

Estimate of Maximum Insurance Loss due to Bushfires

X.G. Lin^a, P. Moran^b, A. Sullivan^c, S. Dunstall^b and G. Carter^d

^a*CSIRO Mathematics, Informatics and Statistics, Canberra*

^b*CSIRO Mathematics, Informatics and Statistics, Melbourne*

^c*CSIRO Ecosystem Sciences, Canberra*

^d*Western Power, Perth*

Email: Xunguo.Lin@csiro.au

Abstract: Bushfire is one of the major natural (and man made) disasters in Australia due to a combination of dense bush/grass land coverage, dry weather and less logging. Recent bushfire events include the Black Saturday fire in Victoria in 2009 and the Canberra fire in 2003. In both cases, many human lives were lost and insurance losses were hundreds of millions of dollars or more.

In this paper, we report on the application of Extreme Value Analysis (EVA) to estimate 50-year and 100-year return levels of insurable bushfires losses. Both the Generalised Extreme Value model and the Generalised Pareto Distribution model were applied to a logarithm-transformed dataset of past claims of only 14 years. A sensitivity analysis was also conducted to test the robustness of the model fitting.

To supplement the EVA work, past fire weather conditions for a study area were analysed and used in simulations to illustrate the damage that might occur due to fires initiated by the electricity distribution network on high fire danger days. The simulated fires burnt through various terrain asset classes including native forest, cropping lands and urban areas. The spatial vegetation and land cover information were acquired based on 2001 and 2003 datasets.

The total burnt area for each asset class was calculated for each fire using a GIS on an hourly basis for the first four hours after ignitions. The per-hectare valuations for each asset class were used to establish an estimate of the total maximum potential insurance claim for each simulated bushfire. The scale of these total claims was in keeping with the return levels estimated by the EVA.

Future work now underway will extend the simulation approach and embed it within a quantitative risk model for estimating maximum foreseeable losses from fires started by electricity distribution networks.

Keywords: Extreme value analysis, Generalised Extreme Value Distribution, Generalised Pareto Distribution, McArthur Forest Fire Danger Index, Bushfire simulation, Bushfire loss

1. INTRODUCTION

Australian bushfires have caused major damage to properties and the loss of human life. Recent large scale bushfire events were the 2009 Victorian Black Saturday Bushfire and the 2003 Canberra Bushfire. The cost of the Black Saturday Bushfire has been estimated to be about \$4.4 billion (see Appendix A of Victorian Bushfire Royal Commission (2010)) with 173 lives lost. Within the \$4.4 billion total cost, about \$1.2 billion was in the general (property, contents and motor vehicle) insurance claims paid.

The 2003 January Canberra Bushfire was estimated to cost about \$300 million with four people killed (McLeod, 2003). About \$250 million was the insured loss (Wuart, 2003).

As the first such study of its kind in Australia, Ashe *et al.* (2009) has estimated the total cost of fire in Australia to be \$12 billion or 1.3% of GDP, based on 2005. Within this total cost, 57% relates to costs in anticipation of fire, 29% for response activities and that the remaining 14% results from the consequences of fire. Therefore, the investment in fire mitigation (57% + 29% = 86%) is about six times more than the actual fire consequences (14% = \$1.67 billion). Ashe *et al.* (2009) raises questions as to whether the current system adopted in Australia is economically efficient. According to Ashe *et al.* (2009), within the \$1.67 billion fire consequences, about \$1 billion was due to property losses in which most will be footed by insurance companies. Our research contributes towards the understanding of the current fire system in two fronts: (1) to better estimate the possible large economic loss by catastrophic fire events such as one in 50-year fire; (2) to better understand the fire spread behaviour under extreme fire weather by means of simulation.

This paper studies the maximum insurance loss due to bushfires caused by electricity transmission and distribution systems. “Maximum” in this context means the return levels such as the one-in-50-year loss (i.e., insurance pay out) or the one-in-100-year loss. A general approach to the problem is to apply the extreme value analysis (EVA) techniques to estimate these return levels. There are two distinctive types of EVA methods: Generalised Extreme Value (GEV) distribution which generally uses the annual maxima in its modelling; and Generalised Pareto Distribution (GPD) which uses those data points which are greater than a given threshold value. An introduction of EVA can be found in Embrechts *et al.* (2001) and Coles (2001).

As a rule of thumb, for reliable GEV modelling a minimum of 20 years of data is normally required. The reality for this study is that we only have 14 years of data on insurance claims related to bushfire damage. To compensate for the shortness of data, a sensitivity analysis is used to test the robustness of the modelling. Furthermore, fire danger weather conditions based on McArthur Forest Fire Danger Index (FFDI) (McArthur, 1967) were analysed for a client-specified study area based on weather data recorded nearby. Extreme fire weather conditions defined by the 95th quantile of FFDI daily values were used as a basis for a fire spread simulation.

2. METHODOLOGY

The classical extreme value theory is based on the analysis of the largest (or smallest) value in an epoch. An epoch is usually taken to be a calendar year.

$$M_n = \max\{X_1, \dots, X_n\}, \quad (1)$$

where X_1, \dots, X_n is a sequence of independent random variables having a common distribution function F . Then

$$\begin{aligned} \Pr\{M_n \leq z\} &= \Pr\{X_1 \leq z, \dots, X_n \leq z\} \\ &= \Pr\{X_1 \leq z\} \times \dots \times \Pr\{X_n \leq z\} \\ &= \{F(z)\}^n. \end{aligned} \quad (2)$$

If we accept that F is unknown then we have to look for approximate models for F^n , which can be estimated based on the extreme data only. This approach is similar to the usual practice of approximating the sample means by the Normal distribution which is justified by the Central Limit Theorem.

To avoid the possibility of distribution of M_n being degenerated to a point mass when $n \rightarrow \infty$, a linear renormalization of the variable M_n can be done by $M_n^* = (M_n - b_n)/a_n$ for appropriate sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ which stabilize the location and scale of M_n^* .

2.1. Generalised Extreme Value (GEV) Theorem (Coles, 2001)

If there exist sequences of constants $a_n > 0$ and b_n such that

$$\Pr \left\{ \frac{M_n - b_n}{a_n} \leq z \right\} \rightarrow G(z) \quad \text{as } n \rightarrow \infty \quad (3)$$

for a non-degenerate distribution function G , then G is a member of the GEV family:

$$G(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (4)$$

defined on $z : 1 + \xi(z - \mu)/\sigma > 0$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$.

μ is called the location parameter; σ is the scale parameter and ξ is the shape parameter. When $\xi > 0$, it is the case of Type II or Fréchet distribution; and if $\xi < 0$, it is the case of Type III or Weibull distribution.

When $\xi = 0$, G has the limit form

$$G(z) = \exp \left\{ - \exp \left[- \left(\frac{z - \mu}{\sigma} \right) \right] \right\} \quad (5)$$

which is Type I or Gumbel distribution.

Note that, when $\xi \leq -0.5$ the maximum likelihood estimator is unlikely to be obtainable or does not have standard asymptotic properties. However, the case of $\xi \leq -0.5$ corresponds to a very short bounded upper tail, which is rarely encountered in applications.

2.2. Peaks over threshold (POT) and Generalised Pareto Distribution (GPD)

A major criticism of traditional extreme value theory (i.e., GEV) is that it only considers a single maximum within each epoch. It ignores other extreme events that may have occurred in each epoch which might even have greater values (i.e., more extreme) than the maxima in other epochs.

An alternative approach, often referred to as the POT approach, is to consider all values greater than a given threshold value, no matter if they are from the same epoch or different ones.

Given a threshold u , the distribution of excess values of X over u is defined by:

$$F_u(z) = \Pr \{X - u \leq z | X > u\} = \frac{F(z + u) - F(u)}{1 - F(u)} \quad (6)$$

which represents the probability that the value of X exceeds u at most by an amount z . It has been shown (Balkema and de Haan (1974) and Pickands (1975)) that for a sufficiently high threshold, u , the distribution function of the excess, $F_u(z)$, converges to the GPD which has a distribution function given by:

$$G(z|\mu, \sigma, \xi) = \begin{cases} 1 - (1 + \xi \frac{z - \mu}{\sigma})^{-1/\xi}, & \xi \neq 0; \\ 1 - e^{-(z - \mu)/\sigma}, & \xi = 0. \end{cases} \quad (7)$$

If $\xi = 0$, the GPD corresponds to an exponential distribution (medium-size tail); when $\xi > 0$, it takes the form of the ordinary Pareto distribution (long tailed); when $\xi < 0$, it is known as a Pareto II type distribution (short tailed), which is also upper bounded by $(\mu - \sigma/\xi)$.

The parameters of a GPD can be estimated with various methods such as the maximum likelihood (ML) method (Davison, 1984) and probability weighted moments (PWM) method (Hosking and Wallis, 1987). An important property of the GPD is that if $\xi < 1$, then the conditional mean exceedance (CME) over a threshold, u , is a linear function of u :

$$E(X - u | X > u) = \frac{\sigma + \xi u}{1 - \xi}. \quad (8)$$

The linearity of the CME plot can thus be used as an indicator of the appropriateness of the GPD model. The shape and scale parameters of GPD may also be estimated using this property. Define the following sample mean excess (SME) function as,

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)^+}{\sum_{i=1}^n 1_{\{X_i > u\}}} \quad (9)$$

with respect to the threshold u , where the '+' sign ensures only the positive values of $(X_i - u)$ will be taken into account. That is, the SME is the sum of the excesses over the threshold u divided by the number of data points which exceed u . The SME is an empirical estimate of the CME, and ξ and σ of the GPD can be determined by the slope and intercept of the SME plot, using those two equations: Slope = $\xi/(1 - \xi)$ and Intercept = $\sigma/(1 - \xi)$.

2.3. Threshold selection

The threshold should be set high enough so that the exceedances approximate the Poisson arrival rate. Otherwise the distribution of selected extremes may not converge to the GPD (Gencay *et al.*, 2001). However, the threshold must be low enough to ensure there is enough data points left for satisfactory determination of the GPD parameters. There are techniques available that may assist in the selection of an optimal threshold. One approach is to use the SME plots defined in (9). If the data is from a GPD, its CME would be linear with respect to u , see (8), if $\xi < 1$ which is our case. Hence, an appropriate threshold can be chosen by selecting the lowest value above which the SME graph is approximately a straight line.

2.4. Quantile (or return level) estimation

The quantile estimate for GEV can be obtained by inverting (4) and using the return period definition:

$$U_T = \begin{cases} \mu - \frac{\sigma}{\xi} \left\{ 1 - \left[-\ln \left(1 - \frac{1}{\lambda T} \right) \right]^{-\xi} \right\} & \xi \neq 0 \\ \mu - \sigma \ln \left[-\ln \left(1 - \frac{1}{\lambda T} \right) \right] & \xi = 0. \end{cases} \quad (10)$$

For GPD, the quantile estimate is obtained by inverting (7):

$$U_T = \begin{cases} \mu' - \frac{\sigma'}{\xi} \left[1 - \frac{1}{(\lambda T)^{-\xi}} \right] & \xi \neq 0 \\ \mu' + \sigma' \ln(\lambda T) & \xi = 0 \end{cases} \quad (11)$$

where σ' and μ' are given by $\sigma' = \sigma(1 - F(u))^\xi$ and $\mu' = u - \sigma'((1 - F(u))^{-\xi} - 1)/\xi$.

It should be pointed out that there are two forms of the GEV and GPD formula in the literature, with the difference in the sign of the shape parameter (ξ). However, the terminology of Types I, II and III is always the same.

3. MODEL VALIDATION

An important question when considering any numerical model of a complex physical process is the reliability and appropriateness of the chosen model. Graphical methods are commonly used in model validation. We used the R "ismev" package which models GEV and GPD using methods given by Coles

(2001). It conveniently provides a group of four graphical diagnostic plots, containing the probability plot, the quantile-quantile (QQ) plot, the return level plot and the density plot.

Both the GEV and GPD modelling were applied to our data after a logarithm base-10 transformation and using a variety of threshold values. For a threshold value of 3 (representing \$1,000 before the logarithm transformation), which represents the best fit, the GPD modelling results are shown in Figure 1.

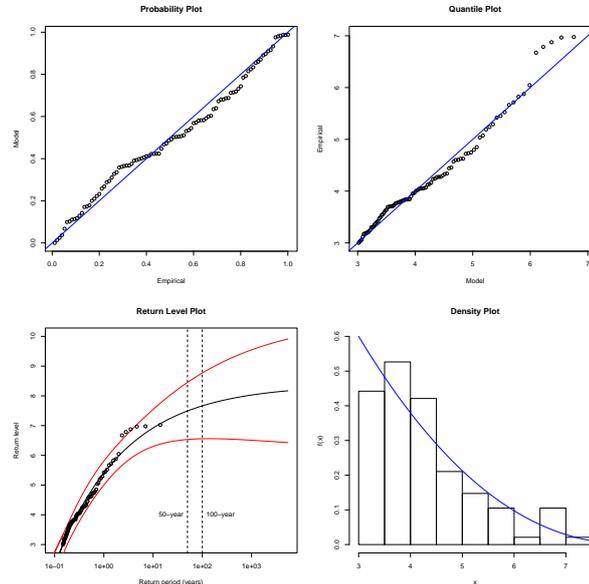


Figure 1. GPD Model fitting diagnostic plots with $u = 3$ (\$1,000).

In the top left of Figure 1 is the probability plot of the data versus the fitted GPD model. Both ends of the data points are on the diagonal line indicating that the fitted model is close to the very smaller and very larger data values. The middle region of data points also close to the diagonal line indicating that the data and the model have similar medians. It also shows the model over-estimates in the region below 0.5 and under-estimates in region above 0.5.

The bottom left is the return level plot with respect to different return periods. The dots correspond to the data points and the black curve is the return level curve from the fitted GPD model with the two red curves being the 95% confidence bounds. The top data point and the other points touch the fitted model curve but not those five data points behind the top one. This means that the fitted model provides a good representation of the data, but not to those five sub-large points. The 50-year return level (dotted vertical line on the bottom left plot) is about \$30 million (dotted line intersecting with the black curve) with the 95% confidence upper bound being below \$300 million (dotted line intersecting with the top red curve). The 100-year return level is about \$50 million with the upper bound being over \$575 million. The interpretation of the confidence bound is that we are 95% confident that the true return level should not be greater than this value. These results are listed in Table 1.

The top right is the QQ-plot which will indicate a good fit if the data points are sitting on the diagonal line. Again, it can be seen that the fitted model is close to the data except those five points. The bottom right plot is the density plot: the histogram plot is from the data and the blue curve is from the fitted GPD model. In three occasions the histogram bars are above the model curve, which means the model is under-estimating the data in these ranges.

To test the robustness of the model fitting, a sensitivity analysis was carried out by artificially inserting a \$50 million value into the claim dataset. The new GPD modelling results are listed in Table 1. The EVA modelling was sensitive to the this large data point but with some discount due to the lack of support from the rest of data. Most notably, the estimated confidence bounds become very wide due to the large gap between the artificial value and the rest of data.

Table 1. Estimated return levels computed by the EVA

Data	Threshold value	Dollar value	50-year return level	95% upper confidence bound	100-year return level	95% upper confidence bound
Original	3	\$1,000	\$31M	\$281M	\$46M	\$583M
Modified	3	\$1,000	\$86M	\$1,274M	\$149M	\$3,475M

4. EXTREME FIRE WEATHER CONDITION ANALYSIS

The fire weather conditions observed for a Perth Hills study area in WA between 1985 and 2010 were evaluated using the McArthur Forest Fire Danger Index (FFDI), see McArthur (1967). The value of FFDI depends on air temperature, relative humidity, wind speed, rainfall and the number of days since last rain. Hourly FFDI values were calculated based on weather observations recorded at stations close to our study area. The daily FFDI value was taken to be the maximum hourly FFDI value of the day.

Analysis of daily FFDI values in the 95th quantile (which were greater than 50, that is Extreme) provided us the characteristics of extreme fire weather conditions, particularly:

- The annual average number of those Extreme days was about four;
- 2010 was a record year within the past 26 years with 10 such Extreme days at one site and 23 days at the other site, that is, 2010 had many times more Extreme days than the long term average;
- On average, there was no rainfall for 19 days prior to an Extreme day;
- The average temperature was at least 34.0°C and the relative humidity was not higher than 11%;
- The average wind speed was at least 19 km/h, mainly coming from either the East or the West.

The statistical correlations between the FFDI at sites on the claim incident dates and their claim values were weak ($\leq 17\%$), which means higher FFDI values were not necessarily leading to larger insurance claims. This indicates the difficulty in estimating the maximum insurance loss based on the extreme weather conditions alone.

5. SCENARIO-BASED BUSHFIRE SIMULATION

A scenario-based bushfire simulation was carried out to illustrate the potential loss associated with a fire burning under the extreme fire weather conditions. Six ignition points on the electricity distribution network within a 150 sq km Perth Hills study area were chosen, with three selected so as to maximize damage under a predominantly easterly wind regime and the other three for a predominantly westerly regime. Fire spread was simulated for a period of four hours and hourly perimeters were plotted. The six simulations were carried out using manual computations of fire spread coupled with various GIS functions and data.

The rate of spread averaged between 2.6 km/h and 3.5 km/h for the four hours. The final perimeter was used to determine the level of likely loss for each fire. The fire perimeters are considered somewhat conservative, although effects of fuel modification, suppression and barriers such as creeks and rivers may reduce the total area affected by any one fire. Areas burnt ranged from approximately 2600 ha to 5200 ha for the six fire simulations within a period of four hours. One simulation is summarized in Figure 2. This had the largest area burnt within the first four hours after ignition.

To estimate the possible fire loss within the burnt area, two vegetation and land use datasets were acquired through Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES) with the first dataset being released in 2001 and the second one in 2003. Using these land cover information and a per-hectare valuation assumption for each land asset class, the possible loss from the largest burnt area was estimated to be between \$300 to \$400 million. This is in keeping with the return levels estimated by the EVA, see Table 1.

6. CONCLUSIONS

Extreme value analyses (GEV and GPD modelling) were carried out on insurance claim data after a log-10 transformation. The estimated 50-year and 100-year return levels were estimated with 95% confidence

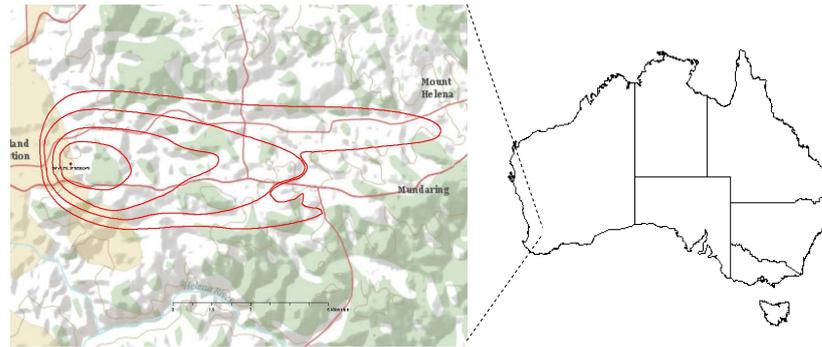


Figure 2. Simulation of fire starting at $116^{\circ}4'1''\text{E}$, $31^{\circ}53'38''\text{S}$ and burning under a predominantly westerly wind. Hourly perimeters are shown expanding from the ignition point.

intervals. A sensitivity analysis was also carried out by artificially adding a large-valued (\$50 million) data point. This addition resulted in different GPD models with much larger estimated return levels and even larger confidence intervals, due to the increased variances and standard errors.

Fire weather conditions at two weather stations near the study area were analysed using the McArthur FFDI. The fire weather characteristics of those days with their FFDI values in the 95th percentile (i.e., extreme) were identified. This information was used for a fire spread simulation. This simulation provided an insight into potential impact of future extreme fires. Work now underway will automate and extend the simulation approach, and embed it within a quantitative risk model for estimating maximum foreseeable losses from fires started by electricity distribution networks.

REFERENCES

- Ashe, B., K. J. McAneney, and A. J. Pitman (2009). Total cost of fire in Australia. *Journal of Risk Research* 12(2), 121–136.
- Balkema, A. A. and L. de Haan (1974). Residual lifetime at great age. *Annals of Probability* 2, 792–804.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.
- Davison, A. C. (1984). Modelling excesses over high thresholds, with an application. In J. T. de Oliveira (Ed.), *Statistical Extremes and Applications*, Reidel, Dordrecht, pp. 461–482.
- Embrechts, P., C. Kluppelberg, and T. Mikosch (2001). *Modelling Extremal Events for Insurance and Finance*. Springer.
- Gencay, R., F. Selcuk, and A. Ulugulyagci (2001). EVIM: A software package for extreme value analysis in MATLAB. *Studies in Nonlinear Dynamics and Econometrics* 5(3), 213–239.
- Hosking, J. R. M. and J. R. Wallis (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29, 339–349.
- McArthur, A. G. (1967). Fire behaviour in eucalypt forests. Forestry and Timber Bureau Leaflet 107, Commonwealth Department of National Development, Canberra.
- McLeod, R. (2003, August). Inquiry into the operational response to the January 2003 bushfires in the ACT. http://www.cmd.act.gov.au/_data/assets/pdf_file/0008/113939/McLeodInquiry.pdf.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* 3(1), 119–131.
- Victorian Bushfire Royal Commission (2010, July). Victorian bushfire royal commission final report. <http://www.royalcommission.vic.gov.au/Commission-Reports/Final-Report>.
- Wuart, P. (2003, January). The January 2003 Canberra bushfires. http://www.extremeweatherheroes.org/media/25782/canberra_bushfires_jan_2003.pdf.