Statistical Methods for Clusters of Extreme Values

Christopher A. T. Ferro, B.Sc. M.Sc.

Submitted for the degree of Doctor of Philosophy

at Lancaster University, September 2003.

I declare that the work presented in this thesis is my own, except where stated otherwise.

Christopher Ferro

21 September, 2003

Statistical Methods for Clusters of Extreme Values

Christopher A. T. Ferro, B.Sc. M.Sc.

Submitted for the degree of Doctor of Philosophy

at Lancaster University, September 2003.

Abstract

Extreme values in sequences of independent random variables tend to occur in isolation; for sequences of serially dependent variables, extremes can occur in clusters. Extreme-value theory is well developed for stationary processes and provides mathematical characterisations of the clustering of extremes. Such characterisations are useful models for the extremal behaviour of physical processes: consider storms, floods and droughts for example. Statistical applications harness extreme-value theory to make inferences about the extremes of a process based on a finite sample of data. This thesis addresses several topics in the analysis of clusters of extreme values from both univariate and multivariate processes. Key developments are the following: a theoretically justified scheme for identifying clusters in a sample; estimators for the extremal index that do not require clusters to be identified; a semi-parametric estimator for multivariate extreme-value densities; estimators for cluster summaries that exploit the asymptotic structure of clusters; and a method for modelling clusters in multivariate processes.

Acknowledgements

The work presented in this thesis was conducted under the expert supervision of Professor Jonathan Tawn. The generosity with which he gave of his time, ideas and encouragement is remarkable and it is a pleasure to record my enduring appreciation. I was fortunate also to benefit from a lengthy and very enjoyable collaboration with Dr Johan Segers, now at Tilburg University. I am grateful for the opportunities afforded me by the Engineering and Physical Sciences Research Council, which funded my period of study at Lancaster University, and for the understanding shown by the Department of Meteorology at Reading University, in particular Dr David Stephenson, who granted me time to complete this thesis. Many other friends and colleagues made their own, multifarious contributions; I hope that they will forgive me for mentioning only Omiros Papaspiliopoulos, for providing a timely courier service.

Contents

List of Figures vi

List of Tables ix

1 Introduction **1**

2 Background Theory & Methods I **4**

2.0 Introduction 4

2.1 Univariate Sequences 5

2.2 Multivariate Sequences 9

3 Background Theory & Methods II **16**

3.0 Introduction 16

3.1 Univariate Sequences 16

3.2 Multivariate Sequences 27

4 Declustering & the Extremal Index **31**

4.0 Introduction 31

4.1 Univariate Sequences 32

4.1.1 Inter-exceedance times 35

4.1.2 Moment estimators 41

4.1.3 Least-squares estimators 46

4.1.4 Maximum-likelihood estimators 48

4.1.5 Declustering and bootstrapping 51

4.1.6 Simulation study 53

4.1.7 Data example 58

4.2 Multivariate Sequences 65

4.A Proof of Theorem 4.1 70

4.B Proof of Theorem 4.2 72

4.C Proof of Theorem 4.3 76

4.D Proof of Theorem 4.4 77

5 Modelling Componentwise Maxima **84**

5.0 Introduction 84

5.1 Parametric Estimators 87

5.2 Non-parametric Estimators 89

5.3 Semi-parametric Estimators 91

5.3.1 Asymptotic motivation 92

5.3.2 Kernel estimator 94

5.3.3 Symmetric kernel estimator 97

5.3.4 Unknown component distributions 98

5.3.5 Asymptotic properties 99

5.4 Simulation Study 101

5.5 Data Example 107

5.6 Discussion 112

6 Estimating Cluster Functionals **114**

6.0 Introduction 114

6.1 Clusters & Strings 115

6.2 SWOP Estimators 118

6.3 SWAP Estimators 119

6.3.1 Modelling the string process 121

6.3.2 Modelling the expectation 124

6.4 Multiple Thresholds 128

6.5 Examples 130

6.5.1 Cluster size 130

6.5.2 Cluster excess 133

6.6 Simulation Study 136

6.7 Data Example 141

7 Inference for Multivariate Clusters **143**

7.0 Introduction 143

7.1 Estimating Failure Probabilities 144

7.2 Modelling Multivariate Clusters 147

7.2.1 Block maxima 147

7.2.2 Strings 148

7.2.3 Failure probabilities 150

7.3 Simulation Study 151

7.4 Data Example 154

7.5 Discussion 158

Bibliography 161

List of Figures

4.1 Blocks and runs estimates of the extremal index 34

4.2 Quantile-quantile plot of normalised inter-exceedance times 38

4.3 Max-autoregressive simulations for runs and intervals estimators of the extremal index 56

4.4 Markov chain simulations for runs and intervals estimators of the extremal index 57

4.5 Simulations for runs, intervals and least-squares estimators of the extremal index 59

4.6 Simulations for the indirect estimator of the cluster-size distribution 60

4.7 Simulations for the indirect, least-squares and direct estimators of the cluster-size distribution 61

4.8 Wooster temperature data 62

4.9 Extremal index estimates for the Wooster temperature data 63

4.10 Cluster-size distribution estimates for the Wooster temperature data 64

4.11 Cluster excess and run length estimates for the Wooster temperature data 65

4.12 Estimates of the multivariate extremal index and cluster-size distribution 70

5.1 Dependence functions of three logistic models 102

5.2 Dependence functions of three Dirichlet models 102

5.3 Sheerness-Lowestoft sea-level data 108

5.4 Dependence function estimates for the Lowestoft-Sheerness data . . 110

5.5 Stability plots for the parameters of the kernel estimator 111

6.1 Simulation results for standard, SWOP and SWAP estimators of the extremal index 137

6.2 Simulation results for standard, SWOP and SWAP estimators of the cluster-size distribution 138

6.3 Simulation results for standard, SWOP and SWAP estimators of the cluster excess 139

6.4 Standard, SWOP and SWAP estimates for the Wooster temperature data 142

7.1 Christchurch wave height and surge time series 155

7.2 Christchurch wave height and surge scatter plot 156

7.3 Exploratory plots for surge 157

7.4 Exploratory plots for significant wave height 158

7.5 Diagnostic plots for surge and significant wave height 159

7.6 Block maxima for surge and significant wave height 160

List of Tables

5.1 Simulation results for estimators of A 105

5.2 Simulation results for estimators of A' 106

5.3 Simulation results for estimators of A'' 106

5.4 Simulation results for estimators of A, A' and A'' under independence 107

5.5 Parameter estimates for the Lowestoft-Sheerness data 109

7.1 Coefficients and cluster shapes for the five M_4 processes 152

7.2 Relative bias 154

7.3 Relative standard deviation 154