한국통계학회논문집 제3권 2호
1996년 8월 pp. 249-273

Modeling Extreme Values of Ground-Level Ozone Based on Threshold Methods for Markov Chains

Seokhoon Yun[1)]

Abstract

This paper reviews and develops several statistical models for extreme values, based on threshold methodology. Extreme values of a time series are modeled in terms of tails which are defined as truncated forms of original variables, and Markov property is imposed on the tails. Tails of the generalized extreme value distribution and a multivariate extreme value distribution are proposed to be used as models for the marginal and the joint distributions, respectively, of the tails of the series. These models are then applied to real ozone data series collected in the Chicago area. A major concern is given to detecting any possible trend in the extreme values.

1. Introduction

Extreme value analysis plays an important role in environmental time series where extreme values are major concerns. In this paper we review and develop some statistical models based on the so-called threshold method which has nowadays become a standard one in analyzing extreme values. Threshold method is a statistical method based on exceedances over a high threshold. Extension to Markov chain models is also studied to take serial dependence of extreme values into account. The models are then applied to the ground-level ozone data collected in the Chicago area. Since excessive levels of ozone are generally taken as an indication of high air pollution, our prime concern is to detect any possible trend, if exists, in the extreme values. The questions to be answered contain estimation problems for the exceedance probability of a high threshold and 100-year return levels.

Classical extreme value analysis is based on the generalized extreme value distribution defined by

$$G(x;\mu,\sigma,\xi)=\exp\left[\ -\left\{1+\frac{\xi(x-\mu)}{\sigma}\right\}_{+}^{-1/\xi}\right]\ ,\ x\in\mathfrak{R}, \tag{1.1}$$

which typically appears as the limit of distributions for normalized maxima in i.i.d. random

1) Full-time Lecturer, Department of Applied Statistics, University of Suwon, Suwon, Kyonggi-do, 445-743, Korea.

variables. Here $x_+=\max\{x,0\}$, and $\mu \in \mathfrak{R}$, $\sigma > 0$ and $\xi \in \mathfrak{R}$ are called location, scale and shape parameters respectively. The case $\xi=0$ is interpreted as the limit $\xi \to 0$, which is often called the Gumbel type. The cases $\xi > 0$ and $\xi < 0$ are also called the Fréchet and Weibull types respectively. Distribution (1.1) was originally introduced by Fisher and Tippett (1928). For a complete derivation of (1.1), the reader is referred to Leadbetter, Lindgren and Rootzén (1983) or Resnick (1987). Statistical application, for instance, is to fit (1.1) to annual maxima of a given time series. This method however requires that the series should be long enough to get a good fit.

A modern approach to extreme value analysis is based on the threshold method which adopts the generalized Pareto distribution defined by

$$H(y;\phi,\xi)=1-\left(1+\frac{\xi y}{\phi}\right)_+^{-1/\xi}, \quad y>0 \tag{1.2}$$

as the model for excesses, the magnitudes of exceedances, over a high threshold. Unlike (1.1), distribution (1.2), which was derived by Pickands (1975), has only two parameters, scale $\phi > 0$ and shape $\xi \in \mathfrak{R}$. The case $\xi=0$ is, as before, interpreted as the limit $\xi \to 0$, which corresponds to the exponential distribution with mean ϕ. For statistical application, see Smith (1984), Hosking and Wallis (1987) and Davison and Smith (1990). Naive application is to fit (1.2) to all excesses over a high threshold assuming they are independent. Under stationarity of observations, Hsing (1987) and Hsing, Hüsler and Leadbetter (1988) showed that high-level exceedances tend to form independent clusters. One may therefore apply (1.2) to cluster maxima instead of all exceedances over a high threshold. In practice, it is not however easy to characterize clusters and it is moreover wasteful to use only the cluster maxima and discard most of the exceedance data which may contain valuable information.

The first attempt to compensate for this was given by Joe, Smith and Weissman (1992), and extended to Markov chains by Smith, Tawn and Coles (1993) and Smith (1993). They derived an appropriate model for the joint distribution of all exceedances within a cluster using the theory of domains of attraction of multivariate extreme value distributions. Galambos (1987) and Resnick (1987) have a good review on this theory. This paper takes the form of a case study, yet proposing and developing variants of existing models in threshold methodology. Specifically, we model extreme values of a given time series in terms of tails of that series and then impose Markov property on the tails. Tails of a time series are defined as truncated forms of original variables, which have all the information about excesses over a high threshold as well as its exceedance probability. Tails of distribution (1.1) and a multivariate extreme value distribution are proposed to be used as models for the marginal and the joint distributions, respectively, of the tails of the series. The proposed models will be tested through actual analysis of the ozone data.

The rest of the paper is organized as follows. Section 2 introduces some recent reports on ozone. Section 3 describes the ozone data we are going to analyze. In Section 4, we review and develop several statistical models in threshold methodology. Section 5 explains a detailed statistical procedure for applying the models discussed in Section 4 to the ozone data. The actual data analysis is performed in Section 6. Section 7 gives estimated exceedance probabilities as well as estimated 100-year return levels. Section 8 contains summary and final comments.

2. Some recent reports on ozone

Ground-level ozone is produced by complex chemical reactions driven by solar radiation as a result of the emissions of hydrocarbons and nitrogen oxides into the atmosphere. As a measure of air pollution, ozone is one of the current important issues of considerable environmental concern. The governmental standard is generally specified in terms of percentiles of the distribution of ozone concentrations. For instance, the U.S. air pollution standard specifies that the number of exceedances by daily ozone maxima of the level of 120 ppb (parts per billion) should not exceed 3 in any 3-year period. The regions which violate this are usually in major cities such as New York, Chicago, Houston and Los Angeles.

There have been a number of ozone studies during this decade. Out of the several recent reports on ozone, we briefly pay attention to the following four papers. First of all, Smith (1989) analyzed the ozone data collected in Houston for 14 years from 1973 to 1986, using the clustering method where the cluster maxima of exceedances are assumed to happen according to a Poisson process and excesses are assumed to follow the generalized Pareto distribution. Here, the year variable was considered as a covariate to examine whether there is a long-term trend in the data. On the other hand, a nonlinear regression technique was used by Graf-Jaccottet (1993) to analyze the average daily ozone concentrations collected in Switzerland for 2 years during 1988 and 1989, who adopted a Box-Cox transformation to take the nonnormality of the data into account. It is noteworthy, here, that the number of hours between sunrise and sunset was introduced as a covariate to model the short-term annual trend. As part of a project carried out at the NISS (National Institute of Statistical Sciences) under a cooperative agreement with the U.S. EPA (Environmental Protection Agency), Bloomfield, Royle and Yang (1993) also developed a nonlinear regression model fitted to weighted averages of ozone concentrations in the Chicago area for 11 years from 1981 to 1991. They used several measured meteorological variables as possible covariates. This data set was again used by Smith and Huang (1993), as part of that project also, where they made an attempt to fit extremal data by a (first order) Markov chain. Here, the meteorological variables were also used as covariates.

The ozone data we are going to analyze is the same data set that was used by Bloomfield et al. (1993) and Smith and Huang (1993). Since we are interested in characterizing what is going on in the tail (or extremes) of the distribution of ozone as the current ozone standard is defined in terms of extremes, we avoid using regression approaches which do not give enough attention to the extremes of the data. We follow, instead, the threshold methodology, as Smith (1989) and Smith and Huang (1993) did, which is accepted as an appropriate technique to capture the tail behavior of ozone. The major concern here is to model the joint distribution of the tails of the given data series using a higher order Markov chain. This is achieved by assuming the domain of attraction of a multivariate extreme value distribution. On the other hand, to simplify the analysis, we do not use explicitly the meteorological variables (such as temperature, wind speed, humidity and so on) as covariates although these are available in the data set. As possible covariates, we use instead the length of the day, the number of hours between sunrise and sunset, to model the short-term annual trend (the seasonal effect) as Graf-Jaccottet (1993) did, and the year itself to examine the long-term trend. This is reasonable since the ozone is produced by chemical reactions driven by solar radiation and so believed to depend on the length of the day. Removing the meteorological covariates from the analysis is particularly useful for the purpose of prediction since meteorological conditions cannot be controlled.

3. The ozone data

The ozone data consist of hourly readings at 45 stations in the Chicago area, collected for 11 years from 1981 to 1991. Detailed description of the stations and their locations is available in Bloomfield et al. (1993). Since we are interested in the tail behavior of ozone, our detailed analysis is concentrated on the three stations, denoted by P, Q and R respectively, with the highest ozone levels. From now on, we therefore focus on three time series of daily ozone maxima from these stations. In addition to these, we also analyze the daily maxima for the whole network, which is referred to as F (or *network maxima*). These four data series are exactly the same as Smith and Huang (1993) used for their detailed analysis.

Table 1 shows some information about these stations, number of observations, number of missing observations due to the measuring equipment being out of service, and the number of exceedances of the daily maxima over the levels 99.5 and 119.5 ppb. Since the U.S. current

Table 1: Numbers of observations and exceedances by station

	P	Q	R	F
station ID	170317002	170971002	181270024	-
latitude	42.1°	42.4°	41.6°	41.8°
no.of nominal obs.	3,956	3,956	2,757	3,866
no.of actual obs.	3,392	3,211	1,053	2,354
no.of missing obs.	564	745	1,704	1,512
no.of exc.over 99.5	94	84	55	321
no.of exc.over 119.5	41	38	23	143

standard is based on 120 ppb, this will be the starting threshold of major concern. The equipment measuring ozone, however, can only measure to the nearest ppb, which means that the actual ozone could be anything between 119.5 and 120.5 if 120 ppb, for example, were measured. Taking this measurement error into account, we start with threshold 119.5 instead of 120, which has an effect of counting 120 as an exceedance. For stations P, Q and R however, the numbers of exceedances over threshold 119.5 are not, as seen in Table 1, believed to be satisfactorily enough for a good fit particularly when complicated models are dealt with. We therefore confine ourselves to only threshold 99.5 for stations P, Q and R, but work with threshold 119.5 for network maxima (F). As introduced before, one of our main covariates is the length of the day, and this can be calculated approximately using the latitude of each station. The explicit formula for this will be given later. For this purpose, the latitude for F should be specified. We use 41.8° , the average latitude of all the 45 stations, for the latitude of network maxima. The number of nominal observations in Table 1 represents the number of days between the first and the last observations of each station. For station R, the observations are almost missing for the first 3 years, for the last year, and for the periods, November~March, of the remaining years. For network maxima, the data are not available either for the period, November~March, of every year.

4. Statistical models in threshold methodology

This section reviews and develops several statistical models in threshold methodology which will be adopted in Section 6 for detailed analysis of the ozone data of each station. The actual statistical estimation procedure is then based on the numerical maximum likelihoods using these models. Major concerns are to model the dependence of extreme values by building up an appropriate model for the joint distribution of the tails of a given time series, detect any possible trend in the extreme values, and predict (or estimate) percentiles such as 100-year return levels using the fitted model. For a somewhat complete analysis of extremes, the statistical analysis will be separated into several steps. The fist step is to consider the exceedance probability of a high threshold. The next is to apply the usual threshold method to exceedances over the threshold, that is, fit the generalized Pareto distribution to excesses. These two steps are then combined to be considered simultaneously by defining truncated variables. Finally, serial dependence is allowed for in the analysis and the tails of a given time series are modeled by a k-th order Markov chain ($k \geq 1$). Throughout the remaining part of the paper, let $\{X_t\}$ denote a given time series of daily ozone maxima of a station, and let u be a fixed high threshold which is either 99.5 or 119.5 in our case. To give a step-by-step motivation for our final model, we start with a simple situation and assume, for the time being, that the observations (daily ozone maxima) are independent of each other.

Exceedance probability of a high threshold

As an initial step, suppose we are interested in modeling the probability of exceedance, as a function of covariates, of the threshold u on each day t. If we consider the indicator variable $Z_t=I(X_t>u)$, then the exceedance probability of u is given by $p_t=P(Z_t=1)=P(X_t>u)$, which is to be estimated. A popular model for p_t is the logit model:

$$\log\left(\frac{p_t}{1-p_t}\right)=\sum_s x_{ts}\beta_s, \tag{4.1}$$

where x_{ts} is the s-th covariate on day t and β_s is the corresponding coefficient. The covariates we are going to use for detailed analysis are given in the following section. Under the independence assumption, the likelihood for $\{Z_t\}$ will be

$$L=\prod_t\{p_t\cdot I(Z_t=1)+(1-p_t)\cdot I(Z_t=0)\}. \tag{4.2}$$

Since $\{Z_t\}$ is a two-state (0 and 1) series, one may easily allow for local dependence between variables by considering it as a higher order two-state Markov chain. This chain, however, does not explain anything about the distribution of the excesses which is one of our major concerns, and therefore we do not pursue this binary Markov chain model here.

Excesses over a high threshold

To see what is happening in the extremes of the data, it is of direct interest to characterize the distribution of excesses over the threshold u. A standard model for this is the generalized Pareto distribution (henceforth GPD) defined by (1.2). This is based on the fact that if a sequence of i.i.d. random variables is given in which their common distribution function is assumed to belong to the domain of attraction of some univariate extreme value distribution, then the excesses over the threshold u follow approximately the GPD. In our case, we extend this to include covariates like:

$$P(X_t-u\le y|X_t>u)\approx 1-\left(1+\frac{\xi_t y}{\phi_t}\right)_+^{-1/\xi_t},\quad y>0, \tag{4.3}$$

where

$$\log\phi_t=\sum_s x_{ts}\zeta_s,\quad \xi_t=\xi.$$

Here, x_{ts} is the s-th covariate on day t and ζ_s is the corresponding coefficient. Since the parameter ϕ_t is necessarily positive, the logarithmic link function may be adequate. In our analysis, we confine ourselves to the constant $\xi_t=\xi$ for the analysis to be simpler, which is not however necessary in general. Finally, if the excesses over the threshold u are observed on days $t_1,t_2,\ldots$ and if the corresponding excesses are denoted by $Y_{t_1}, Y_{t_2},\ldots,$ then their likelihood will be approximated by

$$L=\prod_{r}\left\{\frac{1}{\phi_{t_r}}\left(1+\frac{\xi_{t_r}Y_{t_r}}{\phi_{t_r}}\right)^{-1/\xi_{t_r}-1}\right\} \tag{4.4}$$

under the independence assumption of the original data.

Combining excesses with exceedance probability

Though the previous two models can be applied separately to exceedance probability and excesses, these may be combined through a single rich model to be considered simultaneously. Specifically, from (4.3), one may get

$$P(X_t\le u+y)\approx 1-p_t\left(1+\frac{\xi_t y}{\phi_t}\right)_+^{-1/\xi_t},\quad y\ge 0,$$

or equivalently,

$$P(X_t\le x)\approx 1-p_t\left\{1+\frac{\xi_t(x-u)}{\phi_t}\right\}_+^{-1/\xi_t},\quad x\ge u,$$

where $p_t=P(X_t>u)$. Further, the second formula may be rewritten as

$$P(X_t\le x)\approx 1-\left\{1+\frac{\xi_t(x-\mu_t)}{\sigma_t}\right\}_+^{-1/\xi_t},\quad x\ge u, \tag{4.5}$$

where $\sigma_t=\phi_t p_t^{\xi_t}>0$ and $\mu_t=u-\sigma_t(p_t^{-\xi_t}-1)/\xi_t<u$. Here, the exceedance probability is, of course, given by

$$p_t=\left\{1+\frac{\xi_t(u-\mu_t)}{\sigma_t}\right\}_+^{-1/\xi_t}.$$

If the threshold u is chosen sufficiently high so that p_t becomes very small, then approximation (4.5), by virtue of the Taylor expansion of order 1, may be replaced by

$$P(X_t\le x)\approx \exp\left[-\left\{1+\frac{\xi_t(x-\mu_t)}{\sigma_t}\right\}_+^{-1/\xi_t}\right],\quad x\ge u, \tag{4.6}$$

where the condition $\mu_t<u$ is no longer necessary here. Notice that the right hand side of (4.6) is nothing but the tail of the generalized extreme value distribution (henceforth GEV). This is not surprising since the GPD model (4.3) for excesses is based on the assumption that the distribution function of X_t belongs to the domain of attraction of some extreme value

distribution. Roughly speaking, relations (4.5) and (4.6) therefore imply that taking GPD for the excess is the same in effect as taking GEV for the tail, provided that the threshold u is chosen sufficiently high.

Models (4.5) and (4.6) are referred to as generalized Pareto tail (henceforth GPT) model and generalized extreme value tail (henceforth GEVT) model respectively. When the GPT model is quoted, we will adopt the following covariate structure as

$$\log(u-\mu_t)=\sum_s x_{ts}\gamma_s,\quad \log\sigma_t=\sum_s x_{ts}\eta_s,\quad \xi_t=\xi \tag{4.7}$$

when the GEVT model is used, the following covariate structure will be substituted:

$$\mu_t=\sum_s x_{ts}\delta_s,\quad \log\sigma_t=\sum_s x_{ts}\eta_s,\quad \xi_t=\xi. \tag{4.8}$$

Since model (4.5) is valid only for $x\geq u$ and does not give any information about the remaining part of the distribution of X_t, it is impossible to calculate the likelihood for $\{X_t\}$. Instead, if we define a truncated variable like $Z_t=X_t\cdot I(X_t>u)+u\cdot I(X_t\leq u)$ which has all the information about the tail of X_t, then the right hand side of (4.5) is a valid model for the distribution function of Z_t, which has a jump at the threshold u and is absolutely continuous thereafter. Therefore, the approximate likelihood for $\{Z_t\}$ will be

$$L=\prod_t\left[\ (1-p_t)\cdot I(Z_t=u)+\frac{1}{\sigma_t}\left\{1+\frac{\xi_t(Z_t-\mu_t)}{\sigma_t}\right\}^{-1/\xi_t-1}\cdot I(Z_t>u)\right] \tag{4.9}$$

under the independence assumption. Similarly, one may apply this idea to model (4.6) to find out the corresponding approximate likelihood for $\{Z_t\}$, which is omitted here.

Higher order Markov chain models for the tails of $\{X_t\}$

So far, all the previous models assume the independence of observations. Since the truncated variable $Z_t=X_t\cdot I(X_t>u)+u\cdot I(X_t\leq u)$ has all the necessary information about the tail of X_t, we also work with this $\{Z_t\}$ as before. One of natural ways of allowing for local dependence of variables is to impose Markov property on the series $\{Z_t\}$. Henceforth, $\{Z_t\}$ is assumed to be a k-th order Markov chain for some $k\geq 1$. That is, Z_t is assumed to depend on only the k immediate past values $Z_{t-k},\ldots,Z_{t-1}$. For the marginals, we take model (4.5) as the distribution function of Z_t, or equivalently, as the tail of the distribution function of X_t, which is, of course, based on the previous arguments. One may take instead model (4.6), but the detailed procedure for this is similar to that for model (4.5) and so omitted here. Now, assuming (4.5) is the exact tail of the distribution function of X_t, consider the standardized variable (in some sense) $\widetilde{X}_t$ defined by

$$\tilde{X}_t = \frac{1}{\xi_t} \log \left\{ 1 + \frac{\xi_t(X_t - \mu_t)}{\sigma_t} \right\}_+ ,$$

which is, of course, interpreted as $\tilde{X}_t = (X_t - \mu_t)/\sigma_t$ if $\xi_t = 0$. Then $\tilde{X}_t$ has obviously the standard exponential tail, i.e.,

$$P(\tilde{X}_t \le x) = 1 - e^{-x}, \quad x \ge \frac{1}{\xi_t} \log \left\{ 1 + \frac{\xi_t(u - \mu_t)}{\sigma_t} \right\}_+ > 0.$$

At this stage, we need a kind of stationarity assumption of the tails of the transformed series $\{\tilde{X}_t\}$, which is eventually helpful for characterizing the whole distribution of the chain $\{Z_t\}$. Although not necessary, we now assume that $\{\tilde{X}_t\}$ itself is strictly stationary to make the argument simple. Since the distribution of the chain $\{Z_t\}$ is completely determined by the joint distribution functions of all the k+1 successive variables $Z_t, Z_{t+1}, \ldots, Z_{t+k}$ and hence by the joint distribution functions of all the k+1 successive variables $\tilde{X}_t, \tilde{X}_{t+1}, \ldots, \tilde{X}_{t+k}$ under model (4.5) for the marginal tails, it is a key step to find out an appropriate model for the joint distribution function $\tilde{F}_{k+1}$ of $(\tilde{X}_t, \ldots, \tilde{X}_{t+k})$. By analogy with the univariate case deriving model (4.5), this can be resolved naturally by assuming $\tilde{F}_{k+1}$ belongs to the domain of attraction of some (k+1)-dim. extreme value distribution $\tilde{G}_{k+1}$. We here note that $\tilde{F}_{k+1}$ has the same tails of marginals which are standard exponential so that $\tilde{G}_{k+1}$ must have the equal marginals $\Lambda(x) = \exp(-e^{-x})$, the standard Gumbel distribution. By considering that $\Lambda(x) \approx 1 - e^{-x}$ for large x and that $\tilde{G}_{k+1}$ has the same tail behavior as $\tilde{F}_{k+1}$, our suggestion is to use the tails of $\tilde{G}_{k+1}$ as an approximation for the tails of $\tilde{F}_{k+1}$, i.e.,

$$\begin{aligned} \tilde{F}_{k+1}(x_1, \ldots, x_{k+1}) &\approx \tilde{G}_{k+1}(x_1, \ldots, x_{k+1}) \\ &\approx \tilde{G}_{k+1}(-\log(-\log(1 - e^{-x_1})), \ldots, -\log(-\log(1 - e^{-x_{k+1}}))) \end{aligned}$$

for large $x_1, \ldots, x_{k+1}$. Returning back to the original series $\{X_t\}$, the tail of the joint distribution function $F_{k+1}^{(t)}$ of $(X_t, \ldots, X_{t+k})$ may be, therefore, approximated by

$$F_{k+1}^{(t)}(x_1, \ldots, x_{k+1}) = P(X_t \le x_1, \ldots, X_{t+k} \le x_{k+1})$$

$$\approx \tilde{G}_{k+1}\left(-\log\left(-\log\left(1-\left\{1+\frac{\xi_t(x_1-\mu_t)}{\sigma_t}\right\}_+^{-1/\xi_t}\right)\right),\ldots,\right.$$

$$\left.-\log\left(-\log\left(1-\left\{1+\frac{\xi_{t+k}(x_{k+1}-\mu_{t+k})}{\sigma_{t+k}}\right\}_+^{-1/\xi_{t+k}}\right)\right)\right),$$

$$x_1,\ldots,x_{k+1}\geq u. \tag{4.10}$$

Our derivation for model (4.10) is somewhat heuristic. Those who need more rigorous details are referred to Smith, Tawn and Coles (1993) and Smith (1993). If one starts with model (4.6) for the tail of the distribution function of X_t, one may get the following approximation:

$$F_{k+1}^{(t)}(x_1,\ldots,x_{k+1})=P(X_t\leq x_1,\ldots,X_{t+k}\leq x_{k+1})$$

$$\approx \tilde{G}_{k+1}\left(\frac{1}{\xi_t}\log\left\{1+\frac{\xi_t(x_1-\mu_t)}{\sigma_t}\right\}_+,\ldots,\right.$$

$$\left.\frac{1}{\xi_{t+k}}\log\left\{1+\frac{\xi_{t+k}(x_{k+1}-\mu_{t+k})}{\sigma_{t+k}}\right\}_+\right),\quad x_1,\ldots,x_{k+1}\geq u. \tag{4.11}$$

Once we have model (4.10) or model (4.11) for the tails of $\{X_t\}$, these can be used as valid models for the distribution of the truncated series $\{Z_t\}$. In other words, model (4.10) (or model (4.11)) may be considered as an approximation for the proper joint distribution function of $(Z_t,\ldots,Z_{t+k})$, which has, of course, discrete components in it. Finally, the whole likelihood for $\{Z_t\}$ of sample size n, based on the assumption of Markov structure of $\{Z_t\}$, is easily calculated as

$$L=\frac{\prod_{t=1}^{n-k}L_{k+1}(Z_t,\ldots,Z_{t+k})}{\prod_{t=2}^{n-k}L_k(Z_t,\ldots,Z_{t+k-1})}, \tag{4.12}$$

where

$$L_{k+1}(Z_t,\ldots,Z_{t+k})=F_{k+1}^{(t)}(u,\ldots,u)\cdot I(Z_t=u,\ldots,Z_{t+k}=u)$$

$$+\sum_{\substack{D\subset\{t,\ldots,t+k\}\\ D\neq\emptyset}}\frac{\partial^{|D|}F_{k+1}^{(t)}(Z_t,\ldots,Z_{t+k})}{\prod_{i\in D}\partial Z_i}\cdot I(Z_i>u,i\in D,Z_j=u,j\in D^c)$$

and $L_k(Z_t,\ldots,Z_{t+k-1})$ being defined similarly using the k-dim. distribution function $F_k^{(t)}$ of $(X_t,\ldots,X_{t+k-1})$. For the covariates, we shall use the same structures as (4.7) and (4.8) for models (4.10) and (4.11) respectively.

Since the order k of the chain $\{Z_t\}$ and the $(k+1)$-dim. extreme value distribution $\tilde{G}_{k+1}$ in the above methods are not specified explicitly, some criteria for model selections are needed when the methods are applied to real time series data. If considered models are nested, one may use the standard procedures such as the likelihood ratio test and the t ratio test. The models we use, however, are typically not nested, in which case one may use alternatively the usual time series procedures such as AIC and BIC suggested by Akaike (1974) and Schwarz (1978) respectively. For the context of Markov chains, Tong (1975) and Katz (1981) proposed the use of AIC and BIC procedures respectively to select the order of a chain. In regard to $\tilde{G}_{k+1}$, our detailed analysis is confined to only the logistic model defined by

$$\tilde{G}_{k+1}(x_1,\ldots,x_{k+1}) = \exp\left\{-\left(\sum_{s=1}^{k+1} e^{-x_s/\alpha}\right)^{\alpha}\right\},\quad x_1,\ldots,x_{k+1}\in\mathfrak{R},$$

where $0<\alpha\le 1$. This model was used by several statisticians (e.g., see Tawn (1988) and Smith (1992)) as a basic model for multivariate extreme value distributions because of its simple structure yet to have a full flexibility of controlling any dependence of variables. Applying this to (4.10) and (4.11), we therefore get approximations, respectively,

$$F_{k+1}^{(t)}(x_1,\ldots,x_{k+1}) \approx \exp\left\{-\left[\left(-\log\left(1-\left\{1+\frac{\xi_t(x_1-\mu_t)}{\sigma_t}\right\}_+^{-1/\xi_t}\right)\right)^{1/\alpha}+\cdots\right.\right.$$
$$\left.\left.+\left(-\log\left(1-\left\{1+\frac{\xi_{t+k}(x_{k+1}-\mu_{t+k})}{\sigma_{t+k}}\right\}_+^{-1/\xi_{t+k}}\right)\right)^{1/\alpha}\right]^{\alpha}\right\},$$
$$x_1,\ldots,x_{k+1}\ge u \tag{4.13}$$

and

$$F_{k+1}^{(t)}(x_1,\ldots,x_{k+1}) \approx \exp\left\{-\left[\left\{1+\frac{\xi_t(x_1-\mu_t)}{\sigma_t}\right\}_+^{-1/(\alpha\xi_t)}+\cdots\right.\right.$$
$$\left.\left.+\left\{1+\frac{\xi_{t+k}(x_{k+1}-\mu_{t+k})}{\sigma_{t+k}}\right\}_+^{-1/(\alpha\xi_{t+k})}\right]^{\alpha}\right\},\quad x_1,\ldots,x_{k+1}\ge u. \tag{4.14}$$

5. Statistical procedure

This section describes a detailed statistical procedure for applying the various models discussed in Section 4 to the ozone data. As mentioned before, the covariates we are interested in involve the length of the day and the year. To explain the procedure explicitly, let any day t be expressed as a pair (i,j), where i is the corresponding year and j is the corresponding day number within that year. That is, for example, t=(1981,1) for Jan. 1, 1981, and t=(1991,365) for Dec. 31, 1991. Then, from Graf-Jaccottet (1993), for a day number j, the corresponding length of the day (number of hours between sunrise and sunset) at any station is approximately given by

$$l_j = 12.24 + \frac{24}{\pi}\left(1+\frac{1}{265.25}\right)\arcsin\left(\tan\theta_0 \cdot \tan\theta_1 \cdot \sin\left\{\frac{2\pi}{365.25}(j-80)\right\}\right),$$

where $\theta_0 = 23.45°$ (inclination of the Earth) and θ_1=latitude of the station. Since the covariates are expressed in terms of x_{ts} in every model discussed in Section 4 , it is sufficient to specify only x_{ts} in order to explain our concrete covariate structure of each model. From now on, we adopt the covariates x_{ts} as: for any day t=(i,j), define

$$x_{t0}=1,\ x_{t1}=l_j-12,\ x_{t2}=i-1980,\ x_{t3}=(i-1980)^2,$$

and x_{ts}=0 for $s \geq 4$. In terms of words, x_{t0} explains a constant, x_{t1} a short-term annual trend, x_{t2} a long-term linear trend, and x_{t3} a long-term quadratic trend. Therefore, for instance, the logit model (4.1) becomes

$$\log\left(\frac{p_{(i,j)}}{1-p_{(i,j)}}\right) = \beta_0 + (l_j-12)\beta_1 + (i-1980)\beta_2 + (i-1980)^2\beta_3,$$

which has 4 parameters to be estimated. Similarly, the interpretation of all the other models in Section 4 is straightforward.

With the given covariate structure, the actual parameter estimation procedure is through the numerical maximum likelihood, that is, we choose the parameter estimates so as to maximize the likelihood function L defined in Section 4 for each model. For this, we here use Fortran programs employing the DFPMIN (Davidon-Fletcher-Powell variable metric function minimization) algorithm of Press, Flannery, Teukolsky and Vetterling (1986) to minimize the negative log likelihood -log$L$. Hereafter, the minimized negative log likelihood is referred to as NLLH. In addition to the maximum likelihood estimates, the programs we use here also produce the Hessian matrix of NLLH, the observed information matrix, the inverse of which is generally used as an approximation to the variance-covariance matrix of the parameter estimates. In particular, the square roots of the diagonal entries of the inverse of the Hessian matrix give approximate standard errors for the parameter estimates.

Missing data do not give rise to a problem under independent models for computing NLLH. For dependent models such as the higher order Markov chains, however, it is often a problem. In our case, we handle this by restarting the considered chain whenever there arises a missing observation. This has an effect of separating the whole chain into a number of independent subchains, where the error of the resulting NLLH is hopefully minor.

One final issue is how to discriminate which variables (covariates) are really significant when a specific full model is considered. For example, the GPT model (4.5) has, in fact, 9 parameters in total under our covariate structure, and some of these may be really significant whereas the others may not. A standard method for handling this is a stepwise selection in which the variables are introduced one at a time and the corresponding NLLH is computed. As a rule of thumb, a variable is considered significant if its inclusion in the model reduces the NLLH by more than 2. This is based on the approximate χ_1^2 distribution for the deviance statistic, twice the difference of NLLHs for the two models being compared. Another widely used measure for the test of significance is the t ratio which is defined by the value of the parameter estimate divided by its standard error. A suitable rule of thumb, here, is that any variable with a t ratio greater than 2 in absolute value is considered significant. This is based on the asymptotic normality of the maximum likelihood estimates, which is the case whenever $\xi > -0.5$ (see Smith (1985)). We however recall that the standard errors computed by the programs are only approximate. In our detailed analysis of Section 6, we therefore consult both of the above criteria for the test of significance.

6. The data analysis

We now analyze the daily ozone maxima for stations P, Q, R and F (network maxima), using the various models discussed in Section 4. As mentioned before, we concentrate on the threshold u=99.5 for stations P, Q and R, and the threshold u=119.5 for network maxima.

Exceedance probability of a high threshold

We fit the logit model (4.1) to estimate the exceedance probability of u. For station P, an initial analysis is made using only β_0, which results in NLLH=429.76; adding β_1 to the model gives NLLH=354.06 with the corresponding t ratio=7.75, which means β_1 is significant; adding β_2 to the model results in NLLH=353.99 with the corresponding t ratio=-0.37,which means β_2 is not significant so that the analysis is stopped here. In other words, the long-term trend is not detected for station P. For stations Q and R, the same thing happens and the model containing only β_0 and β_1 turns out to be optimal. For network maxima, however, the initial model gives NLLH=539.11 and the resulting NLLHs by adding β_1, β_2 and β_3 successively are 487.49, 485.36 and 480.25, where the corresponding t ratios turn out to be 6.92, 2.07 and -3.40 respectively. This implies that the long-term quadratic trend should be taken into account significantly for network maxima. The optimal models caught by our analysis are summarized in Table 2. Each entry in the table represents the corresponding parameter estimate together with its standard error in parenthesis. The resulting NLLHs are also shown in Table 2. In network maxima, the negative value of the estimate of β_3 indicates that the long-term trend is eventually downward.

Excesses over a high threshold

For the excesses over a high threshold, we now fit the GPD model (4.3). For station P, an

Table 2: Parameter estimates in logit model

	P	Q	R	F
β_0	-5.20(0.32)	-5.20(0.32)	-5.43(0.61)	-5.29(0.42)
β_1	0.93(0.12)	0.89(0.11)	1.12(0.21)	0.90(0.11)
β_2				0.34(0.12)
β_3				-0.034(0.010)
NLLH	354.06	323.92	186.17	480.25

initial analysis with ξ and ζ_0 produces NLLH=388.40, where the t ratios of ξ and ζ_0 are 0.74 and 16.23 respectively; adding ζ_1 and ζ_2 successively to the model results in NLLH=386.13, 386.13 respectively with the corresponding t ratio=2.28, 0.54, which implies that ζ_1 is significant but ζ_2 is not. Also, the t ratio=0.74 of ξ in the initial analysis raises a question of ξ=0. To test this, we put ξ=0 (i.e., $\xi \to 0$) in model (4.3) with ζ_0 and ζ_1 included, which leads to an exponential distribution for the excesses, and recompute the corresponding likelihood to obtain NLLH=386.16. Comparing this with the former NLLH=386.13, we therefore conclude ξ=0 for station P. For a further investigation on this and for a goodness-of-fit of the fitted model, residual plots will be given later with discussion. For station Q, ξ=0 is rejected by both of the criteria (NLLH and t ratio), but for station R, ξ=0 is not. For network maxima, the two criteria lead to different suggestions: the NLLH suggests the acceptance of ξ=0, whereas the t ratio suggests the rejection of ξ=0. Based on the residual plot which will be given later, we here reject ξ=0 for network maxima. The length of the day turns out to be significant for only P and F. Our further analysis shows that there is no evidence of the long-term trend at any station. Table 3 shows the optimal model selected at each station. In the table, ξ=0 means that an exponential

Table 3: Parameter estimates in GPD model

	P	Q	R	F
ξ	0	-0.22(0.12)	0	-0.14(0.084)
ζ_0	2.04(0.46)	3.34(0.16)	3.02(0.14)	2.17(0.34)
ζ_1	0.40(0.17)			0.35(0.12)
NLLH	386.16	346.47	221.31	568.73

distribution (i.e., $\xi \to 0$ in the GPD model (4.3)) is fitted for the excesses. As before, the parenthesis in the table shows the standard error of the corresponding parameter estimate, and this convention will be adopted throughout this section.

To check the fit of the model, we define a residual $Y_{t_r}/\hat{\phi}_{t_r}$ with respect to the r-th excess Y_{t_r} on day t_r and the associated parameter estimate $\hat{\phi}_{t_r}$. The residuals are then arranged in order and plotted against the corresponding expected values under each of the two models (exponential distribution and GPD). If the model fits the data, then the residuals should be tightly clustered around a straight line of 45° slope through the origin. Figure 1 shows the resulting residual plot for each station. For stations P and R, the GPD does not seem to improve the overall fit, compared with the exponential distribution, and this may justify the failure of rejecting $\xi=0$. For stations Q and F however, the GPD clearly gives a better fit, particularly in the upper tail, which justifies the rejection of $\xi=0$.

Combining excesses with exceedance probability

We now combine the previous two models (logit model (4.1) and GPD model (4.3)) to handle the exceedance probability and the excesses simultaneously. First, consider the GPT model (4.5) with the covariate structure (4.7). After a process of stepwise selection of variables as before, we obtain the optimal models as in Table 4. The residual plots of

Table 4: Parameter estimates in GPD model

	P	Q	R	F
ξ	0	-0.30(0.089)	0	-0.15(0.083)
γ_0	3.76(0.50)	5.30(0.20)	4.72(0.19)	4.25(0.42)
γ_1	0.13(0.18)	-0.15(0.035)	-0.30(0.051)	0.019(0.13)
γ_2				-0.091(0.040)
γ_3				0.0092(0.0033)
η_0	2.04(0.49)	4.26(0.37)	3.02(0.14)	2.80(0.51)
η_1	0.40(0.18)			0.25(0.13)
NLLH	737.13	665.00	409.13	1047.44

excesses, which are omitted here, are also considered to select these optimal models. For stations P and R, the GPT does not give a considerable improvement on the exponential tail ($\xi=0$ in model (4.5)) so that $\xi=0$ is not rejected. The length of the day turns out to be significant in the location parameter μ_t for all stations, but significant in the scale parameter σ_t for only P and network maxima. For network maxima, the long-term quadratic trend is detected in the location parameter and the positive sign of the estimate of γ_3 implies that the trend is eventually downward. Another interesting thing is that for stations P and R, the estimates of η's agree with those of ζ's in Table 3 though they are based on different likelihood functions. This is, however, not surprising because $\sigma_t=\phi_t$ in model (4.5) when $\xi=0$. All the other things do also well agree with the previous two models (4.1) and (4.3). Therefore, the GPT model (4.5) can be used as a single (but rich) unified model if one is interested in both the exceedance probability and the excesses.

Theoretically, the GEVT model (4.6) has the same tail behavior as the GPT model (4.5) as long as the threshold u is chosen sufficiently high. In practice, however, they may lead to slightly different fits because of the level of the threshold u which is usually not so high.

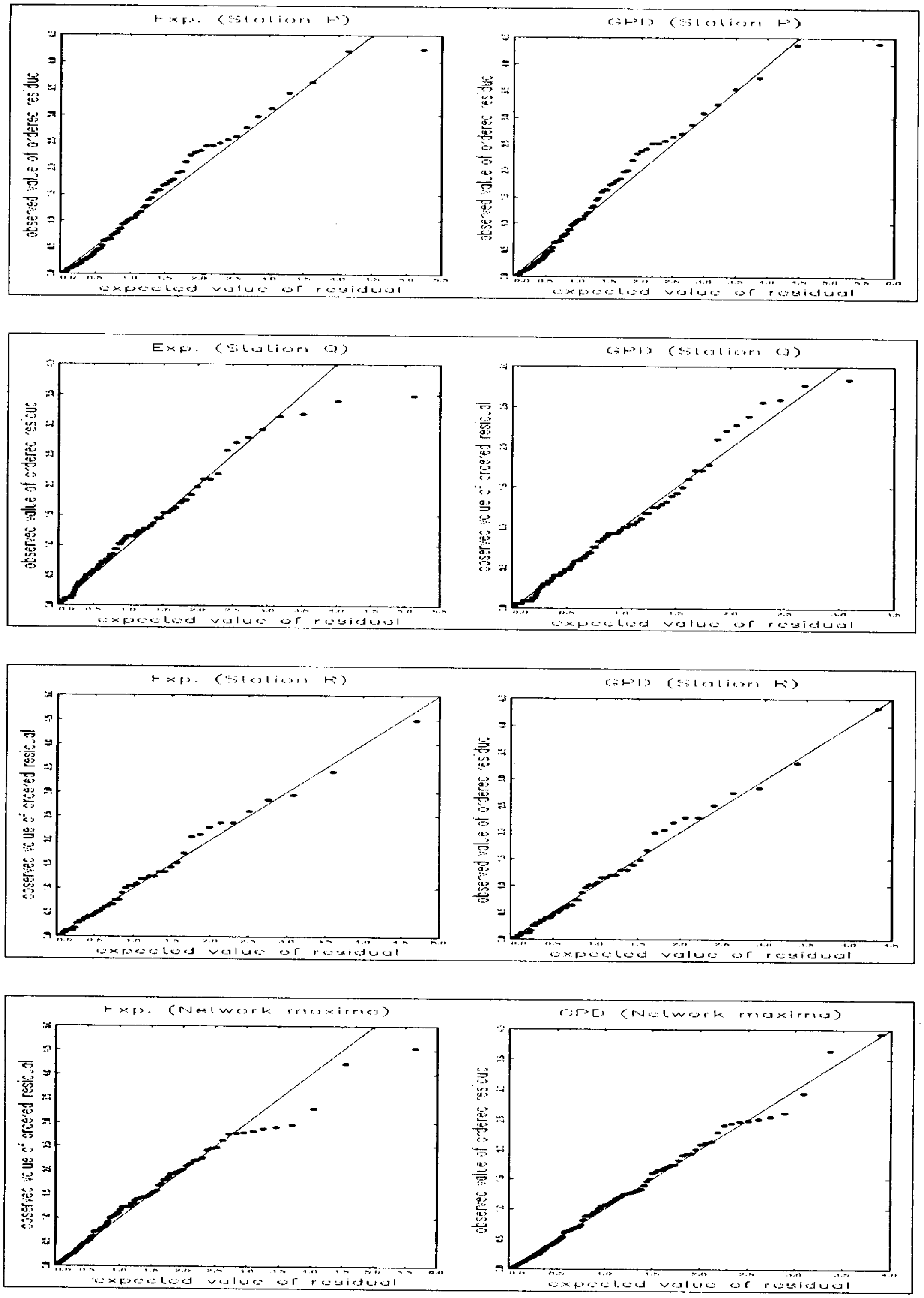

Figure 1: Residual plot under GPD model

Recall that an extremely high threshold tends to cause an insufficient number of exceedances, which may induce a bad fit. One advantage of using model (4.6) is no necessity of the constraint $\mu_t<u$. We now apply the GEVT model (4.6) and obtain the optimal models in Table 5. For each station, the selected variables are exactly the same as in Table 4, but the

Table 5: Parameter estimates in GEVT model

	P	Q	R	F
ξ	0	-0.31(0.042)	0	-0.12(0.072)
δ_0	45.89(1.79)	-96.28(1.81)	-8.86(0.47)	52.02(1.90)
δ_1	-2.00(2.54)	22.11(3.75)	21.85(1.23)	0.12(4.81)
δ_2				6.13(1.50)
δ_3				-0.63(0.14)
η_0	2.26(0.088)	4.29(0.14)	3.00(0.15)	2.68(0.18)
η_1	0.31(0.051)			0.25(0.054)
NLLH	737.27	666.37	407.46	1047.46

estimates of ξ and η's are slightly changed. In the Table 5, $\xi=0$ (i.e., $\xi\to0$ in model (4.6)) corresponds to a Gumbel tail. For station P and network maxima, the difference of two NLLHs from Tables 4 and 5 is minor. For stations Q and R, however, the difference is somewhat bigger, which perhaps results from the smaller number of exceedances. The residual plots of excesses, which are also omitted here, show shapes similar to those for model (4.5). This suggests that there is no big difference, in practice too, between using model (4.5) and using model (4.6), provided that the threshold u is taken reasonably high.

Higher order Markov chain models for the tails of $\{X_t\}$

Although model (4.5) (or model (4.6)) explains both the exceedance probability and the excesses very well, it still does not take into account the serial dependence of variables which usually exists in various environmental data. We now apply the k-th order Markov chain model discussed in Section 4 to the tails of the daily ozone maxima. Our detailed analysis is concentrated on models (4.13) and (4.14) which are based on the logistic multivariate extreme value distribution. The method can, however, be applied to various multivariate extreme value distributions.

First of all, we consider k=1 (i.e., first order Markov chain). Fitting model (4.13) with k=1 produces the optimal model for each station as in Table 6. Since this first order Markov chain model is an extension of the independent GPT model (4.5) (i.e., model (4.13) contains model (4.5) as a special case), we compare the selected optimal models of Table 6 with those of Table 4 to investigate any improvement by the first order Markov chain. For each station, as seen in Table 6, the estimate of α, which is a measure of dependence of variables (α=1 corresponds to independence, which reduces model (4.13) to model (4.5), and $\alpha\approx0$ corresponds

Table 6: Parameter estimates in first order Markov chain model for GPT

	P	Q	R	F
α	0.76(0.034)	0.79(0.045)	0.82(0.052)	0.80(0.035)
ξ	0	-0.22(0.090)	0	-0.16(0.024)
γ_0	4.84(0.46)	5.24(0.20)	4.65(0.079)	4.73(0.031)
γ_1	-0.26(0.014)	-0.17(0.028)	-0.28(0.043)	-0.21(0.021)
γ_2				0.019(0.0087)
η_0	3.16(0.35)	4.05(0.36)	3.00(0.11)	3.60(0.016)
NLLH	697.12	641.21	394.53	1011.74

to a very strong dependence), turns out to be about 0.79 with a standard error of about 0.04. This suggests that the data of daily ozone maxima clearly have a serial dependence which is not so strong though. Comparing the corresponding NLLHs gives a more clear explanation for the existence of serial dependence. The difference between two NLLHs from Tables 4 and 6 varies from 14.6 to 40.01 according to the stations. Taking into account the parameters being rejected due to their insignificancy in Table 6, the difference will be bigger. In fact, when we fit model (4.13) using exactly the same parameters ($\xi, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \eta_0$ and η_1) as in Table 4, the resulting NLLHs turn out to be 696.32 and 1010.20 for station P and network maxima respectively. Based on the likelihood ratio test and even the AIC criterion proposed by Tong (1975), this big drop of NLLH strongly suggests that $\alpha<1$ and hence k=1 in our Markov chain context should be accepted against the null hypothesis $\alpha=1$, or equivalently, the null hypothesis k=0.

Let us look at Table 6 more closely. For stations P and R, the exponential tail ($\xi=0$) is not rejected; in other words, the GPT shows no significant improvement on the exponential tail, which agrees with Table 4. For the long-term trend in the location parameter of network maxima, not a quadratic trend but a linear downward trend turns out to be detected, which does not agree with the results in Table 4. Moreover, the length of the day is no longer significant in the scale parameter at any station, which is not the case in Table 4. The dependent model (4.13) need not, in general, detect exactly the same variables as the independent model (4.5). The inharmonious results of Tables 4 and 6 are not, however, fully understandable to analysts.

Now we fit model (4.14) with k=1, the first order Markov chain model having the GEVT as its marginal tails. The resulting optimal models are summarized in Table 7. Here, the estimate of α turns out to be about 0.80 with a standard error of about 0.06, which is very similar to that of Table 6. The big difference between two NLLHs from Tables 5 and 7, also, justifies the acceptance of the hypothesis $\alpha<1$ and hence the hypothesis k=1. For stations P and R, the Gumbel tail ($\xi=0$) is not rejected against the GEVT. Above all things, the selected variables of Table 7 are consistent with those of Table 5. That is, for network maxima, a quadratic downward trend is detected, and for station P and network maxima, the

Table 7: Parameter estimates in first order Markov chain model for GEVT

	P	Q	R	F
α	0.76(0.064)	0.79(0.047)	0.83(0.076)	0.81(0.038)
ξ	0	-0.24(0.048)	0	-0.13(0.071)
δ_0	33.81(0.80)	-85.44(7.73)	-3.01(2.08)	34.53(1.28)
δ_1	1.57(1.99)	23.48(4.78)	20.03(1.97)	6.64(2.96)
δ_2				3.98(1.18)
δ_3				-0.46(0.085)
η_0	2.50(0.15)	4.10(0.15)	2.97(0.080)	3.10(0.24)
η_1	0.24(0.036)			0.13(0.026)
NLLH	696.34	642.36	393.42	1009.88

length of the day in the scale parameter turns out to be significant. Considering that the dependent model (4.14) is an extension of the independent model (4.6), this harmony is desirable in comparing and interpreting the two models. It is recalled that this harmony is not found between model (4.5) and its extension model (4.13).

The residual plots of excesses, which are omitted here, are also studied under the first order Markov chain models (4.13) and (4.14) for GPT and GEVT respectively. The overall patterns are similar to those of Figure 1. The plots support the GPT (or GEVT) for station Q and network maxima, but do not for stations P and R.

Since the first order Markov chain models (models (4.13) and (4.14) with k=1) improve the likelihood a lot on the independent models (4.5) and (4.6), the next thing is to consider k=2 (i.e., second order Markov chain). The resulting NLLHs of the optimal models selected by fitting model (4.13) with k=2 turn out to be 706.61, 652.24, 398.31 and 1023.07 for stations P, Q, R and F respectively. The corresponding NLLHs for model (4.14) with k=2 are 704.99, 653.17, 397.02 and 1022.97. These NLLHs are clearly greater than those in the first order Markov chain models, even though they are still less than those in the independent models. Based on the likelihood ratio test, we may therefore say that the second order Markov chain models are a significant improvement on the independent models, but not on the first order Markov chain models. Hence, a higher k ($k \geq 3$) is not tried and we finish the detailed analysis here.

Finally, we note that the higher order Markov chain models (4.13) and (4.14), which are based on the logistic multivariate extreme value distribution, are not nested with the change of the order k. This is why the NLLHs in the second order Markov chain models could be greater than those in the first order Markov chain models. Since our higher order Markov chain models highly depend on which model for the multivariate extreme value distribution is chosen, one may resolve this and be successful in finding a better fit with a higher order k by developing and applying an appropriate model for the multivariate extreme value distribution.

7. Exceedance probabilities and return levels

In Section 6 it is shown that the GPT model (4.5) or the GEVT model (4.6), as a natural model combining the excesses with the exceedance probability, gives a satisfactory fit of the daily ozone maxima under the independence assumption of the data. It is also shown in Section 6 that the first order Markov chain models (models (4.13) and (4.14) with k=1), as extensions of the independent models, significantly improve the overall fit of the ozone data. The fitted models are then used to estimate the exceedance probabilities of the threshold u as well as the various return levels. In this section we concentrate on the GEVT model (4.6) and its extension model (4.14) with k=1 because they show a consistency in detecting covariates. The corresponding results of the fitted models are given in Table 5 and Table 7 respectively.

First of all, from models (4.6) and (4.14), the exceedance probability of the threshold u on day t is given by

$$1-\exp\left[\ -\left\{1+\frac{\xi_t(u-\mu_t)}{\sigma_t}\right\}_+^{-1/\xi_t}\right] .$$

Replacing the parameters by the corresponding estimates in Table 5 (or Table 7) through the covariate structure (4.8) gives an estimated exceedance probability whose plot is shown in Figure 2. Since the long-term trend is not detected for stations P, Q and R, only one-year period is shown in the figure for these stations. For network maxima having the long-term quadratic trend in the location parameter, the whole period (1981~1991) is shown in the figure. The day in the figure is started from Jan. 1, 1981 (i.e., 1 for Jan. 1, 1981,..., 4015 for Dec. 31, 1991). In the figure, GUMT represents the Gumbel tail (i.e., ξ=0 in model (4.6)) and MC1 represents the first order Markov chain model (i.e., model (4.14) with k=1). Since the programs used here for minimizing the negative log likelihood also produce the observed information matrix, it is possible to construct an approximate confidence interval for the exceedance probability on each day via the standard delta method, which is not given here. It is also noted that the exceedance probability is based on the marginals only so that the dependence parameter α in model (4.14) does not play an important role in Figure 2. The first order Markov chain model can, however, be used effectively to estimate some quantities which need a dependence structure between variables. For instance, one may easily compute an estimate of the probability of exceedance on day t+1 given the exceedance on day t or the probability of consecutive exceedances on two days.

Return levels are defined by percentiles of the marginal distribution on each day. From models (4.6) and (4.14), the 100(1-p)th percentile (0<p<1) on day t is therefore given by

$$\mu_t+\frac{\sigma_t}{\xi_t}[\ \{-\log(1-p)\}^{-\xi_t}-1]\ ,$$

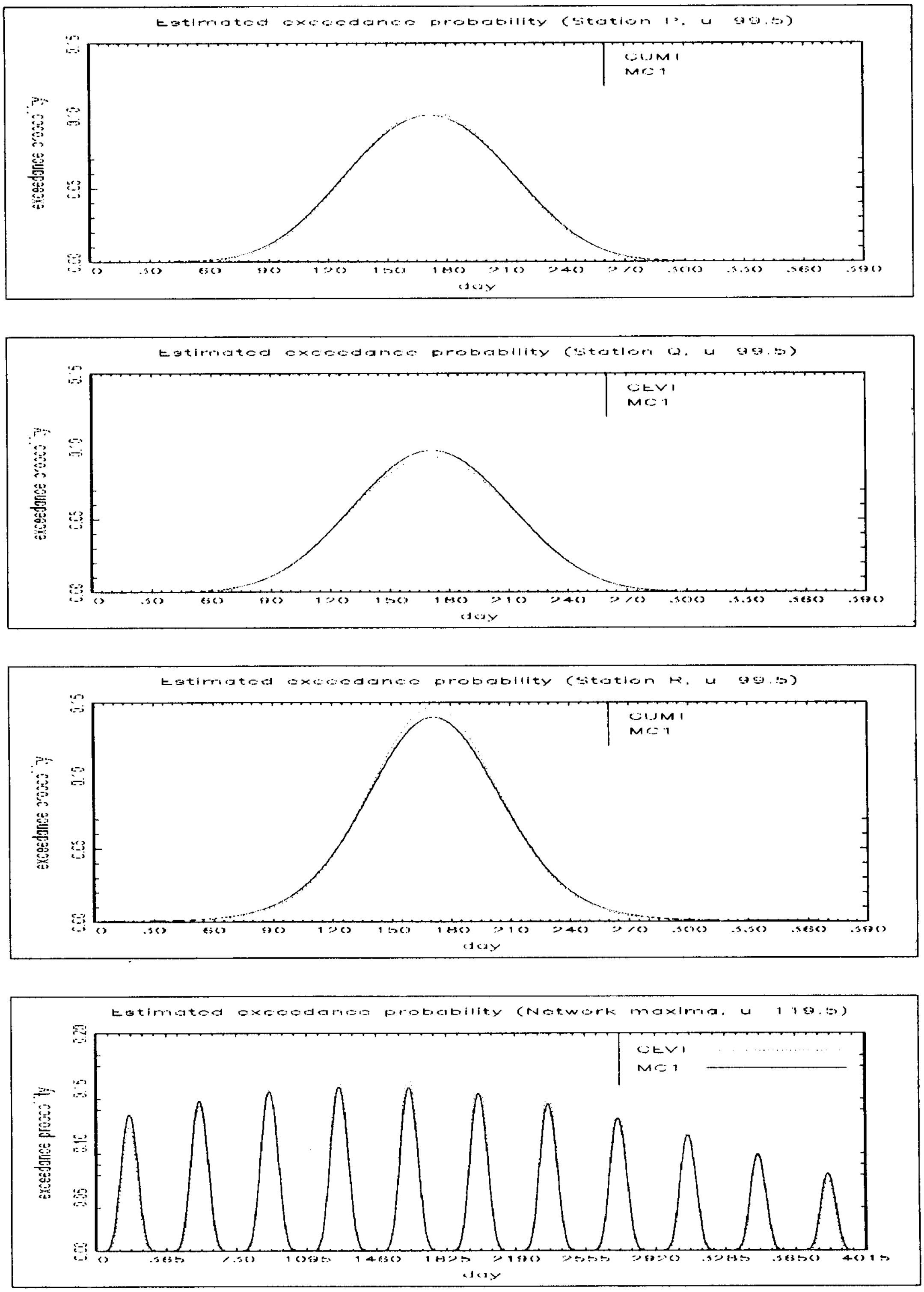

Figure 2: Estimated exceedance probability of a threshold u

which is valid whenever it is not less than the threshold u. This is because models (4.6) and (4.14) are valid only for $x \geq u$. Figure 3 shows the corresponding estimated 99th percentile (100-year return level) on each day together with the original data of daily ozone maxima. Again, one may use the delta method to compute an approximate confidence interval for the return level on each day, which is not shown here.

8. Conclusions and summary

In our ozone study, we have analyzed the time series of daily ozone maxima collected in the Chicago area for 11 years. The main purpose of this study is to examine what is happening in the tails (or extremes) of the series. For this purpose, the detailed analysis has been concentrated on three stations, labeled P, Q and R, with the highest ozone levels, and also the network maxima.

The overall methodology adopted here for a datailed analysis is the threshold technology. Starting from the logit model for the binary data of exceedances, we have extended the classical threshold method to higher order Markov chain models for the tails of the series to take the serial dependence of the data into account. For the tails of the marginals, the GPT model (4.5) and the GEVT model (4.6) turn out to give similar performances under the independence assumption of the data. For higher order Markov chain models, however, only model (4.14) having the GEVT as its marginal tails gave a consistency with the independence model (4.6) in detecting covariates.

The length of the day was used as one of the main covariates and turns out to explain quite well the short-term annual trend. For the long-term trend, the three stations P, Q and R did not show any trend, but a quadratic downward trend was found in network maxima. This downward trend may indicate the success of EPA's efforts to reduce the highest ozone levels over the last 15 years.

For each station, the first order Markov chain model (model (4.13) or model (4.14) with k=1) turns out to be a significant improvement on the independent model (model (4.5) or model (4.6)). For network maxima, our impression from the data analysis indicates that even a higher order might be adequate. This is not, however, verified since the higher order Markov chain model we applied for a detailed analysis is based on the logistic multivariate extreme value distribution so that the corresponding chain is not nested with the change of the order. To resolve this, it is necessary to develop another flexible model for the multivariate extreme value distribution which leads to the nestedness of the chain with the change of the order.

Shortly speaking, our approach in this paper concentrates on the tail (or extremal) behavior of a given time series, and the tails are modeled by (4.10) or (4.11) through a multivariate extreme value distribution together with a k-th order Markov chain. All the data below the

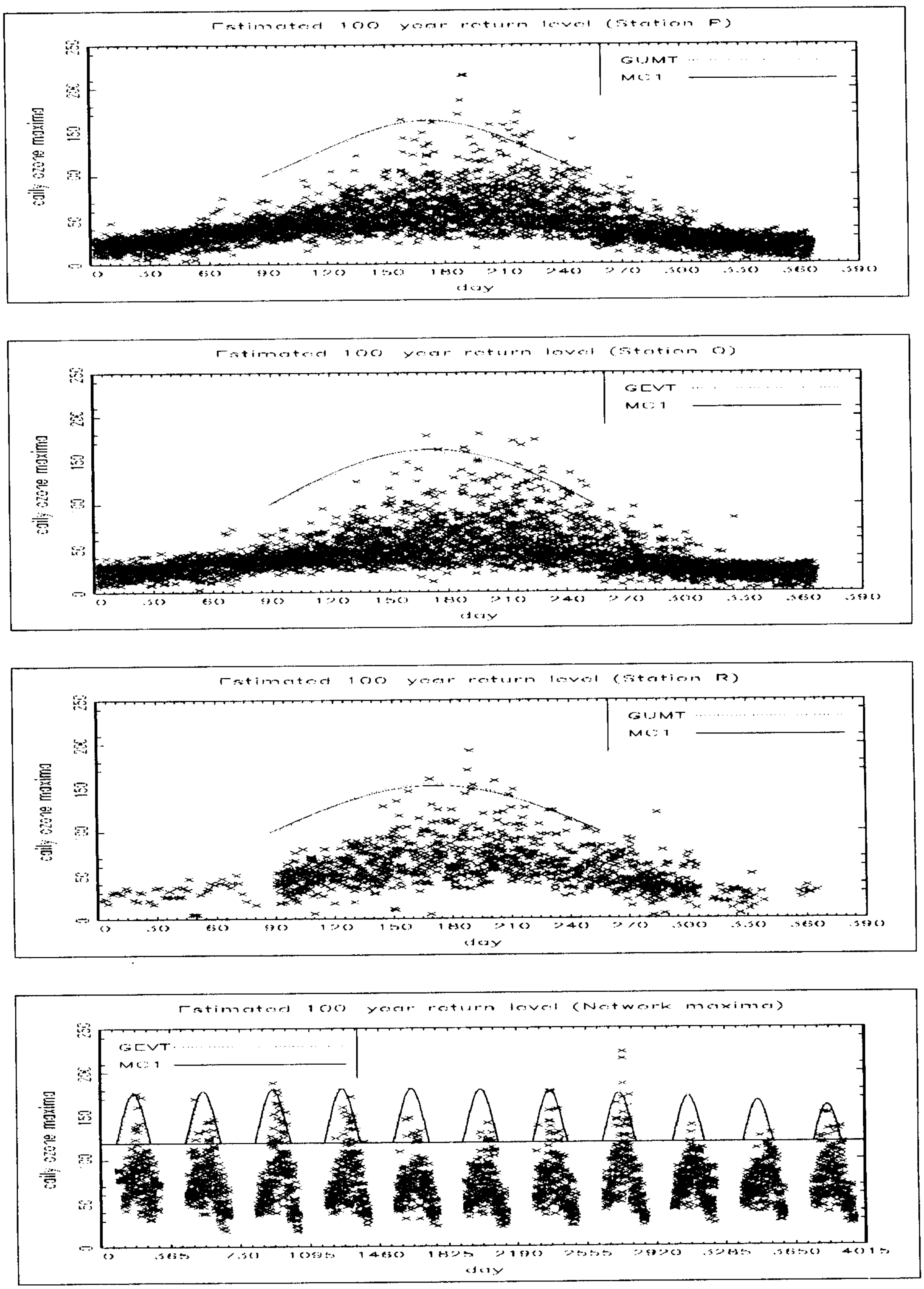

Figure 3: Estimated 100-year return level

threshold u are lumped together to consist of a single state u and are never modeled. A possible extension of this approach is to allow the data below the threshold to have some stochastic structure (e.g., AR(k) model). However, the detailed construction will be much harder.

References

[1] Akaike, H. (1974). A new look at the statistical model identification. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control,* Vol. AC-19, 716-723.

[2] Bloomfield, P., Royle, A. and Yang, Q. (1993). Accounting for meteorological effects in measuring urban ozone levels and trends. Technical Report #1, National Institute of Statistical Sciences.

[3] Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B,* Vol. 52, 393-442.

[4] Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, Vol. 24, 180-190.

[5] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics* (2nd edn.). Krieger, Florida. (1st edn. published 1978 by Wiley, New York).

[6] Graf-Jaccottet, M. (1993). A flexible model for ground ozone concentration. *Environmetrics,* Vol. 4, 23-37.

[7] Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, Vol. 29, 339-349.

[8] Hsing, T. (1987). On the characterization of certain point processes. *Stochastic Processes and their Applications*, Vol. 26, 297-316.

[9] Hsing, T., Hüsler, J. and Leadbetter, M. R. (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, Vol. 78, 97-112.

[10] Joe, H., Smith, R. L. and Weissman, I. (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society. Series B*, Vol. 54, 171-183.

[11] Katz, R. W. (1981). On some criteria for estimating the order of a Markov chain. *Technometrics*, Vol. 23, 243-249.

[12] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

[13] Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, Vol. 3, 119-131.

[14] Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1986). *Numerical Recipes: The Art of Scientific Computing*. Cambridge Univ. Press.

[15] Resnick, S. I. (1987). *Extreme Values, Point Processes and Regular Variation.* Springer, New York.

[16] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, Vol. 6, 461-464.

[17] Smith, R. L. (1984). Threshold methods for sample extremes. In *Statistical Extremes and Applications*, ed. Tiago de Oliveira, Reidel, 621-638.

[18] Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, Vol. 72, 67-90.

[19] Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, Vol. 4, 367-393.

[20] Smith, R. L. (1992). The extremal index for a Markov chain. *Journal of Applied Probability*, Vol. 29, 37-45.

[21] Smith, R. L. (1993). Multivariate threshold methods. Mimeo Series #2304, Institute of Statistics, Dept. Statist., Univ. North Carolina, Chapel Hill.

[22] Smith, R. L. and Huang, L. (1993). Modeling high threshold exceedances of urban ozone. Technical Report #6, National Institute of Statistical Sciences.

[23] Smith, R. L., Tawn, J. A. and Coles, S. G. (1993). Markov chain models for threshold exceedances. Mimeo Series #2303, Institute of Statistics, Dept. Statist., Univ. North Carolina, Chapel Hill.

[24] Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, Vol. 75, 397-415.

[25] Tong, H. (1975). Determination of the order of a Markov chain by Akaike's information criterion. *Journal of Applied Probability*, Vol. 12, 488-497.