Gridded return values of McArthur Forest Fire Danger Index across New South Wales

S.A. Louis
New South Wales Severe Weather Section, Bureau of Meteorology

(Manuscript received September 2013; revised October 2014)

Introduction

Bushfire represents a significant risk within the Australian environment, and providing a spatial estimate of the risk represents a challenge for producing planning guidelines. The primary method for estimating day-to-day fire weather risk within Australian fire fighting operational environments has been the McArthur Forest Fire Danger Index, which is an empirical index taking into account both fire weather and fuel moisture factors. This paper considers a statistical method for estimating long-period return values for the McArthur Forest Fire Danger Index across New South Wales, using Extreme Value theory. This approach is then applied to a spatial reanalysis dataset so that the level of fire danger risk can be evaluated across the landscape.

Bushfire represents a significant risk within the Australian environment. The McArthur Mark V forest fire danger meter is the most commonly used method for measuring bushfire risk in Australian emergency management jurisdictions. The McArthur meter provides an empirical scale that takes into account weather and fuel dryness factors to produce a Forest Fuel Fire Danger Index (FFDI), which is then used as a predictor for fire behaviours such as rate of spread and flame height. The meter as developed by McArthur originally took the form of a circular slide rule (Luke and McArthur, 1978), but the FFDI was later expressed as equations (Noble et al., 1980) and an improved formula for calculating the Drought Factor component was developed (Griffiths, 1999).

To quantify the level of long-term risk attached to fire weather in New South Wales we consider return values of FFDI. This approach has the advantage that these statistics have a long history of use in the context of heavy rainfall and flooding (Kharin et al., 2009; Water Data, 1981), extreme winds (Simiu and Heckert, 1995), extreme temperatures (Kharin et al., 2009; Perkins et al., 2009; Perkins et al., 2012) and large fire sizes (Alvarado et al., 1998), and as a result emergency service agencies have developed a degree of familiarity with this type of analysis.

For risk analysis it is required to calculate FFDI return values over periods that are longer than the available observational and gridded datasets, and so a statistical approach using extreme value theory is considered. Length and quality of observation time series provides a particular challenge when working with FFDI as the index requires measurements of rainfall, wind speed, humidity and temperature, which can lead to an increased number of gaps in the record. Further, changes in instrumentation and siting can introduce inhomogeneities into the observation record. In general, the longer the return period relative to the length of the dataset, the more uncertainty will be introduced into the statistical model estimate of return value.

Care needs to be taken when interpreting extreme values of McArthur FFDI, as the meter was originally designed to be capped at a value of 100, and the actual test fires that went into the empirical design of the meter were conducted under conditions with FFDI lower than 50 (McArthur, 1967). There has, however, been some recent work showing that around 64 per cent of reported house losses in Australia have occurred when the FFDI exceeds 100 (Blanchi et al., 2010).

There are two main statistical approaches considered in the literature. The first is to fit the distribution of observed events measured over a fixed period (e.g. the time series of annual maxima) to the Generalised Extreme Value (GEV) distribution (see e.g. Perkins et al., 2009; Perkins et al., 2012). This approach fails to take into account that the data are not...
identically distributed throughout the year, because there is a pronounced seasonal variation, and as a result will tend to underestimate the return (FFDI) value (Carter and Challenor, 1981) associated with a specified return frequency. In a flood forecasting context, the approach generally used is to fit a time series of annual maxima in stream flow to a GEV distribution, most commonly a log-Pearson type III distribution (Carter and Challenor, 1981). This approach can be used to model fire weather, however multiple extreme events may fall within the same year, with the result that significant data points are not considered by the analysis. Using a time series of maxima over a shorter fixed period, such as monthly or seasonally, can improve the estimate (Carter and Challenor, 1981). This approach would, however, introduce a number of insignificant data points over the cooler months, and could still fail in some cases to capture multiple synoptically independent extreme events.

The second approach is to fit those events exceeding a certain threshold to a Generalised Pareto Distribution (GPD) (see e.g. Begueria, 2005; Frei derichs et al., 2009). This approach is also known as a 'Peaks–Over–Threshold' (POT) or partial duration series approach. This approach has the advantage that it can capture all significant events above the given threshold, even when they fall within the same season, and the statistical model does not need to account for low values sampled from, for example, a La Nina–influenced fire season. For this reason we have chosen to proceed with the POT approach in this work.

The key design decision in a POT model is the choice of threshold value. The effect of threshold choice has been shown, theoretically and using numerical simulation, to modify the return interval distribution, particularly in the case of short return intervals (Santhanam and Kantz, 2008). In the context of hydrologic data, the POT method is subject to a sensitivity to threshold selection, including any tendency for the estimates to become unstable at higher threshold levels due to natural sampling variability (Begueria, 2005). Begueria (2005) also outlined a procedure for assessing the suitability of a threshold choice, and we follow that method in this paper.

There are a number of other criteria that the sample dataset should meet in order to validate the choice of GPD as a statistical model (Coles, 2001). The data fitted should be stationary, which in the present context means that there should not be a statistically significant linear trend in the time series of values exceeding the threshold, or of the frequency of values exceeding the threshold. An L-moments plot should be checked to ensure that the third and fourth moments of the sample data match the theoretical relationship for a given choice of threshold. Finally, the frequency of exceedence event arrivals should be able to be modelled by a Poisson process.

Bushfire risk also varies spatially across the landscape. To evaluate the spatial risk across New South Wales this paper takes a two-stage approach; first the suitability of the statistical models for calculating return values was assessed using observation data, and then the relevant model was applied to each grid point in a model reanalysis dataset of FFDI. Details of the dataset used are outlined in the next section.

**Data**

In this paper we make use of the quality-controlled observational dataset of fire weather parameters developed by Lucas (Lucas, 2010) covering the period June 1972 to December 2009. This dataset contains data from 16 stations across New South Wales, which is a fairly limited sampling for determining a spatial assessment of risk and is also quite a short record for the purposes of estimating risk levels on a 50 to 100 year time frame.

To extend the scope of the analysis we extended the datasets for these 16 stations using as long an observation record as possible, and computed FFDI values for another nine sites where the availability of observations was more limited. Data were extracted from the Australian Data Archive for Meteorology (ADAM). Table 1 contains details of the full set of observation sites used in the study, where the first 16 are those also used in the Lucas dataset. As a number of stations have been relocated and been allocated new station numbers, these are treated as composite sites, and the join dates of these data time series have been included in Table 1 for reference. To calculate daily FFDI values the Noble et al. (1980) computational formulation of the McArthur Mk V forest fire danger meter,

\[
FFDI = 2 \times \exp(0.987 \ln(DF) - 0.45 + 0.0338T + 0.0234W_s - 0.0345RH)
\]

was used, where \(DF\) is the drought factor calculated using the procedure in Griffiths (1999), \(T\) is the daily maximum temperature in °C, \(W_s\) is the 3.00 pm 10-minute average wind speed in km/h and \(RH\) is the 3.00 pm relative humidity. The daily FFDI value was calculated using these values, due to the long availability of 3.00 pm synoptic data in the climate record. It is important to note that these daily FFDI values based on 3.00 pm conditions will be different from the daily maximum FFDI calculated using higher-frequency observations from automatic weather stations, with a recent study focusing on Tasmania showing approximately a difference of 5–15 in 99th percentile values (Fox-Hughes, 2011). For individual days where there is a significant sea breeze or wind change arriving prior to 3.00 pm, causing a marked temperature drop, the difference could be even higher.

Table 1 also shows that there is a large proportion of missing data from a number of the stations, in particular at Hay, Tibooburra, Cessnock and Thredbo. Missing data is a particular problem for FFDI records, as the index is unable to be calculated if any one of the temperature, humidity or wind speed observations are missed on a given day. It should be noted that for those stations where a large number of data points are missing, an underestimate of the return
level is likely, particularly if a significant missed data period coincides with a significant summer of fire weather.

Some limited quality control of the data (additional to that implemented by the Bureau of Meteorology) was undertaken. This was applied by considering the highest few values of FFDI in each dataset and discarding unrealistic values, as these would have a disproportionately large effect on this type of analysis. To assess this the Grubbs test statistic for outliers was computed and used in conjunction with a subjective assessment of the component weather elements in cases where this statistic indicated a potential outlier (Grubbs, 1950). Table 2 shows a list of the data points that were considered to be unreasonable and excluded. For the data points where there was a nearby station recording data (less than 150 km away) for the same day, the nearby conditions are also included in Table 2 for comparison. For the Moruya, Tibooburra and Bourke data points listed, no suitable nearby data was available for comparison.

The choice of which dataset to use is effectively a design trade-off, where using the longer datasets will introduce data quality and inhomogeneity errors, but also reduce the amount of error associated with extrapolation of the statistical model, because of reduced sampling variability.

To determine how much the high level of quality control applied to the Lucas dataset affected the analysis, the estimated return values based on the dataset from Lucas level is likely, particularly if a significant missed data period coincides with a significant summer of fire weather.

Some limited quality control of the data (additional to that implemented by the Bureau of Meteorology) was undertaken. This was applied by considering the highest few values of FFDI in each dataset and discarding unrealistic values, as these would have a disproportionately large effect on this type of analysis. To assess this the Grubbs test statistic for outliers was computed and used in conjunction with a subjective assessment of the component weather elements in cases where this statistic indicated a potential outlier (Grubbs, 1950). Table 2 shows a list of the data points that were considered to be unreasonable and excluded. For the data points where there was a nearby station recording data (less than 150 km away) for the same day, the nearby conditions are also included in Table 2 for comparison. For the Moruya, Tibooburra and Bourke data points listed, no suitable nearby data was available for comparison.

The choice of which dataset to use is effectively a design trade-off, where using the longer datasets will introduce data quality and inhomogeneity errors, but also reduce the amount of error associated with extrapolation of the statistical model, because of reduced sampling variability.
Statistical model

The statistical analysis package ‘R’ with the ‘POT’ library was used to perform the majority of the statistical analysis, including the maximum likelihood fits to the datasets (Ribatet, 2006). The Python (v2.6) programming language was used to apply the statistical model over each data point in the gridded dataset. In the following discussion the example plots are shown for the observation data at Wagga Wagga Airport covering the period 1951 to 2009, with some corresponding plots for the other stations included as appendices.

The first requirement for the statistical fit to be valid is that the number of the exceedance events within a fixed period must follow a Poisson distribution. To test this assumption, we can use the dispersion index (DI) statistic (Begueria, 2005; Cunnane, 1979), which is a measure of the ratio between the variance and mean of a sample. Figure 1 shows a plot of the DI relative to different threshold values for the Wagga Wagga daily fire danger index time series, where the grey shaded region indicates DI values for which the Poisson process holds. For low threshold values (corresponding to a high number of occurrences per year) this assumption does not hold, due to the seasonal correlation. To increase the range of valid threshold choices, a simple de-clustering approach was applied to ensure that consecutive significant events that may be the result of the same weather pattern (and hence not independent) are not all included in the analysis. In this case, the de-clustering was achieved by treating sequences of consecutive days above the threshold as a single event, retaining the maximum value from the time series as the relevant data point and setting the other values to zero.

Table 2. Data points excluded from the recreated time series of daily FFDI observations. Lines without listed FFDI values denote neighbouring sites providing the basis for the exclusion of these FFDI values.

<table>
<thead>
<tr>
<th>Station</th>
<th>Date</th>
<th>Temperature (°C)</th>
<th>Dew Point (°C)</th>
<th>Wind Speed (km/h)</th>
<th>FFDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bourke</td>
<td>10/10/1913</td>
<td>35.1</td>
<td>-2.9</td>
<td>95</td>
<td>272</td>
</tr>
<tr>
<td>Williamtown</td>
<td>04/07/1952</td>
<td>19.4</td>
<td>-5</td>
<td>111.2</td>
<td>164</td>
</tr>
<tr>
<td>Sydney Airport</td>
<td>16/11/1952</td>
<td>36.7</td>
<td>4</td>
<td>83.5</td>
<td>193</td>
</tr>
<tr>
<td>Williamtown</td>
<td>24/11/1957</td>
<td>38.9</td>
<td>4</td>
<td>59.4</td>
<td>117</td>
</tr>
<tr>
<td>Tibooburra</td>
<td>05/11/1965</td>
<td>42.8</td>
<td>5</td>
<td>81.7</td>
<td>235</td>
</tr>
<tr>
<td>Cessnock</td>
<td>23/12/1990</td>
<td>43.5</td>
<td>7</td>
<td>68.4</td>
<td>180</td>
</tr>
<tr>
<td>Paterson</td>
<td>43.1</td>
<td>21</td>
<td>81.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williamtown</td>
<td>40.9</td>
<td>14</td>
<td>33.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tibooburra</td>
<td>42.8</td>
<td>8</td>
<td>27.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paterson</td>
<td>41.1</td>
<td>3.2</td>
<td>40.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

were compared with the return values based on the longer recreated datasets (discussed in a later section).

The value of a Regional Climate Model (RCM) reanalysis in investigating fire weather climatology over Tasmania has been demonstrated (Grose et al., 2011). We consider an RCM reanalysis dataset of FFDI developed by Evans (Evans and McCabe, 2010). These RCM data was generated using a Weather Research and Forecasting (WRF) model reanalysis downscaled to a 10 km grid resolution for the period covering January 1985 to December 2008, with FFDI values calculated as above over the down scaled grid. Overall, the RCM does a good job of reproducing the observed FFDI (Clarke et al., 2013; Evans and McCabe, 2010), and is able to capture aspects of the diurnal cycle (Evans and Westra, 2012). Relative humidity and wind speed introduce the largest component errors to the reanalysis FFDI values, with wind speed particularly influencing the extreme value cases that are crucial for estimating return values (Clarke et al., 2013). Clarke also identified a bias towards higher relative humidities than observed along the New South Wales coast, and hence a bias towards lower FFDI values in these areas.

To investigate whether the WRF reanalysis dataset is able to produce comparable estimates for the 50-year return levels, we calculated return values based on the nearest grid points to each of the observation stations in Table 1.

We identified another issue with the WRF data relating to a land use classification layer. A number of grid cells that included areas of land, but that were classified as water body, were associated with very low and unrepresentative FFDI values in the dataset. This affected two grid cells that included the observation points at Williamtown and Moruya, so for comparison we chose to use the closest grid cell that was not classified as water body.

<table>
<thead>
<tr>
<th>Station</th>
<th>Date</th>
<th>Temperature (°C)</th>
<th>Dew Point (°C)</th>
<th>Wind Speed (km/h)</th>
<th>FFDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bourke</td>
<td>10/10/1913</td>
<td>35.1</td>
<td>-2.9</td>
<td>95</td>
<td>272</td>
</tr>
<tr>
<td>Williamtown</td>
<td>04/07/1952</td>
<td>19.4</td>
<td>-5</td>
<td>111.2</td>
<td>164</td>
</tr>
<tr>
<td>Sydney Airport</td>
<td>16/11/1952</td>
<td>36.7</td>
<td>4</td>
<td>83.5</td>
<td>193</td>
</tr>
<tr>
<td>Williamtown</td>
<td>24/11/1957</td>
<td>38.9</td>
<td>4</td>
<td>59.4</td>
<td>117</td>
</tr>
<tr>
<td>Tibooburra</td>
<td>05/11/1965</td>
<td>42.8</td>
<td>5</td>
<td>81.7</td>
<td>235</td>
</tr>
<tr>
<td>Cessnock</td>
<td>23/12/1990</td>
<td>43.5</td>
<td>7</td>
<td>68.4</td>
<td>180</td>
</tr>
<tr>
<td>Paterson</td>
<td>43.1</td>
<td>21</td>
<td>81.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williamtown</td>
<td>40.9</td>
<td>14</td>
<td>33.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tibooburra</td>
<td>42.8</td>
<td>8</td>
<td>27.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paterson</td>
<td>41.1</td>
<td>3.2</td>
<td>40.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistical model

The statistical analysis package ‘R’ with the ‘POT’ library was used to perform the majority of the statistical analysis, including the maximum likelihood fits to the datasets (Ribatet, 2006). The Python (v2.6) programming language was used to apply the statistical model over each data point in the gridded dataset. In the following discussion the example plots are shown for the observation data at Wagga Wagga Airport covering the period 1951 to 2009, with some corresponding plots for the other stations included as appendices.

The first requirement for the statistical fit to be valid is that the number of the exceedance events within a fixed period must follow a Poisson distribution. To test this assumption, we can use the dispersion index (DI) statistic (Begueria, 2005; Cunnane, 1979), which is a measure of the ratio between the variance and mean of a sample. Figure 1 shows a plot of the DI relative to different threshold values for the Wagga Wagga daily fire danger index time series, where the grey shaded region indicates DI values for which the Poisson process holds. For low threshold values (corresponding to a high number of occurrences per year) this assumption does not hold, due to the seasonal correlation. To increase the range of valid threshold choices, a simple de-clustering approach was applied to ensure that consecutive significant events that may be the result of the same weather pattern (and hence not independent) are not all included in the analysis. In this case, the de-clustering was achieved by treating sequences of consecutive days above the threshold as a single event, retaining the maximum value from the time series as the relevant data point and setting the other values to zero.

To assist with selecting the threshold at which to apply
the clustering and POT analysis, plots of the mean threshold exceedance (MTE) were produced for each set of station data, which show the average amount by which the over-threshold values exceed the threshold for a range of different threshold values. For a given set of data, if the selected threshold is too low, then the bulk of the low end values will dominate the statistical fit process, giving a poor fit for the upper tail of the distribution. If the selected threshold is too high, the statistical process becomes dependent on a small number of events, and the model fit becomes unstable. In the MTE plot, the range of appropriate threshold values shows up as an approximately linear region (Coles, 2001). Figure 2(a) shows the MTE plot for the Wagga Wagga observed FFDI dataset, with a linear trend line over the region where the threshold is between 20 and 50. The dashed blue line on this plot corresponds to the 98th percentile value in the Wagga Wagga dataset of 40.5, indicating that this is an appropriate choice of threshold for this case. Figure 2(b) shows the equivalent plot for the nearest grid point to Wagga Wagga in the reanalysis dataset, with the dashed blue 98th percentile line near the end of, but still within, the linear portion of the plot.

One issue with selecting thresholds is that each station may have differing ranges of suitable thresholds, but we cannot easily check threshold suitability for each cell in a gridded dataset. Hence we selected the 98th percentile level to use for the threshold, as this can be easily calculated for each cell. The dotted green line in Fig. 1 shows that the requirement on frequency of exceedance events also holds for a 98th percentile threshold for Wagga Wagga. Figures A1 and B1, provided in the appendices, show the MTE plots for the remaining stations and for the nearest reanalysis grid cell for each station, showing that in most cases (Coonabarabran and Thredbo are marginal) the 98th percentile does fall within the linear region and does not correspond to an unstable model fit. Other choices of percentile as a threshold were tested, ranging from the 90th percentile to the 99.5th percentile, but tended to be unsuitable for a greater number of stations by this test.

Another diagnostic tool to check the suitability of the GPD and the chosen threshold level is to consider an L-moments plot (Begueria, 2005; Hosking, 1990). If the GPD is suitable, the locus of points should remain close to the line representing the GPD for threshold choices producing stable statistical fits. Good choices for the threshold value should coincide with the line.

Figure 3(a) shows that in the case of the Wagga Wagga observation data the curve does remain close for low threshold values, and that the GPD is a suitable choice in this case. The 98th percentile threshold value of 40.5, gives a ratio of $\tau_3$ to $\tau_4$ that fits the GPD well. To check whether the
98th percentile value is a suitable threshold for a GPD for all of the station datasets, the sample \( r_3 \) and \( r_4 \) were calculated for each case and compared to the theoretical curve. Figure 3(b) shows that the ratio of sample L-moments is close to the theoretical curve for each of the stations, which is consistent with the use of the GPD as a statistical model for each dataset.

One of the assumptions taking a POT approach is that the data to be fitted must be stationary. Figure 4 shows a plot of the estimated trend in the median, as well as a range of other deciles, for the Wagga Wagga FFDI data that exceed the 98th percentile threshold value of 40.5. 95 per cent confidence intervals for the trend are also calculated using the rank inversion method. There is a slight positive trend evident for each decile, however a zero trend is well within the 95 per cent confidence interval bounds, and hence we may conclude that the condition of stationarity may be reasonable for the purposes of this analysis. Although there is no statistically significant trend in the values exceeding the threshold level, there is a statistically significant linear trend when all values in the Wagga Wagga time series are considered (not shown).

Figure C1 shows that there is no evidence for non-stationarity of the threshold exceedances for the majority, but not all of the observation sites considered, with Williamtown, Moruya and Thredbo being the sites for which there is a significant trend in a number of the deciles at the five per cent significance level. A more thorough analysis of FFDI trends within the Lucas dataset also found significant increases in cumulative, 90th percentile and seasonal median FFDIs in New South Wales, particularly for sites in the west of the state (Clarke et al., 2012). The assumption of stationarity is clearly the condition for which the data is most questionable, however for the sake of simplicity further consideration of this point was not undertaken in this work. Future work should look in more detail at the effect of trends within the FFDI datasets, including determining whether return periods for a given FFDI value are decreasing over time.

Having established the suitability of the GPD and the 98th percentile as a threshold value at the majority of locations, we applied a maximum likelihood fit for each of the stations. Using the estimated GPD model parameters we can then calculate the return values using Eqn 2,

\[ z_n = u + b(pmN)^{1/s} - 1/s \]

where \( u \) is the threshold used, \( b \) is the fitted scale parameter, \( s \) is the fitted shape parameter, \( m \) is the period, and \( N \) is the number of data points per period. For the case of a 50-year return period we use \( m = 50 \), with \( N = 365.25 \) indicating that we want the period to be in years. The parameter \( p \) represents the proportion of data points exceeding the threshold value, \( u \).

For comparison we also applied a maximum likelihood fit of the GEV distribution to the time series of annual (July–June) maxima (using Stephenson, 2002), where the July–June window was chosen to encompass the fire season in New South Wales. Figure 5 shows the fitted probability density curve and return level plots for the Wagga Wagga FFDI dataset. This shows that the GPD provides a subjectively better fit to the POT data than the fit of the GEV distribution to the season maximum data. Note that the confidence intervals for the return value estimates in Figs 5–7 are calculated using a profile log likelihood approach, and only cover the component of uncertainty due to sampling error.
Fig. 3. L-moment plots shows the theoretical relationship between the between third and fourth moments ($\tau_3$ and $\tau_4$) of the GPD as a line, with sample $\tau_3$ and $\tau_4$ for a range of different threshold values over plotted as a locus of points. (a) L-moment plot for POT analysis applied to Wagga Wagga observed daily FFDI, where each point represents a different threshold value. Grey line shows the theoretical relationship between $\tau_3$ and $\tau_4$ for the GPD. (b) L-moments for each of the station observation datasets using a 98th percentile threshold.

Fig. 4. Quantile regression plot showing estimated trend of various deciles of the Wagga threshold exceedance events. The grey shaded area shows the 95 per cent confidence interval associated with the estimates.
Fig. 5. (a) Histogram of Wagga Wagga observed daily FFDI values exceeding threshold of 40.5, with fitted GPD. The solid line shows the probability density function of the fitted statistical model, while the dashed line shows the empirical estimate of the probability density function from the sample data. (b) Plot of FFDI return levels for Wagga Wagga based on fitted GPD, with 95 per cent confidence intervals shown as dashed lines. The solid line shows the relationship between return level and return period for the fitted statistical model. (c) Histogram of Wagga Wagga annual maximum observed daily FFDI values (July–June), with fitted GEV distribution. The solid and dashed lines are as in (a). (d) Plot of FFDI return levels for Wagga Wagga based on fitted GEV distribution, with 95 per cent confidence intervals shown as dashed lines. The solid lines are as in (b).
Comparison between observed and gridded estimates

Table 3 shows the estimated 50-year return values based on the Lucas dataset compared to the 50-year return values calculated for the recreated datasets using the same period of record (June 1972 to December 2009) in the first two columns. This shows that the recreated dataset produces estimates quite close to the high quality Lucas dataset, with a mean difference of only 1.1 between the estimates, with the recreated dataset having higher values on average. This difference is certainly much lower than the 95 per cent confidence intervals associated with the estimation of the model parameters.

The third column in Table 3 shows the estimates based on the longest available dataset for each station, as outlined in Table 1. For the case of Bourke and Tibooburra the estimates

Table 3. Estimated 50-year return values from observations. The first two columns use the GPD fit to POT data, and show the difference in estimates between the Lucas dataset and the recreated dataset for June 1972 to December 2009. The third column also shows the GPD fit to POT data, but uses the longest available recreated dataset for each station as outlined in Table 1. The final column shows the estimate from the GEV distribution fit to the time series of seasonal (July–June) maxima.

<table>
<thead>
<tr>
<th>Station</th>
<th>Lucas (GPD)</th>
<th>Bureau dataset (GPD, 1972–2009)</th>
<th>Bureau dataset (GPD)</th>
<th>Bureau dataset (GEV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bourke</td>
<td>100</td>
<td>105</td>
<td>157</td>
<td>148</td>
</tr>
<tr>
<td>Broken Hill</td>
<td>101</td>
<td>103</td>
<td>114</td>
<td>106</td>
</tr>
<tr>
<td>Canberra</td>
<td>92</td>
<td>94</td>
<td>102</td>
<td>104</td>
</tr>
<tr>
<td>Casino</td>
<td>111</td>
<td>108</td>
<td>114</td>
<td>101</td>
</tr>
<tr>
<td>Cobar</td>
<td>109</td>
<td>106</td>
<td>104</td>
<td>104</td>
</tr>
<tr>
<td>Coffs Harbour</td>
<td>90</td>
<td>91</td>
<td>91</td>
<td>86</td>
</tr>
<tr>
<td>Dubbo</td>
<td>102</td>
<td>98</td>
<td>106</td>
<td>105</td>
</tr>
<tr>
<td>Hay</td>
<td>106</td>
<td>110</td>
<td>114</td>
<td>113</td>
</tr>
<tr>
<td>Mildura</td>
<td>136</td>
<td>140</td>
<td>139</td>
<td>128</td>
</tr>
<tr>
<td>Moree</td>
<td>101</td>
<td>107</td>
<td>104</td>
<td>102</td>
</tr>
<tr>
<td>Nowra</td>
<td>101</td>
<td>99</td>
<td>100</td>
<td>105</td>
</tr>
<tr>
<td>Richmond</td>
<td>108</td>
<td>105</td>
<td>110</td>
<td>95</td>
</tr>
<tr>
<td>Sydney</td>
<td>94</td>
<td>93</td>
<td>97</td>
<td>93</td>
</tr>
<tr>
<td>Tibooburra</td>
<td>153</td>
<td>158</td>
<td>193</td>
<td>157</td>
</tr>
<tr>
<td>Wagga Wagga</td>
<td>119</td>
<td>122</td>
<td>122</td>
<td>127</td>
</tr>
<tr>
<td>Williatown</td>
<td>103</td>
<td>110</td>
<td>110</td>
<td>102</td>
</tr>
<tr>
<td>Cessnock</td>
<td>–</td>
<td>–</td>
<td>136</td>
<td>125</td>
</tr>
<tr>
<td>Cooma</td>
<td>–</td>
<td>–</td>
<td>84</td>
<td>80</td>
</tr>
<tr>
<td>Coonabarabran</td>
<td>–</td>
<td>–</td>
<td>84</td>
<td>79</td>
</tr>
<tr>
<td>Grafton</td>
<td>–</td>
<td>–</td>
<td>88</td>
<td>84</td>
</tr>
<tr>
<td>Katoomba</td>
<td>–</td>
<td>–</td>
<td>42</td>
<td>41</td>
</tr>
<tr>
<td>Moruya</td>
<td>–</td>
<td>–</td>
<td>112</td>
<td>110</td>
</tr>
<tr>
<td>Orange</td>
<td>–</td>
<td>–</td>
<td>72</td>
<td>69</td>
</tr>
<tr>
<td>Parkes</td>
<td>–</td>
<td>–</td>
<td>102</td>
<td>95</td>
</tr>
<tr>
<td>Thredbo</td>
<td>–</td>
<td>–</td>
<td>47</td>
<td>53</td>
</tr>
</tbody>
</table>
were far apart, most likely indicating that there may be more systematic quality issues with the data at these sites, and these two sites are generally excluded from further analysis. In all other cases these estimates are subjectively reasonably close, and from this we conclude that it is reasonable to use the estimates based on the full datasets. The mean difference between the estimates for the full length datasets and the 1972–2009 dataset is 3.07, indicating that the shorter dataset may be underestimating the return value, with this difference likely to be partially due to the different periods sampled, however this difference is generally less than the sampling standard error associated with the model parameter estimates.

Table 3 also shows a comparison between the 50-year return value estimates calculated from a GPD fit using the POT series (third column), with the estimates from the GEV distribution using the time series of annual (July–June) maxima (fourth column). From this we can see that the estimate based on the latter time series is generally near to, but a little lower than the estimate from the former. This result is consistent with expectations, given that the seasonal maxima time series does not capture all significant events, and must also try to account for low data points corresponding to below-average fire seasons.

The comparison between 50-year return values from the observations and downscaled reanalysis datasets are shown in Table 4. For this comparison we have used the a subset of the observation data covering the same date range as the reanalysis, and excluded Bourke and Tibooburra due to data quality concerns as well as sites that do not have a long enough observation record (i.e. Casino, Broken Hill, Grafton, Katoomba and Thredbo). The comparison shows that there are some large differences. To some extent this will be the result of representativeness error, as the observation sites are sampling at a point location while the reanalysis data are for 10 km grid cells that are not necessarily centered on the point of the observation. These errors will be introduced where observation sites are at unrepresentative elevations or subject to local-scale meteorological effects.

The majority of locations in Table 4 show that the return value estimate from the observed data is greater than the estimate from the reanalysis data, implying that the reanalysis data provides an underestimate for point locations. The mean difference between the estimates across stations in Table 4, excluding those stations for which stationarity is violated, is 20.14, with the estimated return value from observation data being generally higher than that from the reanalysis data.

To better understand the sources of error, the estimates of 50-year return value along with the 95 per cent confidence intervals are plotted in Fig. 6, with the red points and error bars corresponding to the estimate based on the observation record, and the blue point values and error bars corresponding to the estimate based on the reanalysis dataset. By far the highest difference between estimate

Table 4. Estimated 50-year return values from reconstructed observations and from the nearest (non water classified) grid point from the WRF reanalysis data, covering 1985 to 2008. Estimates are from a GPD fit to POT data. * Indicates stations for which the assumption of stationarity is violated.

<table>
<thead>
<tr>
<th>Station</th>
<th>Observations (1985 to 2008)</th>
<th>Reanalysis</th>
<th>Difference (obs – re)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canberra</td>
<td>98</td>
<td>87</td>
<td>11</td>
</tr>
<tr>
<td>Cessnock</td>
<td>172</td>
<td>85</td>
<td>87</td>
</tr>
<tr>
<td>Cobar</td>
<td>103</td>
<td>102</td>
<td>1</td>
</tr>
<tr>
<td>Cooma</td>
<td>65</td>
<td>70</td>
<td>-5</td>
</tr>
<tr>
<td>Coonabarabran</td>
<td>111</td>
<td>67</td>
<td>44</td>
</tr>
<tr>
<td>Dubbo</td>
<td>94</td>
<td>84</td>
<td>10</td>
</tr>
<tr>
<td>Hay</td>
<td>106</td>
<td>124</td>
<td>-18</td>
</tr>
<tr>
<td>Mildura</td>
<td>136</td>
<td>119</td>
<td>17</td>
</tr>
<tr>
<td>Moree</td>
<td>112</td>
<td>94</td>
<td>8</td>
</tr>
<tr>
<td>Nowra</td>
<td>91</td>
<td>76</td>
<td>15</td>
</tr>
<tr>
<td>Orange</td>
<td>68</td>
<td>52</td>
<td>16</td>
</tr>
<tr>
<td>Parkes</td>
<td>82</td>
<td>86</td>
<td>-4</td>
</tr>
<tr>
<td>Richmond</td>
<td>110</td>
<td>81</td>
<td>29</td>
</tr>
<tr>
<td>Wagga Wagga</td>
<td>129</td>
<td>101</td>
<td>28</td>
</tr>
<tr>
<td>Coffs Harbour</td>
<td>110</td>
<td>68</td>
<td>42</td>
</tr>
<tr>
<td>Moruya*</td>
<td>104</td>
<td>73</td>
<td>31</td>
</tr>
<tr>
<td>Sydney</td>
<td>94</td>
<td>61</td>
<td>33</td>
</tr>
<tr>
<td>Williamtown*</td>
<td>116</td>
<td>78</td>
<td>38</td>
</tr>
</tbody>
</table>
derived from observation and estimate derived from reanalysis from these stations is for Cessnock (87), with the lowest difference being for Hay (~19). Coastal stations, including Sydney (33) and Coffs Harbour (42), are also associated with large reanalysis underestimates, and may be possibly be associated with an identified bias towards higher humidity in the reanalysis dataset. It is likely that the generally lower values from the reanalysis dataset are due to this dataset failing to adequately capture the few high-end events that dominate the statistical fit for such a long-period return value. To investigate whether these underestimates was also present for shorter-period return values, 1-year, 5-year, 10-year, 20-year and 100-year return values with 95 per cent confidence intervals were also calculated.

The plot of estimated 1-year return values and 95 per cent confidence intervals is shown in Fig. 7, and we can see that for this shorter-period the tendency for lower reanalysis estimates has generally disappeared, however there is a wide spread of errors between the observed and reanalysis estimated return values. The mean difference between the estimates across all stations apart from those for which stationarity is violated is 2.0, with the estimated return values from the reanalysis being slightly higher. The plots of the directly calculated 1-year return values from the observed datasets fall within the 95 per cent confidence intervals for the statistical model estimates, with the exception of Cessnock, Coonabarabran and Richmond, indicating that the statistical model is generally adequately estimating these values. The short confidence intervals for the 1-year estimates and the fact that the average error is close to zero indicates that in this case, the error associated with model parameter estimation is small compared to other, site-specific errors such as data quality problems, or station representativeness error.

Producing a gridded map of return value estimates using the reanalysis data presents a number of issues. The non-stationarity of the data and differences between the observation and reanalysis data are likely to vary spatially, and would be difficult to correct systematically. The optimal model threshold value is also likely to vary spatially, and it would be difficult to completely assess the optimal value at each grid cell. We can attempt to make a first-cut map by using the 98th percentile value (which has been assessed as suitable for the station locations) throughout, and bearing in mind that issues with non-stationarity mean that the return value estimates can not necessarily be assumed to be constant.

The reason that there is a systematic difference in error between the observed and reanalysis estimates evident in Fig. 6 but not in Fig. 7, may be that for longer-period return...
values the most extreme few data points in the dataset are
given more weight, and that the reanalysis may be failing
to capture the magnitude of these extreme data points as
well as the rest of the time series. A simple way to make
the estimates from the reanalysis more consistent with the
estimates from observations would be to add a correction
factor that would vary with the length of the return period to
remove this portion of the variance. The remaining variance
would then be chiefly due to representativeness error as in
Fig. 7, and could only be reduced with a better reanalysis
dataset.

To calculate the mean error between the reanalysis and
observed dataset estimates for 5, 10, 20, 50 and 100 years to
apply as a correction, the contributions from Bourke and
Tibooburra (questionable observation dataset quality) and
Sydney, Moruya, Williamtown and Cessnock (humidity
bias near the coast) were discarded. The calculated mean
differences for the remaining stations were then 4.2, 8.0,
12.6, 20.14 and 27.1 respectively for the 5, 10, 20, 50 and 100
year cases. Note that the length of record is much less than
100 years, and so the extrapolation errors associated with
the estimates are very large in this case, and hence should
be treated with caution.

Applying the POT GPD model fit to each of the grid
points in the reanalysis dataset, then adding a correction
factor of 20.14 to the estimated return values results in a
one in 50-year FFDI return value map as shown in Fig. 8.
The grid points that were classified as water bodies in the
underlying reanalysis dataset were excluded, and a simple
spatial interpolation was applied to produce an estimate for
these points. For display purposes, a smoothing algorithm
has also been applied to the map in Fig. 8.

The map appears to reflect the underlying topography
to a large degree, consistent with temperature being one of
the main inputs to FFDI, with generally lower 50-year return
values over the alpine areas, and locally higher values in
valleys such as the Hunter and the Monaro. Temperatures
also indirectly affect the FFDI by affecting the estimated

--

Fig. 7. Estimates of 1-year return FFDI values based on statistical model using (red) the observed dataset and (blue) the reanalysis
dataset. Also plotted in green crosses are the 1-year return values calculated directly using quantile values from the ob-
served dataset, showing a generally good fit to the model estimates.
fuel dryness, with lower temperatures corresponding to slower recovery of the drought factor. The fire danger index return values shown are only valid for forested areas, so the predominately grassland western areas of the State should be interpreted with caution. Maps of 1, 5, 10, 20 and 100 year FFDI return values have also been produced and are shown in Fig. D1.

Conclusions

In this paper we have shown that the distribution of Forest Fire Danger Index (FFDI) values above a certain threshold can be modelled using a Generalized Pareto Distribution (GPD). We have used this approach with data from the climate record to estimate long-period return values of FFDI for a number of stations across New South Wales. We have also applied this approach to a reanalysis dataset of daily FFDI values calculated from maximum temperature, 3.00 pm wind speed and 3.00 pm humidity to produce a map of statistical model estimates of long-period return values across New South Wales. There are, however, some significant caveats that should be attached to these return values.

- FFDI is an index that is designed for use in forested areas and should be used with caution in other fuel types such as grassland or heath.
- There are some identified biases with the humidity and wind fields in the reanalysis dataset used that could impact on the reanalysis FFDI values. This seems most noticeable in the bias towards higher humidity, and hence lower FFDI, along the coastal fringe.
- FFDI is not a measure of overall fire risk, and does not take into account factors like topography, atmospheric stability, building codes, preparedness, fuel load and suppression capacity.
- There are some large differences between the point estimates based on observation records and the grid cell estimates based on the reanalysis, even for short return period estimates, that are most likely due to observation site representativeness and/or data quality issues. These differences are unlikely to reduce without using a higher-resolution reanalysis to better capture local topographic and coastline effects.

Fig. 8. One in 50-year return values for FFDI in New South Wales, with observation locations used in the study also shown.
In addition to these caveats it is important to recognise that the return values of FFDI are subject to the same limitations as McArthur FFDI more generally, most seriously that the FFDI is an empirically derived index that was designed to be used up to a value of 100 only (McArthur, 1967).

There are many future improvements that could be made to this analysis, including investigating the use of other grided analysis such as those produced by the Australian Water Availability Project or the New South Wales and ACT Regional Climate Modelling (NARClIM) project, and further refinement of the statistical analysis, including a more detailed investigation of the sources of error in the observation and reanalysis estimates, as well as using varying thresholds to account for non stationarity in the data. The use of regional climate forecasts to investigate the effects of climate change on the return value estimates could also be investigated.

Acknowledgments

The author would like to thank Hamish Clarke, Sarah Perkins and Fiona Johnson for their comments and advice on early drafts of the paper, as well as Jeff Kepert, Paul Fox-Hughes and Chris Lucas for assisting with internal reviews.

References


Appendices

Appendix A. Mean threshold plots—observation data references

Fig. A1. Mean threshold exceedance plots for observed data for (a) Bourke, (b) Broken Hill, (c) Canberra, (d) Casino, (e) Cobar, (f) Coffs Harbour, (g) Dubbo, (h) Hay, (i) Mildura, (j) Moree, (k) Nowra, (l) Richmond, (m) Sydney Airport, (n) Tibooburra, (o) Williamtown, (p) Cessnock, (q) Cooma, (r) Coonabarabran, (s) Grafton, (t) Katoomba, (u) Moruya, (v) Orange, (w) Parkes, and (x) Thredbo. Dashed blue line shows 98th per cent threshold value and the solid red line shows a linear fitted curve for a region of the mean threshold exceedance plot containing the 98th per cent threshold selection.
Appendices

Appendix B: Mean threshold exceedance plots—reanalysis data

Fig. B1. Mean threshold exceedance plots for the closest reanalysis grid point to (a) Bourke, (b) Broken Hill, (c) Canberra, (d) Casino, (e) Cobar, (f) Coffs Harbour, (g) Dubbo, (h) Hay, (i) Mildura, (j) Moree, (k) Nowra, (l) Richmond, (m) Sydney Airport, (n) Tibooburra, (o) Williamtown, (p) Cessnock, (q) Cooma, (r) Coonabarabran, (s) Grafton, (t) Katoomba, (u) Moruya, (i) Orange, (w) Parkes, and (x) Thredbo. Dashed blue line shows 98th per cent threshold value and the solid red line shows a linear fitted curve for a region of the mean threshold exceedance plot containing the 98th per cent threshold selection.
Appendices

Appendix C: Peaks-over-threshold linear trend plots

Fig. C1. Quantile regression on POT events for (a) Bourke, (b) Broken Hill, (c) Canberra, (d) Casino, (e) Cobar, (f) Coffs Harbour, (g) Dubbo, (h) Hay, (i) Mildura, (j) Moree, (k) Nowra, (l) Richmond, (m) Sydney Airport, (n) Tibooburra, (o) Williambtown, (p) Cessnock, (q) Cooma, (r) Coonabarabran, (s) Grafton, (t) Katoomba, (u) Moruya, (v) Orange, (w) Parkes, and (x) Thredbo. Shaded areas show the 95 per cent confidence intervals on the decile trend estimates.
Appendices

Appendix D. FFDI return value maps

Fig. D1. Maps of 3.00 pm FFDI return values computed from 1985–2008 reanalysis dataset for (a) 1 year, (b) 5 years, (c) 10 year, (d) 20 years, (e) 50 years and (f) 100 years.