Chapter 1

Introduction

Extreme events can have powerful impacts on the world. Storms, floods and droughts are familiar environmental examples; consider also stock-market crashes and catastrophic engineering failures. An extreme event can often be characterised in several ways: as a single, extreme value of some quantity; as a cluster of large values occurring in quick succession; as a rare combination of two or more factors; or as a collection of large values occurring simultaneously at several locations. For example, a flood might be caused by rain from a single storm, by consecutive days of high rainfall, by a strong surge combined with high tide, or by high sea-levels at several sites along a coast. There is considerable interest in understanding and predicting the nature of extreme events such as these. With this insight, action can be taken to prevent extreme events, mitigate their effects, or plan for their impacts. Statistical analysis can help by estimating the rates at which extreme events occur, and their critical features.

But the statistical analysis of extremes is difficult. The amount of data available is usually small and estimates are often required at levels for which there is no data. For example, an estimate might be needed for the probability that a sea-wall is breached although the data record might contain no instance of this happening. A successful statistical analysis must use data efficiently and base extrapolations on justifiable theory. This is the goal of extreme-value theory.

This thesis aims to use extreme-value theory to develop statistical methods for making inferences about extreme events occurring in univariate and multivariate stationary processes. Extreme-value theory is well developed for stationary processes and a review of key results under serial independence and serial dependence is provided in Chapters 2 and 3. This work provides characterisations of extreme events under weak conditions on the underlying process. For example, clusters of large values occur according to Poisson processes in time, the maximum value in a cluster has a distribution that belongs to a certain class, and clusters are independent and identically distributed. For practical application, clusters must be identified in the data, and the distribution of the maximum and other features of clusters must be estimated. These are the principal topics discussed in this thesis.

Cluster identification, known as declustering, is currently a fairly arbitrary procedure that is not governed by any theoretical considerations. Moreover, uncertainty in the identification is not accounted for in any formal way. The declustering scheme introduced in Chapter 4 is justified by theoretical developments and supports a technique for estimating the effect of uncertainty on subsequent analyses.

The maximum value in a cluster from a univariate process has a distribution that belongs to a finite-parameter family, for which estimation is now standard. For multivariate processes, the largest value in a cluster, where 'largest' has a specific meaning, has a distribution that must satisfy certain criteria but does not belong to a finite-parameter family. Non-parametric estimators and parametric models are both in use, but neither approach is ideal: the former estimates are not differentiable; the latter are restricted. A semi-parametric estimator that is differentiable and flexible is introduced in Chapter 5.

Other features of clusters admit simple empirical estimates once clusters have been identified. These estimates rely on clusters being independent and identically distributed. Theoretical results highlight additional structure in clusters that is not exploited by these simple estimators, however, and estimators that do exploit this information are introduced in Chapter 6. An advantage of the new estimators is their ability to estimate features of a cluster at levels outwith the data. One application of this property is estimating the probability that a multivariate observation will fall in some rare 'failure region'. Such probabilities are currently estimated using the distribution of the cluster maxima, but this can lead to over-estimation. A method that avoids this bias is introduced in Chapter 7.

The theory and methods reviewed in Chapter 3 for stationary processes appear with related material and in more detail in Ferro and Segers (2003b). Most of Section 4.1 appears in Ferro and Segers (2003a).