FBMIS Volume 1, 2005, 1-10 ISSN 1740-5955
http://www.fbmis.info/A/5_1_RameshNI_1

SEMI-PARAMETRIC ANALYSIS OF EXTREME FOREST FIRES

N.I. RAMESH

School of Computing and Mathematical Sciences
University of Greenwich, Park Row, Greenwich, London SE10 9LS, UK
N.I.Ramesh@gre.ac.uk

(*Submitted 18th Dec. 2002; Accepted 1st Dec. 2004; Published 18th March 2005*)

Abstract. Forest fires can cause extensive damage to natural resources and properties. They can also destroy wildlife habitat, affect the forest ecosystem and threaten human lives. In this paper incidences of extreme wildland fires are modelled by a point process model which incorporates time-trend. A model based on a generalised Pareto distribution is used to model data on acres of wildland burnt by extreme fire in the US since 1825. A semi-parametric smoothing approach, which is very useful in exploratory analysis of changes in extremes, is illustrated with the maximum likelihood method to estimate model parameters.
Keywords: Forest fire, extremes, semi-parametric smoothing, wildland.

1 INTRODUCTION

The occurrence of forest fire is a common feature in many countries around the world. They cause extensive damage to natural resources, wildlife habitats and properties every year. A fire in the forest may occur either due to natural causes in situations such as lightning strikes, extreme dry weather conditions, accumulation of forest fuel as a result of severe drought etc or may be caused by human. An intense wildfire has the potential to burn over 100,000 acres in just one day. Depending on its strength, geographical location, the local whether conditions, and the efforts of the fire fighters a fire can burn for hours, days or even weeks. Some wildfires are so hard to control that they could burn for months with devastating effects before being put out by natural causes such as a seasonal rain. The effects of extreme wildland fire can retard the natural development of plant communities, alter species diversity, and interact with other physical and biological systems.

Fire also plays a beneficial role in wildlands in that it helps to recycle nutrients, regenerate plants and reduce high concentrations of fuels that contribute to disastrous wildfires. In modern days, land managers recognise the role that wildland fire plays in ecosystems and through careful planning they can manage wildland fires for natural resource benefits. For this reason resource managers sometimes start fires (prescribed fire) or allow naturally occurring fires to burn under very specific conditions. However, a prescribed fire, which is any pre-planned fire with predetermined boundaries to meet specific land management objectives, would not normally cause extreme damage.

In this paper we present a semi-parametric modelling approach to model extreme wildland fire that cause extensive damage to natural resources and properties. Although an extreme wildfire usually last for a few days or weeks, ignoring its duration or by taking its starting time as a point event, the arrival times of extreme fire may be thought of as a point process evolving in time. In this respect one could fit stationary point process models for the occurrence of extreme fires. However, the characteristics of extreme fire over a long period of time may change due to increased land usage, changes in land management policies to reflect modern human needs, better fire prevention practices and education. Changes in global weather cycles will also cause the extremes to change over time. All of these factors will certainly introduce a trend in the behaviour of extreme fire. Point process models for extremes that account for trend in extreme quantities are required to analyse the data. Models of this type will allow us to make better inference of the properties of extreme fire. Models with specified parametric trend, however, would not always describe the features in the data adequately. Therefore, at least in the context of exploratory analysis of trend in extremes, it is desirable to adopt a semi-parametric modelling approach. This can also help us select a more appropriate form of the trend in subsequent models for further analysis.

This paper attempts to apply a semi-parametric smoothing approach, useful in the exploratory analysis of extreme forest fires, using a point process model for extremes for which the parameters vary with time. Section 2 describes the data on extreme wildland fires in US. The statistical model and the smoothing method which uses local likelihood are described in Section 3. Results of the analysis are presented in Section 4 whereas the Section 5 completes the paper with a discussion.

2 FOREST FIRE DATA

We start our analysis by describing the data on wildland fire in America, obtained from the web pages of National Interagency Fire Centre, Idaho, USA. Every year about 100 to 250 thousand wildland fire incidents are recorded throughout the US. These wild fires burnt about four million acres, on average, each year during the second half of the last century. Figures 1 and 2 show the number of wildland fires and acres burnt for each year from 1960 to 1999 in US. The yearly number of fires varies a lot over the period with a notable increase in the seventies and a gradual decline during the eighties. However, the amount of acres burnt seems to be moving about more or less at the same level. A substantial proportion of this damage is due to natural causes such as extreme dry weather conditions, high temperature, lightening. These plots give us some idea of the recurrence nature of the wildfire and also their potential to sustain extensive damage to natural resources. The yearly amount of acres burnt during this 40 year period appears to be a stationary time series regardless of the fluctuations in the annual number of fires. The amount of acres burnt in extreme forest fires, however, tells us a different story.

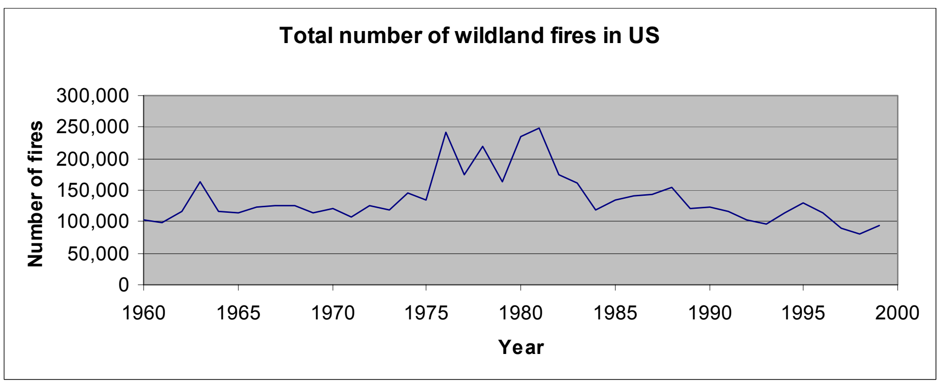

Figure 1. Total number of yearly wildland fires in the US.

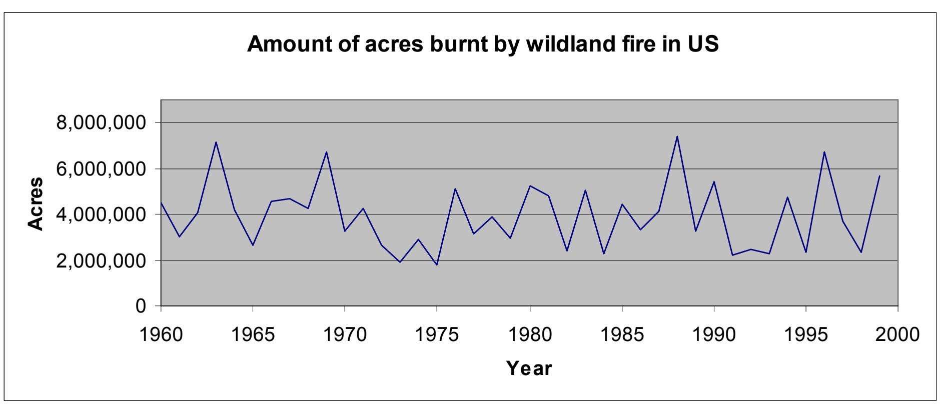

Figure 2. Yearly amount of acres burnt by wildland fires in the US.

Our interest in this paper lies on modelling extreme wildfires. So we will now turn to some of the most serious wildland fires in US, i.e. extreme wildfire. Some were significant because of their size, others because of the value of the resources lost. Some small, but very intense, fires were important because of the loss of lives and property. There have been 32 historically significant wildland fires in US history during the period from 1825 to year 2000. In this paper, we will consider these as extreme wildfires and model their occurrences and the amount of acres they burnt. As the amount of acres burnt by the fire called *Hinckley* in Minnesota in September 1894 is undetermined we exclude this from the analysis - this fire was significant as 418 lives were lost. There have been larger fires than some of those studied here, but few or none with greater impact on lives and resources. The month and year of occurrence of these fires were available for most of them along with the acres burnt. When the month is not available the fire is assumed to have taken place in the middle of the year in our analysis. This will have little effect on the results as our study period covers a long period of 176 years and there were cases where the fire lasted for three months.

Figure 3 displays the times of extreme fire and the amount of acres burnt (in thousands) since 1825. We can immediately notice two aspects of the extreme fires from the figure. Firstly, there appears to be an increasing trend in the occurrence of extreme fire. Secondly, the acres burnt seem to be decreasing with time. These aspects suggest that extreme value models which accommodate changes in time are required to analyse the data. Note however that unlike the data in Figs 1 and 2, which covered only about the last forty or so years of the 20th century, this data on extreme fires cover a longer period going back to 1825. One point for concern here is that the reporting behaviour may have changed over this period. For example, if only large fires were recorded historically (in the early days) and all fires are being recorded in recent years then it would affect not only the rate of occurrence but also the amount of damage caused. This leads to a difficulty in statistical modelling as the observed pattern may have occurred as a result of changes in the level of recording accuracy. In the absence of any information on recording accuracy, we continue our analysis on the assumption of consistent recording practice through out the period covered by the sample.

We now describe smooth trend model for extremes that allow the parameters to vary smoothly with time in the next section.

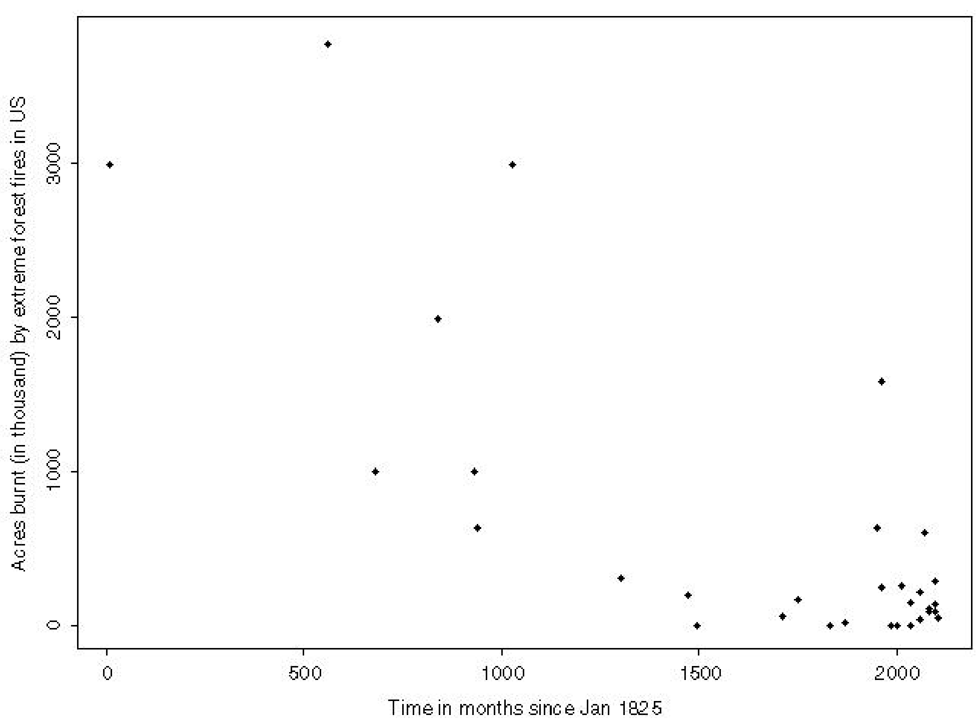

Figure 3. Acres burnt in extreme wildland fire incidences in the US since 1825.

3 MODEL AND METHODOLOGY

In statistical analysis of extremes we usually model the annual maxima using the generalised extreme value (GEV) distribution. However, in some applications data are not observed or recorded in the form of annual maxima. In these cases, a point process characterisation based on threshold exceedances can be used. The parametric forms of the GEV models can also be derived from this point process modelling approach.

3.1 Point process model. Following Smith (1989), we now describe a point process model for extremes that result from a limiting characterisation of exceedances of stationary data over a high threshold.

Let $X_1, \ldots, X_m$ be a stationary series. Consider the pattern in the two dimensional plane with points at (x, y) co-ordinates

$$(\,j/(m+1),\, a_m(X_j - b_m)\,), \quad \text{for } j=1,\ldots, m. \tag{1}$$

Then, for suitable choices of a_m and b_m, the pattern of events above a threshold at $y = y_o$ converges to an inhomogeneous Poisson process with intensity

$$\Lambda\{(t_1,t_2)\times(y_0,\infty)\} = (t_1 - t_2)\left(1+\kappa\frac{y-\mu}{\sigma}\right)_+^{-1/\kappa} \tag{2}$$

where the interval (t_1, t_2) on the x-axis is a subset of $[0,1]$, and $y > y_0$. It follows from this that events in non-overlapping subsets of $[0,1]$x(y_0, ∞) are independent, and a likelihood for the parameters μ, σ, and κ based on such data can be written down. Smith (1989) used this approach to detect changes in ground-level ozone in Texas.

Suppose that we observe extreme events at $(t_1,y_1),\ldots,(t_n,y_n)$ in the region $[0,1]$x(y_0, ∞). Then the likelihood of the data is expressed as

$$\exp[-\Lambda\{[0,1]\times(y_0,\infty)\}]\prod_{j=1}^{n} d\Lambda\{(t_j,y_j)\} = \exp(-\Lambda_0)\Lambda_0^n \times \prod_{j=1}^{n}\frac{d\Lambda\{(t_j,y_j)\}}{\Lambda_0}, \tag{3}$$

where Λ_0 denotes $\Lambda\{[0,1]\text{x}(y_0, \infty)\}$.

The above likelihood expression has two parts. The first part shows the contribution for the number of events N observed, whose distribution is Poisson with mean Λ_0. In the second part, conditional on $N = n$, there are independent contributions for each of the (t_j, y_j). These are generalised Pareto distributions (GPD) with survivor function

$$\Pr(Y \geq y \mid Y \geq y_0) = \left\{1+\frac{\kappa}{\sigma'}(y-y_0)\right\}_+^{-1/\kappa} = 1-G(y) \tag{4}$$

where $\sigma' = \sigma + \kappa(y_0 - \mu)$, and the range of y is that for which this probability is valid. A detailed account of inference based on this distribution is given by Davison and Smith (1990).

In most applications the quantiles of the fitted distribution are of interest as they allow us to make predictions about the level exceeded once every *1/p* years on average. This is called the *return level*. The return level can be calculated by noting that the events in [0,t) follow a Poisson distribution at rate λt and therefore

$$1-p = \text{P}\,\{\text{Maximum over } (0,t) \leq y\} = \sum_{n=0}^{\infty}\frac{e^{-\lambda t}(\lambda t)^n}{n!}\;G(y)^n = \exp\{-\lambda t\,(1-G(y)\;\}. \tag{5}$$

Simple algebra then shows that the *1/p* year return level is given by

$$y_p = y_0 + \left[\left(\frac{-\log(1-p)}{\lambda t}\right)^{-\kappa} - 1\right]\frac{\sigma'}{\kappa}. \tag{6}$$

The usual interpretation of y_p is no longer valid in a non-stationary context and we think of it as a quantile of the distribution of y in the current year.

3.2 Models incorporating trend. One common approach to model the trend in extremes is to allow the model parameters to vary with time, using a specified parametric function. Very often the location parameter or the rate of arrival is taken as

$$\lambda(t) = \beta_0 + \beta_1 t \tag{7}$$

or

$$\lambda(t) = \alpha + \beta \mathrm{e}^{-\gamma t} \tag{8}$$

depending on the application in hand. The first incorporates a linear trend in λ whereas the second accounts for non-linear trend. Similar trend can be included in the parameters σ, κ if necessary. Other forms of parametric functions can also be used. These parametric models have their own drawback, as they restrict the model parameters to certain form, and are not always flexible. For example, in some applications it is not reasonable to take the trend to be linear throughout the period covered by the data. In such circumstances, including our extreme fire data (as noted in Figure 3), a semi-parametric approach would be a better option and we describe this below.

3.3 Semi-parametric modelling of trend. This section describes a semi-parametric approach to model trend in extremes using local likelihood fitting. Local regression models and kernel smoothing have been widely used in data analysis; examples include Cleveland and Devlin (1988), Bowman and Azzalini (1997). Local likelihood fitting is related to kernel density estimation. We now describe this briefly for this point process modelling approach. Davison and Ramesh (2000) give a detailed account of this for generalized extreme value distribution.

Suppose that exceedances of y_0 occur as a inhomogeneous Poisson process with rate $\lambda(t)$. Conditionally on the occurrence of an observation y at time t, the distribution of y is taken to be GPD, with survivor function

$$\left\{1+\frac{\kappa(t)}{\sigma'(t)}u\right\}_{+}^{-1/\kappa(t)} \tag{9}$$

where $u=(y-y_0)$ is the threshold exceedance.

Note that we have re-scaled the time axis to [0,1]. Suppose that the rate, scale and shape parameters vary with time according to smooth functions $\lambda(t)$, $\sigma(t)$ and $\kappa(t)$. Let $w(u)$ be a symmetric kernel which is positive only in the interval $u \in [-1,1]$, for which $w(0) = 1$. In addition to this, we choose the kernel such that $w(u)$ descends smoothly and monotonically to zero as $|u| \rightarrow 1$. An example is the Epanechnikov kernel $w(u) = (1-u^2)_+$, but there are many other possibilities.

In our semi-parametric approach, we assign weights $w_j(t) = w\{(t_j - t)/h\}$ to case j, in estimating the parameter functions $\lambda(t)$, $\sigma(t)$ and $\kappa(t)$ at t. Therefore, only the cases for which j lies inside the interval with end points $(t-h)$ and $(t+h)$ will contribute to estimation at t, and those closer to the end points of this interval will contribute less than those near t. One simple approach to apply these weights is based on the likelihood fitting of a local polynomial model (Fan and Gijbels, 1996), whereby the parameters λ, σ, and κ of the distribution are replaced by low-degree polynomials in $(t_j - t)$.

Since we expect the extremes to change rather slowly, in our application we take, for $j = 1,\dots,n$ and t in [0,1],

$$\lambda(t) = \beta_0 + \beta_1 (t_j - t) \tag{10}$$

$$\sigma(t) = \exp\{ \gamma_0 + \gamma_1 (t_j - t) \} \tag{11}$$

$$\kappa(t) = \delta_0 + \delta_1 (t_j - t) \tag{12}$$

We then use local smoothers to estimate the parameter functions $\lambda(t)$, $\lambda(t)$, $\sigma(t)$ and $\kappa(t)$. Note here that, as pointed out in Section 3.1, the likelihood (3) has two parts and we estimate their parameters separately.

One way to estimate $\lambda(t)$ is to fit Poisson regression model, locally with weights $w_j(t)$, to the number of points in small equal-sized boxes in the time axis. This is easily obtained by using a generalised linear or additive model routine (McCullagh and Nelder, 1989; Hastie and Tibshirani, 1990).

We then use the local polynomial smoothing described above to estimate $\sigma'(t)$ and $\kappa(t)$. Suppose that exceedances of sizes $u_1, \ldots, u_n$ occur at times $t_1, \ldots, t_n$; noting that $0 \le t_j \le 1$. Then the local log likelihood at t is written as

$$\ell(\theta;t) = -\sum_{j=1}^{n} h^{-1} w\left(\frac{t_j - t}{h}\right)\left[\left\{\frac{1}{\kappa(t_j)} + 1\right\}\log\left\{1 + \frac{k(t_j)u_j}{\sigma'(t_j)}\right\}_{+} + \log\sigma'(t_j)\right], \quad (13)$$

where $u_i = (y_i - y_0)$. The local estimates of the parameters $\sigma'(t)$ and $\kappa(t)$ are obtained from this equation using numerical maximisation.

We denote the new parameters by $\theta = (\gamma_0, \gamma_1, \delta_0, \delta_1)$. Then the log likelihood for θ at t is given by

$$\ell(\theta;\ t\ ) = \sum_{j=1}^{n} w_j(t)\ell(\theta; y_j), \quad (14)$$

where $l(\theta;y_j)$ is the log likelihood contribution from the j^{th} observation y_j. In the above expression, the summation need only be over values for j for which $w_j(t)$ is positive. Maximisation of this equation numerically with respect to θ yields the local maximum likelihood estimates of $\theta(t)$, where we suppress the dependence on h and $w(.)$. The fitted parameters of this model at $t = t_j$ are

$$\hat{\sigma} = e^{\hat{\gamma}_0},\ \ \hat{\kappa} = \hat{\delta}_0 \quad (15)$$

4 RESULTS

We now apply the smooth trend model, based on the point process modelling approach, described in Section 3 to model the extreme wildfires displayed in Figure 3. In this study, in the absence of information on all other wildfires, we consider these extreme fires as exceedances of a high threshold in the amount of acres burnt. As mentioned earlier, there appears to be an increasing trend in the occurrence of extreme fires and a downward trend in the amount of acres burnt. Hence we include trend components in our model for both the occurrence of extreme fire events and the amount of acres burnt.

We start our analysis with the occurrence times of extreme fires. The sample data covers a period of 2112 months (176 years) during which 32 extreme wildland fires occurred. As the monthly data are very sparse, we first fit a generalized additive model to the yearly counts. This is shown in the upper right panel of Figure 4 on yearly scale. The fitted intensity is then interpolated to obtain estimates of the monthly rate of occurrence $\lambda(t)$ and this is shown in the bottom left panel. The interpolation is done using the function smooth.spline in Splus. To assess how well the model fits the data, we calculate the corresponding cumulative rate of occurrence, $\Lambda(t)$, from the interpolated monthly rates.

The cumulative rate agrees well with the empirical distribution of occurrence times, normalized to lie in the unit interval, and shown in the lower right panel of Figure 4. These times would be uniform if the process of extreme fire incidence was stationary, but there is a heavy increase in their rate toward the end of the period after a slow start. This departure from uniformity is strongly confirmed by the Kolmogorov-Smirnov test (Cox and Lewis, 1966).

 ISSN 1740-5955

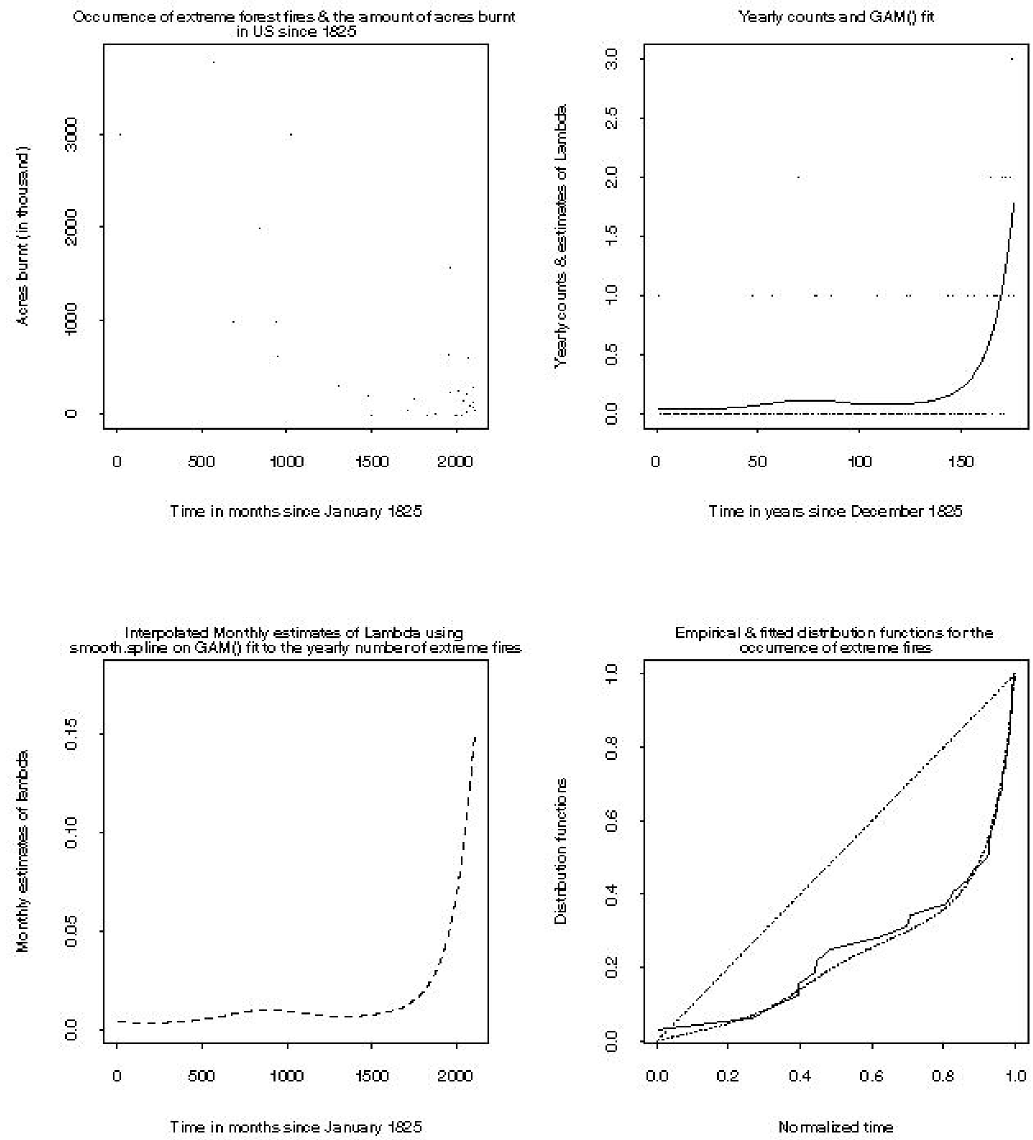

Figure 4. GAM fit for yearly number of extreme wildland fire in the US since 1825.

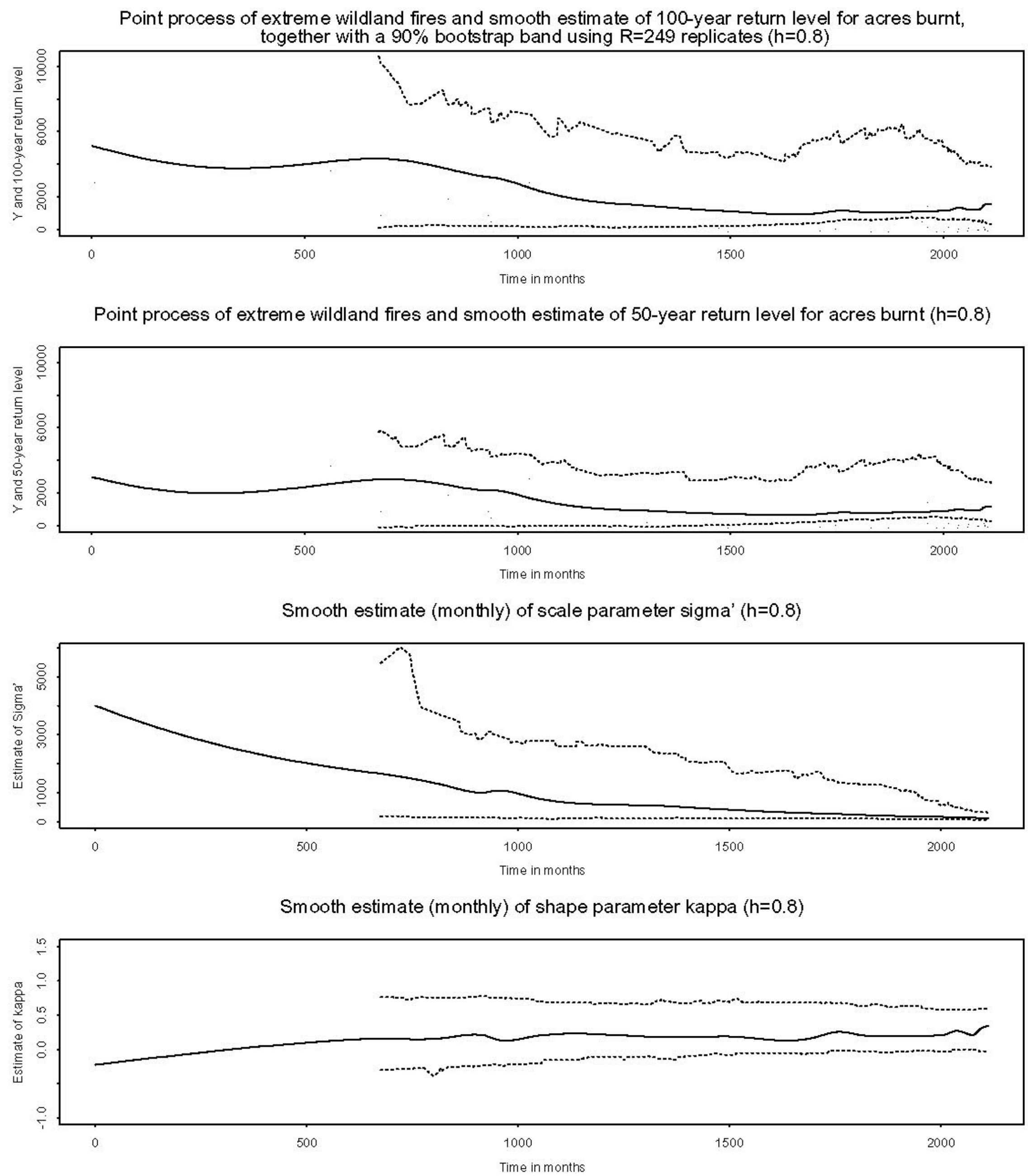

Figure 5. GPD fit for acres burnt in extreme wildland fire in the US since 1825.

We now turn to the fitting of GPD distribution for the amount of acres burnt. Initial analysis showed not much changes in the shape parameter $\kappa(t)$, but decreasing trend in the scale parameter $\sigma(t)$. Hence we fit our smooth model with the following parameter functions $\sigma(t) = \exp\{ \gamma_0 + \gamma_1 (t_j - t) \}$ and $\kappa(t) = \delta_0$ for this data. The bottom panels of Figure 5 show the estimated shape and scale parameters for the amount of acres burnt, using the GPD distribution with smoothing parameter h=0.8. The estimates of the shape parameter $\kappa(t)$ lie more or less at the same level close to 0.25 whereas $\sigma'(t)$ shows a sharp decline in the first half of the period. The variability of these estimates can be assessed by bootstrap confidence bands which could be constructed from a block bootstrap method (Davison and Hinkley, 1997). Unfortunately, the small number of observations especially in the first half of the series makes it difficult to apply this successfully as our procedure is sensitive to data sparseness. We were only able to obtain confidence band from about t = 675. A point-wise 90% bootstrap confidence band obtained from R=249 simulated realizations of the process is also shown in both panels.

The fitted parameter curves depend on the bandwidth *h*. Smaller values of *h* would result in very variable curves whereas larger h gives excessively smooth curves, and there is the usual bias-variance trade-off. In general, the optimal value of *h* may be obtained by cross-validation technique. In this application, however, with just 32 points in the series large values of *h* are needed to get sensible estimates. Experimentation with different values of *h* suggests that *h* =0.7 or 0.8 is a suitable value for this dataset.

The amount of acres burnt and their 100-year and 50-year return levels are shown in the upper panels of the figure, together with a point-wise 90% bootstrap confidence band (from *t* = 675 months onwards) obtained from R = 249 simulated realizations of the process. The return levels seem to decrease in the first half of the data period, before levelling off in the second half. The bootstrap confidence bands, however, suggest that their variability is well within the confidence band.

5 DISCUSSION

We have studied extreme wildland fires using semi-parametric smooth point process model for extremes. Local likelihood method is used to get smooth parameter estimates of the model which incorporates trend in time. Results of the analysis using extreme wildland fire data from US, suggest an increasing trend in the occurrence of wildfires in the second half of the last century. The result also suggests a declining trend in the return levels of the amount of acres burnt. Nevertheless, the bootstrap confidence bands suggest that these trends may be spurious. Although our analysis revealed the presence of slight trend in extreme wildland fire incidences and the amount of acres burnt, to make better inference and to draw firm conclusions a substantive analysis is required based on a larger dataset.

There could be a number of reasons for the trend we noticed in the extreme wildland fires. The increase in number of extreme fire may be due to changes in global weather cycles resulting in frequent dry spells and severe drought which contribute towards heavy buildup of dead vegetation fairly quickly. Besides this, modern land-use practices, urbanization, construction of new settlement areas etc may have altered the landscape which may result in changes such as an increased build-up of dense stands of trees, a shift to species that have not evolved and adapted to fire. All of these or a combination of them may have been the reason, amongst others, for the trend in the occurrence of extreme wildfires. The reduction in the amount of acres burnt could be accounted for by better fire fighting technology, education & prevention programs, more accurate fire weather prediction and forecast, and good wildland resource management practices which help reduce the accumulation of vast amount of forest fuel.

In addition to this, the quality of the data and the consistency of reporting practices are the other factors which might have contributed towards trend. For example, some of the extreme fires may have been historically significant due to the damage they caused to properties and the number of lives lost, along with the amount of acreage burnt. There is also a possibility that, with the advancement in technology, what are being recorded as several separate fires in present days might have been recorded as one larger fire in the past. This reporting inaccuracy would have certainly introduced a trend both in the rate of occurrence and in the size of the fire. In conclusion, although our semi-parametric smoothing approach has revealed a slight trend in extreme quantities of the wildland fire, an extensive analysis based on better quality larger data series is needed to arrive at firm conclusions. Our method, however, is useful as an explanatory tool in the analysis of trend in extremes.

Acknowledgements

The author would like to thank three referees and the editor for helpful comments which improved an earlier version of the manuscript.

6 REFERENCES

Bowman, A. W. and Azzalini, A (1997*).* *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus* Illustrations. Clarendon Press, Oxford.

Cleveland, W. S. and Devlin, S. J.(1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83*, 596--610.

Davison, A. C. and Hinkley, D. V (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge.

Davison, A. C. and Ramesh, N. I (2000). Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society* series B, 62, 191-208.

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

Fan, J. and Gijbels, . (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall. London.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* Chapman & Hall, London.

McCullagh, P. and Nelder, J. A (1989). *Generalized Linear Models. Second edition*. Chapman & Hall, London.

National Interagency Fire Centre, Idaho, USA: http://www.nifc.gov/stats

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