Spatial models for the distribution of extremes

Eric Gilleland, Douglas Nychka, [1] and Uli Schneider
*National Center for Atmospheric Research*

## Summary

Statistical models for the occurrence of extreme or rare events has been applied in a variety of areas, but little work has been done in extending these ideas to spatial data sets. Here, a brief introduction to the theory of extreme-value statistics is given, and a hierarchical Bayesian framework is used to incorporate a spatial component into the model. Two examples are used to motivate the applications of statistics for extremes: high values of ozone pollution and high rates of daily rainfall. In the case of the ozone example, a spatial-temporal model is compared to the extreme-value model, and results for these two very different approaches are remarkably similar. Finally, this tutorial chapter highlights a number of areas of extreme-value statistics that require more research, such as algorithms to compute (or sample) posteriors from a Bayesian analysis.

## 1 Introduction

Although many statistical methods focus on representing the mean tendencies of a process or population, often the connection to a scientific context is best served by considering extreme values of a distribution. By extremes we mean observations or outcomes that have low probability of occurrence but which can be very large. Often such events are said to occur in the tail of the distribution. One surprising aspect of extreme statistics is that there is a well established and growing set of methods to model tails of a distribution. These techniques are often distinct from the usual Gaussian distribution theory, thereby meriting special treatment. A good introduction to this work is Coles (2001) and we adopt this author's notation. This chapter will motivate statistics for extremes using two examples: high values of ozone pollution and high rates of daily rainfall. Although these examples have direct application to air quality standards and climate, they serve a broader purpose because many ecological models and analyses require geophysical variables as input. Understanding the extreme fluctuations in the inputs can often be an important driver of an ecological system.

Apart from a gentle introduction to extremes, this chapter will also reinforce ideas from other chapters in this volume. The ozone pollution case study will contrast extremes derived from a space-time model with an alternative approach using spatial models for the tail of the distributions. Thus, one benefit from this case study is contrasting complex hierarchical models developed from different perspectives. This work is deliberately meant as a tutorial, and the software and data sets for analysis are accessible from the Geophysical Statistics Project homepage (`www.cgd.ucar.edu/stats`). Readers familiar with the R statistical language will be able to reproduce nearly all of the examples, figures and analyses in this chapter.

[1]Corresponding Author: National Center for Atmospheric Research, PO Box 3000, Boulder CO 80307-3000, USA

The chapter is organized as follows. Section 2 introduces extreme value theory with the generalized extreme value (GEV) distribution and its companion model, the generalized Pareto distribution (GPD). Section 3 considers the statistical models for the extreme daily ozone values over the course of a year based on monitoring data in and around North Carolina. In addition to fitting the tails of the ozone field directly, a space-time model is presented for daily ozone data. This model is then used to make inferences about extremes over the year. Section 4 demonstrates a spatial model for the GPD, and the final section has more discussion of this example.

# 2 Models for the extremes

## 2.1 Introduction and quantile plots

Much of extreme value statistics is motivated by several basic limiting results from probability theory. As motivation, consider drawing a sample of size 100 from a standard Gaussian (normal) distribution (mean 0 and variance 1) and finding the maximum. This was done 1000 times and the scaled histogram of the 1000 maxima is given in Figure 1. Note that the shape of the histogram is decidedly asymmetric and not Gaussian.

To probe the tails of this distribution, one can construct a quantile-quantile (Q-Q) plot to compare these outcomes to a standard distribution; and because this diagnostic will be used in subsequent sections, we present it here. To compare the maxima from Figure 1 to a normal distribution, the sorted maxima are plotted against the expected values under the assumption that they come from a normal distribution. If $\Phi$ is the normal cumulative distribution function and $\Phi^{-1}$ is the inverse (known as the quantile function), then the approximate expected value for the k$^{th}$ ordered data point is $\Phi^{-1}(k/(n+1))$. This is true in general for any distribution. Figure 2 (a) is a Q-Q plot of the maxima against a standard normal, and Figure 2 (b) is a Q-Q plot using the fitted GEV distribution. Under the correct distribution, one expects to see a straight-line relationship between the order statistics of the data and the expected quantiles. Clearly, this is not the case for Figure 2 (a), where the curvature indicates that the maxima are not normally distributed. Using the GEV distribution, however, a linear relationship is apparent, suggesting that this is a good approximation to the distribution of the data. Throughout this chapter we will use Q-Q plots in this way to assess the fit of a model. One flexibility of this method is that a linear transformation of the data will not affect the linear relationship (i.e. straight line) between the order statistics of the data and the quantiles of the fitted distribution in the Q-Q plot.

## 2.2 Generalized extreme value distributions

Clearly, it was not serendipity that the GEV distribution was chosen as an example for the second Q-Q plot. A classical result in probability theory is that for a wide class of distributions, the distribution of the maxima will be well approximated by the distribution known as the GEV. The cumulative distribution for the GEV distribution is given by

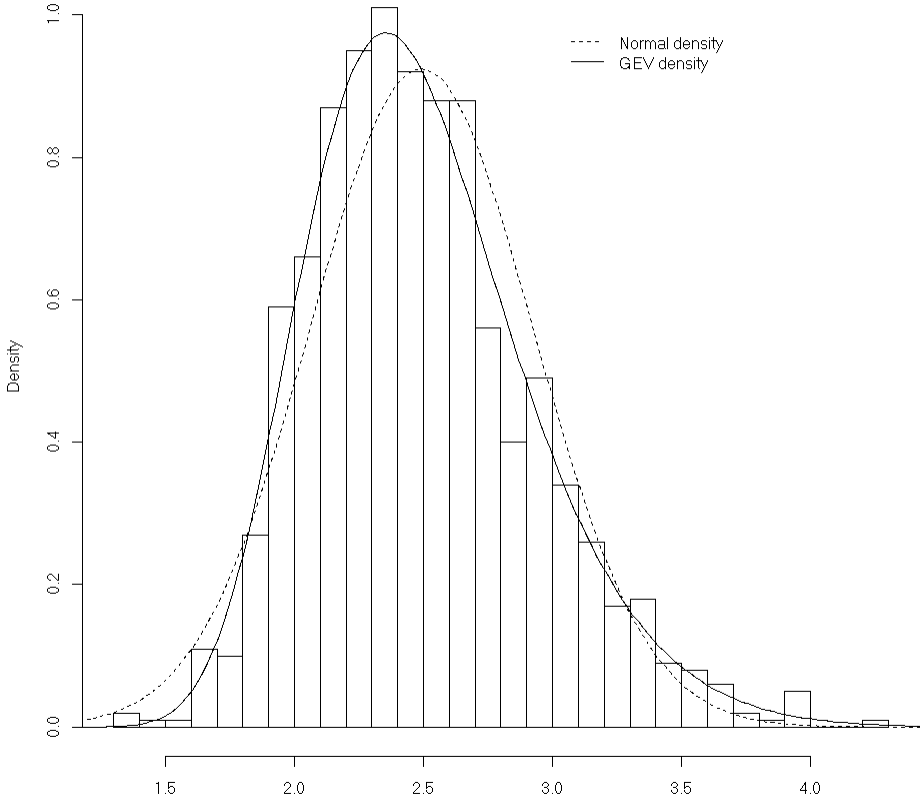


Figure 1: Histogram of maximum values taken from 1000 samples of 100 standard normal (N(0,1)) samples. Associated normal and GEV distributions are superimposed for comparison.

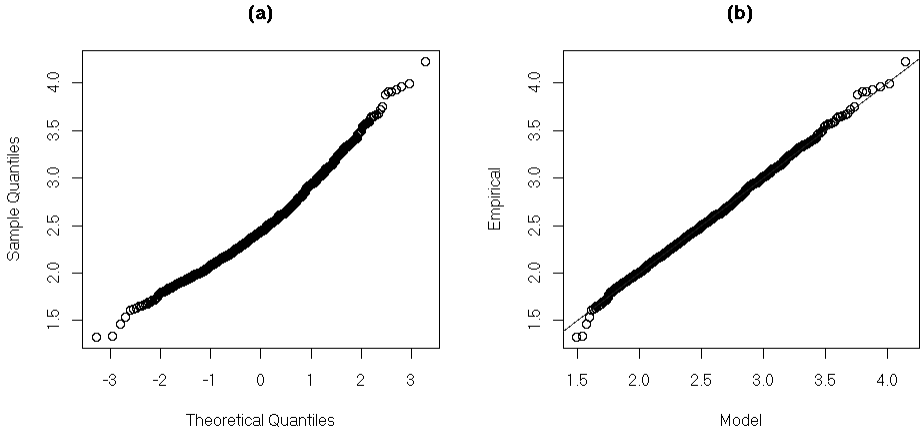


Figure 2: Q-Q plots of maxima sampled from N(0,1) against (a) standard normal (N(0,1)) and (b) GEV distribution.

$$F_{GEV}(z) = e^{-(1+\xi(z-\mu)/\sigma)^{-1/\xi}}. \tag{1}$$

The parameters are location ($\mu$), scale ($\sigma$) and shape ($\xi$). Changing the location and scale parameter of the distribution corresponds to a linear transformation of the underlying variable or data. The shape parameter controls how rapidly the probabilities approach zero for large values of $z$ (Fig. 3). When the shape is negative, the GEV distribution has bounded support with zero probability of exceeding $\mu - \sigma/\xi$.

Formally, if $M_n$ is the maximum of $n$ random variables, then for a suitable choice of the parameters, we have

$$p(M_n < z) \approx F_{GEV}(z),$$

and the error in the approximation goes to zero as the sample size $n$, becomes infinitely large.

The cumulative distribution function in (1) has an important limiting case when $\xi \to 0$, the Gumbel distribution given by

$$F_{\text{Gumbel}}(z) = 1 - e^{-e^{-(z-\mu)/\sigma}}. \tag{2}$$

Theory confirms that this is indeed the limiting approximation for maxima of independent normal random variables.

The GEV distribution also suggests a model for the tail of a distribution even when the statistics of interest are not maxima. This is known as the exceedance-over-threshold method, and involves only considering the distribution of the observations above a (high) threshold. Suppose $X$ is a random variable, such as daily precipitation at a specific location, and fix a threshold, $u$. Now consider the conditional distribution of $X$ given that it exceeds $u$. If $F$ is the distribution function for $X$, then the probability of exceeding $x$ given that $X$ is greater than $u$ is

$$p(X > x | X > u) = \frac{1 - F(x)}{1 - F(u)}$$

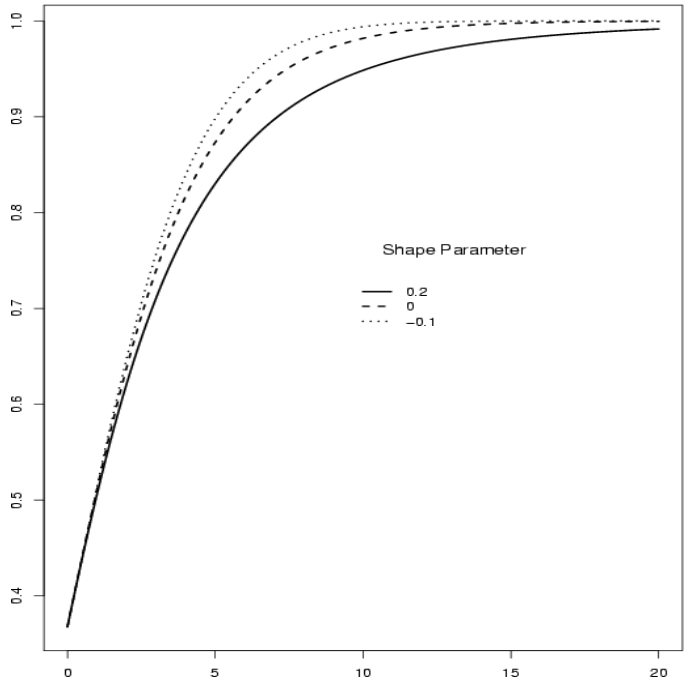


Figure 3: Cumulative distribution function for a GEV under different values of the shape parameter, $\xi$.

The same assumptions that lead to the GEV distribution also imply that this conditional distribution can be approximated by the generalized Pareto distribution (GPD) as $u$ becomes large.

$$\frac{1-F(x)}{1-F(u)} \approx \begin{cases} (1+\xi(x-u)/(\sigma))^{-1/\xi} & \xi \neq 0 \\ e^{-(x-u)/\sigma} & \xi = 0, \end{cases} \tag{3}$$

where the scale and shape parameters depend on $F$ and $u$. One important note is that the GPD approximation has limited range; so, for a given threshold, the GPD can not be extrapolated too far beyond $u$.

Given a GEV or GPD model for the tail of a distribution, a useful transformation is the quantile function. In fact, several parameter combinations may have very similar quantiles, making them more useful for comparing model fits than just the model parameters alone. Additionally, quantiles often have more meaningful physical interpretations. If $q$ is a probability and $F$ a distribution function, the $(1-q)$-quantile is defined as $x_q = F^{-1}(q)$. In other words, $x_q$ is the value such that the probability of $X$ exceeding $x_q$ is $q$. One advantage of the extreme value distributions is that the quantiles have closed form. For the GEV model, setting $y_q = \log(1-q)$, $x_q$ is given by

$$x_q = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - y_q^{-\xi}\right] & \xi \neq 0 \\ \mu - \sigma \log(y_q) & \xi = 0. \end{cases} \tag{4}$$

For the GPD, with $\zeta_u = 1 - F(u)$ denoting the probability of exceeding the threshold $u$, we have

$$x_q = \begin{cases} u + \frac{\sigma}{\xi}\left[(\zeta_u/q)^{\xi} - 1\right] & \xi \neq 0 \\ u + \sigma \log(\zeta_u/q) & \xi = 0 \end{cases} \tag{5}$$

### 2.2.1 Return levels and quantiles

In applications, the quantiles are often phrased in terms of return levels, and it is useful to indicate the connection between these two. Suppose that a GEV distribution represents the maximum over a year. The one-hundred-year return level is the $(1 - 1/100)$-quantile for the GEV, and in general, the $m$-year return level is $x_{1/m}$. In addition to a probabilistic definition, there is a time series interpretation under the assumption that annual maxima are (approximately) independent; the expected time to observe a value of $x_{1/m}$ or higher is $m$ years. Although awkward to define, the return level is common in many fields; such as hydrology and climatology. For example, a "one-hundred-year event" would refer to the .99-quantile of the distribution for annual maxima. The GPD has a similar equivalence, although care must be taken in calibrating the time periods. For example, a GPD exceedance-over-threshold distribution fit from daily observations would use the $(1 - 1/(m * 365.25))$-quantile for the $m$-year return level. Of course, in practice, it is necessary to check the assumption of independence between events, and to understand that there is often a moderate probability that the "one-hundred-year event " will occur before 100 years.

## 2.3 Fitting the GEV and GPD models to data

The most common strategy for fitting extreme value models is by maximum likelihood. For a random sample $\{Z_i, 1 < i < n\}$ assumed to follow the GEV distribution with $\xi \neq 0$, the log-likelihood is

$$l_{GEV}(\mu, \sigma, \xi) = -n \log(\sigma) - (1 + 1/\xi) \sum_{i=1}^{n} \log \left[1 + \xi \left(\frac{Z_i - \mu}{\sigma}\right)\right] - \sum_{i=1}^{n} \left[1 + \xi \left(\frac{Z_i - \mu}{\sigma}\right)\right]^{-1/\xi} \tag{6}$$

subject to

$$1 + \xi \left(\frac{Z_i - \mu}{\sigma}\right) > 0 \tag{7}$$

for all $i$. When $\xi = 0$ the log-likelihood simplifies to

$$l_{Gumbel}(\boldsymbol{Z}, \mu, \sigma, 0) = -n \log(\sigma) - \sum_{i=1}^{n} \left(\frac{Z_i - \mu}{\sigma}\right) - \sum_{i=1}^{n} \exp\left(-\frac{Z_i - \mu}{\sigma}\right). \tag{8}$$

For the GPD, let $i$ $(1 \leq i \leq k)$ index the $k$ observations (out of $n$ total) that exceed the threshold $u$. For this subset of the data, the GPD likelihood has the form

$$l_{GPD}(\boldsymbol{x}, \sigma, \xi) = \begin{cases} -k \log(\sigma) - (1 + 1/\xi) \sum_{i=1}^{k} \log\left(1 + \xi \frac{X_i - u}{\sigma}\right) & \xi \neq 0 \\ -k \log(\sigma) - \sum_{i=1}^{k} \frac{X_i - u}{\sigma} & \xi = 0 \end{cases} \tag{9}$$

In the next section, we will use daily precipitation data from Fort Collins as an example of fitting data to the GPD. The data cover a time period from 1900 to 1999, and is the same dataset analyzed by Katz et al. (2002) and Gilleland et al. (2004). For these data, extreme-value theory is especially of interest because of a destructive flood that happened in Fort Collins on July 28, 1997. The data set consists of daily values, and since there may be precipitation amounts below the annual "maximum" (that are sufficient to cause a flood), an exceedance-over-threshold approach seems to be appropriate here as this method does

not discard those values. Using a threshold of 0.395 inches, maximum likelihood estimates of the GPD for these data are found to be $\hat{\sigma} \approx 0.32$ inches (0.016 inches) and $\hat{\xi} \approx 0.2$ (0.038) (standard errors in parentheses), with associated log-likelihood of about -85.

### 2.4 Inference

Although the maximum likelihood estimator (MLE) does not have a closed form, these likelihoods are simple to maximize numerically, provided some care is taken when $\xi$ is close to zero. Typically, the probability of exceeding the threshold $\zeta_u$ which is needed to compute return levels, is estimated by the proportion $k/n$. The `extRemes` and `ismev` R packages both have functions for finding the MLEs. Based on standard likelihood theory, approximate confidence intervals can be derived for the parameters, but usually return levels are of more interest. Because the quantiles are nonlinear functions of the distribution parameters, their confidence intervals will also involve an approximation, and some details of these approximations can be found in Coles (2001). An alternative strategy for deriving confidence intervals for return levels is to use a profile likelihood that has been re-parameterized so that $x_q$ replaces one of the usual parameters. Although requiring more computation, the profile likelihood approach has the advantage that the statistical approximations for the confidence intervals are more accurate. Unlike the normal approximations based on standard errors, the profile likelihood can give confidence intervals that are asymmetric about the MLE. This feature can be important for accurate inference regarding complicated nonlinear quantities, which is the case for both the shape parameter ($\xi$) or the return level ($x_q$).

Below we give several examples using the profile likelihoods applied to the Fort Collins daily precipitation series. To draw an inference for the shape parameter, the profile likelihood is the recommended method, and it can be computed by fixing $\xi$ in (9), and then maximizing over $\sigma$. That is, find

$$l_{\text{profile}}(\xi) = \max_{\sigma} l_{GPD}(\sigma, \xi),$$

a function of $\xi$ only. In this case, there is only one parameter ($\sigma$), but in general the profiling principle is to optimize over all parameters that are not fixed. $l_{\text{profile}}(\xi)$ is now a "likelihood" measure for $\xi$, and of course, the value that maximizes this profile is the actual MLE. However, parameter values that give profile likelihoods close to the maximum are also reasonable estimates for the parameter, and this is the principle used to generate a confidence interval. Figure 4 is a graph of the profile likelihood of the shape parameter for the Fort Collins precipitation data.

The confidence set is defined implicitly through quantiles of the $\chi^2$ distribution, based on likelihood theory for large samples. A 95% confidence interval for a single parameter is constructed by taking all parameter values whose profile likelihoods differ from the maximum value of the likelihood by less than $\chi^2_{.05}(1)$. In this figure, the horizontal line marks the distance between the maximum and the $\chi^2$ critical value, and the resulting confidence interval ($\approx$ (0.14, 0.29)) is the set of parameter values whose profile likelihood exceeds this threshold; this interval is depicted graphically on the axis in Figure 4 (dashed vertical lines). One advantage of this method is that it is based on the shape of the likelihood surface. Thus, an asymmetric interval, or even a disconnected confidence set, would arise naturally. The

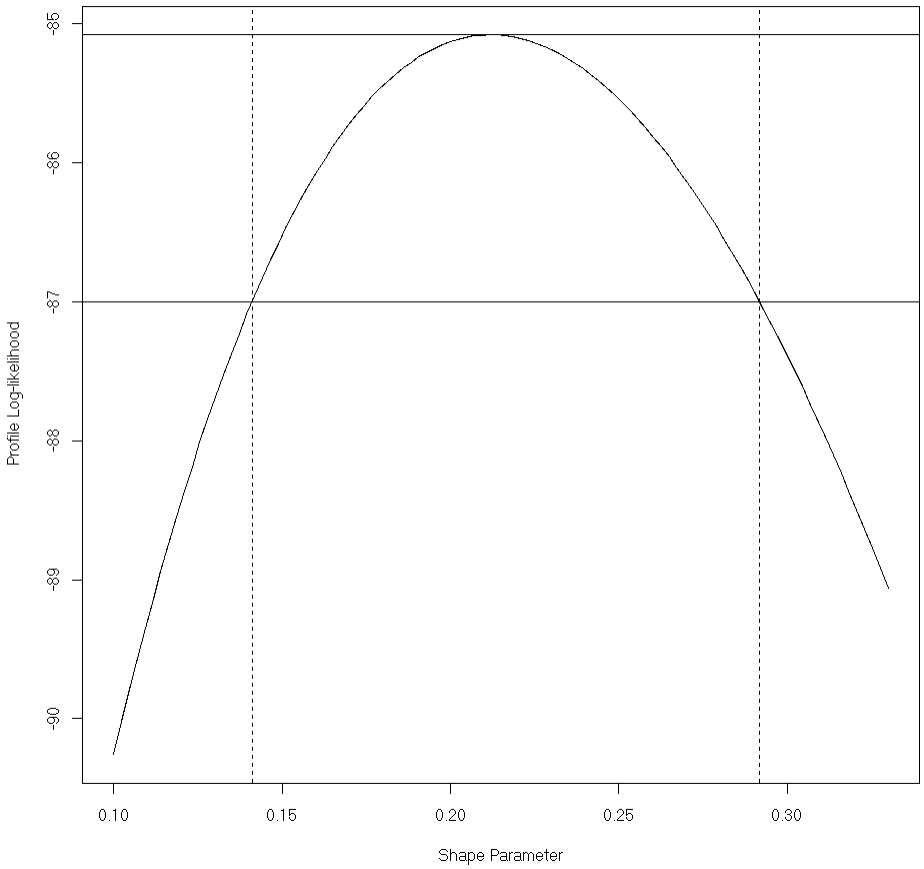


Figure 4: Profile likelihood for the GPD shape parameter ($\xi$) as fitted to the Fort Collins daily precipitation data.

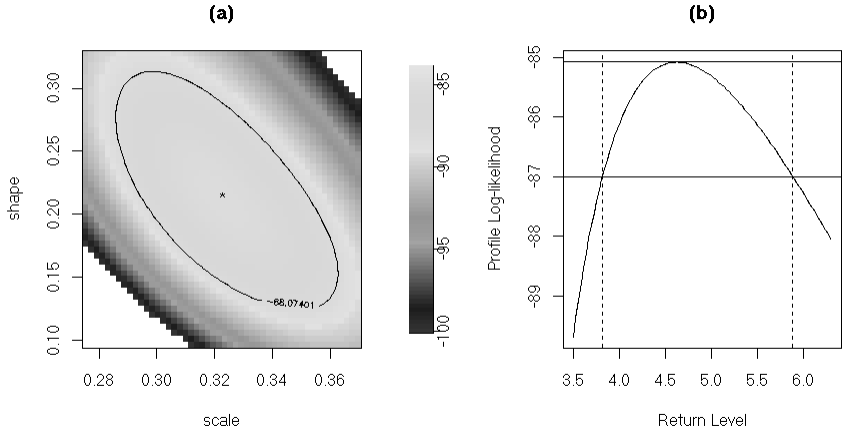


Figure 5: Based on the Fort Collins daily precipitation (inches) series, plot (a) is GPD log-Likelihood surface for scale and shape parameters with 95% confidence set determined by a $\chi^2_{.05}(2)$ critical value. The maximum of the surface is located by an asterisk. Plot (b) is the GPD profile likelihood for 50-year return level for these data. The horizontal lines are drawn at the maximum and at the $\chi^2_{.05}(1)$ critical value.

profile technique can be used in general for more than one parameter to give confidence sets. Like the one dimensional case, the profile is obtained by fixing the parameters of interest, and then maximizing over the remaining parameters. For the threshold that generates the set, the $\chi^2$ quantile must have degrees of freedom equal to the number of parameters. As an example, Figure 5 (a) is the likelihood surface for both $\sigma$ and $\xi$ for the Fort Collins precipitation data using the GPD with a 95% confidence set based on profile likelihood ideas. As a final example, the GPD is re-parameterized in terms of the 50-year return value and based on the profile likelihood we derive a 95% confidence interval (Figure 5 (b)). For this last case one obtains an asymmetric interval about the MLE.

Although we have focused on maximum likelihood in this section, we should note the very close connection with Bayesian methods for drawing inferences. If the prior distribution for the parameters of the GEV distribution or the GPD is sufficiently broad and diffuse, then it may be nearly flat over a large range of the parameters about the MLE. In such a case, the MLE and the posterior mode from a Bayes analysis will be similar, and the region of high posterior probability for the parameters will be similar to regions where the likelihood is large. On a conceptual level, the MLE frequentist intervals and the Bayesian posterior are distinct, but we simply note that the practical differences in analysis, when the prior is sufficiently uninformed, may be fairly small.

## 2.5 Threshold selection and bias

One difficulty in fitting the GPD to data is the choice of threshold. Although it seems natural to estimate this parameter by maximum likelihood along with the others, this approach

is unstable. As $u$ is varied, the number of observations changes, and this effect creates a discontinuous and often unbounded likelihood function. An optimal way to choose the threshold is an area of active research, but there exist several practical approaches based on diagnostic methods (see for example, Coles (2001) and the `extRemes` tutorial of Gilleland et al. (2004)). One simple method for choosing the threshold is based on the approximation of the GPD to the distribution's tail. It can be shown that for a GPD, the shape parameter and a certain linear transformation of the scale parameter ("modified scale") are invariant under thresholding, so in the range where the approximation is effective, one would expect that the GPD parameters would not vary significantly for different thresholds. This reasoning leads to the subjective method of estimating the GPD parameters for different thresholds and trying to discern a point where the parameter estimates stabilize. Figure 6 shows this analysis for the Fort Collins precipitation data, and suggests that a threshold beyond 0.4 inches gives a good fit. Although the choice has some element of subjectivity, this approach has the benefit of considering the sensitivity of the parameters to different thresholds.

A companion problem to selecting the threshold is the bias in the fitted distribution. For arbitrary distributions, the GEV or GPD will only be approximations. For a fixed threshold, and even with large sample sizes, the GPD will have a bias component that will not be zero, and one would expect that the bias would become larger as one extrapolates beyond the range of the observations; especially in estimating return times that are significantly longer than the period of the observation record. However, fitting both the scale and shape parameters gives some finite sample corrections to the bias. For example, even though the Gumbel distribution is the limiting distribution associated with the Gaussian other GEV distributions with a shape different from zero may give a better approximation within a limited range above the threshold. Within the range of the data, the bias in the fit can be checked informally using Q-Q plots. Assessing the validity in extrapolating the GEV or GPD beyond the data range remains an open research issue.

# 3 Surface level ozone

This section applies extreme value modeling to extreme daily ozone measured in and around North Carolina from April 1 through September 30 in 1997, and results are compared to those of a space-time model for daily ozone. The data consist of 72 stations and 184 days of observations. Each daily observation represents the maximum of the 8-hour (running) averages for that day measured in parts per billion (ppb), and of particular interest is the fourth-highest daily average (FHDA). The regulation of surface ozone pollution focuses on the FHDA being below 80 ppb for the year. For reference, Figure 7 plots these stations and, for descriptive value, superimposes a surface of the interpolated FHDA statistics.

## 3.1 Fitting high daily ozone values to the GPD

This section demonstrates fitting the GPD model to univariate data. First, a GPD distribution is fit at a single station in North Carolina to illustrate some details of assessing

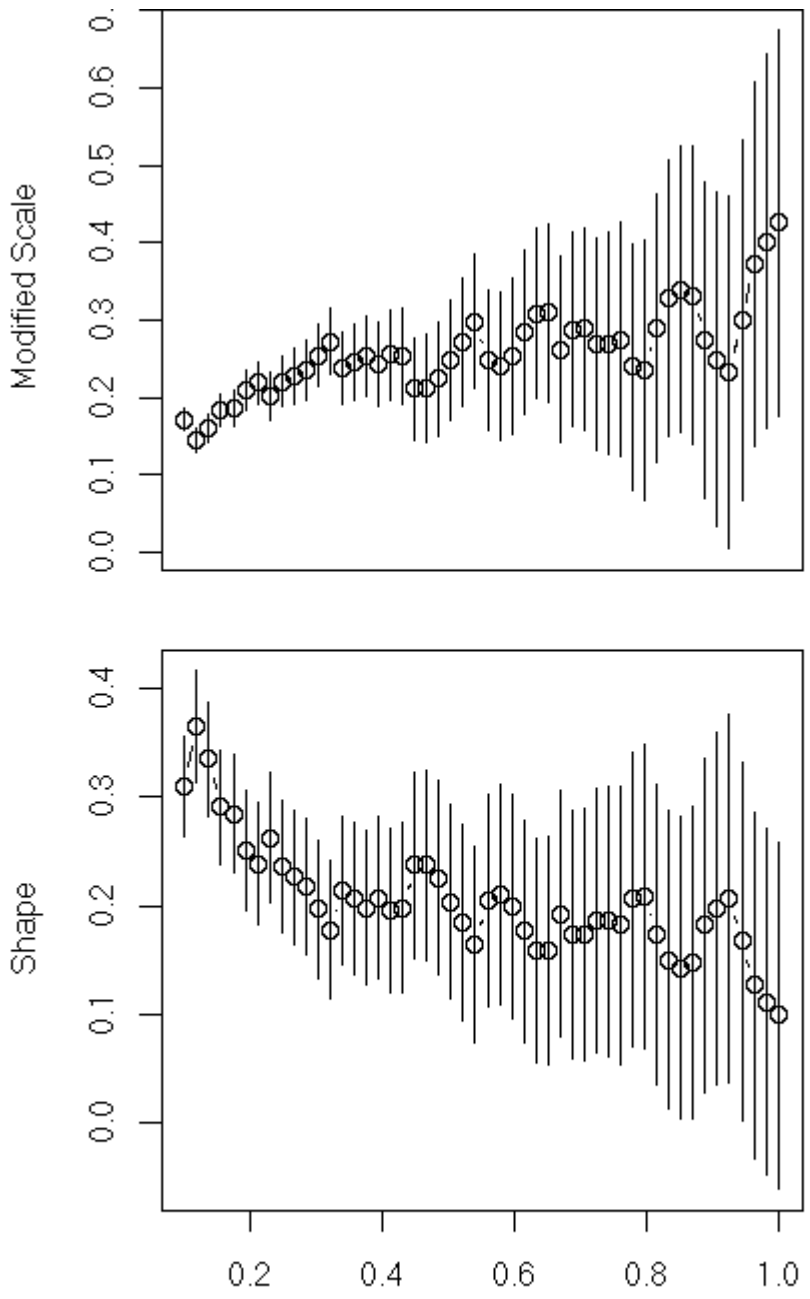


Figure 6: Maximum likelihood estimates with standard errors for GPD parameters over a range of thresholds.

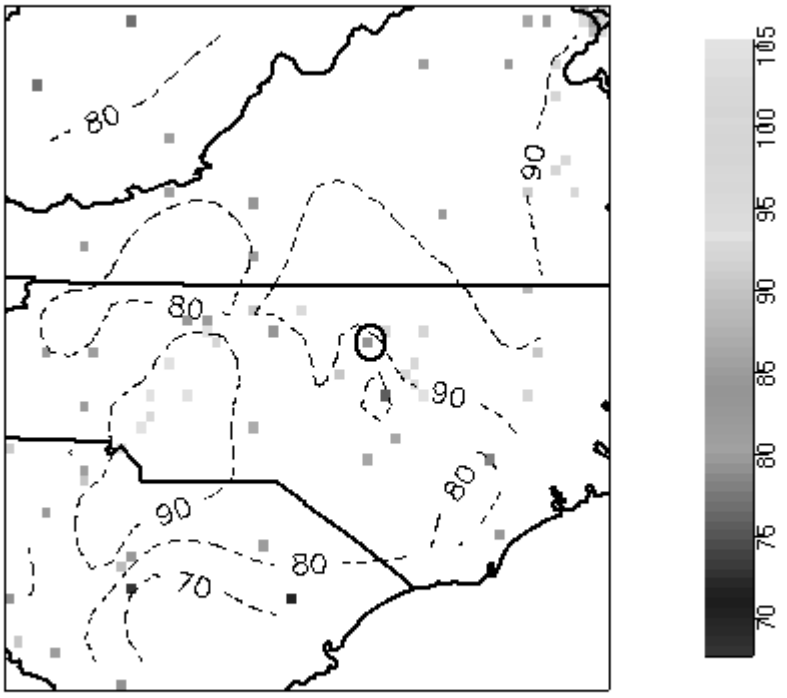


Figure 7: Locations of the 72 ozone monitoring stations used for analysis. Shadings indicate the FHDA statistic for each station for 1997 and for exploratory benefit the contours are an interpolating surface fit to the station statistics. The circled location is the station treated in depth in Section 3.

the fit. Next, separate GPD models are fit for the 72 locations to investigate the spatial variation of GPD parameter estimates.

A single station was chosen from the Raleigh-Durham-Chapel Hill metropolitan region for an in-depth analysis of a GPD fit (see Figure 7). For these data, a threshold of 60 ppb is chosen to ensure an adequate number of data points exceeding the threshold and a reasonably small variance. MLE estimates, using this threshold, are (standard deviations in parentheses) $\hat{\sigma} \approx 13.12$ ppb (2.296 ppb) and $\hat{\xi} \approx -0.35$ (0.123). The maximum log-likelihood is about -184. The Q-Q plot for this fit (Figure 8 (a)) shows that the GPD is a reasonable approximation to the exceedance-over-a-threshold distribution. The upper tail does not appear to be as good of a fit, but considering the lack of data (only 57 exceedances), we judge this to be acceptable. It also gives credence to the notion that the GPD does not fit well to data that is far from the threshold. The histogram of data exceeding 60 ppb, shown in Figure 8 (b), further supports the appropriateness of the GPD for these data because the general shape follows closely to that of the Pareto density (super-imposed).

Fitting the GPD to each of the 72 stations yields reasonably similar parameter values. Observed FHDA, computed form the daily data, and estimated FHDA, derived from the fitted GPDs, return levels are in good agreement (Figure 9). Although estimated FHDA across stations range from about 67 ppb to about 105 ppb, the interquartile range is much smaller; from about 83 ppb to about 94 ppb. The five-year return levels (Figure 9 (c)) vary much more than those of the observed or estimated FHDA; but this is to be expected for such a short return period, and the interquartile range is reasonably tight – from about 91 ppb to about 105 ppb.

### 3.2 A space-time model for daily ozone

Another method to infer extreme behavior of ozone is to sample from the desired distribution indirectly by using a spatial-temporal model developed for the daily measurements. Given a complete statistical specification for daily ozone, this model implies a specific distribution for the FHDA. Although the FHDA distribution is complicated and does not have a closed form, it can be determined empirically by Monte Carlo sampling (details of this approach are in Gilleland and Nychka (2004)). In this section, we give a short description of this model as a contrasting analysis to the extreme value analysis in Section 4.

The daily ozone data is modeled by a spatial AR(1) model. Let $y(\boldsymbol{s}, t)$ denote the daily ozone values at station $\boldsymbol{s}$ and time $t$. The spatial AR(1) model is

$$y(\boldsymbol{s}, t) = u(\boldsymbol{s}, t)s(\boldsymbol{s}) + m(\boldsymbol{s}, t), \tag{10}$$

where $s(\boldsymbol{s})$ are the marginal standard deviations at each location, $\boldsymbol{s}$, and $m(\boldsymbol{s}, t)$ account for station means and spatially cohesive seasonality. The AR(1) process $u(\boldsymbol{s}, t)$ has a mean of 0 and standard deviation of 1, and is modeled explicitly as follows.

$$u(\boldsymbol{s}, t) = \rho(\boldsymbol{s})u(\boldsymbol{s}, t-1) + \varepsilon(\boldsymbol{s}, t), \tag{11}$$

where $\varepsilon(\boldsymbol{s}, t)$ are independent over time, but follow a spatially correlated stationary, isotropic Gaussian process. The fitted spatial covariance function for the AR(1) shocks and the covariances for related spatial fields are plotted in Figure 10.

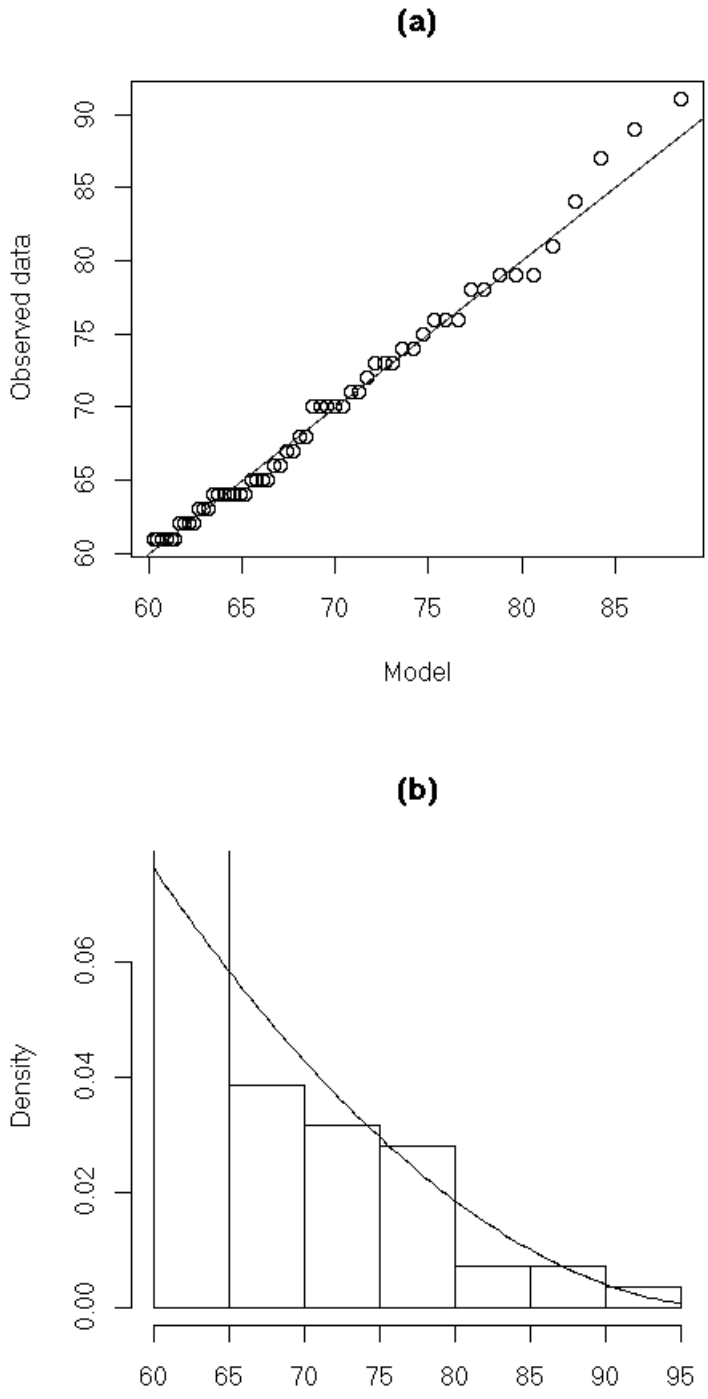


Figure 8: GPD fit for daily ozone from one monitoring station in the Raleigh Durham Chapel Hill Metropolitan region using a threshold of 60 ppb with (a) Q-Q plot for GPD and (b) histogram with associated GPD density.

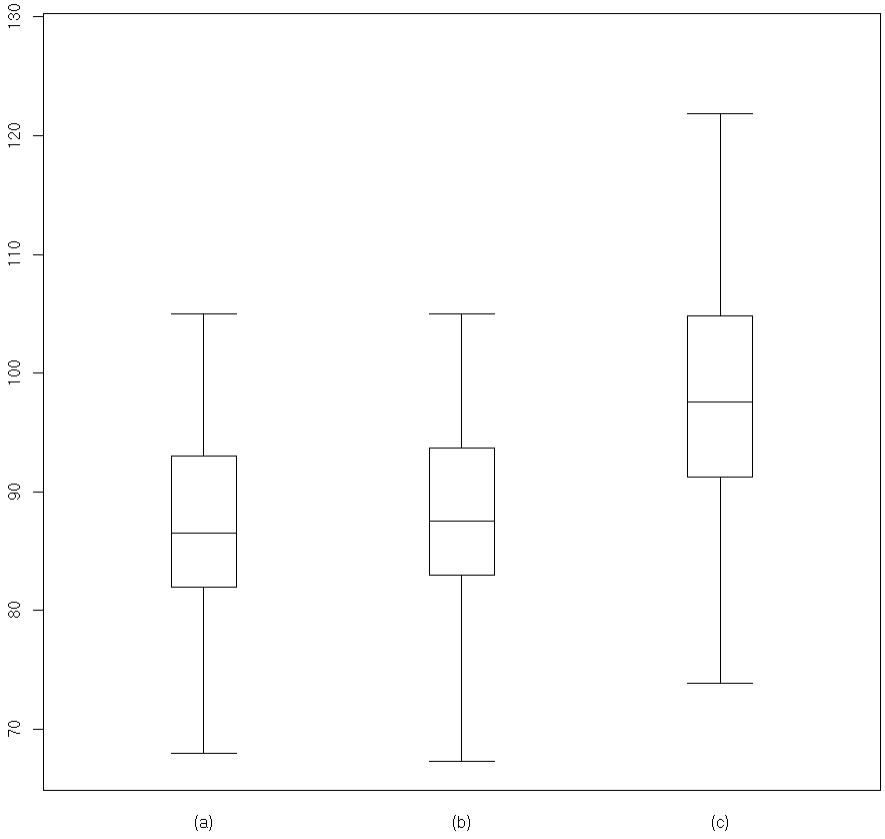


Figure 9: Box plots for 1997 (a) observed FHDA, (b) estimated FHDA return levels and (c) five-year return levels.

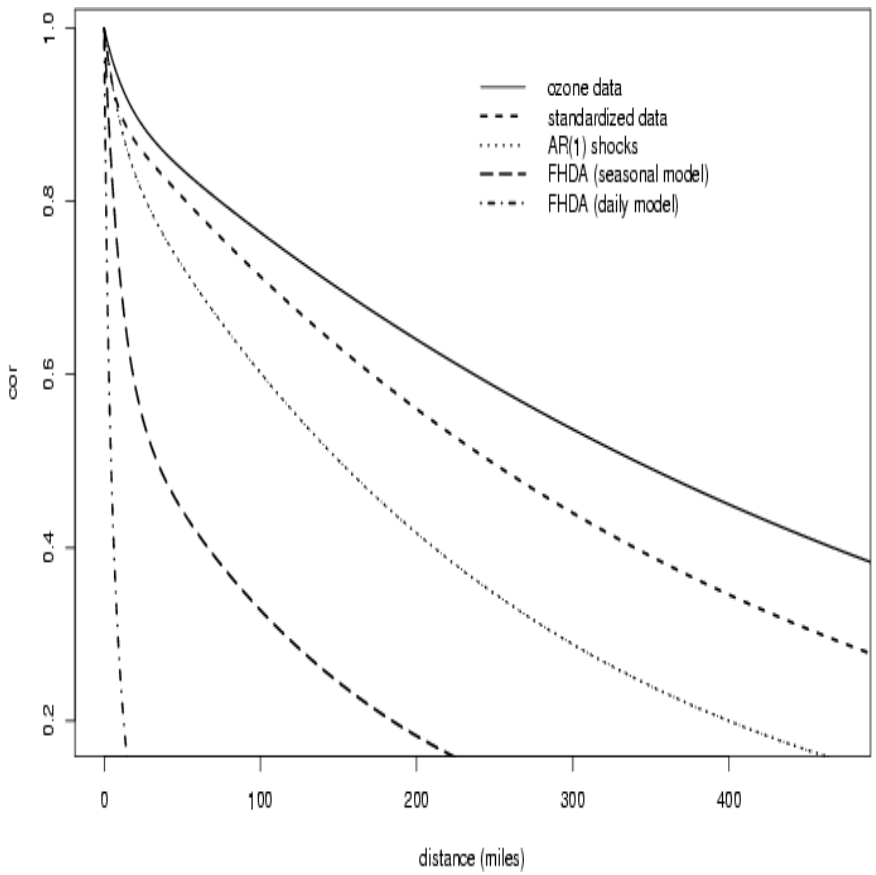


Figure 10: Fitted empirical correlation functions for original daily maximum 8-hour average ozone measurements, the standardized daily values, the estimated spatial AR(1) shocks ($\varepsilon(\boldsymbol{s}, t)$) and unconditional (seasonal model) and conditional (daily model) simulations of the FHDA field.

### 3.3 Space-time model results

One result from Gilleland and Nychka (2004) that is of importance when analyzing the extreme behavior of daily ozone is that the spatial correlations are much weaker in the tails, as is evident in Figure 10. The longest correlation ranges are present in the raw ozone measurements. The spatial dependence is reduced by standardizing the daily data as in Equation (10), and further diminished by transforming to the AR(1) shocks. The spatial covariance for the FHDA field from this model has even weaker spatial coherence, with correlations falling to .5 below 10 miles. This drastic difference between the correlation of the daily fields and the FHDA field implied by the spatial-temporal model is partially confirmed by estimating the spatial correlations of the FHDA field directly. We also used a simple geostatistical model applied to the FHDA statistics from the station. The resulting correlation function is referred to in Figure 10 as FHDA (seasonal model). The weakness of the spatial dependence of the FHDA statistic (graphed in Figure 7) will be used in formulating an approximate likelihood for the GPD model in Section 4.

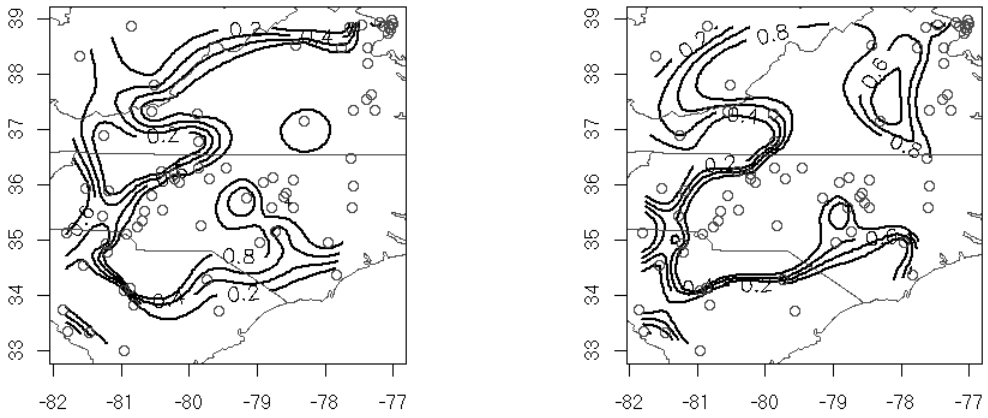


Figure 11: Probability of FHDA ozone exceeding 84 ppb estimated by (a) daily model simulations and (b) spatial GPD model.

In order to relate our analysis to the regulatory standard for ground-level ozone, we look at the probability of FHDA ozone exceeding 84 ppb. The daily space-time model computed from 720 Monte Carlo iterations (Figure 11 (a)) shows high probabilities of exceedance for most of the study region; with smaller probabilities in rural areas.

# 4 Spatial models for extremes

The analysis from Section 3.2 arrives at estimates of extreme properties of the ozone monitoring data by a space-time model based on daily ozone measurements. Most of the fitting is focused on mean and variance properties and the distribution of the tails is partly constrained by the multivariate assumptions made for the daily model. An alternative approach is to model the tail behavior of the station measurements directly. This can be achieved by fitting each station separately using the GPD exceedance over threshold model from Section 2.3. However, the number of observations is small for any one station, and one would expect significant uncertainty in the estimates because of too few observations (exceeding the threshold). Also, the individual models provide no obvious way to extrapolate to locations where ozone is not measured. One strategy to improve the accuracy and provide for spatial prediction is to include a spatial component that links the distribution for different stations. In this section, a hierarchical component is added that treats the parameters of the GPD as a smooth surface. This device is not only reasonable given the spatial dependence of the surface ozone, but also combines strength across the stations to give a more stable estimate of the tail parameters. We should note that in working through this example, there are several places where we have made simplifying assumptions in the model that may have dubious justification. Many of these could be avoided at the cost of more

complex models, but for this tutorial example, we prefer to emphasize some of the basic concepts using simpler approaches.

### 4.1 Elements of a hierarchical model

With this introduction, we assume that $\boldsymbol{s}_k$ are the locations for the ozone stations, and $y(\boldsymbol{s}, t)$ denotes the daily ozone at location $\boldsymbol{s}$ and day $t$ as in Section 3.2. The goal is to estimate the surfaces $\sigma(\boldsymbol{s})$ and $\xi(\boldsymbol{s})$ that describe how the GPD parameters change as a function of location. Based on these surfaces, the probability of an extreme ozone event can be evaluated at any point in the region.

In terms of a hierarchical model, we assume that; conditional on the values of $\sigma(\boldsymbol{s})$, $\xi(\boldsymbol{s})$ and threshold $u$; the exceedances of ozone at location $\boldsymbol{s}$ follow a GPD. It is useful to denote the probability density function (pdf) of this conditional distribution as

$$p(y(\boldsymbol{s}, t)|\sigma(\boldsymbol{s}), \xi(\boldsymbol{s}), u). \tag{12}$$

In this format, the quantities on the left side of the bar ( | ) are taken to be random, but have a distribution that is constrained (i.e., conditioned) by fixing the quantities on the right side of the bar.

The next level in the hierarchy is a statistical model for $\sigma(\boldsymbol{s})$, $\xi(\boldsymbol{s})$ and $u$. We denote the pdf for these components as

$$p(\sigma(\boldsymbol{s}), \xi(\boldsymbol{s}), u|\boldsymbol{\theta}) \tag{13}$$

In general, $\boldsymbol{\theta}$ is a vector of hyperparameters controlling the distributions for $\sigma(\boldsymbol{s})$, $\xi(\boldsymbol{s})$ and $u$, and the final stage in the hierarchy is a prior distribution on $\boldsymbol{\theta}$ (denoted $p(\boldsymbol{\theta})$).

Multiplying these pieces together gives the joint pdf

$$p(y(\boldsymbol{s}_i, t)|\sigma(\boldsymbol{s}), \xi(\boldsymbol{s}), u) \quad p(\sigma(\boldsymbol{s}), \xi(\boldsymbol{s}), u|\boldsymbol{\theta}) \quad p(\boldsymbol{\theta}). \tag{14}$$

Here $t$ $(1 \leq t \leq T)$ indexes the $T$ days and $i$ $(1 \leq i \leq M)$ indexes the $M$ station locations.

For a formal Bayesian analysis, the specification of (14) is a complete recipe for inference on the parameters. Using Bayes Theorem, the posterior for $\sigma(\boldsymbol{s})$ and $\xi(\boldsymbol{s})$ given the data $(p(\sigma(\boldsymbol{s}), \xi(\boldsymbol{s})|y(\boldsymbol{s}, t), \theta))$ can, in principle, be computed. In particular, a useful summary of the posterior distribution is the combination of parameters that has the highest probability given the observed data. This combination is known as the posterior mode. It is an elementary fact that the posterior mode can be found by maximizing the joint density in (14), and we use this equivalence to find estimates of the surfaces of the parameters. Although the basic outline of the Bayes analysis is clear, the details of the model are important. Most practical applications require a balance among the full richness of the hierarchical model,the limitations of the data and a lack of detailed prior knowledge concerning hyperparameters. This is also true of our analysis of the ozone data given in the next section.

### 4.2 Modeling assumptions for the ozone application

Under the assumption that the observations are conditionally independent over both time and space, the joint distribution of parameters and data is

$$\prod_{t=1,i=1}^{T,M} p(y(\boldsymbol{s}_i, t)|\sigma(\boldsymbol{s}), \xi(\boldsymbol{s}), u) \quad p(\sigma(\boldsymbol{s}), \xi(\boldsymbol{s}), u|\theta) \quad p(\theta). \tag{15}$$

The assumption of conditional independence is a strong one, but can be justified because extreme values tend to be less correlated than more central parts of a distribution. In particular, the results from the daily model of Section 3.3 suggest that the spatial correlation of the fourth-highest ozone value for the year is much weaker than the correlation among daily ozone measurements.

Finally, in order to give the specific form for the model in (14), we will need several additional assumptions. We assume that $u$ is specified, $\xi(\boldsymbol{s}) \equiv \xi$ is a constant, and $\sigma(\boldsymbol{s})$ is assumed to be a Gaussian random field with the form

$$\sigma(\boldsymbol{s}) = P(\boldsymbol{s}) + e(\boldsymbol{s}); \tag{16}$$

where $P$ is a fixed linear function, and $e(\boldsymbol{s})$ is a mean zero spatial process related to a Matérn covariance (Stein (1999)). $P$ is known as the spatial drift; and as a linear function, has three parameters that will be denoted by the vector $\boldsymbol{\beta}$. Creating a matrix with the constant and linear terms for the observed locations, $\boldsymbol{X}$, the spatial drift contribution to the scale parameter at the stations is the vector $\boldsymbol{X\beta}$.

The Matérn family of covariance functions has three parameters: $\sigma$, $\nu$ and $\rho$. The full set of parameters would be difficult to identify with the ozone data set, however, because we have little prior knowledge of their values, and the data set is small. Given these constraints, we will restrict $\nu$ to 2, and estimate the combination of the scale and range parameters that describes how spatial correlations vary for small distances. This function is referred to as the principle irregular term (Stein (1999) page 32), and the coefficient for this term is a combination of $\sigma^2$ and $\rho$. Here, we denote this term by $\lambda$, and note that it is also the smoothness parameter commonly used in penalized likelihood problems. This approximation matches the spatial process model associated with a second order thin plate spline (see Green and Silverman (1994) for more on thin plate splines), and there is both heuristic and theoretical support that the approximation provided with just this single parameter is adequate.

The last component of the model is the specification of prior distributions for $\xi$ and the hyperparameters $\boldsymbol{\theta}$, which includes the spatial drift ($\boldsymbol{\beta}$), and $\lambda$. A prior for these hyperparameters would lead to another level of the hierarchy that is only indirectly related to the observed data. Again, we make some simplifying assumptions based on practicality and the limitations of the data. Specifically, we take an empirical Bayes approach by not specifying priors; or, equivalently, assuming them to be improper and constant. With this simplification, finding the posterior mode can be interpreted as applying maximum likelihood to determine these parameters.

With all the assumptions included, the logarithm of the joint distribution from (15) is given by

$$\sum_{i=1}^{M} l_{GPD}(\boldsymbol{Y}_i, \sigma(\boldsymbol{s}_i), \xi) - \lambda(\boldsymbol{\sigma} - \boldsymbol{X\beta})^T(K^{-1})(\boldsymbol{\sigma X\beta})/2 - \log(|\lambda K|) + C. \tag{17}$$

Note that the terms are now additive–because of properties of logarithms–, and conspicuously absent are priors for $\xi$, $\boldsymbol{\beta}$ and $\lambda$. Here, the log-likelihood, $l_{GPD}$, is exactly the GPD log-likelihood with a threshold of 60 ppb that is developed in Section 2.3. $\boldsymbol{Y}_i$ is the vector of ozone measurements for the i$^{th}$ station, $\boldsymbol{\sigma}$ is the vector of scale parameters with i$^{th}$ element $\sigma(\boldsymbol{s}_i)$, $K$ is the covariance for the scale parameters among the station locations and $C$ is a

constant independent of the parameters. Because the data is conditioned on $\boldsymbol{\sigma}$, it is sufficient to find the maximum over this vector of parameters. The posterior mode for $\sigma(\boldsymbol{s})$ at an arbitrary location can be approximated as the conditional expectation of $\sigma(\boldsymbol{s})$ given $\boldsymbol{\sigma}$ at the observed locations, and based on the Matérn covariance for this surface. This estimate is not exact because this simplification fixes the parameters of the covariance at their mode values. For multivariate normal distributions, this conditional expectation is also the well known kriging estimate from geostatistics, and it is common practice to condition on the covariance parameters when forming a spatial estimate.

### 4.3 Posterior modes for the GPD

As an initial analysis and benchmark, the individual MLEs for each station were found under the constraint that the shape is constant, but the scale parameter can vary. For this case, $\hat{\xi} = -0.343$ and the posterior mode surface for the scale is plotted in Figure 12 (a). As mentioned above, the surface for the scale can be recovered using a spatial statistics estimate that extrapolates from the estimates of $\sigma$ at the observed locations. The spatial model we have assumed for $\sigma$ is equivalent to a thin plate spline, so the estimate of the surface simply involves an interpolation using standard spline algorithms. These results can be interpreted as the limiting case when the parameter $\lambda$ in (17) becomes very small. Although the surface in Figure 12 (a) is a useful visual benchmark, it is not believable because the surface interpolation assumes that each station's GPD parameters are known without error. In fact, we know there is substantial uncertainty in these individual estimates because of the small number of exceedances measured for each station.

The spatial analysis based on maximizing (17) combines information about the GPD scale parameters across stations. Specifically, the combination depends on the value of the smoothing parameter, $\lambda$. Because the mode is sensitive to the value of $\lambda$, we examine the estimates for some fixed choices of this parameter; and to be precise, let $\hat{\sigma}_\lambda(\boldsymbol{s})$ and $\hat{\xi}_\lambda$ be the parameter values that maximize (17) for a fixed value of $\lambda$. The plots in Figure 12 (b)-(d) show the estimates of the $\hat{\sigma_\lambda(\boldsymbol{s})}$ surface for different values of $\lambda$–$\hat{\xi}_\lambda$ does not vary significantly as a function of lambda. The sequence of surfaces illustrate why this parameter controls the smoothing. As $\lambda$ increases the surface tends to be smoother; with fewer sharp features and less resolution. Because of the linear spatial drift in the model, as $\lambda$ increases, the surface will simplify to a linear function; a plane.

Figure 13 is a plot of the profile likelihood of $\log \lambda$, and can be used to draw inferences about values for this parameter. The increasing profile indicates that the posterior is maximized for very large values of $\lambda$; and, in the limit, describes a surface for $\sigma$ that is simply a plane. The fitted plane in this case has a small gradient and little variability over the data region.

Surprisingly, this result suggests that there is little evidence for spatial structure in the scale parameter surface. These results can be contrasted with an *ad hoc* approach of smoothing the GPD MLEs directly. A simpler approach, though lacking a rigorous statistical model, is to smooth the individual estimates of the scale parameters at the

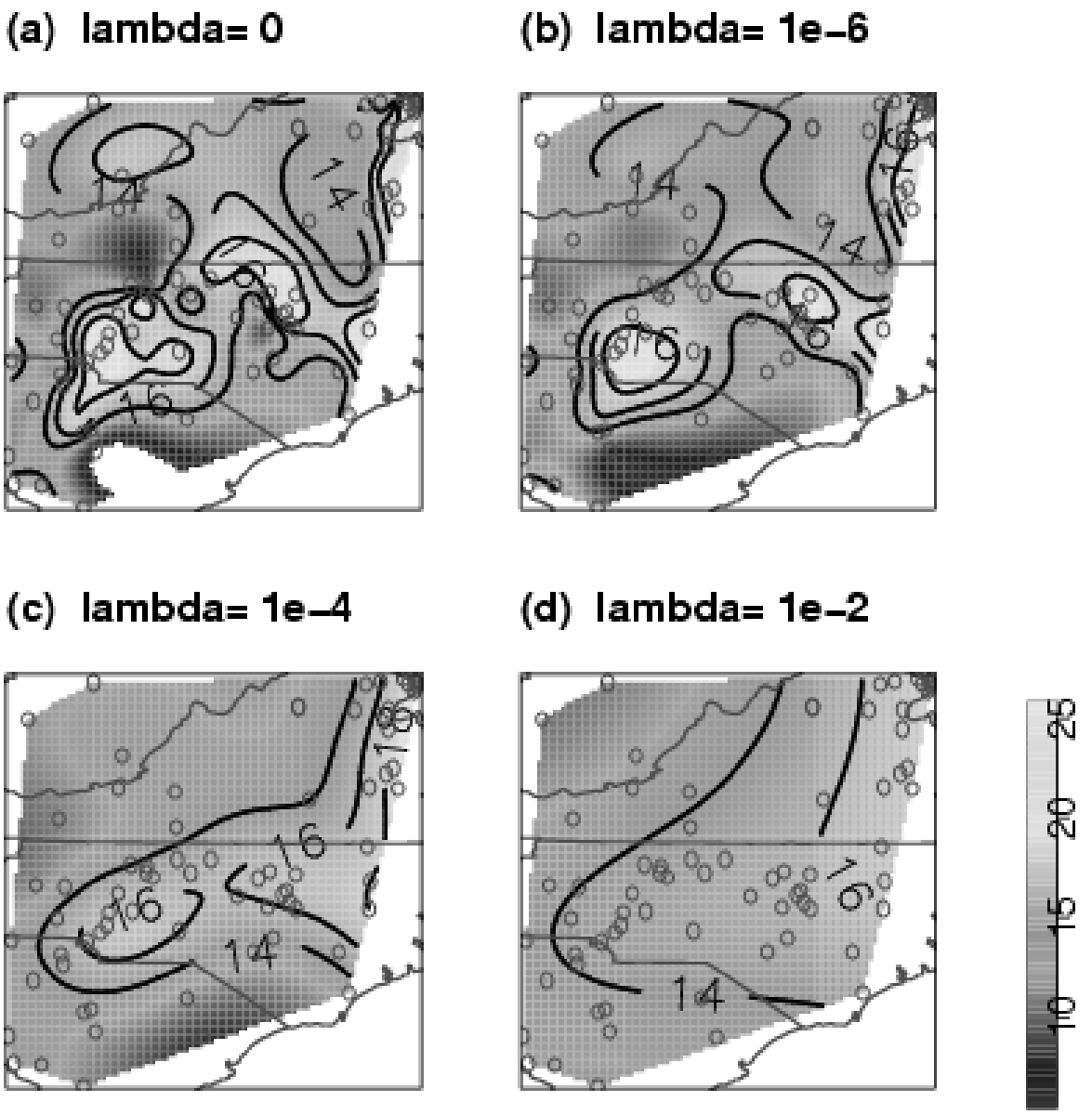


Figure 12: Estimated surfaces of the GPD scale parameter, $\sigma(\hat{\boldsymbol{s}})$, for different values of the smoothing parameter, $\lambda$.

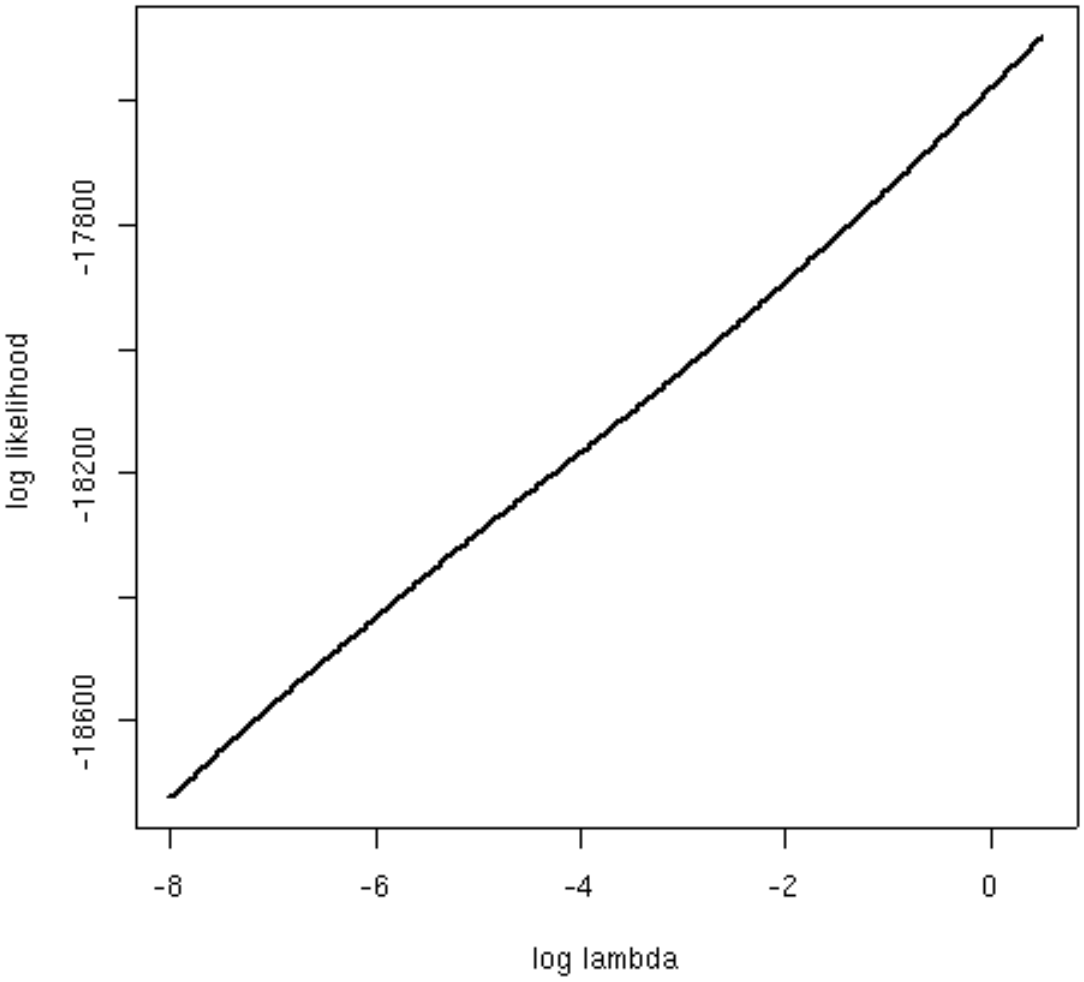


Figure 13: Profile likelihood for $\lambda$, the smoothing parameter for the surface of scale parameters.

locations using a thin plate spline. Generalized cross-validation, a frequentist criterion for estimating $\lambda$, is used to select a value for smoothing that gives a surface similar to Figure 12 (d). Note that the surface in (d) exhibits higher levels along the urban corridor through North Carolina and Virginia, and levels are lower in more rural areas such as Western Virginia. Although this interpretation is reasonable, we are unable to reconcile this with the profile likelihood that suggest that $\sigma$ has little spatial dependence.

Assuming that the intermediate value of $\lambda$ depicted in Figure 12 (d) gives a useful summary of the monitoring data, we consider a more interpretable functional of the GPD distribution. Recall that for meeting proposed EPA air quality standards it is important that the FHDA fall below 84 ppb. Figure 11 (b) is the probability of the FHDA exceeding 84 ppb estimated from the spatial GPD model. Here, the probability of a location exceeding the threshold of 60 ppb is estimated from a thin plate spline fit to the empirical probabilities from each station. This quantity is $\zeta_u$ from Section 2. The surface of probabilities, $\zeta_u(\boldsymbol{s})$, is combined with the surface of scale and shape parameter estimates to estimate the probability that daily ozone exceeds 84 ppb. Under the assumption of independence between daily exceedances, the binomial distribution is used to calculate the probability that the FHDA exceeds 84 ppb (i.e., four or more events out of 184). Despite one small area where the daily model predicts a very high probability of exceeding 84 ppb contrary to the extreme value model, which predicts very low probability for this area, the surfaces in Figures 11 (a) and (b) are surprisingly similar indicating a broad region across the Southeastern United States in 1997 where there is high probability the FHDA will exceed 84 ppb. Areas of lower probability tend to be in more rural areas or at the edges of this region.

### 4.4 Extensions and discussion

There are several extensions to this work that could improve the model, or give more accurate estimates. With more data, one could consider a more flexible covariance model for $\sigma$, and add a spatial component for $\xi$. One way to accumulate more observations is to extend the analysis over multiple years; there are five years of data that we consider for the daily model. One difficulty with multiple years is that ozone levels would need to be adjusted by covariates such as meteorology and time trends. Another extension is to include a link function for $\sigma$, such as the exponential, in order to preserve positivity of $\sigma$, and possibly to give a better approximation of a Gaussian field. Finally, by adding proper priors to this analysis, it may be possible to sample from the posterior to obtain a Monte Carlo approximation to the posterior distribution. A Markov Chain Monte Carlo approach to sample the posterior is an area of our current research.

## 5 Discussion

This chapter has provided an introduction to extreme value statistics, and then applied these ideas to a spatial example drawn from air quality. Part of our interest in this analysis using extremes was to compare the results with a more conventional space-time model that focuses on the central part of the distribution of ozone measurements. For interest only in the regulatory statistic, the probability that the FHDA exceeds 80 ppb in a year is similar whether one uses the extremes or space-time modeling approach–at least for our study region in 1997. One interesting feature of this correspondence is that the approaches are

very different in character, and involve very different assumptions. The extremes approach largely ignores temporal and spatial dependence conditional on the parameters of the GPD, but is more flexible in representing the larger values of observed ozone. The space-time approach is a hierarchical model that represents the daily dependences of ozone over time and its correlation over space, but it relies on normal distributions for the daily distributions. The agreement between the surfaces in Figure 11 suggests that for both approaches the assumptions are reasonable.

As highlighted throughout this chapter, there are many areas of extreme value statistics that need more statistical research; including algorithms to compute (or sample) posteriors from a Bayesian analysis. A key step would be the ability to sample the surface of GPD parameters from a posterior such as that developed in Section 4. This would allow for quantifying the uncertainty in the estimated parameters and the subsequent quantities based on the GPD; such as return times and exceedance probabilities. Despite many open methodological questions, we feel that there is much benefit from an extremes perspective. In particular, if one is interested in extreme events, it may be possible to avoid some of the complexity of the spatial and temporal dependence that is ordinarily associated with the majority of the measurements.

## References


Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, London.

Gilleland, E., Katz, R. W., Young, G., 2004. Extremes Toolkit: Weather and Climate Applications of Extreme Value Statistics – R software and accompanying tutorial. `http://www.assessment.ucar.edu/toolkit/index.html`.

Gilleland, E., Nychka, D., 2004. Statistical models for monitoring and regulating ground-level ozone. Evironmetrics (Special Issue), submitted.

Green, P.J. and Silverman, B.W. Nonparametric Regression and Generalized Linear Models. Chapman and Hall, 2-6 Boundary Row, London SE1 8HN, UK, 1994.

Katz, R. W., Parlange, M. B., Naveau, P., 2002. Statistics of extremes in hydrology. Advances in Water Resources 25, 1287–1304.

Stein, M., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.