

## CHAPTER 4

# STATISTICAL ANALYSIS OF LARGE WILDFIRES

Thomas P. Holmes, Robert J. Huggett, Jr.,  
and Anthony L. Westerling

### 1. INTRODUCTION

Large, infrequent wildfires cause dramatic ecological and economic impacts. Consequently, they deserve special attention and analysis. The economic significance of large fires is indicated by the fact that approximately 94 percent of fire suppression costs on U.S. Forest Service land during the period 1980-2002 resulted from a mere 1.4 percent of the fires (Strategic Issues Panel on Fire Suppression Costs 2004). Further, the synchrony of large wildfires across broad geographic regions has contributed to a budgetary situation in which the cost of fighting wildfires has exceeded the Congressional funds appropriated for suppressing them (based on a ten-year moving average) during most years since 1990. In turn, this shortfall has precipitated a disruption of management and research activities within federal land management agencies, leading to a call for improved methods for estimating fire suppression costs (GAO 2004).

Understanding the linkages between unusual natural events, their causes and economic consequences is of fundamental importance in designing strategies for risk management. Standard statistical methods such as least squares regression are generally inadequate for analyzing rare events because they focus attention on mean values or typical events. Because extreme events can lead to sudden and massive restructuring of natural ecosystems and the value of economic assets, the ability to directly analyze the probability of catastrophic change, as well as factors that influence such change, would provide a valuable tool for risk managers.

The ability to estimate the probability of experiencing a catastrophic event becomes more advantageous when the distribution of extreme events has a heavy-tail, that is, when unusual events occur more often than generally anticipated. Heavy-tail distributions have been used to characterize various types of catastrophic, abiotic natural phenomena such as Himalayan avalanches (Noever 1993), landslides, and earthquakes (Malamud and Turcotte 1999). Several studies also indicate that wildfire regimes have heavy-tails (discussed in section 2 below). For decades, economists have been interested in heavy-tails appearing in the distribution of income (Mandelbrot 1960), city sizes (Gabaix 1999, Krugman 1996), commodity prices series (Mandelbrot 1963a, Mandelbrot

1963b), financial data (Fama 1963, Gabaix et al. 2003), and insurance losses (Embrechts et al. 2003).

Despite the fact that heavy-tail distributions have been used to characterize a variety of natural and economic phenomena, their application has been limited due to the fact that heavy-tail distributions are characterized by infinite moments (importantly, mean and variance). Reiss and Thomas (2001) define a distribution function  $F(x)$  as having a heavy-tail if the  $j^{\text{th}}$  moment  $\int_0^{\infty} x^j dF(x)$  is equal to infinity for some positive integer  $j$  (p. 30). Note that a moment is infinite if the integral defining the statistical moment is divergent (it converges too slowly to be integrated)—therefore, the moment does not exist.

Recognizing that standard statistical tools such as the Normal distribution and ordinary least squares regression are not reliable when moments are infinite, Mandelbrot (1960, 1963a, 1963b) suggested that the Pareto distribution be used to analyze heavy-tail phenomena. The Pareto distribution is extremely useful because, in addition to the capacity to model infinite moments, it has an invariant statistical property known as stability: the weighted sum of Pareto-distributed variables yields a Pareto distribution (adjusted for location and scale). Other commonly used long-tail distributions, such as the log-normal, do not share this stability property. More recently, Mandelbrot (1997) refers to distributions with infinite variance as exemplifying a state of randomness he calls “wild randomness”.

Over the past few decades, special statistical methods, known as extreme value models, have been developed for analyzing the probability of catastrophic events. Extreme value models utilize stable distributions, including the heavy-tailed Pareto, and have been applied to problems in ecology (Gaines and Denny 1993, Katz et al. 2005), finance, and insurance (Reiss and Thomas 2001, Embrechts et al. 2003). The goals of this chapter are to: (1) show how extreme value methods can be used to link the area burned in large wildfires with a set of explanatory variables, and (2) demonstrate how parameters estimated in the linkage function can be used to evaluate economic impacts of management interventions. In doing so, we provide a brief, somewhat technical overview of the statistical analysis of extreme events and discuss previous applications of these models to wildfire analysis (section 2). A major contribution of this chapter is the discussion of how extreme value models can be parameterized to include covariates such as climate or management inputs as explanatory variables (section 3). To clarify the presentation, the statistical methods are applied to an empirical analysis of nearly a century of fire history in the Sierra Nevada Mountains of California (section 4). A summary of the major points, and implications of the empirical analysis for risk managers, are discussed (section 5).

## 2. HEAVY-TAIL DISTRIBUTIONS AND WILDFIRE REGIMES

The idea that much can be learned about economic costs and losses from wildfires by recognizing the special significance of large fires can be traced to an article published by Strauss and colleagues (1989) titled "Do One Percent of the Forest Fires Cause 99 Percent of the Damage?" In that study, the authors provided a statistical analysis of wildfire data from the western United States and Mexico that showed the underlying statistical distribution of fire sizes was consistent with the heavy-tailed Pareto distribution. Several subsequent studies, spanning a wide array of forest types in the United States (Malamud et al. 1998, Malamud et al. 2005), Italy (Ricotta et al. 1999), Canada (Cumming 2001), China (Song et al. 2001) and the Russian Federation (Zhang et al. 2003), also concluded that wildfire regimes are consistent with the heavy-tailed Pareto distribution. The Pareto wildfire distribution may be truncated (Cumming 2001) or tapered (Schoenberg et al. 2003) to account for the finite size that can be attained by fires within forested ecosystems.

To fix ideas regarding the nature of the heavy-tailed Pareto distribution and the consequence of such a data generation process for the analysis of large wildfires, it is necessary to introduce some notation. To begin, a cumulative distribution function of the random variable  $X$ , denoted by  $F(x) = P(X \leq x)$ , is said to be heavy-tailed if  $x \geq 0$  and

$$\lim_{x \rightarrow \infty} P(X > x + y | X > x) = \lim_{x \rightarrow \infty} \frac{\bar{F}(x + y)}{\bar{F}(x)} = 1, \quad y \geq 0 \quad (4.1)$$

where  $\bar{F} = 1 - F(x)$ , referred to as the "tail distribution" (Sigman 1999) or "survivor function" (Miller, Jr. 1981). Intuitively, equation (4.1) states that if  $X$  exceeds some large value, then it is equally likely that it will exceed an even larger value as well. The Pareto distribution is a standard example of a heavy-tailed distribution:  $\bar{F}(x) = x^{-\alpha}$  where  $x \geq 1$  and  $\alpha > 0$ . If  $\alpha < 2$ , then the distribution has infinite variance (the distribution converges so slowly to zero that it cannot be integrated), and if  $\alpha > 1$ , the distribution has infinite mean.

Extreme value models focus attention on the tail of a statistical distribution of events rather than imposing a single functional form to hold for the entire distribution. It is important to understand that the family of extreme value statistical models does not impose a heavy-tail upon the data. Rather, the extreme value parameter estimates indicate whether the data have a light, moderate or heavy-tailed distribution (Coles 2001). The classical method used in the statistics of extremes, known as the Generalized Extreme Value (GEV) method focuses attention on the statistical behavior of the maximum value attained by some random variable during each time period (or "block"):

$$M_n = \max(X_1, X_2, \dots, X_n) \quad (4.2)$$

where  $X_1, \dots, X_n$  is a sequence of independent random variables each having an underlying distribution function  $F$ . If  $n$  represents the number of wildfire observations recorded in a year, then  $M_n$  is the largest wildfire recorded that year. Classical extreme value theory shows that there are three types of distributions for  $M_n$  (after linear renormalization): the Gumbel (intermediate case), Fréchet (heavy-tail) and Weibull (truncated at a maximum size) families. These three families are described by Coles (2001).

Using extreme value theory, Moritz (1997) fitted a GEV distribution using wildfire data from two geographic divisions within the Los Padres National Forest in southern California. He found that the percentage of years in which the single largest fire burned more than one-half the annual total was 65 percent and 81 percent for the two study areas, and that the size distribution of the largest annual wildfires between the years 1911 and 1991 was heavy-tailed. This result is important because it is consistent with empirical studies showing that the entire range of fire sizes is Pareto distributed. Further, based on graphical evidence, he speculated that "extreme weather" might create conditions such that large wildfires are "immune to suppression" (p. 1260). Thus, a possible linkage between very large wildfires, environmental conditions, and fire suppression technology was suggested.

Although the GEV model provides a theoretical foundation for the analysis of extreme events, data use is inefficient in model estimation because only a single observation per time period is utilized. A second approach to extreme value analysis overcomes this limitation by using observations which exceed a high threshold value, often referred to as the "peaks over threshold" method. Again let  $X_1, X_2, \dots$  represent a sequence of independent and identically distributed random variables with distribution function  $F$ , and let  $u$  represent some high threshold. The stochastic behavior of extreme events above the threshold is given by the conditional probability

$$\Pr(X > u + y | X > u) = \frac{\bar{F}(u + y)}{\bar{F}(u)}, \quad y > 0 \quad (4.3)$$

which clearly bears a strong resemblance to equation (4.1). It can be shown that, by taking the limiting distribution of equation (4.3) as  $u$  increases, the distribution function converges to a Generalized Pareto distribution  $G_{\xi\sigma}(y)$  (Coles 2001):

$$G_{\xi\sigma}(y) = \Pr(X - u \leq y | X > u) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma}} & \text{if } \xi = 0 \end{cases} \quad (4.4)$$

where  $y = x - u$ . The parameter  $\xi$  is called the shape parameter and  $\sigma$  is the scaling parameter. When  $\xi < 0$ , the distribution has a finite upper endpoint at  $-\sigma/\xi$ ; when  $\xi = 0$ , the distribution is an exponential (light-tail) distribution with mean  $\sigma$ ; when  $\xi > 0$ , the distribution has a heavy-tail (or Fréchet distribution) with mean

$\sigma/(1-\xi)$ , given that  $\xi < 1$  (Smith 2003). If  $\xi \geq 1$  the mean of the distribution is infinite, and if  $\xi > 1/2$  the variance is infinite (“wildly random”).

Parameters of the Generalized Pareto model were estimated by Alvarado and colleagues (1998) for large wildfires between 1961 and 1988 in Alberta, Canada. Using alternative threshold values (200 hectares and the upper one percentile of fire sizes) they concluded that the data were Fréchet (heavy-tail) distributed. In fact, the fire data were so heavy-tailed that the fitted distributions were found to have both infinite means and infinite variances.

The various findings reported above—that wildfire size distributions are heavy-tailed—represent an important statistical regularity. However, economists are generally interested in conditional probabilities, that is, factors that induce non-stationarity in statistical distributions (Brock 1999). In the following section, we describe how covariates can be introduced into models of heavy-tailed statistical distributions and show how hypotheses about covariates can be tested in a “regression-like” framework. These methods provide a powerful tool for researchers to investigate factors that influence the generation of large wildfires.

### 3. INCLUDING COVARIATES IN EXTREME VALUE THRESHOLD MODELS

As mentioned above, Generalized Pareto models are more efficient in the use of data than classical extreme value models because they permit multiple observations per observational period, such as fire year. The main challenge in the Generalized Pareto model is the selection of a threshold for data inclusion. Statistical theory indicates that the threshold  $u$  should be high enough to be considered an extreme value, but as  $u$  increases less data is available to estimate the distribution parameters. Although rigorous methods for determining the appropriate threshold are currently receiving a great deal of research attention, graphical data exploration tools are typically used to select an appropriate value for  $u$  using a plot of the sample mean excess function (Coles 2001). In particular, the threshold is chosen where the sample mean excess function (i.e., the sample mean of the values that exceed the threshold) becomes a linear function when plotted against the threshold value.

Having determined a threshold value, parameters of the Generalized Pareto distribution can be estimated by the method of maximum likelihood. For  $\xi \neq 0$ , the likelihood function is

$$L(\xi, \sigma) = \prod_{i=1}^m \frac{1}{\sigma} \left[ 1 + \xi \frac{(x_i - u)}{\sigma} \right]^{-\left(1 + \frac{1}{\xi}\right)} \quad (4.5)$$

and the log-likelihood is

$$\ln L(\xi, \sigma) = -m \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \ln \left[ 1 + \frac{\xi}{\sigma} (x_i - u) \right] \quad (4.6)$$

where  $m$  is the number of observations,  $x_i$  is the size in acres of fire  $i$ , and  $u$  is the threshold fire size in acres. Note that equation (4.6) can only be maximized when  $(1 + \sigma^{-1}\xi(x_i - u)) > 0$  for all  $i = 1, \dots, m$ . If this is untrue, it is necessary to set  $\ln L(\xi, \sigma) = -\infty$  to assure convergence. For the special case where  $\xi = 0$ , the log-likelihood is

$$\ln L(\sigma) = -m \ln \sigma - \sigma^{-1} \sum_{i=1}^m (x_i - u) \quad (4.7)$$

and the model is a member of the exponential (non-heavy-tailed) family of distributions.

If the underlying stochastic process is non-stationary, then the simple Generalized Pareto model can be extended to include covariates such as time trends, seasonal effects, climate, or other forcing variables. Non-stationarity is typically expressed in terms of the scale parameter (Smith 2003). For example, to test for a time trend, the scale parameter could be expressed as a function of time, where the scale parameter for observation  $i$  is  $\sigma_i = \beta_0 + \beta_1 t$ , where  $t$  represents time. More generally, a vector of covariates can be included in the model by expressing the scale parameter as a linear function of the product of a vector of explanatory variables and parameters ( $\beta$ ) to be estimated:

$$\sigma = [1, z_1, \dots, z_n] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \quad (4.8)$$

where  $n$  is the number of covariates included in the model.

The Generalized Pareto model is asymptotically consistent, efficient, and normal if  $\xi > -0.5$  (Coles 2001, Smith 2003), allowing for the derivation of standard errors for the parameter estimates using either the bootstrap method or the inverse of the observed information matrix (Smith 2003). Having obtained estimates of standard errors, hypotheses regarding the statistical significance of the covariates can be tested.

The statistical model can be used to estimate the expected value (average size) of large fires during a fire season given values for the set of covariates and estimates of the parameter vector  $[\beta_0, \dots, \beta_n]$ . In the simplest case, the value for a covariate may represent an updated value for a time trend. Or, the value may represent the forecasted value of a covariate such as a climate indicator. For the Generalized Pareto model, the expected value of an event that exceeds the threshold has a simple expression:

$$E(Y) = \frac{\sigma(Z)}{1 - \xi} \quad (4.9)$$

given that  $\xi < 1$  (recall that if  $\xi > 1$  the mean is infinite),  $Y$  is the amount by which an observation exceeds the threshold ( $Y - \mu > 0$ ), and  $Z$  is a vector of covariates. In terms of wildfire sizes,  $E(Y) + \mu$  provides an estimate of the expected (or average) size of a large wildfire given that a wildfire size has exceeded the threshold value.

Economic metrics can be calculated using information on the economic values associated with the expected area burned. For example, the expected value of timber at risk of loss to a large wildfire could be estimated by multiplying the expected number of acres burned in a large wildfire by an average per acre estimate of stumpage value. Expected suppression costs associated with large wildfires could be estimated in a similar fashion. Or, if information were available on the non-market economic values of resources related to recreation, watersheds or wildlife habitat, then economic estimates of non-market values at risk could be computed as well. If statistically significant covariates associated with management interventions are identified that alter the production of large wildfires, then the parameter estimates on the covariates can be used to estimate the economic benefits of interventions. An illustration is presented in the following empirical example.

#### 4. LARGE WILDFIRES IN THE SOUTHERN SIERRA NEVADA MOUNTAINS

The Southern Sierra Nevada Mountains (SSNM) provide a useful case study for illustrating the application of extreme value analysis to wildfire modeling. Nearly a century of fire data are available for land management units located in this region, allowing us to investigate factors influencing wildfire production over short, medium and long time scales. The fire data analyzed in this chapter come from the Sequoia National Forest (SQF) which sits at the southern extension of the SSNM and comprises 5,717 km<sup>2</sup>, or 27 percent, of the federally managed lands in the SSNM (fig. 4.1). The northern and western reaches of SQF have the most forest cover, with substantial area at lower elevations in the southwest in grassland and in the southeast in chaparral. Giant sequoia groves are a small, but important, component of the fire-adapted ecosystems in SQF.

Fire history data for SQF were derived from fire perimeter records (fig. 4.2) for the years 1910–2003, obtained from the California Department of Forestry Fire and Resource Assessment Program. A histogram of the fire size distribution for SQF (fig. 4.3) clearly shows that the distribution is not normal or log-normal, is highly skewed, and has a long right-hand-side tail (note that fire sizes above 10,000 acres have been combined for graphical convenience). At a first approximation, the distribution of fire sizes for SQF appears as though it may be Pareto distributed (heavy-tailed) or, perhaps, distributed as a negative exponential (light-tailed). Fortunately, statistical methods can be used to test whether the distribution is light- or heavy-tailed (Reiss and Thomas 2001).

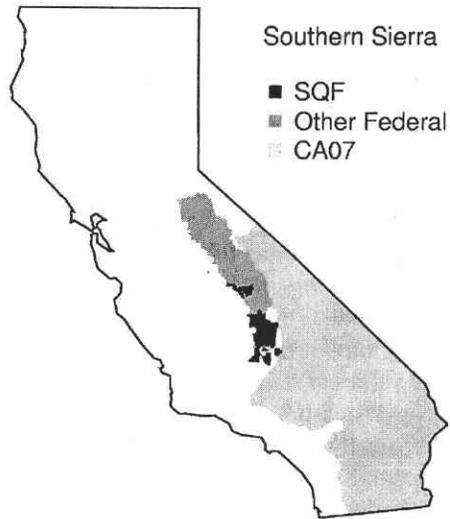


Figure 4.1. Location of Sequoia National Forest (SQF, black) in relation to other Federal Forest and Park land in the Southern Sierra Nevada Mountains (grey) and California Climate Division number 7 (CA07, light grey).

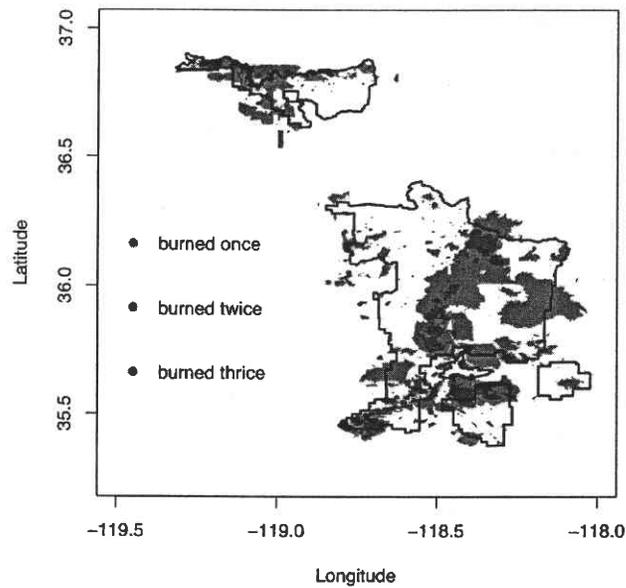


Figure 4.2. Map showing areas burned since 1910 (shaded) and Sequoia National Forest boundary.

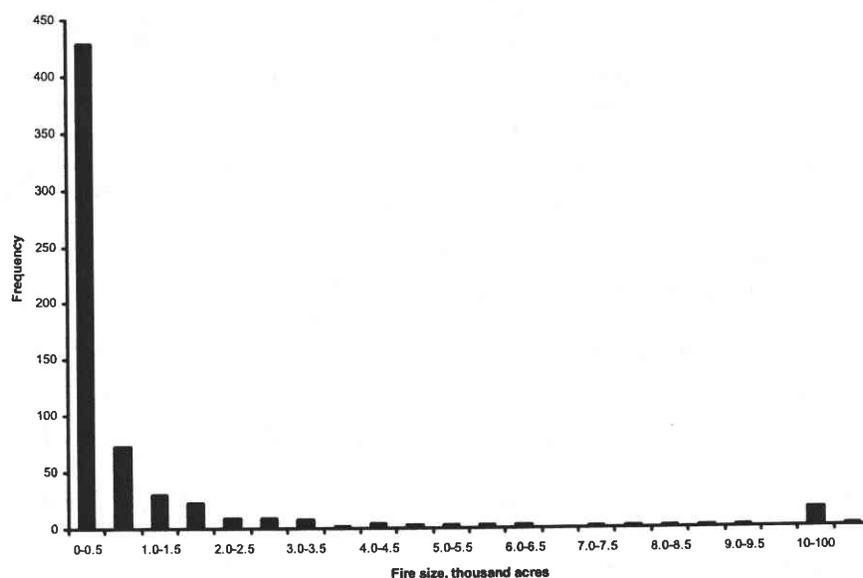


Figure 4.3. Fire size distribution, Sequoia National Forest, 1910-2003

#### 4.1 Model Specification

The fire history for SQF permits us to test a variety of hypotheses, including whether or not a long-term trend can be identified in the occurrence of large fires. Additionally, the time-series allows us to investigate whether shorter-run trends, such as changes in fire suppression technology, and seasonal influences, such as climatic effects, have influenced the production of large wildfires. Although the covariates discussed below are included in the model specification, the results should be viewed as illustrative. Because this model is the focus of ongoing research, it should be understood that alternative model specifications may (or may not) yield somewhat different results.

##### 4.1.1 Time trend

In the SSNM, the combined influence of livestock grazing during the nineteenth century and fire suppression during the twentieth century have changed tree species composition and increased the density of forest stands (Vankat and Major 1978). As early as the late 1800's, foresters in California were arguing for fire exclusion to protect timber resources for the future, and by the early twentieth century fire reduction was occurring (Skinner and Chang 1996). Suppression of low and moderate severity fires has caused conifer stands to become denser, especially in low- to mid-elevation forests, and shade tolerant, fire-sensitive tree species have become established. In turn, these vegetative changes have led to a profusion of wildfires that burn with greater intensity than in the past, with crown

fires becoming more common (Skinner and Chang 1996). Further, the proportion of the annual acreage burned by the largest wildfire on National Forest land has trended upwards during the twentieth century (McKelvey and Busse 1996).

We hypothesize that a positive trend may be identified in the probability of observing large wildfires in the SSNM which may reflect these long-term changes in forest composition. A trend variable,  $time_i$ , was created by setting  $time_i = 1$  for the first year of the data record,  $time_i = 2$  for the second year, and so forth up to the final year of the data record.

#### 4.1.2 Fire suppression technology

The use of air tankers for fighting wildfires began in California. The first air drop was made on the Mendenhall Fire in the Mendocino National Forest in 1955 in a modified agricultural biplane. These early aircraft had roughly a 100 gallon capacity and dropped about 124,000 gallons of water and fire suppressants during that year. By 1959, heavier air tankers with as much as a 3,000 gallon capacity were in operation and dropped nearly 3.5 million gallons in 1959 (Anon. 1960). Aircraft are now commonly used in fire suppression and their expense is a major component of suppression costs on large wildfires (Mangan 2001).

Although historical aircraft fire suppression cost data are not available for the SSNM, an aircraft variable was specified for use in our large wildfire probability model by creating a dummy variable,  $air\_dummy$ , to approximate the effective use of air tankers for fire suppression in California. In particular, we set  $air\_dummy = 0$  for years prior to 1960 and  $air\_dummy = 1$  for subsequent years.

#### 4.1.3 Climate

The moisture available in fuels is a critical factor in wildfire spread and intensity. Climatic effects are specified in our model using PDSI, which is an index of combined precipitation, evapotranspiration, and soil moisture conditions. PDSI has been used successfully in previous studies of climate-fire relationships (Balling et al. 1992, Swetnam and Betancourt 1998, Mitchner and Parker 2005, Westerling et al. 2003, Westerling, chapter 6). The index is negative when inferred soil moisture is below average for a location, and positive when it is above average. In this chapter, we investigate the relationship between climate and large fire sizes using observations on July values for PDSI for California region 7 (fig. 4.1). PDSI values from the U.S. Climate Division Data set were obtained from NOAA for 1895-2003. July PDSI calculated from monthly climate division temperature and precipitation is used here as an indicator of inter-annual variability in summer drought.

We note a pronounced trend toward drier summer conditions over the entire period of analysis, with a highly significant trend in July PDSI (fig. 4.4). This tendency toward drier summers is probably a function of both lower precipitation and higher temperatures. There has been a trend toward lower precipitation throughout the entire Sierra Nevada Mountain range over the period of analysis,

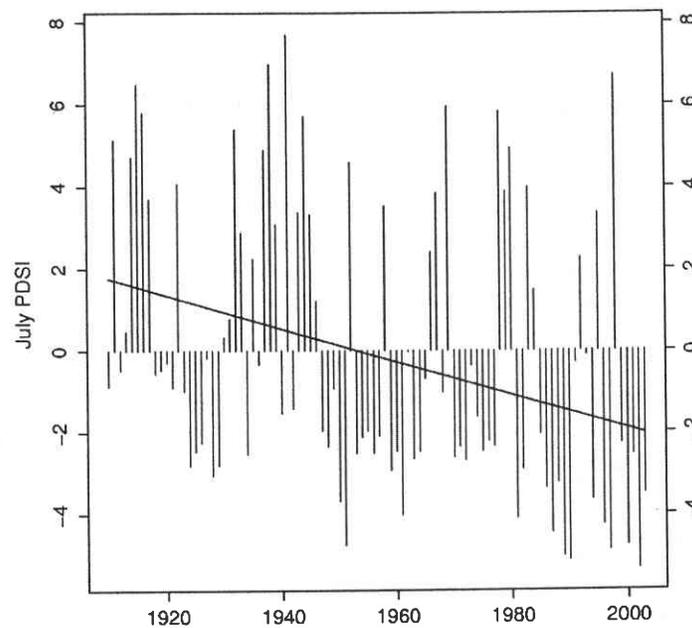


Figure 4.4. July PDSI index for California Climate Division number 7, 1910–2003. Diagonal line is ordinary least squares regression fit to a time trend.

while spring and summer temperatures have been much warmer since the early 1980s. Warmer springs in particular, combined with less precipitation, result in an earlier snow melt at mid and higher elevations, which in turn implies a longer, more intense summer dry season and fire season (Westerling et al. 2006).

## 4.2 Empirical Models

Two Generalized Pareto models were estimated that fit historical fire size data for Sequoia National Forest: (1) a basic model with a constant scale parameter, and (2) a covariate model that specified the scale parameter as a linear function of a time trend ( $time_i$ ), an air tanker dummy variable ( $air\_dummy$ ), and climate effects ( $PDSI$ ). The models were estimated using the Integrated Matrix Language (IML) programming code in the SAS statistical software.

Prior to estimating either model, it was necessary to choose the threshold fire size  $u$  above which large or “extreme” fires would be modeled. Mean excess plots were created to identify the location of the fire size threshold. As explained in Coles (2001), the Generalized Pareto distribution will be a valid representation of the distribution of exceedances above  $u$  if the plot is linear past that point. Visual inspection of the mean excess plot indicated that a threshold of 500 acres

would be appropriate as the plot became generally linear beyond the  $u = 500$  acre fire size. Reference to figure 4.3 also suggests that the rather long tail of the fire size distribution may be initiated at a threshold of 500 acres.

Although a relatively small proportion (30 percent) of the total number of fires in SQF exceeded 500 acres, they accounted for nearly all (94 percent) of the total area burned during the fire record (fig. 4.5). After eliminating observations with total burned area of 500 acres and less, 181 observations remained for estimation of the SQF large fire distribution. Summary statistics for the 181 fires, and the set of covariates included in the model, are given in table 4.1.

Maximum likelihood techniques were used to estimate the parameters in equation (4.6), where the scale parameter was specified using covariates as shown in equation (4.8). That is, the scale parameter was specified as:  $\sigma_i = \beta_0 + \beta_1 time_i + \beta_2 air\_dummy_i + \beta_3 PDSI_i$ . Standard errors for the parameter estimates were derived from the inverse of the observed information matrix, and allowed us to test whether the parameter estimates were significantly different than zero.

### 4.3 Results

In the simple model with no covariates, both the shape and scale parameter estimates were significantly different than zero at the 1 percent level. Since the parameter estimate for the shape parameter  $\xi$  is greater than 0, the distribution has a heavy-tail (Fréchet). This result is consistent with the studies in the literature reviewed above. Further, since  $\xi > 1$ , the distribution has an infinite mean and variance, which is consistent with the findings reported by Alvorado and others (1998). Although forest extent is finite and, therefore, average wildfire size must be finite, the finding of an infinite (divergent) mean and variance for

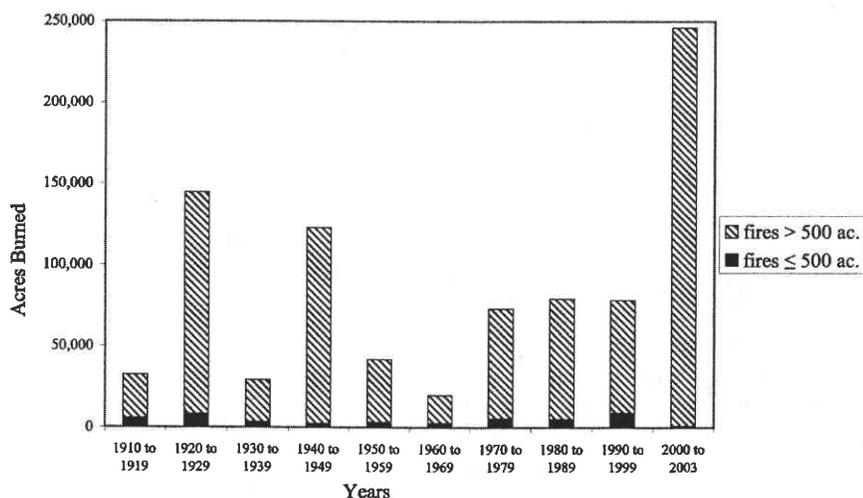


Figure 4.5. Acres burned by fire size class, Sequoia National Forest, 1910-2003

**Table 4.1. Descriptive statistics for large fires and covariates included in the model.**

	Mean	Std. Dev.	Min.	Max.
fire size (ac.)	4,533.49	13,365.25	504.00	149,470.00
time (trend)	47.96	29.75	1	94
air_dummy	0.47	0.49	0	1
PDSI	-1.00	3.13	-5.35	7.70

**Table 4.2. Parameter estimates of the basic Generalized Pareto extreme value model.**

Parameter	Value	Std. Error	t-statistic
shape $\xi$	1.02	0.15	6.84
scale $\sigma$	780.23	115.68	6.75
N = 181			
log likelihood	-1,570.74		

**Table 4.3. Parameter estimates of the Generalized Pareto extreme value model with covariates.**

Parameter	Value	Std. error	t-statistic
shape $\xi$	0.91	0.14	6.49
scale $\sigma$			
constant	593.47	199.12	2.98
time	15.19	7.88	1.93
air_dummy	-983.44	420.01	-2.34
PDSI	-62.70	21.52	-2.91
N = 181			
log likelihood	-1,564.25		

fires exceeding the 500 acre threshold implies that fires greatly exceeding fire sizes included in the historical record are possible. This finding is important because it indicates that large wildfire production is extremely variable despite the constraints imposed by physical conditions. In turn, extreme variability in the production of large wildfires makes fire program planning and budgeting difficult, especially if the variables driving the stochastic fire generation process cannot be identified or reliably forecast.

The parameter estimate on  $\xi$  in the Generalized Pareto model with covariates indicates that large fires in SQF have a heavy-tail, or Fréchet distribution, with infinite variance, which is similar to the basic model. However, because the

parameter estimate  $\xi < 1$  in the covariate model, the mean of the distribution is finite (in contrast with the basic model). This result is important, as it allows us to estimate the size of the average fire for wildfires exceeding the threshold (500 acres), and evaluate how changes in management inputs influence the average size of large wildfires.

Including covariates in the Generalized Pareto model suggests that the distribution of large fires in SQF has been non-stationary over the recorded fire history. The positive parameter estimate on the time trend is significant at the 10 percent level and indicates that the probability of observing a large wildfire has increased over the 94 year fire record. This result should be viewed as illustrative, not definitive, and a more fully specified model (considering, for example, non-linear effects and other covariates) may alter this finding. None-the-less, this result is consistent with the idea that land use history and fire suppression have contributed to altered tree species composition and density which, in turn, have contributed to forest conditions with greater flammability. We note further that this effect may be confounded to some degree by the increased development of roads and trails in SQF over the 94 year period, and the concomitant increase in the number of people visiting the forest may have contributed to the increasing trend in large wildfires.

Consistent with our *a priori* hypothesis, drier fuel conditions (as measured using PDSI) were found to be related with larger fires. The parameter estimate on PDSI was negative and significant at the 1 percent level. Recall that negative values of PDSI correspond with the driest conditions, while positive values correspond with wet conditions. Consequently, the model results indicate that very dry conditions are associated with an increased probability of large wildfires.

The parameter estimate on the air tanker dummy variable is negative and significant at the 5 percent level and suggests that the deployment of air tankers since 1960 has decreased the probability of observing large wildfires. This result is consistent with the finding reported by Moritz (1997) who concluded that air tankers have aided the containment of large wildfires in California's Los Padres National Forest. Again we note that these results are provisional and might change with improved model specifications.

Given the parameter estimates, various scenarios can be constructed to demonstrate the effect of the covariates on the expected large fire size and to evaluate the impact of management interventions (table 4.4). For example, the scenarios shown in table 4.4 indicate that the expected large fire size is quite sensitive to the use of air tankers for fire suppression. Under average drought conditions in the year 2002, the use of air tankers reduces the expected large fire size from 24,700 acres to 13,690 acres, a reduction of about 45 percent. Given data on fire suppression costs, this relationship could be used to estimate the expected benefits (reductions in cost) due to the use of air tankers.

Because they represent averages, expected large fire sizes may not be sensitive to extreme conditions experienced during a single fire year. The fire year 2002 provides an instructive example, as the July PDSI for that year was the driest

**Table 4.4. Scenarios depicting the expected size of large fires (thousand acres) under alternative conditions.**

Scenario	Year 2002	Year 2010	Year 2025	Year 2050
1. Average drought w/ tankers	13.69	15.05	17.60	21.85
2. Average drought w/o tankers	24.70	26.05	28.6	32.86
3. Extreme drought w/ tankers	16.74	18.10	20.66	24.91
4. Extreme drought w/o tankers	27.76	29.12	31.67	35.92

since 1910. The expected large wildfire size for SQF for 2002, incorporating the time trend and the effect of PDSI, was computed to be roughly 15,000 acres if air tankers were used in suppression and roughly 26,000 acres if air tankers were not used for fire suppression. During the summer of 2002, the largest fire for SQF since 1910 was recorded (the McNally fire) which burned nearly 150,000 acres. Our estimate of average fire size for that year is much too low and suggests either that our model has omitted some important variables (such as wind speed) or that other unobserved factors create the extraordinary variance observed in large wildfire regimes.

Although an annual estimate of expected large wildfire size may be inaccurate under extreme climatic conditions, averaging expected large wildfire sizes over time improves the precision of expected values. For example, the average large fire in SQF between 1994 and 2003 was 9,625 acres. Using the parameters in our Generalized Pareto covariate model, we estimated that the average large fire would be 12,247 acres, which is within 1 standard deviation of the sample average. Therefore, temporal averaging can smooth out the estimate of expected large wildfire size even during periods of extreme climatic conditions. In turn, this suggests that estimates of the benefits of large wildfire management interventions should likewise be temporally averaged and that confidence intervals should be reported.

## 5. SUMMARY AND CONCLUSIONS

Although extreme value statistical models are not widely used in wildfire modeling, the literature review, results, and analysis reported in this chapter suggest that further development of these models is warranted for four principal reasons:

- Wildfire production often does not follow a "light-tail" distribution such as a normal or log-normal distribution. Rather, fire size distributions reported for several regions around the globe have heavy-tails characterized by infinite moments.
- Standard statistical techniques, such as ordinary least squares regression, may produce very misleading parameter estimates under conditions of infinite variance (second moment).

- Extreme value models focus attention on the tail of the distribution which, in fire modeling, is where most of the ecological and economic impacts occur. These statistical models are stable under conditions of infinite moments and allow probabilities of catastrophic events to be rigorously estimated.
- A set of covariates can be included in extreme value models providing the ability to test hypotheses regarding variables that influence the production of large wildfires. Parameter estimates on covariates can be used to evaluate the impacts of management interventions on the production of large wildfires.

A major conclusion of this chapter is that large wildfires are intrinsic to fire-adapted ecosystems and that memorable events such as the Yellowstone fires of 1988 (Romme and Despain 1989) and the McNally fire of 2002 in SQF cannot be simply dismissed as catastrophic outliers or anomalies. Rather, the underlying fire generation process operates in a fashion such that wildfires greatly exceeding those represented in local or regional fire histories may occur sometime in the future. Infinite variance in wildfire production, or wild randomness, greatly complicates planning operations for large fires. For example, moving average models of acres burned in large fires likely provide poor forecasts of the size of future large fires because the first moment converges very slowly to its true value in a wildly random state. The development of decision-making strategies for resources exposed to the state of wild randomness remains a challenge for risk managers in the finance and insurance sectors as well as for wildfire managers.

The second major conclusion of this chapter is that the ability to include covariates in a model of large wildfires characterized by infinite variance provides a robust method for evaluating the impact of management interventions. For example, the impact of deploying air tankers in fire suppression (captured using a dummy variable) was illustrated. The parameter estimate on the air tanker dummy variable was shown to have a substantial effect on the expected size of large fires. Given the size of this effect, data on large wildfire suppression cost could be used to estimate the expected benefits (cost savings) attributable to air suppression. In turn, the expected benefits of air suppression could be compared with air suppression costs. This modeling approach can be more generally used to evaluate the costs and expected benefits of other management interventions on large wildfires. Future research will be directed at identifying and testing alternative extreme value covariate models on an array of large wildfire regimes and management interventions with the goal of understanding how large fire costs might be better managed.

## 6. REFERENCES

- Alvorado, E., D.V. Sandberg, and S.G. Pickford. 1998. Modeling large forest fires as extreme events. *Northwest Science* 72:66-75.

- Anonymous. 1960. Air tanker capabilities and operational limitations. *Fire Control Notes* 21(4):129-131.
- Balling, R.C., G.A. Myer, and S.G. Wells. 1992. Relation of surface climate and burned area in Yellowstone National Park. *Agricultural and Forest Meteorology* 60:285-293.
- Brock, W.A. 1999. Scaling in economics: A reader's guide. *Industrial and Corporate Change* 8:409-446.
- Coles, S. 2001. *An introduction to statistical modeling of extreme values*. Springer-Verlag, London, UK. 208 p.
- Cumming, S.G. 2001. A parametric model of the fire-size distribution. *Canadian Journal of Forest Research* 31:1297-1303.
- Embrechts, P., C. Klüppelberg, and T. Mikosch. 2003. *Modelling extremal events for insurance and finance*. Springer-Verlag, Berlin. 648 p.
- Fama, E.F. 1963. Mandelbrot and the stable Paretian hypothesis. *The Journal of Business* 36(4):420-429.
- Gabaix, X. 1999. Zipf's law for cities: An explanation. *Quarterly Journal of Economics* CXIV:739-767.
- Gabaix, X., P. Gopikrishnan, V. Plerou, and H.E. Stanley. 2003. A theory of power-law distributions in financial market fluctuations. *Nature* 423:267-270.
- Gaines, S.D., and M.W. Denny. 1993. The largest, smallest, highest, lowest, longest, and shortest: Extremes in ecology. *Ecology* 74:1677-1692.
- Government Accountability Office (GAO). 2004. *Wildfire suppression: Funding transfers cause project cancellations and delays, strained relationships and management disruptions*. Available online at [www.gao.gov/cgi-bin/getrpt?GAO-04-612](http://www.gao.gov/cgi-bin/getrpt?GAO-04-612). Last accessed April 20, 2007.
- Katz, R.W., G.S. Brush, and M.B. Parlange. 2005. Statistics of extremes: Modeling ecological disturbances. *Ecology* 86:1124-1134.
- Krugman, P. 1996. *The self-organizing economy*. Blackwell, Cambridge, MA. 122 p.
- Malamud, B.D., and D.L. Turcotte. 1999. Self-organized criticality applied to natural hazards. *Natural Hazards* 20:93-116.
- Malamud, B.D., J.D.A. Millington, and G.L.W. Perry. 2005. Characterizing wild-fire regimes in the United States. *Proceedings of the National Academy of Science* 102(13):4694-4699.
- Malamud, B.D., G. Morein, and D.L. Turcotte. 1998. Forest fires: An example of self-organized critical behavior. *Science* 281:1840-1842.
- Mandelbrot, B. 1960. The Paveto-Lévy law and the distribution of income. *International Economic Review* 1(2):79-109.
- Mandelbrot, B. 1963a. The variation of certain speculative prices. *The Journal of Business* 36(4):394-419.
- Mandelbrot, B. 1963b. New methods in statistical economics. *The Journal of Political Economy* 71:421-440.

- Mandelbrot, B.B. 1997. States of randomness from mild to wild, and concentration from the short to the long run. In: *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, B.B. Mandelbrot, R.E. Gomory, P.H. Cootner, and E.F. Fama (eds.). Springer, New York, NY. 551 p.
- Mangan, R.J. 2001. Issues in reducing costs on large wildfires. *Fire Management Today* 61(3):6-10.
- McKelvey, K.S., and K.S. Busse. 1996. Twentieth-century fire patterns on Forest Service lands. Chapter 41, *Sierra Nevada Ecosystem Project: Final Report to Congress, vol. II, Assessments and scientific basis for management options*. Davis: University of California.
- Miller, Jr., R.G. 1981. *Survival analysis*. John Wiley & Sons, New York, NY. 238 p.
- Mitchner, L.J., and A.J. Parker. 2005. Climate, lightning, and wildfire in the national forests of the southeastern United States: 1989-1998. *Physical Geography* 26:147-162.
- Moritz, M.A. 1997. Analyzing extreme disturbance events: Fire in Los Padres National Forest. *Ecological Applications* 7:1252-1262.
- Noever, D.A. 1993. Himalayan sandpiles. *Physical Review E* 47(1):724-725.
- Reiss, R.-D., and M. Thomas. 2001. *Statistical analysis of extreme values*. Birkhäuser Verlag, Berlin. 443 p.
- Ricotta, C., G. Avena, and M. Marchetti. 1999. The flaming sandpile: Self-organized criticality and wildfires. *Ecological Modelling* 119:73-77.
- Romme, W.H., and D.G. Despain. 1989. Historical perspective on the Yellowstone fires of 1988. *Bioscience* 39:695-699.
- Schoenberg, F.P., R. Peng, and J. Woods. 2003. On the distribution of wildfire sizes. *Environmetrics* 14:583-592.
- Sigmon, K. 1999. A primer on heavy-tailed distributions. *Queueing Systems* 33:261-275.
- Simon, H. 1955. On a class of skew distribution functions. *Biometrika* 42:425-440.
- Skinner, C.N., and C.-R. Chang. 1996. Fire regimes, past and present. *Sierra Nevada Ecosystem Project: Final Report to Congress, vol. II, Chapter 38*. Davis: University of California.
- Smith, R.L. 2003. Statistics of extremes with applications in environment, insurance and finance. In: *Extreme Values in Finance, Telecommunications and the Environment*, B. Finkenstädt and H. Rootzén (eds.). Chapman & Hall/CRC Press, Boca Raton, FL. 405 p.
- Song, W., F. Weicheng, W. Binghong, and Z. Jianjun. 2001. Self-organized criticality of forest fire in China. *Ecological Modelling* 145:61-68.
- Strategic Issues Panel on Fire Suppression Costs. 2004. Large fire suppression costs—Strategies for cost management. Available online at <http://www.fireplan.gov/resources/2004.html>. Last accessed April 20, 2007.
- Strauss, D., L. Bednar, and R. Mees. 1989. Do one percent of the forest fires cause ninety-nine percent of the damage? *Forest Science* 35:319-328.
- Swetnam, T.W., and J.L. Betancourt. 1990. Fire-Southern Oscillation relations in the southwestern United States. *Science* 249:1017-1020.

- Vankat, J.L., and J. Major. 1978. Vegetation changes in Sequoia National Park, California. *Journal of Biogeography* 5:377-402.
- Westerling, A.L., T.J. Brown, A. Gershunov, D.R. Cayan, and M.D. Dettinger. 2003. Climate and wildfire in the Western United States. *Bulletin of the American Meteorological Society* 84(5):595-604.
- Westerling, A.L., H.G. Hidalgo, D.R. Cayan, and T.W. Swetnam. 2006. Warming and earlier spring increases Western United States forest wildfire activity. *Science* 313:940-943.
- Zhang, Y.-H., M.J. Wooster, O. Tutubalina, and G.L.W. Perry. Monthly burned area and forest fire carbon emission estimates for the Russian Federation from SPOT VGT. *Remote Sensing of Environment* 87:1-15.

F408