Extreme Datamining

V.C. Chavez-Demoulin[1], S.A. Jarvis[2], R. Perera[3],
A.S.A. Roehrl[4], S.W. Schmiedl[4] & M.P. Sondergaard[4]

Department of Mathematics, Swiss Federal Institute of Technology[1], DMA,
1015 Lausanne, Switzerland. e-mail: valerie.chavez@epfl.ch
Department of Computer Science, University of Warwick[2], UK
Departamento de Estadistica, Instituto Tecnológico Autónomo de México[3]
Approximity GmbH[4], Loreleystr. 5, 94315 Straubing, Germany

June 15, 2001

Abstract

In recent years there have been a number of developments in the datamining techniques used in the analysis of terrabyte-sized logfiles resulting from Internet-based applications. The information which these datamining techniques provide allow knowledge engineers to rapidly direct business decisions. Current datamining methods however, are generally efficient only in the cases when the information obtained in the logfiles is close to the average. This means that in cases where non-standard logfiles (extreme data) are being studied, these methods provide unrealistic and erroneous results. Non-standard logfiles often have a large bearing on the analysis of web applications, the information which they provide can impact on new or even well established services. In this paper aspects of the recent Extreme Value Theory methodology are discussed. Particular emphasis is made to its application; a unique toolkit is provided with which to describe, understand and predict the non-standard fluctuations as discovered in real-life Internet-sourced log data.

1 Introduction

With the arrival of the Internet and the ability to store vast amounts of information by the millisecond, comes the need for sophisticated tools to examine and interpret this data. It is probable that extreme events form component parts of this data and, by their nature, hold potentially useful business information. Extreme events on the Internet can give rise to huge economic fallouts and for that reason they can determine the success or failure of many web-based applications. The precise estimation of such event probabilities therefore, is essential for the proper managing of web-based resources (including features such as quality of service, request routing and also security).

Analysis of application logfiles makes it is possible to estimate extreme quantiles and therefore measure the probabilities of rare events. These events have particular impact on the following:

1. Load-testing – It is important to be able to estimate target load levels for web-based applications. Estimations are often sought for the overall growth in the amount of site traffic, the peak load level which can occur within the overall traffic, the numbers of users which might ramp-up to that peak load level and how long that peak load level is expected to last.

2. Reliability – The optimal design of new web applications require precise estimations of extreme quantiles of both the operating load and the physical properties of the system architecture.

3. Marketing – Extreme events will effect marketing. Product trends and one off events can have a huge effect on the traffic and the patterns of users. By describing the fluctuations caused by these events and by providing precise estimation of their probabilities, marketing professionals can expose and exploit these events.

4. Finance – Extreme data analysis is crucial in this area. Until recently volatility has been assumed constant in most models, nevertheless new models suggest that this is wrong. By analysing extreme events it should be possible to determine causes of high volatility (which will cause the extreme events) as well as predicting the time at which these events may occur.

Extreme Value Theory (EVT) offers a model in which these factors can be considered. In a number of cases it has provided useful views on aspects of extreme data and allowed the derivation of associate solutions. See Embrechts, Klüppelberg and Mikosch (1997) for detail of the mathematical theory of EVT and for a discussion of its application to financial and insurance risk management[1]. For further information the authors also recommend the edited volume Embrechts (2000), in which various papers highlight the current state of the art on EVT modelling in Integrated Risk Management (IRM).

The traditional approach to EVT is based on extreme value limit distributions. Here, a model for extreme log data is based on the possible parametric form of the limit distributions of maxima over independent, identically distributed (iid) (or weakly dependent) data; Whereas the original data may not be iid, by considering maxima over blocks of data within certain periods, one might hope to reduce the data to uncorrelated maxima observations (for details of this so–called annual maxima method, see Embrechts, Klüppelberg and Mikosch (1997), p. 317). A more flexible model is based on a so–called point process characterisation. The resulting Peaks Over Threshold (POT) method considers exceedances over a threshold u. Mathematical theory (see Davison 2001, Chapter 6) supports the condition of a possibly inhomogeneous Poisson process with intensity λ for the number of exceedances combined with independent excesses over the threshold. Given u, the excesses are treated as a random sample from the generalized Pareto distribution (GPD), with scale parameter σ and shape parameter κ. An advantage of the threshold method over the method of annual maxima is that since each exceedance is associated with a specific event, it is possible to let the scale and shape parameters depend on covariates. For instance, website log data can be of different types; a news website typically may belong to various subclasses (sport, news, finance etc.) and their occurrence shows a non-constant intensity, possibly depending on factors such as hit cycles, exceptional events etc. Logfile sales data from an e-shop like Amazon will typically be a function of product prices, user preferences (clusters), time and other information. Extreme log data may become more or less frequent over time and may become more or less severe. It is also the case that

[1] Other related articles can be found at `www.math.ethz.ch/finance` and `www.risklab.ch`.

in general, they will show cyclic behaviour.

In this paper we discuss some of the more recent EVT methodology which may be useful in handling the presence of such covariates and the resulting modelling of extremal events.

The natural variability of the exceedances tends to mask any trends or other dependence on time. While variation due to the different covariates such as type of customers, type of logfiles or server locations could be summarized parametrically, changes in time need not have a specific parametric form. In this paper we therefore propose to combine the point process for exceedances with smoothing methods to give a flexible exploratory approach to model changes in large values for logfiles data. In doing so, we closely rely on the methodology developed by Chavez-Demoulin (1999), Chavez-Demoulin and Davison (2001) in the environmental context and Chavez-Demoulin and Embrechts (2001) in the financial and insurance context.

A possible model might consist of an inhomogeneous Poisson process for the number of exceedances, with intensity of the form $\lambda(t) = \exp\left\{x^T\alpha + f(t)\right\}$, combined with the generalized Pareto distribution for the sizes of exceedances (the excesses) with a parameterisation of the form $\kappa(t) = x^T\beta + g(t)$ and $\log\sigma(t) = x^T\gamma + s(t)$ where α, β and γ are vectors of parameters and f, g and s are smooth functions (see Section 2 for the basic POT notation). The vector of covariates x can also depend on time, in particular taking into account possible discontinuities in λ, κ and σ, due for example to events such as a worldwide Internet crisis (such as that seen in April 2000), or a stock market crash (such as that of March 2000). Other reasons for discontinuous effects include one-off events such as (in a sporting context) the Olympic games or the football world championship.

Problems can arise when statistically identifying the functions g and s. These can be avoided by working with so–called orthogonal parameters. We might use either the re-parameterisation $\{\kappa, \nu(\kappa,\sigma)\}$ such that the parameters κ and ν are orthogonal with respect to the Fisher information metric or the re-parameterisation $\{\zeta(\kappa,\sigma),\sigma\}$ such that the parameters ζ and σ are orthogonal. As κ is hard to estimate and physically more stable than σ, we prefer to use the parameterisation (κ,ν). Following the orthogonalisation technique described in Cox and Reid (1987), we find the parameter $\nu = \sigma(1+\kappa)$ to be orthogonal to κ. Below, we use $\nu(t) = \exp\left\{x^T\eta + s(t)\right\}$. Whatever statistical estimation method we use, we are faced with a mixture of a finite dimensional problem (parameters α, β, η) and an infinite dimensional one (functions f, g, s). In order to handle the latter, some smoothness assumptions must typically be made. Estimation algorithms carry a penalty component which is a function of the amount of smoothness required for the functions f, g, s. We could also restrict these functions to finitely parameterized classes of functions; we prefer however to allow the data to define this crucial time dependence and hence provide a general, versatile model. The construction of such a model requires semi-parametric techniques. Having observed $w_1, \ldots, w_n$, we might estimate α, β, η, f, g and s using maximum likelihood estimation based on penalized log-likelihood criteria. A motivation for the use of a procedure based on penalized log-likelihood is that it treats the entire dataset as a single entity. We use a Fisher scoring algorithm which has a clear justification through the penalized log-likelihood and furthermore allows for the incorporation of different smoothing methods.

The paper is organized as follows: In Section 2 the stochastic techniques underlying the threshold (POT) method are reviewed. In Section 3 a smoothing methodology is proposed, this provides a new tool for practical extreme value exploration of Internet-style log data. The new method is applied and the result are documented in Section 4.

2 The Threshold Method

The approach based on the threshold method considers a characterisation of all observations which are extreme in the sense of having exceeded a high threshold u. Consider a sequence of independent and identically distributed random variables $Z_1, \ldots, Z_q$ from a distribution $F(z)$ in a wide class of continuous distribution functions. The number of exceedances over the level u has a Poisson distribution with mean λ and conditional on n exceedances, the excesses $W_j = Z_j - u$ are a random sample of size n from the generalized Pareto distribution (GPD)

$$G_{\kappa,\sigma}(w) = \begin{cases} 1-(1-\kappa w/\sigma)_+^{1/\kappa}, & \kappa \neq 0, \\ 1-\exp(-w/\sigma)\,, & \kappa = 0\,. \end{cases} \tag{1}$$

As $\kappa \to 0$, $G_{\kappa,\sigma}(w)$ tends to the exponential distribution with mean σ. Equation (1) can be used as the basis for a likelihood for σ and κ which is

$$l(\sigma,\kappa) \doteq -n\log\sigma - (1-1/\kappa)\sum_{j=1}^{n}\log\left(1-\kappa w_j/\sigma\right)_+ ,$$

and the Poisson Process log-likelihood in term of λ, σ, κ is then

$$l(\lambda,\sigma,\kappa) \doteq n\log\lambda - \lambda - n\log\sigma - (1-1/\kappa)\sum_{j=1}^{n}\log(1-\kappa w_j/\sigma)_+\,. \tag{2}$$

In deriving (2), the (asymptotic) independence of the frequency and sizes of the losses over a high threshold u are used. Maximum likelihood estimation of the parameters κ and σ of a generalized Pareto random variable is non–regular in the sense that the score statistic is not asymptotically normal if $\kappa > 1/2$ (Davison (1984a, 1984b), Smith (1985)). For $\kappa > 1$ the GPD has infinite mean and so, whereas the usual Taylor expansions can be made, they do not yield a consistent estimator. Typically, in most applications the value of κ is close to zero and consistency and asymptotic efficiency of the maximum likelihood estimator hold. The generalized Pareto distributions yield a practical family for statistical estimation, provided that the threshold is taken sufficiently high.

The choice of the threshold is important. Smith (1987) proposes a graphical technique to get an aid for choosing the threshold and to assess the fit of the model; a "mean residual life plot" (see Yang (1978)) in which the mean excess over a threshold u is plotted against u, for a wide range of values u. See Davison and Smith (1990) and Embrechts, Klüppelberg and Mikosch (1997) for an extensive discussion of this approach.

The level exceeded on average once in $1/p$ years (or any other relevant time period), called the $1/p$–year return level, is often a quantity of interest. Based on the threshold model it's value is

$$y_{1-p} = u - \frac{\sigma}{\kappa}\left\{(\lambda/p)^{-\kappa} - 1\right\}\,. \tag{3}$$

which may be estimated by replacing σ, κ and λ by their maximum likelihood estimates. Interval estimates may be obtained by the delta method or by a re-parameterisation in terms of $(y_{1-p}, \lambda, \kappa)$, treating κ and λ as nuisance parameters, and solving (3) for σ. This method is also referred to as the profile likelihood approach.

Independence of widely separated extremes seems reasonable in most applications, but they almost always display short–range dependence in which clusters of extremes occur together. Serial dependence will typically imply clustering of large values: for example, high sales of a bestseller book tend to occur together and maxima visits to the sports pages of a news website occur during the Olympic games. In these cases, it seems unrealistic to assume independence within each period (some weeks). In the threshold method, the usual solution is to fit the point process model to cluster maxima, as the use of the GPD for the peak excess in each cluster is justified. An important practical problem is the identification of clusters from data, provided that the cluster size is random and its distribution depends on the local correlation of the Z_i. The identification of clusters has been influenced by earlier work including Leadbetter *et al.* (1983), and is a topic of much current research. Following Davison and Smith (1990), Robinson and Tawn (2000) propose a run approach for both the choice of a suitable high threshold and of a method to identify independent clusters, for more details and alternative estimation procedures see Embrechts, Klüppelberg and Mikosch (1997, Section 8.1).

3 A Methodology for Extreme Datamining

We present a methodology which applies smoothing techniques to extreme values in the framework of Internet logfiles. In this context there is a need for exploratory data analysis to gain an understanding of the structure of the data. The goal of the proposed method is to offer more advanced exploratory data analysis tools for website management (marketing and operational) in the presence of extremal events. Theoretical results and details of the approach are described in Chavez-Demoulin (1999) and Chavez-Demoulin and Davison (2001).

We see below that the modelling of exceedance times in terms of a Poisson process combined with independent excesses over a threshold facilitates the procedure. It does so by permitting separate modelling of the number of exceedances and their sizes. Moreover, when modelling the two parameters of a GPD, it is straightforward to reparameterize the problem using orthogonal parameters. Chavez-Demoulin (1999) gives an example where a non–orthogonal parameterisation leads to computational difficulties and non–convergence of the algorithm. This could arise when smoothing is performed directly on the three parameters of the generalized extreme value distribution (GEV) $H_\kappa\left(\frac{x-\theta_1}{\theta_2}\right)$ where

$$H_\kappa(x) = \begin{cases} \exp\left\{-(1-\kappa x)_+^{1/\kappa}\right\}, & \kappa \neq 0\,, \\ \exp\{-\exp(-x)\}\,, & \kappa = 0\,, \end{cases}$$

for which it is difficult to find an orthogonal reparameterisation. For the link between the GPD and the GEV, see Embrechts, Klüppelberg and Mikosch (1997, Section 3.4).

Suppose a sequence of excesses results from the threshold method for which we recall the basic properties; for detail see Leadbetter (1991):

i) the excesses over a high threshold u occurs at the times of a Poisson process with intensity λ;

ii) the corresponding sizes over u are independent and have a GPD(κ, σ) distribution;

iii) exceedance sizes and exceedance times are independent of each other.

The way in which i), ii), and iii) are formulated allows for an immediate application to the construction of likelihood functions. Hence the resulting process is of the so–called compound Poisson type. For this process, the overall log-likelihood in terms of the parameters λ, κ, σ is

$$l(\lambda,\kappa,\sigma) = \sum_{i=1}^{T}\left[n_i \log\lambda_i - \lambda_i - n_i\log\sigma_i - (1/\kappa_i + 1)\sum_{j=1}^{n_i}\log\{1+\kappa_i w_j/\sigma_i\}\right],$$

where T is the observed number of time periods (the number of years, say), n_i is the number of exceedances during the ith year, and $w_k = z_k - u$, $k = 1, \ldots, N$, are the excesses, having observed a total number of $N = \sum n_i$ data $z_1, \ldots, z_N$ over a threshold u. Due to the following "cut" of the log-likelihood

$$l(\lambda,\kappa,\sigma) = l(\lambda) + l(\kappa,\sigma)\,,$$

estimation can be performed separately for the point process of exceedance times and for the excesses. Based on Leadbetter's results, semi-parametric models for exceedance times and excesses can be proposed.

Assuming an inhomogeneous Poisson process for the exceedance times, with intensity $\lambda(t) = \exp\left\{x^T\alpha + f(t)\right\}$, the point process part defines a semi-parametric generalized linear model within the Poisson family. Following the approach of Green and Yandell (1985), we use the Fisher scoring algorithm to maximize the penalized log-likelihood

$$l(\lambda) - \frac{1}{2}\gamma_\lambda \int f''(t)^2\, dt\,,$$

where γ_λ is a smoothing parameter. Estimation and inference procedures are customary for the exponential family (Green and Silverman, 1994).

Consider the semi-parametric generalized Pareto model. Our aim is to fit $\kappa(t) = x^T\beta + g(t)$, $\nu(t) = \exp\{x^T\eta + s(t)\}$, and for that reason, attempt to maximize the penalized log-likelihood

$$l(\kappa, \nu) - \frac{1}{2}\gamma_\kappa \int_a^b g''(t)^2\, dt - \frac{1}{2}\gamma_\nu \int_a^b s''(t)^2\, dt\,, \tag{4}$$

where γ_κ and γ_ν control the degree of smoothing applied to the shape and scale parameters respectively. Here,

$$l(\kappa, \nu) = \sum_{i=1}^{T} \left[n_i \log(1+\kappa_i) + n_i \log \nu_i - (1/\kappa_i + 1) \sum_{j=1}^{n_i} \log\left\{1 + \kappa_i\left(1+\kappa_i\right) w_j \nu_i\right\} \right].$$

Details of the estimation process are given in Chavez-Demoulin (1999). A procedure can be implemented to estimate simultaneously the parameters κ and ν. Inference in semi-parametric models for the GPD uses results for semi-parametric generalized linear models or generalized additive models.

For the non-parametric component, we justify the use of simultaneous tests based on deviances for the two models by the orthogonality of the parameters κ and ν. That is, applications show that deviance test results for one model do not depend on the other parameter model form. Furthermore, for smoothing parameter selection we recommend the use of criteria such as AIC; see McQuarrie and Tsai (1998). The behaviour of AIC curves for one parameter remains unchanged by a modification of the smooth function for the other parameter.

Use of appropriate degrees of freedom in conjunction with deviances allows for the assessment of model adequacy. A possible general approach to assess uncertainty for semi-parametric GPD estimates is to use bootstrap methods. A bootstrapping strategy is based on the result that, given the model is correct, the residuals

$$R_j = \hat{\kappa}_j^{-1} \log \{1 + \hat{\kappa}_j W_j (1 + \hat{\kappa}_j) \hat{\nu}_j\}, \qquad j = 1, \ldots, N, \tag{5}$$

are distributed approximately as independent, unit exponential random variables. We define simulated responses by

$$W_j^* = \frac{1 + \exp\left\{\hat{\kappa}_j \epsilon_j^*\right\}}{\hat{\kappa}_j (1 + \hat{\kappa}_j) \hat{\nu}_j}, \qquad j = 1, \ldots, N, \tag{6}$$

where $\epsilon_1^*, \ldots, \epsilon_N^*$ is a random sample from the residuals R_j defined in (5). This leads to basic bootstrap confidence intervals for $\hat{\kappa}$ and $\hat{\nu}$.

To summarize, the following general guidelines should be considered when applying the methodology:

a) Decide upon a model form for each parameter λ, κ and ν. In practice the Poisson process for the exceedance times is not necessarily homogeneous and hence a smoothing of $\lambda = \lambda(t)$ over t is appropriate. Additionally, the parameter κ which controls the weight of the tail of the extremal distribution is generally hard to estimate and a fully parametric form $\kappa = x^T\beta$ often suffices. The scale parameter σ (and hence ν) may vary much more with external explanatory variables.

b) The choice of the smoothing parameters depends on the aim of the analysis. If the aim is purely exploratory, the smoothers can be modified to suit the situation by changing the degrees of freedom. If an automatic procedure for selecting the smoothing parameters is required, we recommend the use of criteria such as AIC rather than cross–validation, which becomes computationally costly when the size of the data increases.

c) Informal inference based on deviance tests is also useful to assess models and their differences separately for each parameter. Uncertainty about the parameter estimates is assessed by constructing confidence intervals. For the non-parametric component, plots of pointwise confidence bands around the fitted curve yield useful visual tools. Procedures based on residuals are also useful

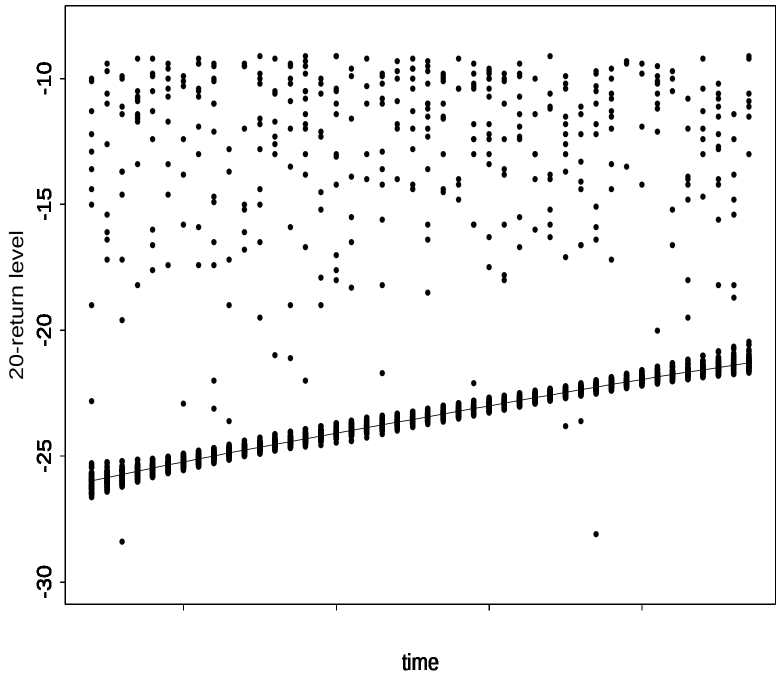

Figure 1: *A simulation study. 20-return level estimates from 100 simulated samples (points around the solid line) and minimal exceedances (points). The solid line represents the "true" 20-return levels.*

to assess the goodness–of–fit of the model. A possible graphical diagnostic is based on the result that the residuals (5) are distributed approximately as independent unit exponential variables when the model is correct.

For details on accuracy of the method see Chavez-Demoulin, 1999 where a number of simulations are presented. Figure 1 provides an example of the simulation results showing 100 estimated 20-return levels for minimal exceedances in an environmental context. The solid lines representing the "true" 20-return level, providing credible values for the bias and variance which one can expect when estimating the GPD parameters.

4 An Application

The capability of the extreme value methodology is considered. We present (in Figure 2) the changing quantitative behaviour in extreme observations above some given threshold. A typical data set might include the number of hits of the k pages of a website, each recorded as an excess above some threshold value u_i,

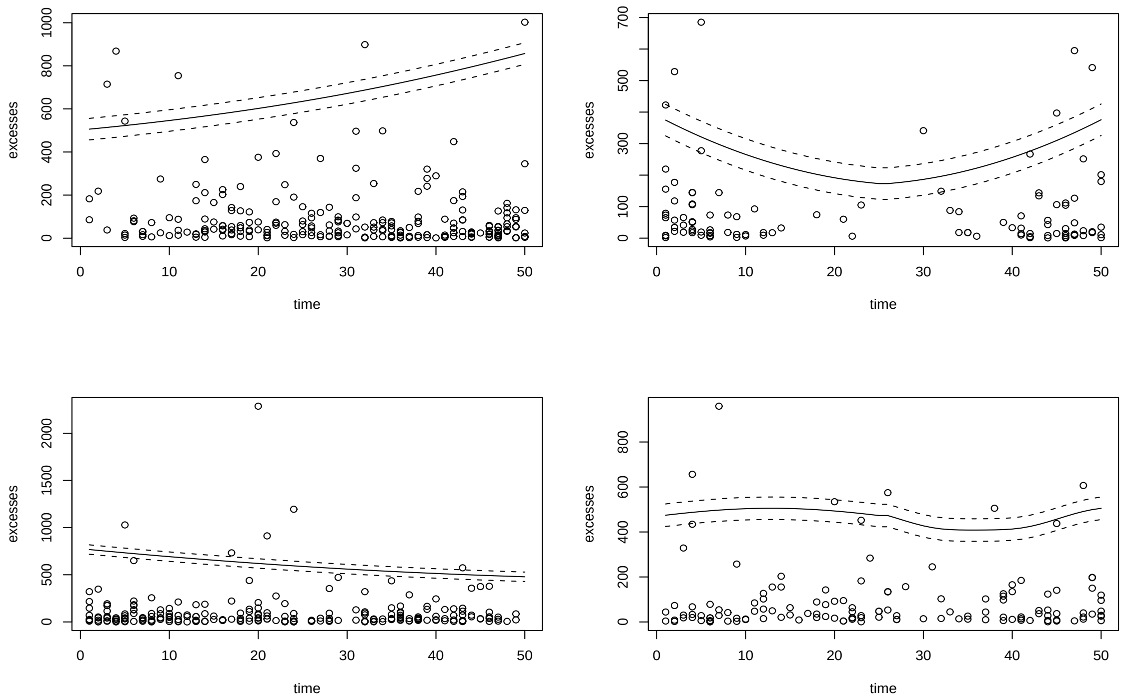

Figure 2: *e-shop data. The points are the values of the excesses of sales (big sales exceeding a certain limit minus the threshold) for four types of product of a well-known e-shop against time (50 weeks). The straight lines are the 8-weeks return levels and their 90% confidence interval (dotted lines).*

$i = 1, \ldots, k$. One could also consider the e-shop value of sales over a given time period for k different product types, or indeed losses above given retention limits (thresholds). In each case, a measure of the loading risk or risk can be given; the $1/p$ return level is then considered. Objectively, we want to model y_{1-p} as a function of time: i.e. is y_{1-p} constant or changing in time, and if the latter is the case, how does y_{1-p} change with time? We might also consider whether y_{1-p} depend on the page or type of product which the site presents?

As an example, Figure 2 represents the sales (in US dollars) across $k = 4$ types of product of a well-known website (after some preprocessing) over a period of 50 weeks. The points correspond to the values of the excess of sales, that is, the large sales figures exceeding a certain limit minus the threshold. For each product, we chose a non-parametric threshold form such that about 5% of the original data are excesses. Following the methodology developed in the previous section, we fit different models for λ, κ and ν and compare them using tests based on the likelihood ratio statistics. After an extreme value analysis, we obtain the estimated quantile of interest as a function of time for each product type. The figure shows (in straight lines) the 8-weeks return level, that is the level crossed on average once in 8 weeks. The dotted lines are the corresponding 90%-bootstrap confidence intervals based on the residuals.

There are a number of ways in which these results can be used.

- Marketing - Understanding the extreme values which the datamining identifies allows marketing strategies to highlight and respond to future extreme periods. This might include seasons, campaigns, geography, time-zones and product segments etc.
- Managing Extreme Values - It is important to understand the impact of extreme values on the business. For example, the change in corresponding productivity of other business areas during a period of extreme sales. This will be beneficial to choosing the right fit between strategy, customer expectations and the capabilities of the organisation.
- Security – By performing multi-scale analysis on a number of different time-frames it is possible to quickly and effectively detect *denial of service attacks* (automatic page requesting at a rate which is designed to crash the server). This analysis is possible both at the packet level and also at a level higher in the protocol stack.
- Future Planning - EVT provides the skilled strategist with a tool for making future assumptions and for deducing the strengths and opportunities of the organisation at some point in time. The identification of previous extreme values allows greater certainty in price positioning and the selection of entire value propositions.

5 Conclusion

Popular web services are now generating huge quantities of log data. The example e-shop application documented in this paper currently produces approximately 5 gigabytes of logfile-data per hour. Results on this scale are difficult to analyse with conventional tools; it is also the case that current datamining methods are generally only efficient in the cases when the information obtained in the logfiles is close to the average. Studying non-standard logfiles (extreme data) can provide important additional information to a business.

A new model for extreme datamining is presented. This is based on the recent extreme value theory methodology. Our motivation for the use of a semi-parametric procedure based on penalized log-likelihood is that it treats the entire dataset as a single entity. This provides a natural approach to the analysis of Internet-sourced log data. We provide a supporting tool-kit based on this model with which a large real-life case study is tested and documented.

The identification of extreme values can have a large bearing on web-based applications. Extreme datamining offers precision to marketing strategies, allows the management of extreme values, provides a mechanism for monitoring security and a basis on which future business plans can be honed.

References

Chavez-Demoulin, V. (1999) *Two Problems in Environmental Statistics: Capture-Recapture Analysis and Smooth Extremal models.* Ph.D. thesis. Department of Mathematics, Swiss Federal Institute of Technology, Lausanne.

Chavez-Demoulin V. and Davison A. C. (2001) Generalized Additive Models for Sample Extremes. *Submitted to Applied Statistics.*

Chavez-Demoulin V. and Embrechts P. (2001) Smooth Extremal Models in Finance *In preparation*

Cox, D. R., Reid, N. (1987) Parameter orthogonality and approximate conditional inference (with Discussion). *Journal of the Royal Statistical Society, B,* **49**, 1–39.

Davison, A. C. (1984a) *A statistical model for contamination due to long-range atmospheric transport of radionuclides.* Ph.D. Thesis. Department of Mathematics, Imperial College of Science and Technology, London.

Davison, A. C. (1984b) Modelling excesses over high thresholds, with an application. *Statistical Extremes and Applications.* Ed. J. Tiago de Oliveira, 461–482.

Davison, A. C. (2001) Statistical Models. Cambridge University Press. To appear.

Davison, A. C., Smith, R. L. (1990) Models for exceedances over high thresholds (with Discussion). *Journal of the Royal Statistical Society, B*, **52**, 393–442.

Embrechts, P. (Ed.) (2000) *Extremes and Integrated Risk Management.* Risk Books, Risk Waters Group, London.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Berlin: Springer.

Green, P. J., Silverman, B. W. (1994) *Non-parametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* Chapman and Hall, London.

Green, P. J., Yandell, B. (1985) Semi-parametric generalized linear models. *Proceedings 2nd International GLIM Conference.* Lancaster, Lecture Notes in Statistics, **32**, 44–55. New York: Springer-Verlag.

Leadbetter, M.R. (1991) On a basis for 'peaks over threshold' modelling. *Statist. Probab. Letters* **12**, 357–362.

McQuarrie, A. D. R., Tsai, C.-L. (1998) *Regression and Time Series Model Selection.* Singapore: World Scientific.

Robinson, M. E., Tawn, J. A. (2000) Extremal analysis of processes sampled at different frequencies. *Journal of the Royal Statistical Society, B* **62**, 117-136.

Smith, R. L. (1985) Maximum likelihood estimation in a class of non-regular cases. *Biometrika,* **72**, 67–90.

Smith, R. L. (1987) Estimating tails of probability distributions. *Annals of Statistics*, **15**, 1174–1207.

Yang, G. L. (1978) Estimation of a biometric function. *Annals of Statistics*, **6**, 112–116.