129. Jahrgang (2012), Heft 3-4, S. 228 – 247

Austrian Journal of Forest Science
Centralblatt für das gesamte Forstwesen

Fire size/frequency modelling as a means of assessing wildfire database reliability

Analyse der Zuverlässigkeit von dokumentierten Waldbranddaten anhand von Häufigkeit und Flächengröße

Eastaugh CS [1,*] and Vacik H [1]

Key words: forest fire, power law, size, trend, Austria

Stichworte: Waldbrand, power law, Fläche, Trend, Österreich

Abstract

Many jurisdictions around the world have recently begun compiling databases of wildfire records, in an effort to determine patterns, quantify risks and detect possible changes in fire regimes. Such datasets, if valid and comprehensive, could be used for fire hazard model validation, detection of trends and risk modelling under current and future climatic conditions. It may be however that data quality issues can hinder these efforts. In particular, older records may be less comprehensive, and smaller fires may have a greater chance of being unrecorded. A database of Austrian wildfires has been compiled, based on historic documentary records from a variety of sources that cover different time periods or geographical regions. The non-comprehensive and non-random nature of such datasets (both spatially and temporally) makes the direct analysis of wildfire patterns impossible, ne-

[1] Institute of Silviculture, Department of Forest and Soil Sciences, BOKU University of Natural Resources and Life Sciences Vienna, Peter Jordan Str. 82, A-1190 Wien, Österreich
* Corresponding author

cessitating the use of models to identify trends and patterns. It is likely however that small fires are substantially underreported, particularly in early decades. We test this proposition by examining the fire size/ frequency distribution of all fires with recorded areas. The thesis behind the work is that we may compare the fire size/frequency relationships in the data across different time periods and that anomalies in the fire size/frequency distribution may indicate weak parts of the dataset. Our results lead us to suspect that data for smaller fires the current database is incomplete and imparts a bias to the size/frequency relationship in periods prior to the mid 1990s.

Zusammenfassung

Weltweit haben viele Institutionen in der letzten Zeit begonnen, Datenbanken zu Waldbränden aufzubauen, um mögliche Muster und Änderungen beim Auftreten von Waldbränden und die damit verbundenen Gefahren zu quantifizieren. Vorausgesetzt solche Datensätze sind vollständig und valide, dann können sie für die Validierung von Modellen zur Abschätzung der Waldbrandgefahr sowie für Vorhersagen unter aktuellem und zukünftigem Klima verwendet werden. Eine schlechte Datenqualität kann diese Anstrengungen mindern. Besonders ältere Datensätze könnten unvollständig sein oder kleinere Waldbrände könnten möglicherweise nicht dokumentiert worden sein. Eine Waldbrand-Datenbank für Österreich basierend auf historischen Aufzeichnungen von unterschiedlichen Quellen, für verschiedene Zeitpunkte und räumliche Bezugsebenen ist kürzlich erstellt worden. Die Eigenschaften von Waldbranddaten (nicht zufällige Verteilung, nicht vollständige Dokumentation) machen eine direkte Analyse der Trends und Muster schwierig, wenn keine Modelle zum Testen der Verteilung angewendet werden. Es ist sehr wahrscheinlich, dass kleinere Waldbrandereignisse aus früheren Jahren in den Daten unterrepräsentiert sind. Wir testen diese Hypothese durch eine eingehende Analyse der Häufigkeit und Größe des Auftretens von dokumentierten Waldbränden. Durch die Analyse sollen mögliche Abhängigkeiten und Unregelmäßigkeiten beim Auftreten von Waldbränden in bestimmten Zeitperioden festgestellt werden, um die Validität der Waldbranddatenbank zu prüfen. Unsere Ergebnisse lassen vermuten, dass die Daten zu kleineren Waldbränden in der aktuellen Datenbank unvollständig sind und für die Perioden bis in die mittleren 1990er die Beziehungen zwischen Größe und Häufigkeit verzerren.

Introduction

Many national jurisdictions around the world have recently begun compiling databases of wildfire records, in an effort to determine patterns, quantify risks and detect possible changes in fire regimes. Examples are available from Switzerland (Conedera et al. 1996; Conedera 1999), the USA (Brown et al. 2002), Canada (Stocks et al. 2003) and Europe (EC 2008). Such datasets, if valid and comprehensive, could be used for model validation, detection of trends and quantitative risk analyses. In some cases, historic fire databases have been used to demonstrate an apparent increase in fire danger (UN 2002; EC 2008; Lorz et al. 2010; Seidl et al. 2011), but as is often pointed out (Podur et al. 2002; Schelhass et al. 2003; Larjavaara et al. 2005; San-Miguel and Camia 2009), data quality issues can hinder these efforts. In particular, older records may be less comprehensive, and smaller fires may have a greater chance of being unrecorded (Hall and Harwood 1989). This is not unique to wildfires, similar issues have been raised in the broader context of historical natural disaster records (Kron et al. 2012).

If fire databases are to be used for practical or research purposes every effort must be made to ensure the integrity of the dataset. Missing data from earlier periods may result in an apparent increase in the chance of fire ignition (a rising trend in fire occurrence), where no such change has occurred. Similarly, if unreported fires are smaller in size than those recorded this may bias fire hazard models and risk assessments, as the overall possibility of ignition would be underestimated but the mean size of fires overestimated. It is difficult however to assess how many (if any) fire events are truly missing from a dataset in any given period, and what size they may have been.

One common approach to fire modelling involves determining fire size/frequency relationships. It is well accepted that smaller fires occur much more often than larger fires, and great effort has been put into defining the mathematical function of this relationship. Some previous work (i.e. Ricotta et al. 1999; Malamud et al. 2005) suggests that the fire size/frequency relationship of larger fires might comply with a power law distribution, although this is not universally accepted (i.e. Reed and McKelvey 2002). Regardless of the form of this relationship, in most previous work it is assumed to be temporally stable (although this is rarely tested). Malamud et al. (2005) did test this assumption and were unable to find a statistically significant difference in their power-law parameters across different time periods. We are not aware of any studies that show these parameters to vary in time.

Our aim in this study is to develop methods to assess historical fire datasets and determine whether or not they contain anomalies that may adversely impact future work by giving a false picture of the historical fire situation (e.g., they would comprise a false validation set for risk models). To illustrate these methods we apply them to the database of Austrian wildfires compiled by the Institute of Silviculture at the University of Natural Resources and Life Sciences in Vienna (Vacik et al. 2011; Müller et al. 2012). The rationale of our approach is that we compare the fire size/frequency relationships in the data across different time periods and that anomalies in the fire size/frequency distribution may indicate weak parts of the dataset that should be treated with especial caution. Our hypothesis is that changes in the observed size/frequency relationships will be consistent with increasingly better reporting of smaller fires in more recent years. We also test the dataset for its compliance with a power law distribution in the upper tail using the techniques of Clauset et al. (2007), but the bulk of our analysis relies on non parametric methods.

Methods

Data

As part of the AFFRI and ALP FFIRS projects (Valese et al 2010; Vacik and Gossow 2011), a database of Austrian wildfires has been compiled, based on historic documentary records. The database has been assembled with information from a variety of sources that cover different time periods or geographical regions. Records were collated from the public online fire news platforms 'www.wax.at' and 'www.feuerwehr-news.at', from regional fire brigade records and through personal contact with the various Austrian municipalities and Federal government departments. 2455 records are available after processing to remove repeated observations, 1870 of which pertain to forest fires. 1012 of these have a value recorded for burnt area (the earliest in the year 1903), with values ranging from one square metre to 200 hectares, distributed as shown in figure 1.

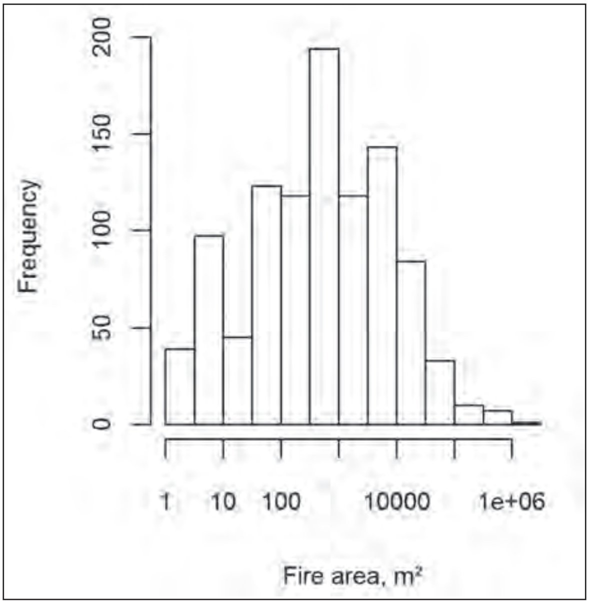


Figure 1 Frequency distribution of forest fires in Austria 1903-2010.

Methods

A power law distribution is one where $\Pr[X \geq x] = Cx^{-\alpha}$, C and α both >0. Newman (2005) demonstrated that where $\alpha > 1$,

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha} \qquad (1)$$

α and x_{min} being constants. The power law parameters α and x_{min} have commonly in the past been estimated graphically or with a least-squares regression, although these methods can introduce significant bias to power law parameter estimates (Goldstein et al. 2004; White et al. 2008). Newman (2005) gives a derivation for a Maximum Likelihood Estimation (MLE) estimation of α, arriving at Eq. 2.

$$\alpha = 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}} \right]^{-1} \qquad (2)$$

To determine x_{min} we follow Clauset et al. (2007) and test the fit of the modelled distribution to the empirical data. Power law models are constructed with a range of possible values of x_{min}, and a distance statistic determined for each model's fit to the empirical data. We then select the x_{min} from the model with the smallest distance between the observed and modelled cumulative probability distributions that leaves a useful number of records above the cut-off. The modelled cumulative probability curve (Pm_i) is calculated as

$$Pm_i = \left(\frac{x_i}{x_{\min}}\right)^{-\alpha+1}, x_{min} \leq i \leq \max(x) \qquad (3)$$

and the observed curve (Po_i) as the normalised ranks (R) of each observation

$$Po_i = \frac{Rx_i}{\max(Rx_i)}, x_{min} \leq i \leq \max(x) \qquad (4)$$

The maximum difference between Pm_i and Po_i is the distance for the particular xmin tested.

Standard errors and 90 % confidence limits around the estimates of α are estimated with a non-parametric bootstrapping procedure (Efron 1987). Estimates are made in the 'R' statistical package (R Development Core Team 2011), using code kindly made public by Clauset et al. (2009), available via http://tuvalu.santafe.edu/~aaronc/powerlaws/.

The database is divided into 6 subsets based on year of record, with each subset containing a comparable number of fires above x_{min} and (with one exception) at least 30 fires per period. The slope of the fire size/frequency relationship is calculated for each subset (using the global x_{min}), and compared graphically with the slope for all records subsequent to that period. This is done in an effort to identify a possible 'change point' in the time series, where the size/frequency relationship becomes concordant with later periods, without having to a priori assume that any particular period is correct. Formal 'goodness of fit' test results are calculated and consist of the standard error and bias of the slope estimate and the Kolmogorov-Smirnov (KS) D statistic and p-values for the fit of the data to the estimated power law model.

If record-keeping is consistent and the fire size/frequency distribution is temporally stationary, then there should be no significant difference between the size/frequency slope in any period and that from the combined records from subsequent periods. Note that there is no need for the periods under comparison to be of the same length, as the length of period per se will not affect the size/frequency distribution (although greater confidence can be obtained with larger data subsets). The significance of any apparent trend differences between each period and all subsequent periods is examined with a t-test.

Given that a power law distribution may fail formal goodness of fit tests when applied to some data subsets, we also compare data from each subset with subsequent periods with a non-parametric KS test. As there are a large number of 'ties' in the data, we use the bootstrapping ks.boot algorithm in the Matching package (Sekhon 2011). This procedure compares the KS statistic from the original pair of datasets with that from a large number of samples drawn from a resampled pooled dataset. The p value reported is the proportion of times that the KS statistic from the generated samples is greater than that from the original (Abadie 2002). A low p value suggests that the Null hypothesis that the datasets are drawn from the same distribution should be rejected.

Fires smaller than x_{min} are divided into three magnitude classes; from 1 to 100m², from 101 to 1000m² and from 1001 to the x_{min} previously determined. Fires of less than 100m² are possibly not 'wild' fires (in the sense that these may be deliberately lit and controlled fires such as campfires or trash burning) and so may exhibit different statistical behaviour. No parametric testing is performed as these data subsets are arbitrarily truncated and parametric statistics are unlikely to be meaningful. Differences between the data distributions from each period and its subsequent periods are analysed with the KS testing procedure outlined above, and detailed graphical representations supplied.

Results and discussion

The optimal value for x_{min} was found to be 4500 m² (fig 2), with this value giving a power law model for the fire database that most closely matches observations. 260 fires are recorded with a size of at least 4500 m² since 1903. Dividing these into 6 time periods gives between 29 and 55 fires per period, with between 0 and 32 fires recorded per year (fig 3a). Other possible subsetting resulted in periods with too few fires for meaningful analysis.

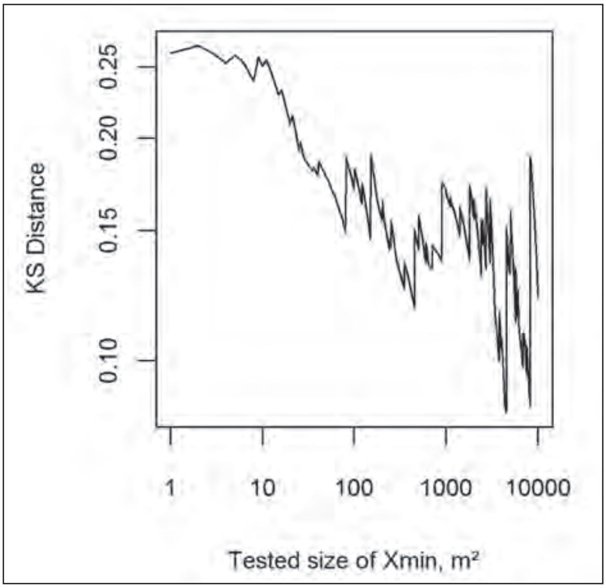


Figure 2 Test of possible values of xmin.
Kolmogorov Smirnoff distances for possible values of xmin to fit a power law relationship to fires in the Austrian fire database. The smallest distance between the theoretical and observed distributions suggests the best fit, in this case 4500 m².

Table 1 Goodness of fit of power law relationships for each period.

Period	1903-2010	< 1993	> 1992	1993-1994	> 1994	1995-2001	> 2001	2002-2004	> 2004	2005-2007	> 2007
n	260	29	231	46	185	45	140	55	85	47	38
alpha	1.784	1.619	1.811	1.945	1.784	1.796	1.780	1.739	1.808	1.877	1.737
SE	0.049	0.124	0.054	0.147	0.057	0.123	0.067	0.104	0.090	0.137	0.126
Bias	0.003	0.023	0.003	0.019	0.004	0.016	0.006	0.013	0.009	0.019	0.020
KS	0.165	0.109	0.178	0.160	0.184	0.204	0.178	0.191	0.170	0.163	0.182
p	0.000	0.697	0.000	0.041	0.000	0.005	0.000	0.002	0.000	0.032	0.033

n: number of fires in each period

alpha: best-fit power law exponent

SE: standard error of exponent estimate

Bias: bias of exponent estimate

KS: bootstrapped Komolgorov -Smirnoff distance

p: p value for KS test. p values less than 0.1 suggest that the power law hypothesis should be rejected (Clauset et al. 2009).

Table 2 Results of comparisons for fires > 4499 m²

1st period		2nd Period		Slope difference	Slope difference significance		
years	n	years	n		t statistic	df	p
< 1993	29	> 1992	231	0.193	1.415	256	0.158
1993-1994	46	> 1994	185	-0.161	1.033	227	0.303
1995-2001	45	> 2001	140	-0.017	0.119	181	0.905
2002-2004	55	> 2004	85	0.069	0.509	136	0.612
2005-2007	47	> 2007	38	-0.140	0.763	81	0.448

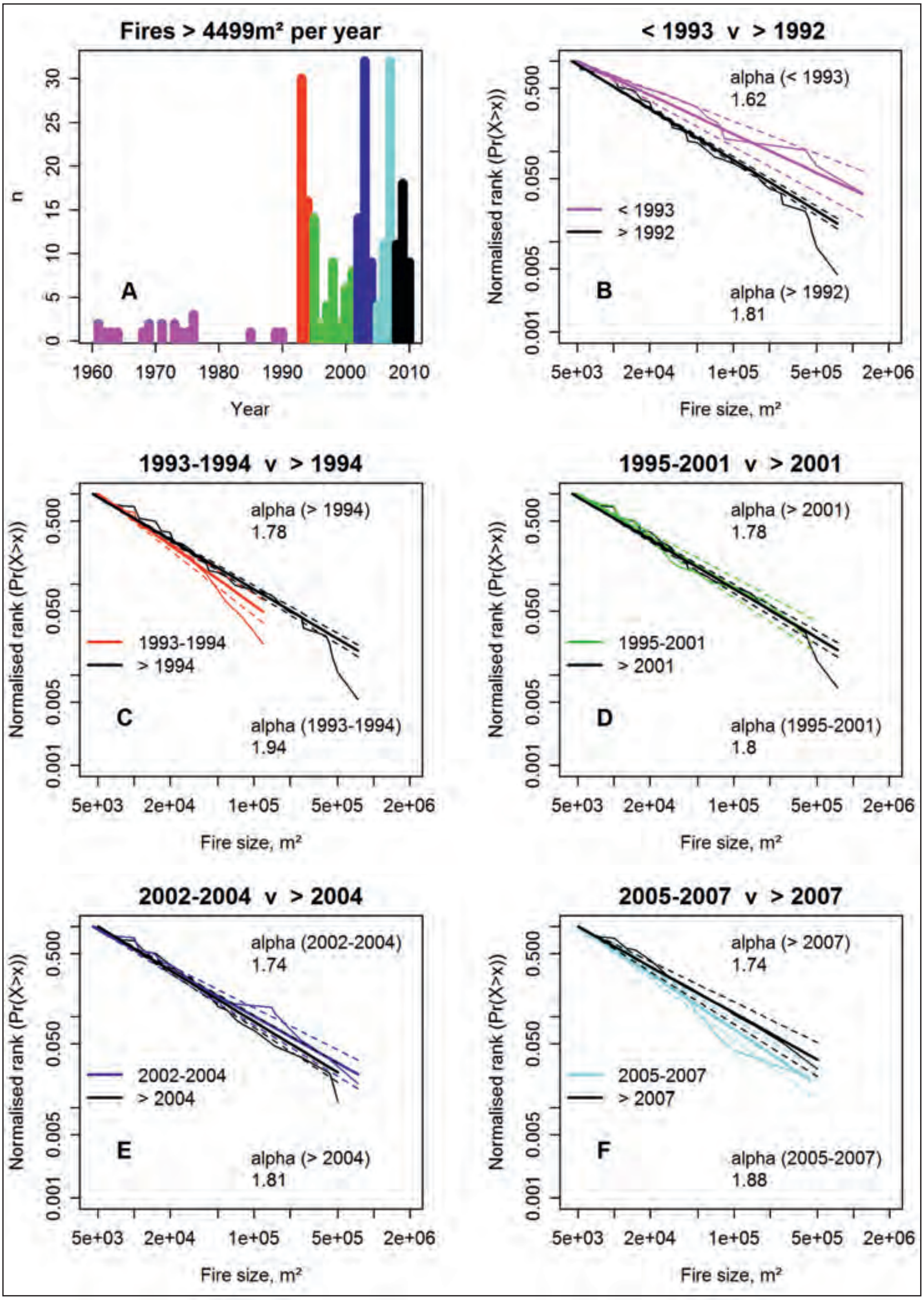
The estimated power law exponent (slope) for each period is tested against that for the years following that period. The differences in the slopes are tested for significance with t-tests.

Table 1 gives a summary of the parameters and the formal goodness of fit statistics for each period. Results in table 1 suggest that a power law is a poor model for fire size/frequency relationships of the documented fire records in Austria. Clausett et al. (2009) suggest that a p value of 0.1 should be considered a minimum standard, which was achieved only in the pre 1993 period. This poor fit makes it difficult for the parametric t-test to find significant differences between periods (table 2). To interpret these results we consider the Null hypothesis that each pair of data subsets are in fact from distributions with the same parameters. H_0 cannot be rejected with greater confidence than the p_t values in table 2.

Graphically summarising the data as the slopes of best-fit power law relationships (figure 3) is however still a valid method of data exploration. If a 'true' size/frequency model does exist it would plot on these axes with a slight curve, but visually it would probably be indistinguishable from the straight lines that we show. A stronger model however would have smaller standard errors and hence the t-test would have greater power to distinguish differences between periods.

For fires in the largest size class (fig 3), two anomalies are apparent. In the pre-1993 period (fig 3b) there appears to be a flatter than normal size/frequency distribution, which is consistent with the thesis that the largest early fires are more likely to appear in the database. The 1993-1994 period

Figure 3 Recorded occurrence and size/frequency relationships of large fires.
Panel 'A' shows the number of fires each year in the database of at least 4500m² (the figure is truncated before 1960 for clarity). Panels 'B' to 'F' show the fire size/frequency relationships for each time period, and for all fires subsequent to that period. 'alpha' denotes the best-fit power law exponent for each curve. Note that in panels 'D', 'E' and 'F' there is less difference in the respective alpha values than is apparent in panels 'B' and 'C'.


Fires > 4499m² per year
n
Year
A
< 1993 v > 1992
alpha (< 1993)
1.62
< 1993
> 1992
B
alpha (> 1992)
1.81
1993-1994 v > 1994
alpha (> 1994)
1.78
1993-1994
> 1994
C
alpha (1993-1994)
1.94
1995-2001 v > 2001
alpha (> 2001)
1.78
1995-2001
> 2001
D
alpha (1995-2001)
1.8
2002-2004 v > 2004
alpha (2002-2004)
1.74
2002-2004
> 2004
E
alpha (> 2004)
1.81
2005-2007 v > 2007
alpha (> 2007)
1.74
2005-2007
> 2007
F
alpha (2005-2007)
1.88
Normalised rank (Pr(X>x))
Fire size, m²

(figure 3c) shows an α of 1.94, markedly higher than that across ensuing periods (1.78). This is especially curious considering that this period recorded a particularly high number of large fires in both 1993 and 1994 (fig 3a). At face value, it seems that this period experienced many fires in the lower end of the ≥ 4500 m^2 class, but few at the upper end (hence the high α). This difference in α however would imply that the size/frequency relationship is extremely non-stationary, which would be at odds with earlier work on the stability of these relationships (Malamud et al. 2005). The parametric tests in table 1 suggest that this difference is not formally significant, and the KS results in table 2 show that we cannot reject the hypothesis that the current data available from 1993-1994 are drawn from the same distribution as subsequent periods with greater than 53.1 % confidence. Nevertheless, it is striking that in each of the three periods after 1994 the p value from the KS test is over 0.95. From 1995 onwards, the size/frequency relationship of large fires in all periods appears to be consistent, with overlapping 90 % confidence limits for the estimates of α (figures 3d,e,f).

Table 3 Komolgorov Smirnoff statistics for comparing the size/frequency distributions of fires in each time period with that from subsequent periods

	Period 1	< 1993	1993-1994	1995-2001	2002-2004	2005-2007
Size	Period 2	> 1992	> 1994	> 2001	> 2004	> 2007
Large	n1	29	46	45	55	47
	n2	231	185	140	85	38
> 4449 m^2	D	0.185	0.133	0.079	0.080	0.111
	p	~~ 0.342	0.531	0.983 ##	0.983 ##	0.958 ##
Medium	n1	34	21	18	22	25
	n2	102	81	63	41	16
> 1000 m^2, < 4500 m^2	D	0.500	0.111	0.164	0.125	0.105
	p	~~ 0.000	0.986 ##	0.863 ##	0.978 ##	1.000 ##
Small	n1	61	28	32	52	77
	n2	251	223	191	139	62
> 100 m^2, < 1001 m^2	D	0.263	0.203	0.252	0.178	0.111
	p	~~ 0.002	~~ 0.257	~~ 0.062	~~ 0.180	0.791 ##
Tiny	n1	6	28	9	44	91
	n2	298	270	261	217	126
< 101 m^2	D		0.674		0.172	0.100
	p		~~ 0.000		~~ 0.230	0.672

For each size class, the size/frequency distributions are compared using a bootstrapped Komolgorov Smirnoff test.

Statistics are not reported where less than ten fires occurred in either period.

n1, n2: number of fires in periods one and two; D: Komolgorov Smirnoff D statistic

p: p values of the given D statistic. Low p values suggest rejection of the Null hypothesis of equal distributions.

. p values over 0.75; ~~ p values under 0.5

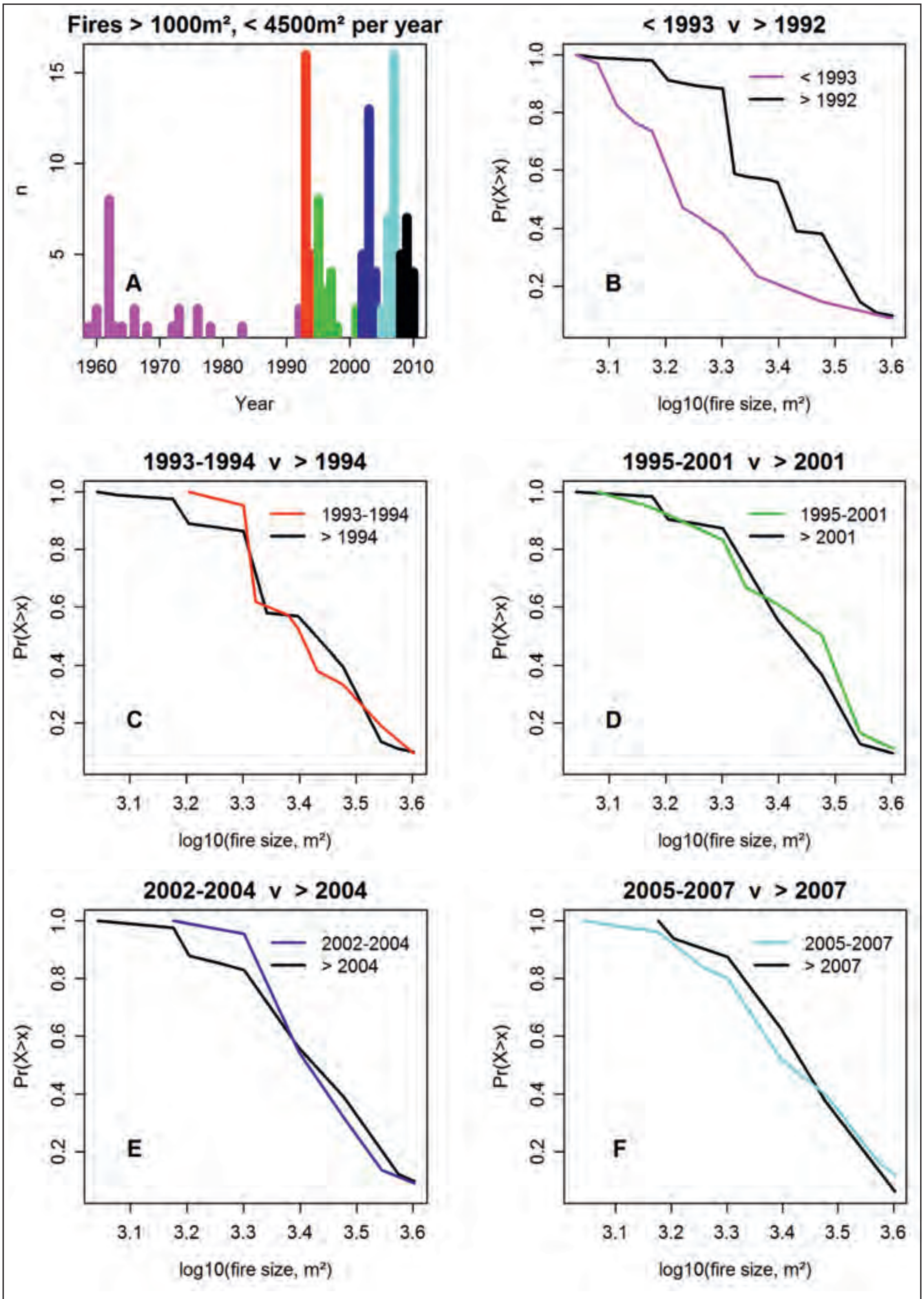
In table 3 are the results from the non-parametric bootstrapped KS tests. The p values are the probability that a D statistic this high would be attained if the data subsets being compared were drawn from the same distribution. Following our working hypothesis that data quality has improved with time, it is clear that in general, the conformity of the data increases as the periods examined are more recent (larger p values towards the right of the table), and as the size class of fires examined increases (larger values towards the top of the table). Fires in the < 100 m² class showed very little concordance until 2005, and even then the p value is much smaller than that attained in larger size classes. None of the size classes prior to 1993 show similar distributions to later periods. Fires of up to 1000 m² show very poor agreement until 2005. This is consistent with the thesis that fire reporting has shown continual improvement, and that underreporting is a greater problem for smaller fires than for larger ones.

The 1993-1994 period is anomalous, in that concordance for fires in the medium class appears good (0.986), but less so for larger fires (0.531). The data for the smallest fires is also clearly not concordant ($p = 1.909 \times 10^{-10}$). Unlike in all other periods, in this smallest size class no fires were recorded of less than 10 m² (figures 4,5,6). There is no apparent physical reason as to why 1993-1994 should have had no fires of less than 10 m², very many fires near the lower end of the ≥5000 m² class and yet few at the higher end. A highly plausible explanation is that fires of less than 10m² have at some stage of the data reporting/recording/transcribing process been mistakenly listed with a size in hectares, rather than m². Unfortunately this is impossible to confirm without access to long-lost original source documents. Although the concordance for medium fires in 1993-1994 is high, the far lower concordance for larger fires makes the value of the medium fire data moot, as it is unlikely that it would be useful if the larger fire data was discarded.

Following pages (240-241):
Figure 4 Recorded occurrence and size/frequency relationships of medium fires.
Panel 'A' shows the number of fires each year in the database of 1001 – 4499m² (the figure is truncated before 1960 for clarity). 136 fires are recorded with a size greater than 1000m² and less than 4500m² since 1960. Dividing these into 6 time periods gives between 16 and 34 fires per period, with between 0 and 16 fires recorded per year. Panels 'B' to 'F' show the fire size/frequency relationships for each time period, and for all fires subsequent to that period.

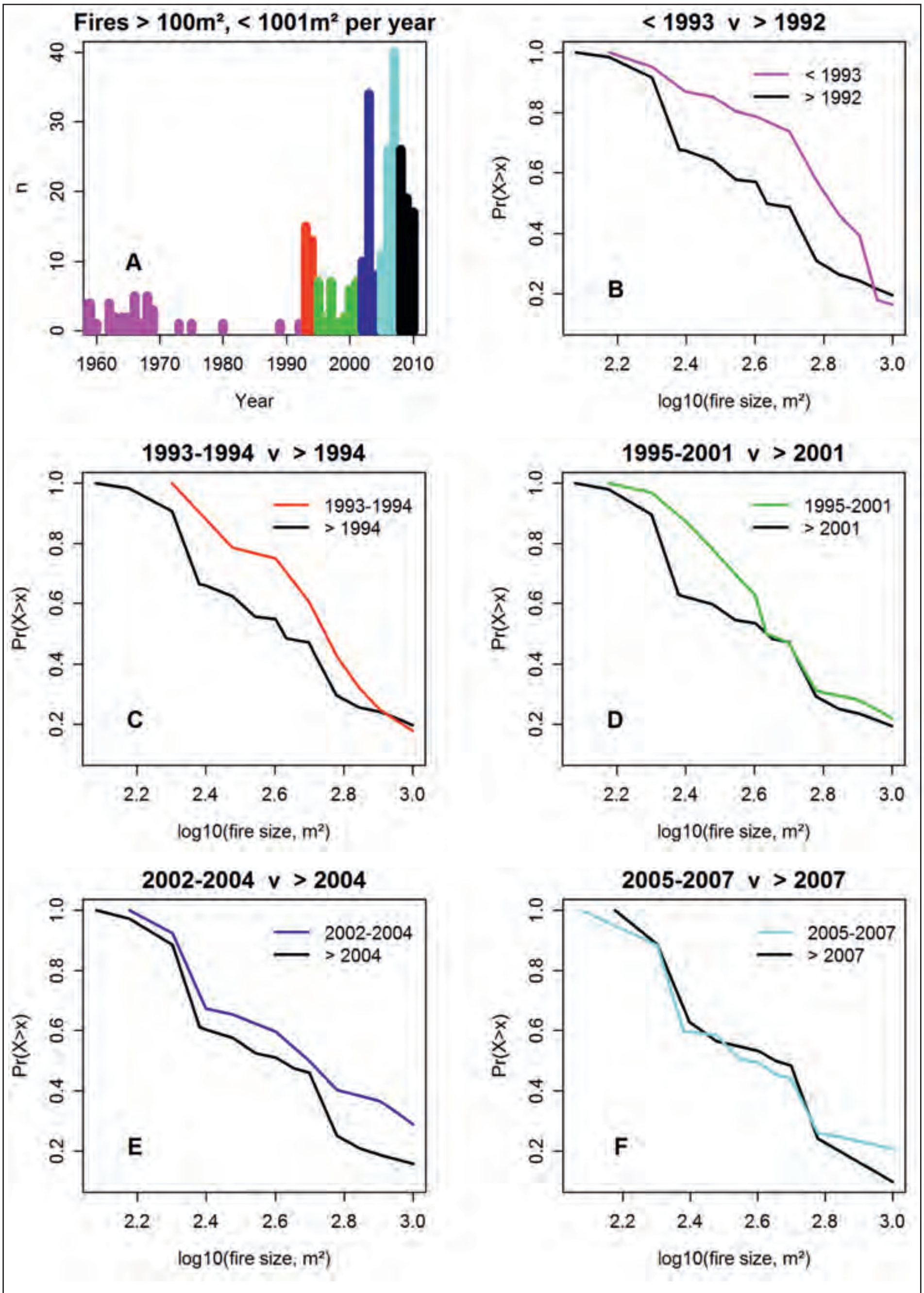
Figure 5 Recorded occurrence and size/frequency relationships of small fires.
Panel 'A' shows the number of fires each year in the database of 101 – 1000m² (the figure is truncated before 1960 for clarity). 312 fires are recorded with a size greater than 100m² and less than 1001m². Dividing these into 6 time periods gives between 32 and 77 fires per period, with between 0 and 40 fires recorded per year. Panels 'B' to 'F' show the fire size/frequency relationships for each time period, and for all fires subsequent to that period.

Figure 4



Fires > 1000m², < 4500m² per year
n
A
Year
< 1993 v > 1992
< 1993
> 1992
Pr(X>x)
B
log10(fire size, m²)
1993-1994 v > 1994
1993-1994
> 1994
C
1995-2001 v > 2001
1995-2001
> 2001
D
2002-2004 v > 2004
2002-2004
> 2004
E
2005-2007 v > 2007
2005-2007
> 2007
F

Figure 5


Fires > 100m², < 1001m² per year
n
Year
A
< 1993 v > 1992
< 1993
> 1992
B
1993-1994 v > 1994
1993-1994
> 1994
C
1995-2001 v > 2001
1995-2001
> 2001
D
2002-2004 v > 2004
2002-2004
> 2004
E
2005-2007 v > 2007
2005-2007
> 2007
F
Pr(X>x)
log10(fire size, m²)

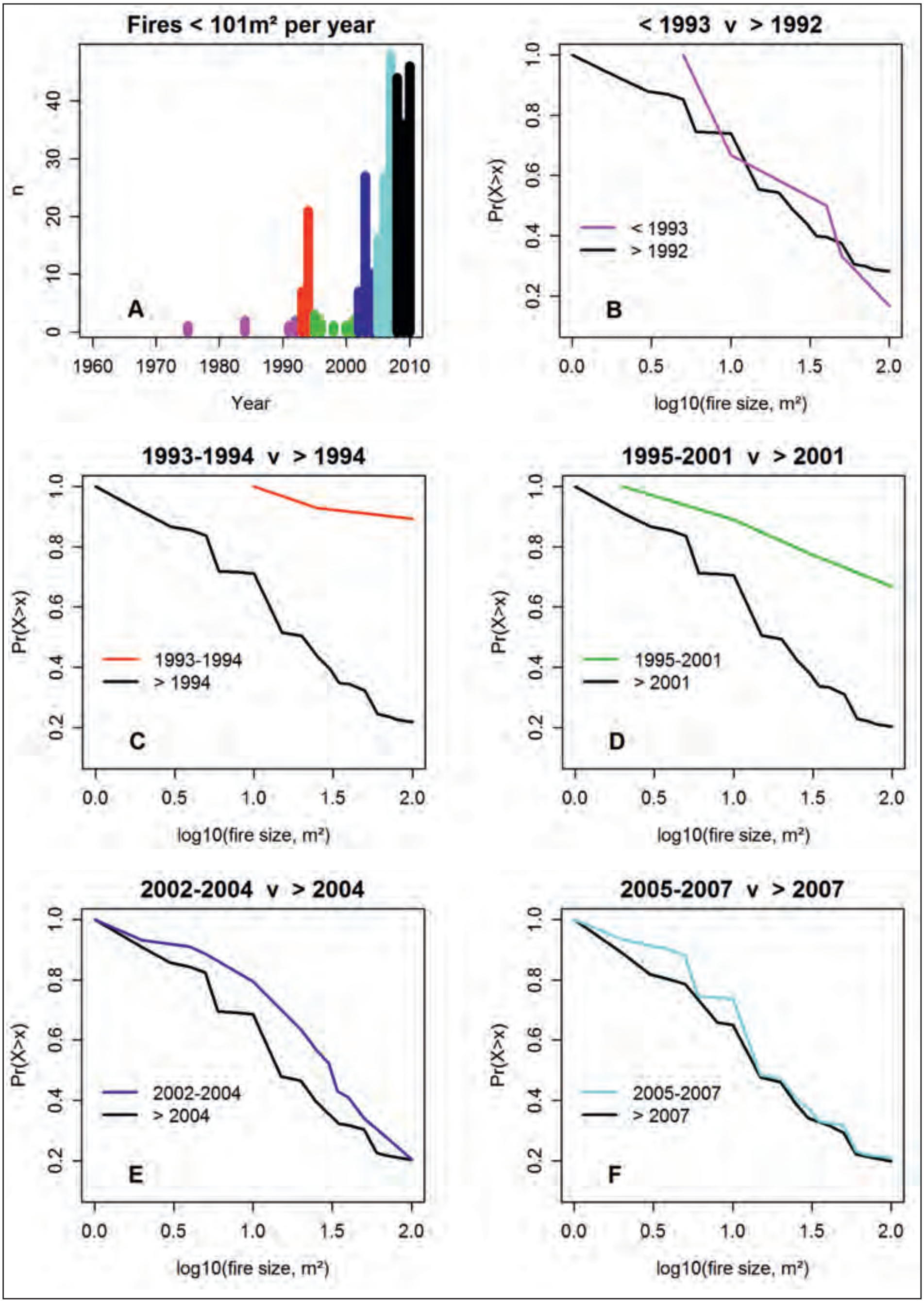
Our attempt to fit Austrian forest fires to a power law relationship was not successful. We present the results here both as a convenient means of summarising the data for larger fires in each period (in terms of the best-fit slope on logarithmic axes) and as support for earlier work that has questioned the assumption of power law behaviour. To the best of our knowledge, in all cases where claims of power law distributions were rigorously tested the relationship failed (i.e. Reed and McKelvey 2002; Newman 2005; Clauset et al. 2009). Although it may be possible to find relationships that provide a better fit, this is beyond the scope of this paper. It should be noted however that figure 3 visually suggests a close fit to a power law distribution over a range of magnitudes similar to that found by Ricotta et al. (1999).

In any case, an accurate model must be based on accurate data, so a pre-screening of the type we demonstrate here should be a prerequisite. Our non-parametric testing has not conclusively proven data flaws in the 1993-1994 period to scientific standards, but enough suspicion is raised to advise not using this data in applications where the sizes of fires is an important factor. The p value cut off levels we show in the table (0.5 and 0.75) are arbitrary, but may be considered useful as an indication of the confidence we have in the integrity of the data subsets.

We have demonstrated here that the Austrian fire database displays temporal non-stationarity, but we cannot conclusively attribute this to either physical causes or data quality issues. Nevertheless, we have presented a plausible explanation for the observed departures from stationarity which are consistent with our hypothesis of improving data quality over time. Our methods in this paper rely on the assumption of temporal stationarity in the fire size/frequency distribution. Given the relative stationarity of larger fires post 1994, this assumption seems reasonable, and agrees with the findings of Malamud et al. (2005). It is not impossible that the assumption is flawed, and that physical or anthropogenic reasons for substantial non-stationarity exist. It is notable however that Europe's most serious summer heatwave in living memory was in 2003, and it seems reasonable that if fire size/frequency distributions were not temporally stable over the past few decades (and were climate dependent) then this is the period where we would expect to see the most evidence of this. The increasing trend in average national tem-

Figure 6: Recorded occurrence and size/frequency relationships of tiny fires.
Panel 'A' shows the number of fires each year in the database of 101 – 1000 m² (the figure is truncated before 1960 for clarity). 304 fires are recorded with a size up to 100 m². Dividing these into 6 time periods gives between 6 and 126 fires per period, with between 0 and 48 fires recorded per year. Panels 'B' to 'F' show the fire size/frequency relationships for each time period, and for all fires subsequent to that period.

Figure 6


Fires < 101m² per year
A
n
0
10
20
30
40
1960
1970
1980
1990
2000
2010
Year
< 1993 v > 1992
B
< 1993
> 1992
1993-1994 v > 1994
C
1993-1994
> 1994
1995-2001 v > 2001
D
1995-2001
> 2001
2002-2004 v > 2004
E
2002-2004
> 2004
2005-2007 v > 2007
F
2005-2007
> 2007
Pr(X>x)
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
log10(fire size, m²)

peratures in Austria (Eastaugh et al. 2010) does not appear to be affecting the size/frequency distribution either. We suggest then that the conservative course is to exclude that data we are not confident in where it may bias the results of future studies.

Where the data is not used to make distinctions regarding trends in fire occurrence or sizes over time, the full database can be used. If it is clear that neither the year of occurrence nor the fire size are important to the conclusions then it is not invalid to consider the database as an effectively random subset of all fires that truly occurred. No inferences regarding temporal trends or patterns can be made, as these may simply be an artefact of a possibly incomplete database. The collection of historical data is however ongoing, and as more sources become available it may be possible to extend the period of data confidence further back in time.

Cloppet and Regimbeau (2011) noted that fire events databases are usually incomplete and inhomogeneous. Their approach to assessing past temporal trends in fire ignition danger in France was to model the danger with the Canadian Fire Weather Index, validating against fires from a recent period. The validated danger model could then be applied to earlier periods, overcoming the likelihood that apparent trends of increasing fire danger were a result of missing data from earlier periods. Although a fuller reporting of historic fire occurrences would perhaps allow for the construction of a more-skilled model, the strength of the model inter-comparison (Arpaci et al. 2013; Eastaugh et al. 2012) is not affected by the location or timing of unreported fires.

Conclusions

We have presented here a novel method of examining historical forest fire records in order to estimate the reliability of records from different time periods. The attempted fit to a power law distribution of the fire size/frequency relationship was very poor, but non-parametric methods were found to be sufficient to raise concerns about using the current database, particularly in earlier time periods and for smaller fires. For applications where historic Austrian fire size data are important, we suggest that based on the current database data for fires of greater than 100 m² only be used from 2005 onwards, or data for fires greater than 1000 m² from 1995 onwards. Although our analysis is limited to the Austrian forest fire database, the methods we have applied are easily transferable to any environment, and can give indications of where possible flaws may exist in any historic fire dataset.

Acknowledgements

The dataset used in this analysis was jointly collated by Dipl.- Ing. Mortimer Müller and Dipl.- Ing. Natalie Arndt of the Institute of Silviculture, University of Natural Resources and Life Sciences, Vienna. The authors acknowledge those efforts to compile a comprehensive dataset. This research has been conducted partly within the frame of the Austrian Forest Fire Research Initiative (AFFRI), which is funded by the Austrian Science Fonds (FWF) with the reference number L539-N14, the European Project ALP FFIRS (Alpine Forest Fire Warning System), which is funded by the European Regional Development fund of the Alpine Space Program with the reference number 15-2-3-IT.

References

Abadie A (2002) Bootstrap tests for distributional treatment effects in instrumental variable models. Journal of the American Statistical Association 97(457): 284-292.

Arpaci A, Eastaugh CS, Vacik H (2012) Selecting the best performing Fire Weather Indices for Austrian Ecozones. Theoretical and Applied Climatology (in review).

Brown TJ, Hall BL, Mohrle CR, Reinbold HJ (2002) Coarse Assessment of Federal Wildland Fire Occurrence Data. Reno, Desert Research Institute. pp. 31

Clauset A, Young M, Gleditsch KS (2007) On the frequency of severe terrorist events. Journal of Conflict Resolution 51(1): 58-88.

Clauset A, Shalizi CR, Newman MEJ (2009) Power Law distributions in empirical data. SIAM Review 51(4): 661-703.

Cloppet E, Regimbeau M (2011) Fire Weather Index: From high-resolution climatology to Climate Change impact study. International Conference on current knowledge of Climate Change Impacts on Agriculture and Forestry in Europe COST-WMO Topolcianky, SK, 3-6 May 2011. Online at. http://www.shmu.sk/File/akcie/Cloppet%20Emmanuel,%20Impact%20of%20climate%20change%20on%20fire%20weather%20index%20from%20high%20resolution%20climatology%20to%20Climate%20Change%20impact%20study.pdf Accessed 2012 14 March.

Conedera M (1999) Forest Fires. In: Minor H, editor, Coping study on disaster resilient infrastructure. Zurich, VAW, Laboratory of Hydraulics, Hydrology and Glaciology of the Swiss Federal Institute of Technology. pp 71-76.

Conedera M, Marcozzi M, Jud B, Mandallaz D, Chatelain F et al. (1996) Incendi boschivi al Sud delle Alpi: passato, presente e possibili sviluppi futuri. NRP 31 report. Zurich, vdf Hochschulerverlag and ETH. 143 pp. [in Italian]

Eastaugh CS, Petritsch R and Hasenauer H (2010) Climate characteristics across the Austrian forest estate from 1960 to 2008. Austrian Journal of Forest Science 127:133-146.

Eastaugh CS, Arpaci A, Vacik H (2012) A cautionary note regarding comparisons of fire danger indices. Natural Hazards and Earth System Sciences 12:927-934.

EC (2008) Forest fires in Europe 2007. Report No8. European Commission, Joint Research Centre, Institute for Environment and Sustainability. Ispra, Italy. 80 p.

Efron B (1987) Better bootstrap confidence intervals. Journal of the American Statistical Association 82(397): 171-185.

Goldstein ML, Morris MA, Yen GG (2004) Problems with fitting to the power-law distribution. European Physics Journal B 41: 255-258.

Hall JR Jr, Harwood B (1989) The national estimates approach to U.S. fire statistics. Fire Technology 25(2): 99-113.

Kron W, Steuer M, Löw P, Wirtz A (2012) How to deal properly with a natural catastrophe database – analysis of flood losses. Natural Hazards and Earth System Sciences 12: 535-550.

Larjavaara M, Kuuluvainen T, Rita H (2005) Spatial distribution of lightening-ignited forest fires in Finland. Forest Ecology and Management 208: 177-188.

Lorz C, Fürst C, Galic Z, Matijasic D, Podrazky V, Potocic N, Simoncic P, Strauch M, Vacik H, Makeschin F (2010)GIS-based Probability Assessment of Natural Hazards in Forested Landscapes of Central and South-Eastern Europe. ENVIRON MANAGE. 2010; 46(6): 920-930.

Malamud BD, Millington JDA, Perry GLW (2005) Characterizing wildfire regimes in the USA. Proceedings of the National Academy of Sciences 102(13): 4694-4699.

Müller M, Vacik H, Diendorfer G, Arpaci A, Formayer H, Gossow H (2012) Analysis of lightening induced forest fires in Austria. Theoretical and Applied Climatology (in press) doi 10.1007/s00704-012-0653-7

Newman MEJ (2005) Power laws, Pareto distributions and Zipf's law. Contemporary Physics 46(5): 323-351.

Podur JJ, Martell DL, Knight K (2002) Statistical quality control analysis of forest fire activity in Canada. Canadian Journal of Forest Research 32(2): 195-205.

R Development Core Team (2008) R: A language and environment for statistical computing. Vienna, R Foundation for Statistical Computing.

Reed WJ, McKelvey KS (2002) Power-law behaviour and parametric models for the size-distribution of forest fires. Ecological Modelling 150: 239-254.

Ricotta C, Avena G, Marchetti M (1999) The flaming sandpile: self-organised criticality and wildfires. Ecological Modelling 119: 73-77.

San-Miguel J, Camia A (2009) Forest fires at a glance: Facts, figures and trends in the EU. In: Birot Y, editor. Living with wildfires: What science can tell us. Joensuu, European Forest Institute. pp. 11-18.

Schelhass MJ, Schuck A, Varis S (2003) Database on forest disturbances in Europe (DFDE) – Technical description. European Forest Institute internal report 14, Joensuu, EFI. pp. 44.

Sekhon JS (2011) Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R. Journal of Statistical Software 42(7): 1-52.

Seidl R, Schelhaas M-J, Lexer MJ (2011) Unraveling the drivers of intensifying forest disturbance regimes in Europe. Global Change Biology 17(9): 2842-2852.

Stocks BJ, Mason JA, Todd JB, Bosch EM, Wotto BM et al. (2003) Large forest fires in Canada, 1959-1997. Journal of Geophysical Research 108(D1): 5.1-5.1

UN (2002) Forest fires in Europe 1961-1998. International Forest Fire News 27: 76-80.

Vacik H, Gossow H (2011) Forest Fire Research and Management Options in Austria: Lessons Learned from the AFFRI and the ALPFFIRS Networks. In: Borsdorf A, Stötter J, Veulliet E, editors. Managing Alpine Future II conference, 21-23 November 2011, Innsbruck, Verlag der Österreichischen Akademie der Wissenschaften. pp 203-211

Vacik H, Arndt N, Arpaci A, Koch V, Müller M, Gossow H (2011) Characterisation of forest fires in Austria. AUSTRIAN J FOR SCI. 2011; 128(1): 1-31.

Valese E, Beck A, Comini B, Conedera M, Cvenkel H, Di Narda N, Ghiringhelli A, Japelj A, Lemessi A, Mangiavillano A, Pelfini F, Pelosini R, Ryser D, Vacik H, Wastl C (2010) The Alpine Forest Fire Warning System (ALP FFIRS) project. In: D. X. Viegas (Ed.), Proceedings of the VI International Conf. on Forest Fire Research, Coimbra, Portugal.

White EP, Enquist BJ, Green JL (2008) On estimating the exponent of power-law frequency distributions. Ecology 88: 905-912.