideally reflects a threshold relevant in the context of the problem. There are also three models to which these random variables, commonly referred to as exceedences, will converge in distribution. For completeness, these distributions are listed below, but without location and scale parameters. Note that the simple analytic relationship:

$$W(x) = 1 + \log G(x), \quad \text{if } \log G(x) > -1$$

links the three extreme value densities G , from THM1 with the following three exceedence limiting distributions, W .

Exceedance Limiting Distributions:

$$W_I(x) = 1 - e^{-x}, \quad x \geq 0$$

$$\begin{aligned}
W_{II}(x) &= 1 - x^{-\alpha}, \quad x \geq 1 \\
W_{III}(x) &= 1 - (-x)^{-\alpha}, \quad -1 \leq x < 0
\end{aligned}$$

These distributions are the standard exponential, standard Pareto, and a subclass of the full Beta family of distributions, respectively. These three separate models can also be unified, via a reparameterization, into a three-parameter model called the Generalized Pareto Distribution (GPD). The parametric form is given in the following theorem, which also highlights the relationship between the GEVD and the GPD.

Exceedance Types Theorem (THM3):

Suppose F satisfies THM2 so that $M_{n:n}$ converges to a member of the GEVD, $G(x)$ indexed by some μ, σ, ξ . Then for large enough u , the distribution of the exceedance random variables $X - u | X > u$ converges to

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}$$

defined on $\{y : y > 0 \text{ and } (1 + \frac{\xi y}{\tilde{\sigma}}) > 0\}$ with $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

Just as in the case of the GEVD, the shape parameter ξ determines the behaviour of the GPD. That is, by taking the limit of $H(y)$ in THM3 as $\xi \rightarrow 0$, one obtains the unbounded exponential distribution $W_I(x)$. The cases $\xi > 0$, $\xi < 0$ correspond respectively to the unbounded Pareto distribution $W_{II}(x)$, and the Beta sub-family $W_{III}(x)$ with an upper endpoint of $u - \tilde{\sigma}/\xi$.

Once again, note that knowledge of the parent distribution F is not assumed. Were F directly estimable then Equation 1.2 could be used to calculate the exact distribution of exceedences. Instead one employs THM3, using the data to estimate the parameters of the appropriate GPD model, which proves to be the only possible limiting distribution of exceedences extracted from a very wide range of population distributions.

Section 1.3 delves deeper into the theoretical foundations of the GEVD and GPD models, so those reading to obtain an algorithm for the modeling of extreme events may skip to Section 1.4, wherein the discussion of useful estimation procedures and inferences begins.

1.3 DEEPER THEORETICAL FOUNDATIONS OF EV MODELS

That the three distributions in THM1 constitute the only possible limiting distributions in the extreme value context completes the analogy to central limit theory. The sequences $\{a_n > 0\}$ and $\{b_n\}$ are analogous to the sequences of means and standard deviations used to normalize data and thus obtain a member of the Gaussian model as the limiting distribution of sample means (or other measures of a distribution's central tendency). Unlike central limit theory, the stabilizing sequences in an extreme value analysis are not unique. Appropriate sequences $\{a_n > 0\}$ and $\{b_n\}$ have been identified for a wide array of underlying distributions, leading to the aforementioned list of distributions residing within each extremal domain of attraction. Recall however, that in most applications knowledge of the underlying distribution is negligible at best. Two problems are now apparent: identifying appropriate norming sequences from a list of candidates; and, relating the distribution of the transformed maxima to that of the maxima themselves. Both problems are easily resolved for all practical purposes.

First, suppose that there exist sequences $\{a_n > 0\}$, $\{b_n\}$ and $\{a'_n > 0\}$, $\{b'_n\}$ such that for a common parent distribution F , THM1 is satisfied with each set of sequences. Then we denote the two non-degenerate limiting ($n \rightarrow \infty$) distributions of the two series of random variables by:

$$\frac{M_{n:n} - b_n}{a_n} \rightarrow G \quad \text{and} \quad \frac{M_{n:n} - b'_n}{a'_n} \rightarrow G'$$

By a classic result of Khintchine (1938) on sequences of random variables, we then know:

$$\frac{a_n}{a'_n} \rightarrow \alpha \quad \text{and} \quad \frac{b_n - b'_n}{a_n} \rightarrow \beta$$

where $G'(\alpha x + \beta) = G(x)$, and hence G and G' are identical but for location and scale parameters. That is, the limiting distributions of the random variables obtained by transforming maxima with either set of sequences will have a common shape parameter, with possibly different location and scale parameters. The issue of non-unique norming sequences is thus resolved, as every appropriate sequence will lead to the same family of extreme value model (G_I , G_{II} , or G_{III}) given a parent distribution, F .

Next, assume that THM1 is satisfied, so the transformed maxima converge in distribution to one of the three extreme value distributions, and for large n we have:

$$P\{(M_{n:n} - b_n)/a_n \leq x\} \approx G(x)$$

This in turn implies that:

$$P\{M_{n:n} \leq x\} \approx G\{(x - b_n)/a_n\} = G^*(x)$$

where G^* is another member of the same extreme value model family. Thus, we see that when the distribution of the transformed extremes can be approximated by a GEVD model, the distribution of the extremes themselves is also estimable by a different member of the same GEVD model. Because the parameters of these limiting distributions are to be estimated, their equivalence for large enough n makes moot the fact that the parameters of G and G^* differ. Consequently, the apparent problems associated with the stabilizing sequences in the extreme value context can be ignored in practical applications.

We now consider an informal justification of the extremal types theorem that highlights a characterizing feature of the GEVD. Once again assume that the limiting

distribution of $(M_{n:n} - b_n)/a_n$ is the non-degenerate extreme value distribution G , so that:

$$P\{(M_{n:n} - b_n)/a_n \leq x\} \approx G(x)$$

Now consider the random variable $M_{nk:nk}$, $k \in N$ and n sufficiently large, representing extremes extracted from the same parent distribution arranged in blocks of size nk . Since nk is also large it must be that these extremes can also be rescaled such that:

$$P\{(M_{nk:nk} - b_{nk})/a_{nk} \leq x\} \approx G(x) \quad (1.3)$$

Furthermore, since each $M_{nk:nk}$ is also the maximum of k random variables each having the same distribution as $M_{n:n}$, the theory of order statistics permits:

$$\begin{aligned} P\{(M_{nk:nk} - b_{nk})/a_{nk} \leq x\} &= [P\{(M_{n:n} - b_n)/a_{nk} \leq x\}]^k \\ &= \left[P \left\{ \left(\frac{M_{n:n} - b_n}{a_n} + \frac{b_n - b_{nk}}{a_n} \right) \left(\frac{a_n}{a_{nk}} \right) \leq x \right\} \right]^k \\ &= \left[P \left\{ \frac{M_{n:n} - b_n}{a_n} \leq \left(\frac{a_{nk}}{a_n} \right) x - \frac{b_n - b_{nk}}{a_n} \right\} \right]^k \\ &\approx G^k(\alpha_k x + \beta_k) \end{aligned} \quad (1.4)$$

where $a_{nk}/a_n \rightarrow \alpha_k$ and $(b_{nk} - b_n)/a_n \rightarrow \beta_k$.

From Equations 1.3 & 1.4 one arrives at the following approximate equivalence:

$$G\left(\frac{x - b_{nk}}{a_{nk}}\right) \approx P\{M_{nk:nk} \leq x\} \approx G^k\left(\alpha_k \left[\frac{x - b_n}{a_n}\right] + \beta_k\right)$$

and concludes that G and G^k are once again identical apart from location and scale parameters. Note that α_k and β_k provide the new location and scale respectively.

This property, where the operation of extracting sample maxima from differently sized blocks of a single parent distribution has lead to limiting distributions from the same GEVD family (though they may differ in scale and location), is known as max- stability. The heuristic argument above demonstrates that max-stability is naturally expected in the extreme value context. In fact, this property characterizes

the GEVD and it is well-known that a distribution is max-stable if, and only if, it is a generalized extreme value distribution. An example involving the Frechet model $G_{II}(x)$, is illustrative.

Consider $G_{II}(x) = \exp\{-x^{-\alpha}\}$ for $x > 0$. Then it is known that $\{a_n > 0\} = \{n^{1/\alpha}\}$ and $\{b_n\} = 0$. Then:

$$\begin{aligned}\frac{a_{nk}}{a_n} &= \frac{(nk)^{1/\alpha}}{n^{1/\alpha}} \rightarrow k^{1/\alpha} = \alpha_k \\ \frac{b_{nk} - b_n}{a_n} &= 0 \rightarrow 0 = \beta_k\end{aligned}$$

so that according to Equation 1.4:

$$\begin{aligned}G^k(\alpha_k x + \beta_k) &= G^k(k^{1/\alpha} x) = [\exp\{(-k^{1/\alpha} x)^{-\alpha}\}]^k \\ &= [\exp\{-k^{-1} x^{-\alpha}\}]^k \\ &= \exp\left\{-\frac{k}{k} x^{-\alpha}\right\} = G_{II}(x)\end{aligned}$$

Thus we see that the Frechet model is indeed max-stable, as one expects in the extreme value context. The Gumbel and Weibull distributions are similarly seen to be max-stable, though the proof of the converse (that these are the only max-stable distributions), involves functional analysis beyond the scope of this study.

There is an analogous result called peaks-over-threshold (POT) stability that characterizes the GPD models. A full discussion of POT stability can be found in Reiss & Thomas (1997). Suffice it for our purposes to roughly summarize the work of Pickands (1975), who demonstrated that $F \in D(G)$ necessitates a good approximation for the conditional distribution:

$$F^{[u]}(x) = \frac{F(u+y) - F(u)}{1 - F(u)}$$

by a shifted GPD, assuming that u is large. That is, a parent distribution F , lies in the domain of attraction of a specific GEVD, G if, and only if, its upper tail can be appropriately approximated by a shifted GPD. The GPD models essentially zoom

in on the tail behaviour of the underlying distribution, while the GEVD models seek to describe only block extremes. It is the application of these models to real data, estimation procedures and inference to which we now turn our attention.

1.4 PARAMETER ESTIMATION IN THE EV DISTRIBUTIONS

The arguments of Sections 1.2 & 1.3 suggest the following algorithm for modelling the extremes from a series of independent observations. First, the data are blocked into contiguous sequences of equal and sufficiently large length, n . Then, a series of block maxima are extracted and a member of the GEVD family is fit to this series. More will be said in Section 2.3 with regard to the selection of an appropriate block length, though typically the block length is chosen to correspond to a certain fixed time period so the resulting maxima may, for example, represent monthly or annual extremes. Alternatively, one may define extreme events by identifying a high threshold u , and denoting by $x_{(1)}, \dots, x_{(k)}$ those $\{x_i : x_i > u\}$. Then, the random variables (called threshold exceedances):

$$y_j = x_{(j)} - u, \quad j = 1, \dots, k$$

are regarded as independent observations from a distribution likely to converge to, and hence well estimated by, a member of the GPD family.

The history of estimation procedures for fitting GEVD and GPD models to data parallels that of most any other statistical estimation problem, and is outlined extensively in the monograph by Kotz and Nadarajah (2000). Ongoing efforts have included method of moments procedures; simple, unbiased and asymptotically unbiased linear estimators; ranked set estimation; and a method employing probability-weighted moments. With more recent advances in computing power, there is strong preference for the maximum likelihood estimation (MLE) procedure. This is especially true when seeking to estimate all three unknown parameters in the GEVD

model from a complete dataset. While each method has its niche, the overall adaptability of MLE methods to complex model-building make this technique very useful. The fact that the GEVD model endpoints are functions of the parameters actually violates regularity conditions that ensure the asymptotic validity of the MLEs. However, the range of shape parameter values for which MLEs are unobtainable ($\xi < -1$) correspond to distributions with a very short bounded upper tail that are very rarely encountered in contexts calling for extreme value modelling (Smith, 1985). This theoretical limitation proves no obstacle in the modelling of extreme behaviour in the SREL deer dataset discussed in Section 3.1.

Stuart Coles has implemented MLE estimation procedures in S-Plus, and his text (Coles, 2001) outlines in detail the form of the log-likelihood functions for the separate cases of $\xi = 0$ and $\xi \neq 0$ in both the GEVD model of THM2, and the GPD model of THM3. With a fairly straightforward optimization routine, the MLEs $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ are obtained. Subject to the limitations alluded to on the permissible values of ξ , the MLEs have an approximately multivariate normal distribution with mean (μ, σ, ξ) and variance-covariance matrix obtained by inverting the observed information matrix evaluated at the MLEs. This leads directly to calculations of confidence intervals, profile likelihood confidence intervals on any combination of parameters, inference on return levels, and the construction of tolerance limits for the extreme value distributions. Since both kinds of confidence interval are fundamental to any maximum likelihood procedure, we focus instead on inference involving return level and tolerance limit estimation, as they are indeed particular to the extreme value modelling context.

1.5 RETURN LEVEL PLOTS AND DIAGNOSTICS FOR GEVD MODELS

For convenience, assume that a GEVD model has been successfully fit to a series of annual maxima. We seek, to a reasonable degree of accuracy, the value z_p that is exceeded, on average, once every $1/p$ years. $1/p$ is called the return period associated with the return level z_p . In other words, z_p is the data value that is exceeded by an annual maxima in any particular year with probability p . In general, long return periods (i.e., large $1/p$ and hence small p) are of the most interest because they are associated with the most extreme observations and model predictions. Again, caution is necessary when making such extrapolations. It is generally accepted that in extrapolations beyond about four times the span of the data being modelled, the compromised asymptotics of the GEVD model parameters break down. It is important also to distinguish between the return level z_p , where we are requiring $G(z_p) = 1 - p$ with $0 < p < 1$, and the usual definition of quantiles where we'd have $G(z_p) = p$.

Estimates of the extreme quantiles z_p are obtained by inverting the parametric form of the GEVD in THM2:

$$G(z_p) = 1 - p \quad \Rightarrow \quad z_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{-\log(1 - p)\}^{-\xi}], & \xi \neq 0 \\ \mu - \sigma \log\{-\log(1 - p)\}, & \xi = 0 \end{cases}$$

Since we've assumed that the MLE's $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ have been obtained, we substitute these into this expression and make the simplifying definition $y_p = -\log(1 - p)$ to yield:

$$z_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} [1 - y_p^{-\hat{\xi}}], & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log y_p, & \hat{\xi} = 0 \end{cases}$$

Now, it is apparent that if z_p is plotted against $\log y_p$ the resulting plot is: linear in the case $\xi = 0$; convex with asymptotic limit (as $p \rightarrow 0$) at $\mu - \sigma/\xi$ in the case $\xi < 0$; and concave with no finite bound in the case $\xi > 0$. Because the logarithmic

scale condenses the tail of the distribution, the effect of extrapolation is highlighted. Furthermore, Coles adds empirical quantile estimates and confidence intervals to account for sampling error in these return level plots. If the model is suitable for the data, then the empirical and model based quantile estimates should be in agreement, with all empirical points falling within the model's confidence bands. There are two methods for obtaining confidence interval estimates for z_p . First, the delta method permits the approximation:

$$Var(\hat{z}_p) \approx \Delta z_p^T \times V \times \Delta z_p$$

where V is the variance-covariance matrix of $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ and

$$\begin{aligned} \Delta z_p^T &= \left[\frac{\delta z_p}{\delta \mu}, \frac{\delta z_p}{\delta \sigma}, \frac{\delta z_p}{\delta \xi} \right] \\ &= \left[1, \frac{-(1 - y_p^{-\xi})}{\xi^{-1}}, \frac{\sigma(1 - y_p^{-\xi})}{\xi^{-2}} - \frac{\sigma y_p^{-\xi} \log y_p}{\xi^{-1}} \right] \end{aligned}$$

evaluated at $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$. While all necessary estimates are provided by Coles S-Plus procedures, a simple spreadsheet or Matlab program is required to perform the actual calculation of Δz_p^T , and hence obtain the approximate standard error of z_p .

An alternative, and usually slightly more accurate method of return-level confidence interval estimation is via the profile likelihood method. Coles has implemented the minimization of the negative log-likelihood function expressed as a function of (z_p, σ, ξ) , and obtains a direct estimate of z_p . A 100%(1 - α) confidence limit estimate for the return period is obtained by noting the parameter values on the x-axis corresponding to points of intersection with the horizontal line drawn at a height of $0.5 \times c_{1,\alpha}$ below the maximum, where $c_{1,\alpha}$ is the 100%(1 - α) quantile of the χ_1^2 distribution.

For shorter return periods, confidence interval estimates obtained by the profile likelihood method are similar to those from the delta method. For longer return periods the methods do not produce similar results. The asymmetry of the profile

log-likelihood surface increases with the return period, and is responsible for this discrepancy. Since the data provide only limited information about high levels of the physical process being modelled, this asymmetry is to be expected. Although the GEVD model is supported by mathematical argument, parameter estimates and measures of precision are based on the assumption that the model is correct. Strictly speaking, all model inference and particularly that associated with long return periods, should be considered lower bounds since measures of precision would increase if model uncertainty were accounted for. Here is a potential application of Bayesian techniques where uncertainty due to model correctness is quantifiable in some manner.

The profile likelihood estimates generally produce more accurate results than the delta method. For the practitioner interested in obtaining results quickly, the profile likelihood method permits a shortcut to comparing nested models via the likelihood ratio test. The calculation of the deviance statistic and its comparison for significance against the χ^2_1 distribution is operationally equivalent to observing whether or not 0 lies in the profile likelihood interval for the relevant model parameter. Hence, testing the Gumbel model (e.g., $\xi = 0$) which is nested in the Frechet model (e.g., $\xi = -0.2$) motivated by data-based MLE is achieved simply by noting if 0 lies in the estimated profile likelihood confidence interval for ξ in the Frechet model. While Coles' S-Plus functions produce graphs of the profile log-likelihood function, a more user friendly implementation is provided in the EVIS, Version 4 S-Plus library by Alexander J. McNeil (2001). The EVIS library includes a function that returns the confidence interval endpoints, eliminating the subjectivity in reading coordinates from Coles' plots. This function borrows Coles' optimization routine, so the equivalence of the two implementations is assured.

The Coles S-Plus functions also include other diagnostic plots: the standard quantile and probability plots; and a visual comparison of the fitted model's den-

sity function with a data histogram. The quantile plot compares the ordered data values and the fitted model's quantile function. Departures from linearity between the two sets of quantiles indicate a lack of model fit. The probability plot checks for linearity between the empirical and fitted distribution functions; however, observe that as extreme values increase, both distribution functions necessarily approach 1, assuring an increasingly linear plot. In the extreme modelling context, we are most concerned with model accuracy for these large extreme values. Hence, the probability plot yields the least information in the data region of greatest interest. Finally, the density/histogram plot is actually the least informative. A histogram provides an empirical estimate of the density function; however, different choices of grouping intervals when constructing the histogram can lead to significantly different plots, so that comparisons with the model-based density estimate are subjective at best.

1.6 TOLERANCE LIMIT ESTIMATION

In ecological applications, one may be concerned with identifying a limit associated with dangerous concentrations of substances in biological systems. For example, excessive ozone concentrations contribute to global warming, and extreme radiocesium concentrations in the SREL deer population could lead to a hunting moratorium.

A direct method for calculating tolerance limits for extreme value distributions from fitted parameter estimates is presented by Dasgupta and Bhaumik (1995). They deal with the three extreme value models separately. Below, the upper tolerance limits for the Frechet and Gumbel distributions are developed using the unified GEVD model. For the Frechet distribution G , we seek an upper bound such that a high percentage ($100\beta\%$) of future observations from the parent distribution F , will fall below that bound with a very high probability, γ . This β -content tolerance

limit is denoted $M_{(n)}\delta$, where $M_{(n)}$ is the maximum of the i.i.d. variables from F (i.e., the maximum observed value used to fit the Frechet model). For some future observation $Y \sim G$, we require for $\delta > 0$:

$$P_{M_{(n)}}\{P_G[Y \leq M_{(n)}\delta] \geq \beta\} = \gamma.$$

From this equation, we have:

$$\begin{aligned} \Rightarrow P_{M_{(n)}}\{G(M_{(n)}\delta) \geq \beta\} &= \gamma \\ \Rightarrow P_{M_{(n)}}\{M_{(n)} < G^{-1}(\beta)/\delta\} &= 1 - \gamma \\ \Rightarrow G^n[G^{-1}(\beta)/\delta] &= 1 - \gamma \\ \Rightarrow G^{-1}(\beta)/\delta &= G^{-1}[(1 - \gamma)^{1/n}] \\ \Rightarrow \delta &= \frac{G^{-1}(\beta)}{G^{-1}[(1 - \gamma)^{1/n}]} \end{aligned}$$

Now, for the Frechet distribution we use the inverse as defined in Section 1.5 to obtain:

$$\delta(\mu, \sigma, \xi) = \frac{\mu - \frac{\sigma}{\xi}[1 - (-\log \beta)^{-\xi}]}{\mu - \frac{\sigma}{\xi}[1 - (-\frac{1}{n} \log(1 - \gamma))^{-\xi}]}$$

Finally, substituting the MLE estimates $M_{(n)}\delta(\hat{\mu}, \hat{\sigma}, \hat{\xi})$, provides the desired upper tolerance limit for the Frechet distribution and predetermined percentages β and γ . Similarly, the upper tolerance limit for the Gumbel distribution is provided by:

$$\delta(\mu, \sigma) = \frac{\mu - \sigma \log(-\log \beta)}{\mu - \sigma \log[-\frac{1}{n} \log(1 - \gamma)]}$$

and the subsequent estimate is again given by $M_{(n)}\delta(\hat{\mu}, \hat{\sigma})$.

These point-estimates of the β -content tolerance limits do not yet reflect the uncertainty associated with the selection of a specific EV model. The delta method again permits the estimation of model uncertainty and leads to β -content tolerance limit confidence intervals. In the Frechet model, let $\hat{d} = \delta(\hat{\mu}, \hat{\sigma}, \hat{\xi}) \approx \delta(\mu, \sigma, \xi)$, and approximate the variance of $\hat{d}$ by:

$$Var(\hat{d}) \approx \Delta \hat{d} \times V \times \Delta \hat{d}$$

where V is again the variance-covariance matrix of the model estimates $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ and

$$\Delta \hat{d} = \left[\frac{\delta d}{\delta \mu}, \frac{\delta d}{\delta \sigma}, \frac{\delta d}{\delta \xi} \right]$$

is the vector of partial derivatives of the tolerance limit multiplier $\delta(\mu, \sigma, \xi)$, evaluated at $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$.

To simplify the evaluation of this approximation, first let

$$\begin{aligned} a &= 1 - (\log \beta)^{-\xi} \\ b &= 1 - \left(-\frac{1}{n} \log(1 - \gamma) \right)^{-\xi} \\ T &= \mu - \frac{\sigma}{\xi} \times a \\ B &= \mu - \frac{\sigma}{\xi} \times b \end{aligned}$$

so that $\hat{d} = \frac{T}{B}$. Thus we may proceed:

$$\frac{\delta \hat{d}}{\delta \mu} = \frac{B - T}{B^2} = \frac{\sigma(a - b)}{\xi \cdot B^2}$$

then

$$\begin{aligned} \frac{\delta \hat{d}}{\delta \sigma} &= \frac{-\frac{1}{\xi} \cdot a \cdot B + \frac{1}{\xi} \cdot b \cdot T}{B^2} \\ &= \frac{-1}{\xi \cdot B^2} \left[\left(a\mu - ab\frac{\sigma}{\xi} \right) - \left(b\mu - ab\frac{\sigma}{\xi} \right) \right] \\ &= \frac{-\mu(a - b)}{\xi \cdot B^2} \\ &= \frac{-\mu}{\sigma} \cdot \frac{\delta \hat{d}}{\delta \mu} \end{aligned}$$

and finally

$$\begin{aligned} \frac{\delta \hat{d}}{\delta \xi} &= \frac{1}{B^2} \left\{ - \left[\frac{-\sigma a}{\xi^2} + \frac{\sigma}{\xi} \cdot \frac{\delta a}{\delta \xi} \right] B + \left[\frac{-\sigma b}{\xi^2} + \frac{\sigma}{\xi} \cdot \frac{\delta b}{\delta \xi} \right] T \right\} \\ &= \frac{1}{B^2} \left\{ -B \cdot \frac{\sigma}{\xi} \left[(-\log \beta)^{-\xi} \cdot \ln(-\log \beta) - \frac{a}{\xi} \right] \right. \\ &\quad \left. + T \cdot \frac{\sigma}{\xi} \left[\left(-\frac{1}{n} \log(1 - \gamma) \right)^{-\xi} \cdot \ln \left(-\frac{1}{n} \log(1 - \gamma) \right) - \frac{b}{\xi} \right] \right\} \\ &= \frac{1}{B^2} \cdot \frac{\sigma}{\xi} \left\{ T \left[(1 - b) \ln \left(-\frac{1}{n} \log(1 - \gamma) \right) - \frac{b}{\xi} \right] - B \left[(1 - a) \ln(-\log \beta) - \frac{a}{\xi} \right] \right\} \end{aligned}$$

Next the estimated variance is used to quantify model uncertainty in the tolerance limit point estimates $M_{(n)}\delta(\hat{\mu}, \hat{\sigma}, \hat{\xi})$, by constructing confidence intervals. Observe that $Var(M_{(n)}\delta(\hat{\mu}, \hat{\sigma}, \hat{\xi})) = M_{(n)}^2 \cdot Var(\delta(\hat{\mu}, \hat{\sigma}, \hat{\xi}))$, so that

$$M_{(n)}\delta(\hat{\mu}, \hat{\sigma}, \hat{\xi}) \pm 1.96 \cdot M_{(n)} \cdot \sqrt{Var(\delta(\hat{\mu}, \hat{\sigma}, \hat{\xi}))}$$

provides a 95% confidence interval for $M_{(n)}\delta(\hat{\mu}, \hat{\sigma}, \hat{\xi})$. A simple Matlab program is used to calculate these confidence intervals for a given model and plot them over a range of β and γ in Section 3.2.

1.7 FINDING THE DOMAIN OF ATTRACTION FOR NEARLY EVERY EXTREME

To gain insight into the highest feasible contamination levels in a population, one naturally looks to the upper tail of the distribution of observed levels. In fact, the distributional properties of extremes are governed by the underlying distribution. It is well known in Extreme Value Theory that if the underlying population distribution yielding the maxima under consideration satisfies mild regularity conditions, then the distribution of exceedences over a specified threshold in the data will converge (as the threshold approaches the endpoint of the parent distribution) to one of the three types of the generalized Pareto distribution (GPD). In this case, the population density is said to be in the domain of attraction of the specified GPD. Similarly, the distribution of the maximum value extracted from a sample will converge (as the sample size approaches infinity) to one of the three types of generalized extreme value distributions (GEVD), and the population density is said to be in the domain of attraction of the specified GEVD.

The problem of identifying the domain of attraction of the parent distribution has received much attention in the literature, and three strategies have been developed for statistical inference in the extreme value context. The first strategy selects the type of EV model to be fit (Frechet, Weibull or Gumbel) based on evidence gleaned

from hypothesis tests and/or expert knowledge, and then determines the best fitting model within the assumed type. Wang (1995) offers a modified Greenwood statistic based on the largest order statistics in a sample to test for each domain of attraction separately. The second approach determines the best model from the unified GEVD (or GPD) family, whose shape parameter (ξ) determines the specific extreme value model. Dietrich, De Haan and Husler (2002) formalize a test based on an asymptotic expansion of the tail version of a statistic comparing the inverse population and empirical distribution functions, and determine the domain of attraction without first specifying the shape parameter. Their method is motivated by the familiar Cramer-von Mises statistic. The first approach makes no effort to capture the uncertainty inherent in the initial selection of the EV model to be fit, and the second suffers from the reduction of the Gumbel model to a single point ($\xi = 0$) in continuous parameter space (μ, σ) , meaning that the probability of selecting the Gumbel model via maximum likelihood techniques is zero. However, the unified parametric models do lend themselves nicely to likelihood ratio testing techniques, which can be used to give satisfactory consideration to all possible models. A third approach championed by Stephenson and Tawn (2001) applies Bayesian inference to account for each EV model in an extreme value analysis, and thus addresses the deficiencies of the previous methods.

The curvature method for identifying the EV domain of attraction was formalized by Castillo, Galambos and Sarabia (1989). The upper tail of the empirical distribution is shown to imitate the shape of that EV limiting distribution that attracts it, and this result used to develop both a visual selection method and a test for significance based on the curvature of the empirical distribution function in the tail. As emphasized by Castillo et. al., it is important to consider only the upper quantiles of the empirical distribution function (the 50th and up are suggested) when determining the curvature in these plots, because the convergence of the empirical

tail to the tail of the limiting distribution which attracts it is established for only the upper quantiles. This is another example of the second approach to selecting the domain of attraction, but the visual selection method does allow for the consideration of the Gumbel model. Several other visual methods are commonly employed to allow the data to suggest the extremal domain of attraction of the population distribution. These generally include identifying characteristic behaviour in plots of estimated mean excess, survival, hazard or loss functions. Coles' (2001) development of return level plots as discussed in Section 1.5 is yet another example of a graphical method to assess model fit.

CHAPTER 2

GETTING TO KNOW THE SRS DEER DATA

2.1 PRELIMINARY ANALYSIS LEADS TO INVESTIGATION OF EXTREME DEER

The data used in this extreme value analysis are a univariate series ($n = 29519$) of radiocesium concentrations in single muscle plugs taken from each deer harvested from the Savannah River Site (SRS) in Aiken, South Carolina during the fall deer-hunting seasons between 1965 and 1995. The number of observations varies from year to year. Measurement units are Becquerels per kilogram (Bq/kg) and have a minimum value of 37 occurring 2452 times, a maximum value of 3626 occurring twice, and increments occurring in multiples of 37 except for data from two years when measurements were taken at a finer scale and increments occur in multiples of 18.5. To reduce the scale of subsequent results, all radiocesium measurements were divided by 37 and thus converted to picocuries per kilogram (pCi/kg). The resulting data range from 1 to 98 with increments of either 0.5 or 1 and are used in all subsequent analyses. Primarily, interest lies in characterizing the frequency and magnitude of the maximum pCi/kg radiocesium body burden in the deer data.

It should be noted that the nature of the data implies a true maximum body burden measurement must exist in the SRS deer population. It is unreasonable to expect that radiation concentration can be infinite in an organism. In the interest of achieving the most accurate model, which incorporates scientific knowledge in a sound statistical analysis, consideration should be given to restricting the domain of attraction to those models that restrict upper quantiles to reasonable values. In

fact, was one able to quantify a reasonable range of upper limits (i.e., assign a prior distribution to more and less likely parameter values in the GEVD or GPD), the SRS deer data could provide a nice application of Stephenson and Tawn's (2001) Bayesian inference scheme, designed to reduce uncertainty in parametric estimates by integrating expert knowledge with the Extreme Value Theorems.

With the goal of identifying a population density model for which the extreme value domain of attraction is known, the pCi/kg sample density is examined in Figures 2.1 and 2.2. It has been observed that radiation exposure in animal populations frequently conforms to a lognormal distribution, implying that maximum values taken from these populations converge to a Gumbel distribution. Clearly, Figure 2.1 shows that the body burden histogram does not resemble a typical lognormal density (the best fitting density is superimposed), nor does the $\ln(\text{pCi/kg})$ histogram conform to a suitable normal density as it should if the body burden data were indeed lognormal. Specifically, the familiar QQ-Plot in Figure 2.2 illustrates that the empirical distribution has a lighter left tail (due to the concentration of minimum values at 1) and a heavier right tail than any normal distribution. Here a tail is heavy (light) in the sense that more (less) values are more (less) extreme in that tail's direction. Since the density of the deer data is not easily identified, further analysis using the tools of Extreme Value Theory is carried out in Chapter 3.

2.2 IMPLICATIONS OF TIME TAKEN SERIOUSLY, BUT RULED OUT

Each measurement in the dataset is accompanied by a date; however, the dates mark when measurements were recorded in the dataset and *not* when each measurement was physically collected. Data entries are grouped every 3 or 4 days (indicating twice weekly compilation of daily harvest measurements) during the fall deer-hunting

seasons from 1965 to 1995 (either September or October through mid December). Though the individual measurements cannot be considered a true time series, the inherent and somewhat arbitrary data grouping justifies sub-dividing the data into less arbitrary and more informative blocks, the maxima of which will be treated as a traditional time series to check for any underlying time-dependent variability.

Prior to delving into the methods used to determine suitable extreme value models for the maxima, a closer look at the deer data will highlight any serial structure. Figures 2.3 and 2.4 show a new graphical technique to summarize and visually analyze extreme values taken from contiguous blocks of the entire dataset. The first graph blocks the data according to the year in which observations were recorded, while the second graph blocks the data into lengths of $n = 984$, the average number of observations per year indicated on the first graph by a solid straight line. In each graph: the top three broken lines indicate the magnitude and frequency of the first, second and third highest order statistics in each block; the three dashed red lines and the gray line at bottom track three quantiles (98th, 95th & 50th and the mean of the radiocesium measurements in each block, respectively; finally, the bar graph illustrates the number of observations in each block.

Figure 2.3 highlights a few oddities worth mentioning. Data collection was sparse in the first five years, yet 1969 yields the second highest maximum value of 3071 Bq/kg despite having the fifth lowest number of observations overall. Similarly, several other years with below average block sizes give rise to high maximum values (e.g., 1970-71, 1978-80, 1987-88). Conversely, those years with above average block size do not seem to give rise to higher maximum values. Finally, the mean exposure line does not indicate any significant trend across blocks. These observations indicate that neither time related, nor block-size dependent variability exists in the magnitude of extreme values in the body burden data. A potential ecological hypothesis to explain the seeming lack of any trend in the extreme value summaries follows.

Perhaps a sub-population of deer with elevated (i.e. far above average) radiocesium exposure does exist on site, and is as likely to be harvested as any other deer on site. If harvest numbers in general were suitably small compared to the total deer population on site, then you would not necessarily expect increased (decreased) occurrence of elevated exposures from blocks of slightly above (below) average size.

Figure 2.4 conforms more closely to traditional extreme value theory in that blocks of equal size are considered. The plots across blocks of each statistic under consideration show at least slight improvement in smoothness over the unequal block case discussed above. Further, no significant trend or seasonality becomes apparent in the series of maximum values, nor in the mean exposure. However, this second graph does highlight one potentially interesting data anomaly. In the 18th block, exposures seem to decrease markedly for the remaining 13 blocks. This block begins with observation number 16728, collected in mid 1983: the same year in the previous graph exhibits a somewhat similar drop. It may be of interest to build separate models for the first 17 and last 13 blocks (or years) to quantify differences and seek an ecological explanation.

Finally, two additional methods of assessing the serial dependence structure of the series of maxima from each block size are available. First, the partial autocorrelation functions are graphed for each series of maxima (block sizes of 500, 987, 1000), and checked for any indication that univariate autoregression (AR) models are appropriate. Diagnostic tools produced using the S-Plus **acf** functions show that AR models do not fit well to sequences of maxima extracted from the SREL deer data. Had a time-dependant trend been indicated, standard likelihood ratio testing techniques would assess if non-stationary EV models achieved an improvement in model fit sufficient to justify increasing model complexity by including extra parameters to capture any serial tendencies in the maxima.

2.3 BLOCKING SCHEMES CAREFULLY CONSIDERED

It is evident from the statement of the EV Theorem that the limiting distribution to which extreme values converge is dependent on the sample, or block size (n) through the standardizing series $\{a_n\}$ and $\{b_n\}$. As demonstrated in Section 1.3, one need not be concerned with identifying the appropriate standardizing series since they are essentially absorbed in the procedure for estimating the GEV and GPD model parameters. The fixed block size is necessary for accurate estimation of return level plots and tolerance limits. In experimentally driven data collection, it is likely that the block size in an extreme value analysis will correspond to the repeated sample size in the experiment, or to some inherent time-frame (e.g., minimum breaking strengths of materials from multiple batches of fixed size; maximum monthly wind speeds or atmospheric ozone concentrations). Figure 2.3 demonstrates that no natural block size is evident in the SRS deer data - such is frequently the case when considering data collected in an observational, as opposed to experimental, manner. However, the relative constancy of the quantiles depicted in Figure 2.3 does suggest that arranging the data into equally sized contiguous blocks is not inappropriate. On average there are 984 observations in each of the 30 years of data; hence, contiguous blocks of size $n = 984$ (obtaining 30 extreme observations, with the last being the maximum of 983 observations) are used in the initial fitting of the GEV distribution via Coles' maximum likelihood estimation routines in S-Plus. While this blocking method is a nicely intuitive approach, basing a model on blocks of size 1000 will facilitate the communication of statistical results to the scientists to whom this discussion is targeted. In the interest of simplifying discourse, the initial model is compared to the GEV model fit to 29 maxima obtained from contiguous blocks of size $n = 1000$.

The arbitrary manner in which block sizes must be established in any similar observational data analysis raises questions as to the effect on model parameters

when the block size is changed. The choice of block size given a fixed total sample size is essentially a trade-off between bias and variance. The larger the block size from which maxima are extracted, the better the estimate of truly extreme values in the total sample, and hence the lower the bias when estimating extremes; however, this coincides with fewer observed extremes, and hence a higher variance results from fitting a model to fewer data points. As the number of blocks is increased and the variance of estimates is reduced, truly extreme values are diluted, thereby increasing the bias of these estimates. At present, this bias-variance trade-off has not been the subject of much formal investigation, but it remains an important consideration when devising an intuitive blocking method for observational data.

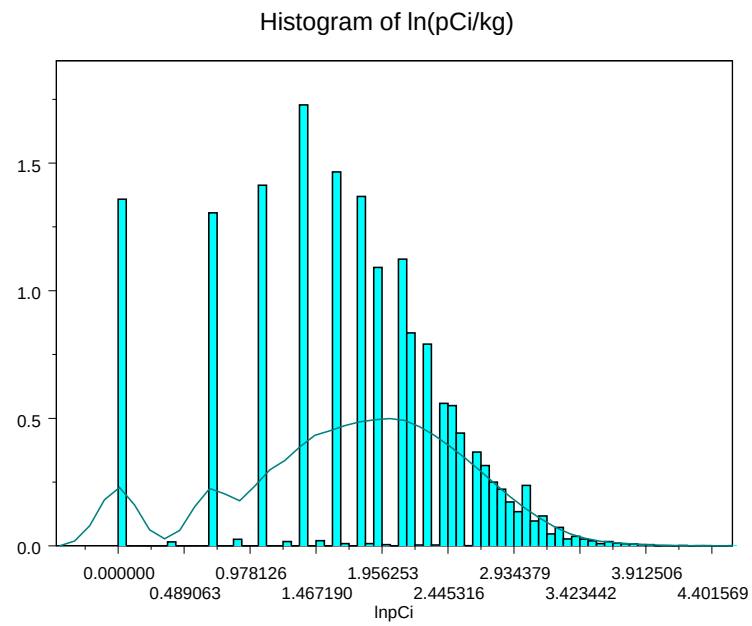


Figure 2.1: Histogram of $\ln(\text{pCi/kg})$.

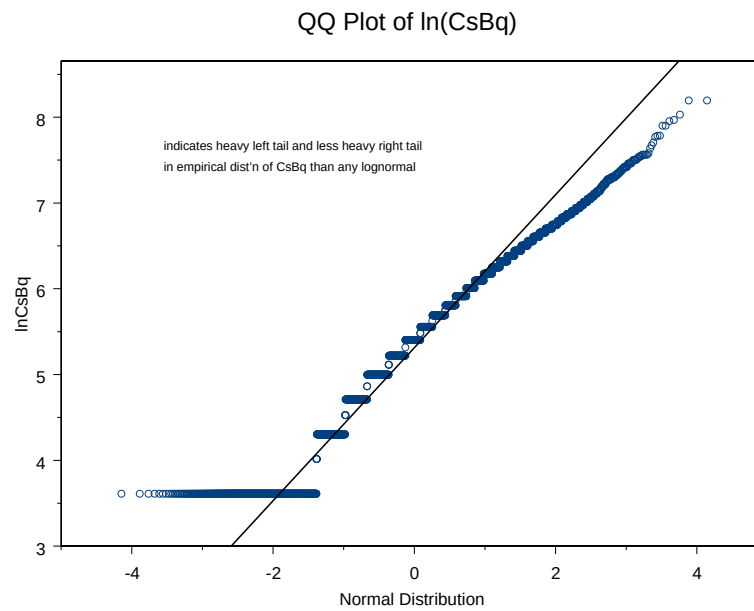


Figure 2.2: QQ-plot of $\ln(\text{pCi/kg})$.

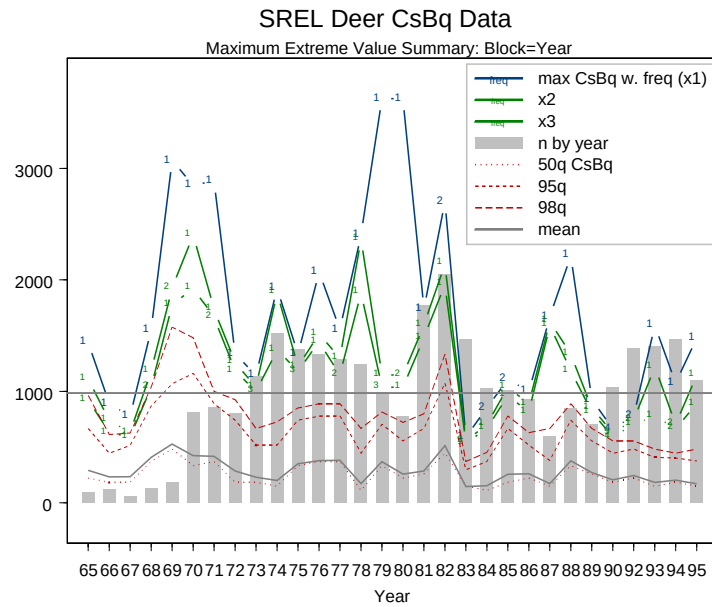


Figure 2.3: Extreme value summary plot of raw body-burden dataset.

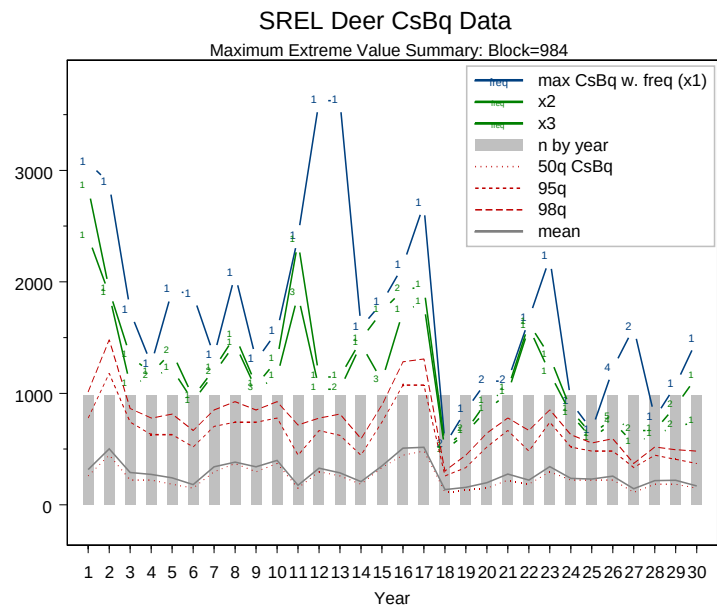


Figure 2.4: Extreme value summary plot of body-burden dataset arranged in contiguous blocks of 984 observations.

CHAPTER 3

RESULTS AND CONCLUSION

3.1 MAXIMUM RADIOCESIUM BODY-BURDENS IN THE SREL DEER DATA

As discussed earlier, we assume the pattern of variation in the body-burden measurements is relatively constant over the 30-year observation period, and consequently model the series of maxima extracted from different block sizes as independent observations from the GEVD distribution.

The first model is fit to the maximum observations from contiguous blocks of size $n = 984$, the average number of observations in each year. Maximization of the GEVD log-likelihood for these data result in the parameter estimates:

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (36.74, 16.676, 0.0322)$$

with approximate standard errors given respectively by $(3.530, 2.643, 0.165)$, and corresponding log-likelihood of -132.454 . From these results, we obtain approximate 95% confidence intervals for each parameter via the familiar formula $\text{MLE} \pm 1.96 \times \text{s.e.}$:

$$\hat{\mu} \mapsto (29.82, 43.66)$$

$$\hat{\sigma} \mapsto (11.50, 21.86)$$

$$\hat{\xi} \mapsto (-0.29, 0.36)$$

Note that since $\hat{\xi} > 0$, these data suggest that an unbounded Frechet model is appropriate. However, the $\hat{\xi}$ confidence interval is very nearly centered on zero,

indicating that the data do not provide strong evidence in support of the Frechet extreme value distribution.

Return levels are calculated based on the block size. In the theoretical development of return levels, we assumed a model built from truly *annual* maxima. In the current example, we've modelled block maxima as opposed to annual maxima, so return times are expressed as multiples of the block size. For example, when interpreting a “3-year” return level from this model, one is actually estimating the level expected to be exceeded on average once every 3 blocks, or $3 \times 984 = 2952$ observations. Clearly, these results are neither intuitive nor easy to communicate and when possible, one would prefer either a model based on annual maxima, or a nice round block size. To that end, we now consider a second model fit to the maximum observations from contiguous blocks of size $n = 1000$.

Maximization of the GEVD log-likelihood for these data result in the parameter estimates:

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (35.90, 16.287, 0.0669)$$

with approximate standard errors given respectively by $(3.496, 2.668, 0.178)$, and corresponding log-likelihood of -132.335 . Again, we obtain approximate 95% confidence intervals for each parameter:

$$\hat{\mu} \mapsto (29.05, 42.75)$$

$$\hat{\sigma} \mapsto (11.06, 21.52)$$

$$\hat{\xi} \mapsto (-0.28, 0.42)$$

Figure 3.1 depicts histograms overlayed with the fitted density curves for both these models. It is clear from the similarity between both sets of parameter estimates and between the panels of Figure 3.1 that the two models are almost indistinguishable. Hence, the $n = 1000$ model is adopted as the basis for calculating return times and tolerance limits and is used in hierarchical model testing against the reduced

Gumbel model and models allowing for time trends in either the location or scale parameters.

Recall also that an alternative formulation of confidence intervals is available from a profile likelihood analysis. In fact, increased accuracy usually results from examining the profile log-likelihood function. Coles' S-plus routines produced Figure 3.2, from which the approximate 95% confidence interval for ξ is obtained: $(-0.22, 0.49)$. This interval is only a little different from the previous calculation, being shifted subtly to the right. Hence, slightly stronger evidence is garnered in support of the Frechet model for these maxima.

Note that the best estimate of the confidence interval for ξ still includes zero, and extends well into negative values of ξ . Thus, the appropriateness of replacing the Frechet model with the Gumbel model should be assessed. To do this, the log-likelihood of the Gumbel model ($\xi = 0$) is maximized, leading to the parameter estimates $(\hat{\mu}, \hat{\sigma}) = (36.495, 16.75)$ with standard errors given by $(3.216, 2.46)$. The corresponding maximized log-likelihood is -132.411 , which is slightly worse than that of the Frechet model. Consequently, no ratio testing is required to determine that the Gumbel model is in fact inferior to the Frechet in capturing the behaviour of extremes from this dataset. Observe that the approximate standard errors in the Gumbel model also exceed those of the Frechet model. However, the two models are certainly comparable, and the Gumbel model may still seem advantageous given its simpler form. The most important difference between these two models is evident from a comparison of the diagnostic plots for each (see Figures 3.3 & 3.4). Observe that the return levels for the Gumbel model have significantly narrower confidence bands than for the Frechet model. While this reduction of uncertainty is desirable, and the data do suggest the plausibility of the Gumbel model, this does not imply that other members of the GEVD family should be disregarded. Due consideration must be given to the maximum likelihood estimate within the Frechet family, and

given the uncertainty evident in this discussion of the two models, the conservative and safe option is to give preference to inference based on the Frechet model. In so doing, we are more realistically quantifying the inherent uncertainty of extrapolated predictions. This, of course, is reflected in the Frechet model's wider confidence intervals for return levels.

Return level inference is obtained by determining return periods of interest and setting p accordingly in the calculations outlined in Section 1.5. The Frechet model is used to estimate the body-burden level that is expected to be exceeded on average once every 10 blocks ($n = 1000$) in the SREL deer population by setting $p = 1/10$ and calculating $\hat{z}_{p=1/10} = 75.457$. Confidence intervals are obtained by estimating the variance via the delta or profile-likelihood method. The delta method yields $\text{Var}(\hat{z}_{0.1}) = 80.733$ and hence the 95% confidence interval is estimated by $75.457 \pm 1.96 \times \sqrt{80.733} = [57.846, 93.067]$. More accurate estimates come from the profile-likelihood method which yields the confidence interval of $[62.881, 108.747]$ for $\hat{z}_{0.1}$. The most precise and accurate interpretation of this return level result follows: the block maximum in any particular block will exceed $\hat{z}_{0.1} = 75.457$ with probability $p = 10\%$, and 19 times out of 20 $\hat{z}_{0.1}$ will fall in the interval $[62.881, 108.747]$. Table 3.1 includes estimates from both methods for return periods of 30, 45 and 60 blocks. These correspond to extrapolations up to twice as far into future observations as the SREL dataset provides information for (i.e., 29 blocks of size $n = 1000$). Figures 3.5 to 3.8 highlight that profile log-likelihood surfaces will increase in asymmetry as the return period increases, explaining the difference between confidence interval estimates from the two methods.

3.2 TOLERANCE LIMITS FOR THE SREL DEER DATA

The delta method is applied to obtain 95% confidence intervals for upper β -content tolerance limits in the Frechet ($n = 1000$) model of the SREL deer dataset. Figure 3.9 summarizes two sets of tolerance limit estimates and confidence intervals for the range $75\% \leq \beta \leq 99\%$. The solid line on the left (right) maps the upper boundary of body-burden measurements that $(100\beta)\%$ of future observations will fall below with probability $\gamma = 90\%$ ($\gamma = 99.5\%$). To interpret these results, identify a β -content level on the x-axis and read the body-burden tolerance limit estimates from the y-axis. For example, with probability $\gamma = 90\%$ at least $\beta = 95\%$ of future radiocesium 137 observations will fall below 55.477 pCi/kg, and the interval (25.004, 85.949) forms a 95% confidence band to capture model uncertainty (these points are indicated by arrows in Figure 3.9). Similarly, as γ increases to 99.5% at most 5% of future measurements will register above 63.438 pCi/kg, and 19 times out of 20 this value will fall in the interval (38.545, 88.332).

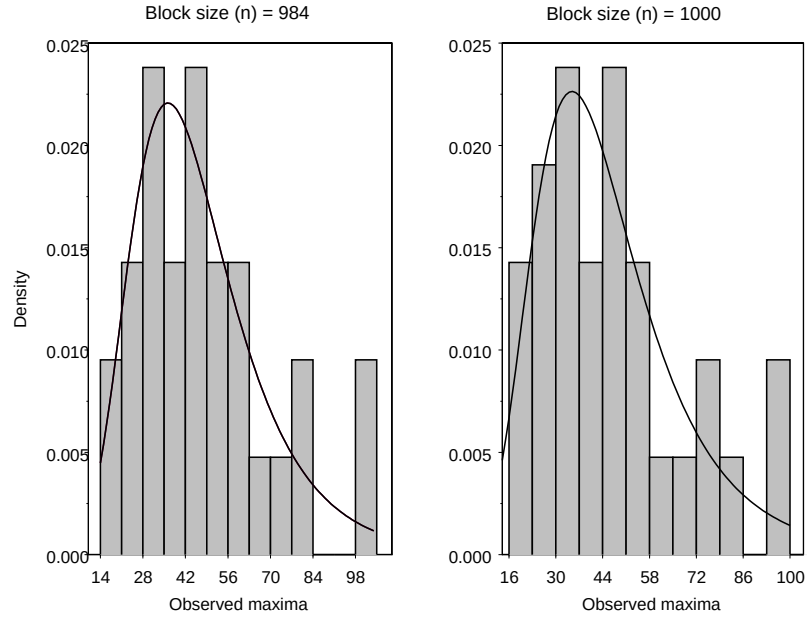


Figure 3.1: Comparison of GEVD densities for different block sizes in the SREL body-burden data. For $n = 984$ (left), parameter estimates are $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (36.74, 16.676, 0.0322)$. For $n = 1000$ (right), estimates are $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (35.90, 16.287, 0.0669)$.

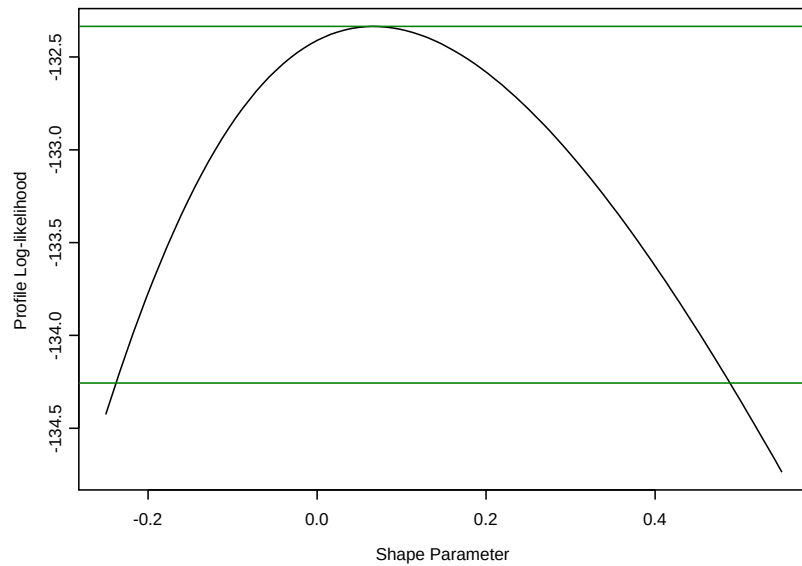


Figure 3.2: Profile likelihood for ξ in the $n = 1000$ Frechet model.

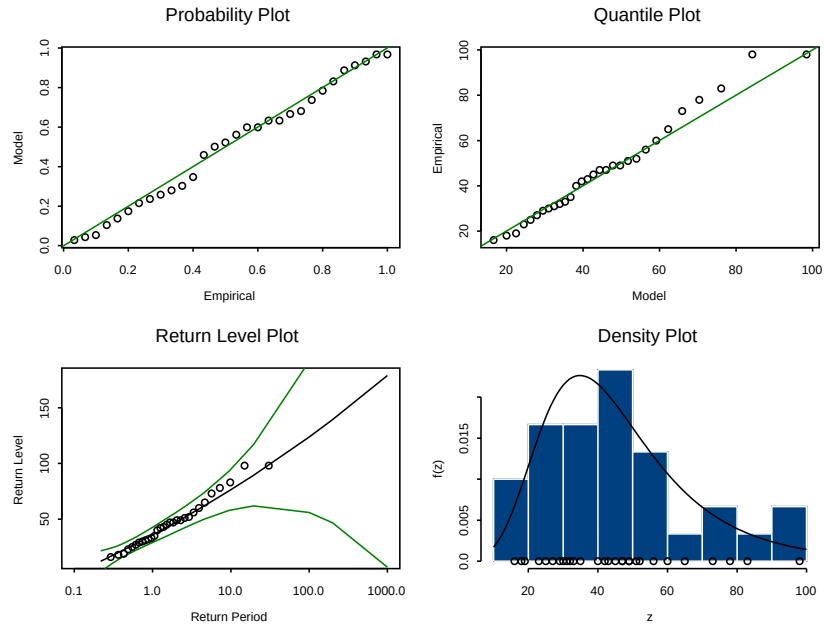


Figure 3.3: Diagnostic plots for GEVD fit to SREL body-burden data ($n = 1000$).

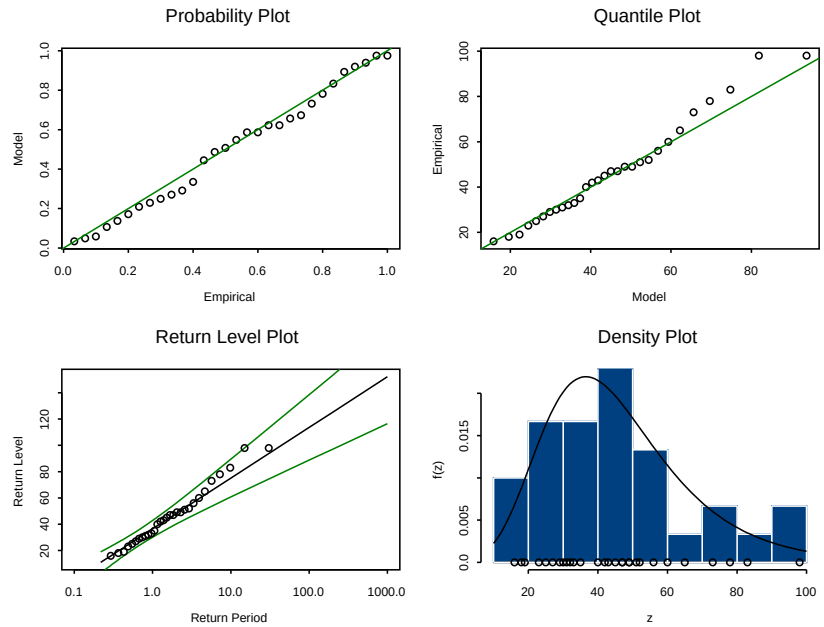


Figure 3.4: Diagnostic plots for Gumbel fit to SREL body-burden data ($n = 1000$).

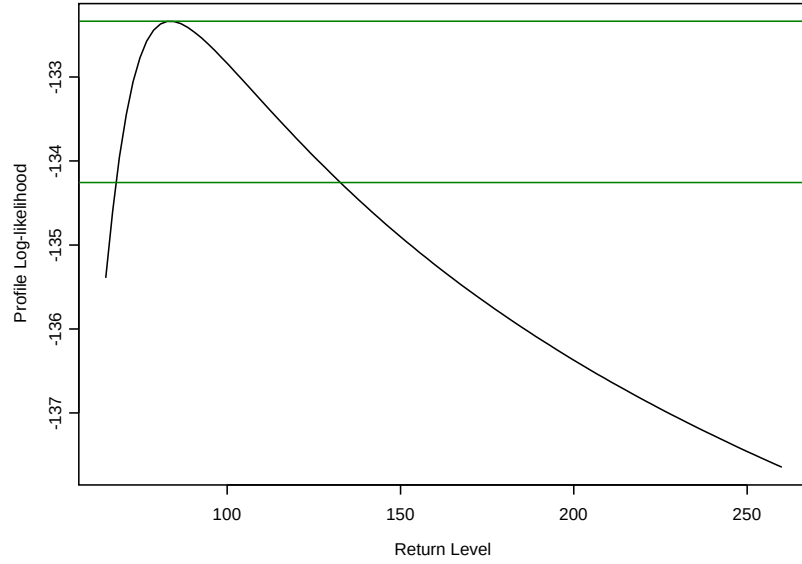


Figure 3.5: Profile likelihood for 15-block return level: $\hat{z}_{p=1/15} = 83.588$ in the $n = 1000$ Frechet model.

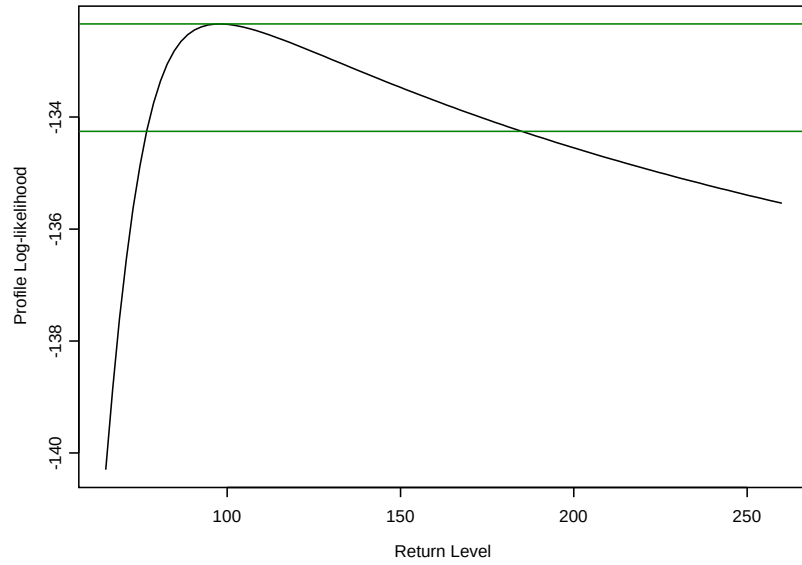


Figure 3.6: Profile likelihood for 30-block return level: $\hat{z}_{p=1/30} = 97.764$ in the $n = 1000$ Frechet model.

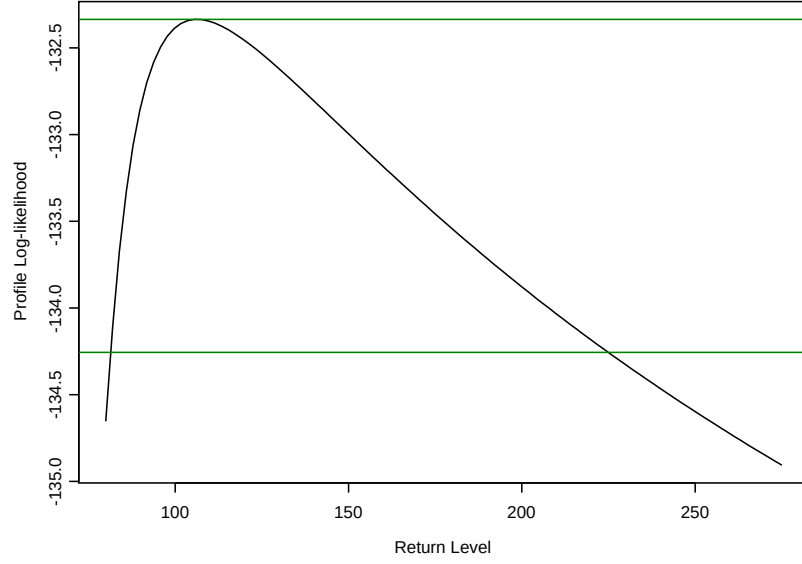


Figure 3.7: Profile likelihood for 45-block return level: $\hat{z}_{p=1/45} = 106.280$ in the $n = 1000$ Frechet model.

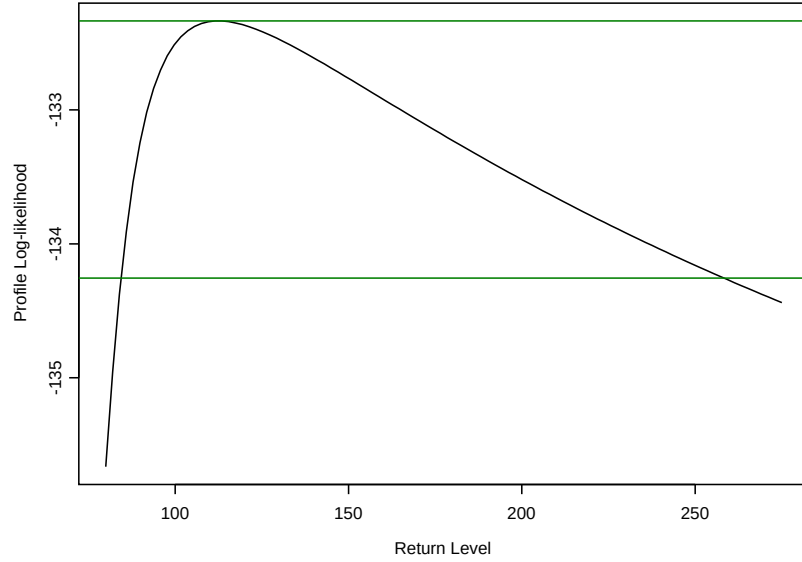
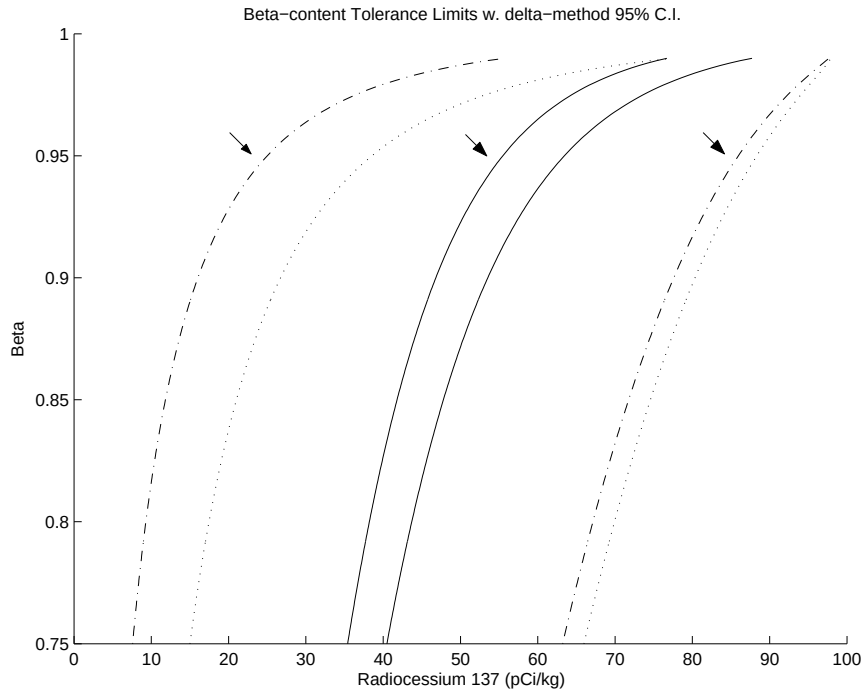


Figure 3.8: Profile likelihood for 60-block return level: $\hat{z}_{p=1/60} = 112.440$ in the $n = 1000$ Frechet model.

Return Period $1/p$	Method	Lower 95% c.i.	$\hat{z}_p$	Upper 95% c.i.
10	Delta	57.846	75.457	93.067
10	Prof lkh	62.881		108.747
15	Delta	60.689	83.588	106.488
15	Prof lkh	68.960		131.813
30	Delta	62.422	97.764	133.107
30	Prof lkh	78.211		184.248
45	Delta	61.620	106.280	150.940
45	Prof lkh	82.899		224.824
60	Delta	60.229	112.440	164.651
60	Prof lkh	85.455		257.315

Table 3.1: Return Level Inference from the $n = 1000$ Frechet modelFigure 3.9: Delta-method confidence intervals for $\gamma = 90\%$ (indicated by arrows) and $\gamma = 99.5\%$ β -content tolerance limits.

CHAPTER 4

FUTURE CONSIDERATIONS

4.1 NEW THEORY ON THE BLOCK

It is clear that our assumption of equal block sizes is one of convenience. While return levels and tolerance limits are expressed in terms of thousands of observations, it is evident from Figure 2.3 that it may take as little as half a hunting season or as many as two hunting seasons to actually collect the next 1000 observations. Suppose that it is known that in the next hunting season the deer harvest from the SRS will be fixed at some level $n_q \neq 1000$ via a system of restricted hunting licenses or imposed quotas. Then it becomes of interest to determine the best way to use the historical data to predict the highest contamination levels observed in the predicted sample of size n_q . An immediate approach is to simply reorganize the data into contiguous blocks of the appropriate size and proceed to fit any EV model. This method is a little labor intensive if several estimated sample sizes are being considered, and is also subject to the bias-variance trade-off.

Instead, the ratio n_q/n may be incorporated in the maximization routine to obtain estimates for the sample size n_q , from the model developed assuming blocks of size n . In fact, Coles' S-Plus functions may be recoded to accept the ratio as an input and provide return-level estimates adjusted for a block size under question that differs from the block size upon which the model is based.

4.2 SREL'S GOOD DATA

The Savannah River Site is a former U.S. Department of Energy nuclear production facility. The continual monitoring of, and data collection from resident wildlife populations provide extensive datasets to assess the effects of radiological concentrations on animal species. In particular, when compared to similar datasets from non-exposed populations, statistically and scientifically significant differences between the populations may be identified and studied. For example, in addition to radiocesium concentration, data is compiled on the age, sex, mass, lactation (for does) and antler points (for bucks) of each harvested deer in the SRS dataset. If a suitable model were found to predict radiocesium concentration from these potential explanatory variables, it could be compared to a similar model fitted to non-exposed populations and any differences highlighted for study.

CHAPTER 5

REFERENCE LIST

- Beirlant, J. et al. (1996). *Practical Analysis of Extreme Values*. Leuven, Belgium: Leuven University Press.
- Castillo, E., Galambos, J. & Sarabia, J.M. (1989). "The Selection of the Domain of Attraction of an Extreme Value Distribution from a Set of Data," in *Extreme Value Theory*, eds. J. Husler and R.-D. Reiss, New York: Springer-Verlag, 181-190.
- Coles, Stuart. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London, G.B.: Springer-Verlag.
- Dasgupta, R & Bhaumik, D. K. (1995). "Upper and lower tolerance limits of atmospheric ozone levels and extreme value distributions," *Sankhya B*, 57, 182-199.
- Dietrich, D., De Haan, L & Husler, J. (2002). "Testing Extreme Value Conditions," *Extremes*, 5:1, 71-85.
- Falk, M. (1989). "Best Attainable Rates of Joint Convergence of Extremes," in *Extreme Value Theory*, eds. J. Husler and R.-D. Reiss, New York: Springer-Verlag, 1-9.
- Falk, M., Husler, J., & Reiss, R.-D. (1994). *Laws of Small Numbers: Extreme and Rare Events*. Boston: Birkhauser Verlag.

- Fisher, R.A. & Tippett, L.H.C. (1928). "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Proc. Camb. Phil. Soc.*, 24, 180-190.
- Hasofer, A. M., & Wang, Z. (1992). "A Test for Extreme Value Domain of Attraction," *JASA*, 87, 171-177.
- Husler, J. & Reiss, R.-D. (Eds.). (1989). *Extreme Value Theory*. New York: Springer-Verlag.
- Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. London: Imperial College Press.
- Pickands, J. III (1975), "Statistical inference using extreme order statistics," *Ann. Statist* 3, 119-131.
- Reiss, R.-D. & Thomas, M. (1997). *Statistical Analysis of Extreme Values (1st ed.)*. Boston: Birkhauser.
- Smith, R.L. (1985). "Maximum likelihood estimation in a class of non-regular cases," *Biometrika* 72, 67-90.
- Stephenson, A. & Tawn. J. (2001). "Bayesian Inference for Extremes: Accounting for the Three Extremal Types," Dept. of Math. & Stats., Lancaster University.
- Wang, Julian Z. (1995). "Selection of the k Largest Order Statistics for the Domain of Attraction of the Gumbel Distribution," *JASA*, 90, 1055-1061.