



**HAL**  
open science

## Statistical modelling of wildfire size and intensity: a step toward meteorological forecasting of summer extreme fire risk

Charles Hernandez, C. Keribin, Philippe Drobinski, Solène Turquety

### ► To cite this version:

Charles Hernandez, C. Keribin, Philippe Drobinski, Solène Turquety. Statistical modelling of wildfire size and intensity: a step toward meteorological forecasting of summer extreme fire risk. *Annales Geophysicae*, 2015, 33 (12), pp.1495-1506. 10.5194/angeo-33-1495-2015 . hal-01260501

**HAL Id: hal-01260501**

<https://hal.sorbonne-universite.fr/hal-01260501v1>

Submitted on 22 Jan 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Statistical modelling of wildfire size and intensity: a step toward meteorological forecasting of summer extreme fire risk

C. Hernandez<sup>1</sup>, C. Keribin<sup>2,3</sup>, P. Drobinski<sup>1</sup>, and S. Turquety<sup>4</sup>

<sup>1</sup>LMD/IPSL, CNRS and École Polytechnique, Université Paris-Saclay, Palaiseau, Paris, France

<sup>2</sup>Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay, France

<sup>3</sup>Inria Saclay-Île-de-France, 91405 Orsay, France

<sup>4</sup>LMD/IPSL, UPMC Univ. Paris 06, Sorbonne Universités, PSL Research University, CNRS, ENS, École Polytechnique, Université Paris-Saclay, Palaiseau, Paris, France

Correspondence to: C. Hernandez (charles.hernandez@lmd.polytechnique.fr)

Received: 31 July 2015 – Revised: 3 November 2015 – Accepted: 19 November 2015 – Published: 15 December 2015

**Abstract.** In this article we investigate the use of statistical methods for wildfire risk assessment in the Mediterranean Basin using three meteorological covariates, the 2 m temperature anomaly, the 10 m wind speed and the January–June rainfall occurrence anomaly. We focus on two remotely sensed characteristic fire variables, the burnt area (BA) and the fire radiative power (FRP), which are good proxies for fire size and intensity respectively. Using the fire data we determine an adequate parametric distribution function which fits best the logarithm of BA and FRP. We reconstruct the conditional density function of both variables with respect to the chosen meteorological covariates. These conditional density functions for the size and intensity of a single event give information on fire risk and can be used for the estimation of conditional probabilities of exceeding certain thresholds. By analysing these probabilities we find two fire risk regimes different from each other at the 90 % confidence level: a “background” summer fire risk regime and an “extreme” additional fire risk regime, which corresponds to higher probability of occurrence of larger fire size or intensity associated with specific weather conditions. Such a statistical approach may be the ground for a future fire risk alert system.

**Keywords.** Meteorology and atmospheric dynamics (synoptic-scale meteorology)

## 1 Introduction

In order to better manage fire risk, several methods have been investigated. Among the first are the fire risk indices, such as the Canadian Fire Weather Index (Van Wagner, 1974, 1987; Van Wagner and Pickett, 1985). This index relates to the expected intensity of the fire line, expressed in energy output rate per unit length of fire front. It is currently used as a fire risk indicator by the European Forest Fire Information System (EFFIS) of the Joint Research Center (JRC) of the European Commission. The Haines Index (Haines et al., 1983) is another indicator of dangerous fire development that focuses on atmospheric stability. It can be used in conjunction with the Canadian Fire Weather Index but is deemed less informative. These indices are empirically calibrated for predicting whether the atmospheric and hydrological conditions are prone to fire development. However, one of their main drawbacks is that they lack temporal contrast: they identify correctly fire-prone seasons but fail to provide short-term variability in fire risk (e.g., San-Miguel-Ayanz et al., 2013, Figs. 7, 8, 12 and 15). Other approaches exist, based on different criteria of fire risk. Using probabilistic cellular automata fire propagation models, simulations of multiple starting points can lead to risk maps than can be helpful for fire suppression forces deployment (Russo et al., 2014). The main weak point of this method is the lack of strong validation for the calibration of the propagation model. More in-depth simulations, using fully physical models such as FIRETEC (Linn et al., 2002), can provide accurate predictions of the propagation of a fire. This method can be very demanding computation-wise

and requires a precise knowledge of the initial and boundary conditions. Using a probabilistic framework, a preliminary risk assessment study was conducted (Preisler et al., 2004). The aim of the study was to reconstruct the probabilities of fire occurrence and large fire propagation using meteorological and geographical covariates. The results, although encouraging, gave only mitigated quality in the estimation of monthly fire occurrence. Modelling accumulated seasonal burnt area time series using meteorological predictors gave satisfying results, with adjusted  $R^2$  of 68 % for the July–August time period and northwestern region of Iberia (Sousa et al., 2015). Besides fire size or fire occurrence, another important factor of risk regarding wildfires is the intensity of the fire front. The propagation of particularly intense wildfires is indeed very hard to control and can trigger very severe pollution episodes. However large data sets do not exist for this quantity, so we focus instead on the fire radiative power (FRP), a remotely-sensed variable strongly linked with the fire intensity. The general framework of this study is the estimation of fire size and intensity of individual fires in the Mediterranean Basin using parametric statistical methods. Several studies focusing on the estimation of fire size exist, proposing to derive this quantity based on meteorological and geographical covariates. Their authors mainly use statistical learning techniques in order to give a quantitative or qualitative insight on fire size (Alonso-Betanzos et al., 2003; Cortez and Morais, 2007; Sakr et al., 2011). In some cases this analysis is extrapolated to future weather in the context of climate change (Amatulli et al., 2013). However one can reproach to these studies their lack of performance. An examination of Cortez and Morais (2007) lead to the observation that the estimation of fire size done by the best tested method was only very marginally better than the mean of the observations. For fire intensity no studies of this kind were conducted. Our approach will be to provide parametric estimations of both single-event fire size and intensity distribution functions conditionally to weather covariates. We take a multi-timescale approach for the choice of our weather covariates, with seasonal and immediate weather information. Using these conditional distribution estimations we can then compute probabilities that a given fire grows particularly large or becomes very intense. Because of our methodology, these probabilities would be both sensitive to seasonal trends and immediate weather. These estimations would be much more informative than a conditional mean of fire size of intensity with respect to weather. In Sect. 2 we describe the data we use. After presenting our fire variables, we show our weather covariates and explain their relevance. In Sect. 3 we find an adequate parametric distribution to model fire size and intensity of individual events. Using this result, we develop in Sect. 4 a methodology of fire risk assessment that focuses on the use of probabilities of large and/or intense wildfires.

## 2 Data

### 2.1 Fire variables

The detection of fires is performed using the fire products from MODIS (Moderate Resolution Imaging Spectroradiometer), an instrument carried on board of the Aqua and Terra polar heliosynchronous orbiting satellites. The recorded fire variables are the burnt area (BA) and the fire radiative power (FRP) which can be seen as a proxy of the fire intensity. We focus on the Mediterranean Basin. We therefore select the fires occurring within the box [35, 50° N] and [−10, 50° E]. We keep only individual wildfire events occurring during the months of July and August in order to focus on the core of the fire season in the study area (Ganteaume et al., 2013). However summer is not the only season when fires occur. For example in Northern Iberia and Galicia the month of September also exhibits strong fire activity (e.g. Pereira et al., 2005, Fig. 2). Winter and early spring fires can also occur in Portugal and the Balkans (e.g. Moriondo et al., 2006, Table 1). However, we focus our analysis on the summer period to avoid seasonal changes in the driving factors, especially at the scale of the Mediterranean basin. Such a generalization of our approach is left for future work. There are 5821 and 4840 wildfires in our two BA data sets and 24 273 wildfires in our FRP data set. The FRP is retrieved by using measured radiance of the 4 and 11  $\mu\text{m}$  channels at nadir. Other spectral bands are used for assessing cloud masking, glint, bright surface and other sources of false alarms and disturbances. FRP is provided at 1 km resolution by the MOD14 product. BA is retrieved from the observed changes in land cover. Indeed, the albedo is modified by the deposition of charcoal and ash, the loss of vegetation and the change in fuel bed characteristics. Albedo alteration produces changes in surface reflectance which are processed to produce daily burnt area at a 500 m resolution in the MDC64A1 product (Giglio et al., 2010). Only the fraction of the detected burning pixel covered by vegetation is burned following Turquety et al. (2014). The FRP and BA products are then regridded at 10 km resolution which was chosen to be a good trade-off in order to keep detailed enough information on the fire location and facilitate the comparison with the ERA-Interim meteorological data. We use the first 10 years (2003–2012) of MODIS data. It should be noted that there are important uncertainties on the date of beginning of wildfires taken individually. The incertitude can be as large as 5 days and is caused by several factors such as cloud cover impairment of remote sensing and lack of detection of wildfires at the beginning of their development. As we deal with statistics on a large number of such wildfires, the uncertainty is reduced. Additionally, since the time period of study is mostly cloud-free, the uncertainty on the day of detection should be low (Giglio et al., 2010). This is confirmed by the strong link between fire and synoptic weather dynamics observed using the same methods in Her-

andez et al. (2015b). In the following sections we will call these BA and FRP data sets  $BA_M$  and FRP respectively. We also use the EFFIS Rapid Damage Assessment system, provided by the JRC of the European Commission (European Commission, 2010). This set is built using 250 m MODIS images. A first step of automated classification is used to isolate fire events and a post-processing using human visualization of the burnt scar is performed. A cross-analysis using the active fire MODIS product, land-cover data sets as well as fire event news collected in the EFFIS News module is finally done to ensure a low number of misclassifications (<http://forest.jrc.ec.europa.eu/effis/>). The system records burnt areas of approximately 40 ha and larger (Sedano et al., 2013). It also contains smaller wildfires, but is less complete below this 40 ha threshold. The JRC provided the data for the 2006–2012 time period. We call this BA data set  $BA_E$  in the following. A 3-D (latitude, longitude and time) connected component algorithm is used to determine what are the distinct fire events in the  $BA_M$  data set. This algorithm aggregates the adjacent fire spots into larger fire events. The main interest of this method is that it allows for the detection of wildfires larger than 10 000 ha which are those expected to be most influenced by weather conditions (e.g. Pereira et al., 2005). The main weakness is that it does not take into account cloud cover impairment of remote sensing. Indeed an absence of detection of 1 day between two detections could be caused by clouds. Another problem is that two independent fire events taking place close to one another (less than 20 km of distance and less than a day between the end of the first event and the beginning of the second) are considered the same by this method. “Megafire” events, such as those defined by San-Miguel-Ayanz et al. (2013), could also be grouped in clusters with this method of analysis. The processing of the  $BA_E$  data set is simpler. The data set provides the shape and time of beginning of all detected wildfires. We take as location the centroid of this shape. Detection of smaller wildfires being quite hard with remote-sensing techniques, we choose to eliminate < 25 ha wildfires from our burnt area data sets. They correspond to wildfires burning less than one pixel in the  $BA_M$  data set and the authors have doubts about the completeness of the BA data sets below this value. In the following sections it should therefore be stressed that the obtained results only hold for such wildfires. Our fire data sources and pre-processing methods are identical to that of Hernandez et al. (2015a, b). After these preprocessing steps we retain 5821 observations for the  $BA_M$  data set, 4840 for the  $BA_E$  data set and 24 273 wildfires for the FRP data set.

## 2.2 Meteorological covariates

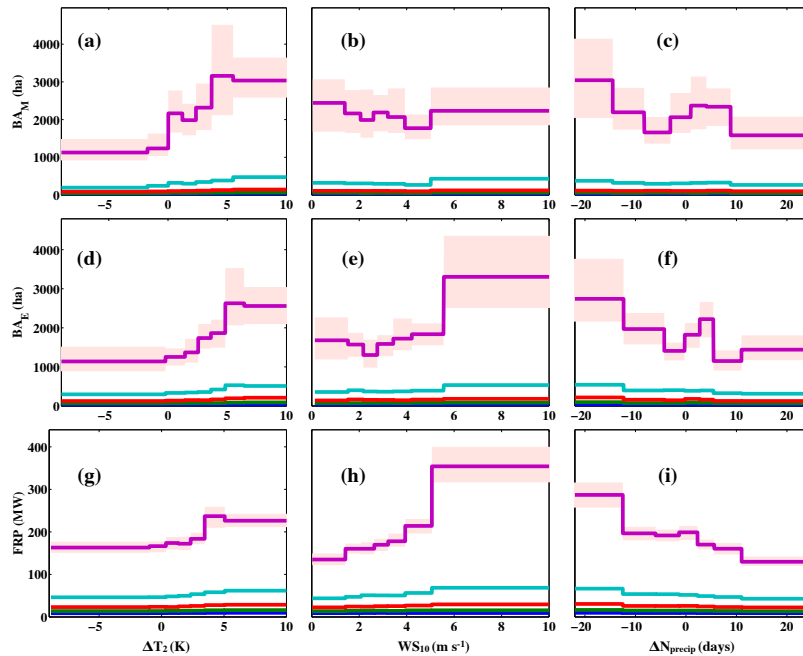
Our weather database was built upon the ERA-Interim reanalysis of the European Center for Medium-range Weather Forecast (ECMWF) (Dee et al., 2011). The horizontal resolution of the reanalysis does not allow the derivation of the small-scale weather conditions in the immediate vicinity of

the fire. To link the weather data to the fire data, we take the ERA-Interim grid point nearest from the detected fire event. We then associate to this event the weather recorded at 12:00 UTC the day of first detection. We extract the following meteorological covariates:

- $\Delta T_2$  (in K): the 2 m air temperature anomaly, the difference between the 12:00 UTC 2 m air temperature and its climatological daily mean;
- $WS_{10}$  (in  $\text{m s}^{-1}$ ): the 10 m wind speed;
- $\Delta N_{\text{precip}}$  (in days): the anomaly with respect to the climatology of the number of days when precipitations  $\geq 0.5$  mm occur during the January–June time period preceding of the year of the fire.

$\Delta N_{\text{precip}}$  is mostly impacted by spring drought occurrence but positive winter precipitation anomaly has also been linked to the 2003 Portugal megafire event (Trigo et al., 2006). However, as shown in Vautard et al. (2007) and Stéfanon et al. (2012a), anomalies of precipitation during spring are favourable to summer heatwave conditions. Stéfanon et al. (2012b) have also shown that deficit of precipitation during spring, can trigger early vegetation growth, providing abundant fire fuels in summer. Positive winter precipitation anomalies may amplify this mechanism. Our choice of covariates was done to retain a broad range of timescales. We go from the hourly to daily timescales ( $\Delta T_2$ ,  $WS_{10}$ ) to seasonal timescales ( $\Delta N_{\text{precip}}$ ). We also settled on covariates with proven impact on wildfire activity. Wind speed accelerates the propagation of the fire (Rothermel, 1972) in the direction of the wind and blocks back propagation. The temperature anomaly  $\Delta T_2$  is an indicator of heatwave occurrence. Pereira et al. (2005) showed that in Portugal wildfires often co-occurred with synoptic blockings and heatwaves. In Sardinia, Cardil et al. (2014) showed that large fire occurrence, daily burnt area and daily number of fires were higher on high temperature days. Hernandez et al. (2015a, b) further this work by showing that heatwaves and surface wind control wildfire size and duration strongly. Dimitrakopoulos et al. (2011) emphasized the link between drought and wildfire activity (wildfire occurrence and area burnt) in Greece. We chose  $\Delta N_{\text{precip}}$  as an indicator of drought occurrence preceding the wildfire. Low values of  $\Delta N_{\text{precip}}$  indicate both low precipitation amount and low overall cloudiness in the January–June time period. Intuitively, we could say that more arid preceding seasons could lead to lower values of soil and fuel moisture during summer. Zampieri et al. (2009) showed that this quantity is linked to drought occurrence in summer. Additionally Vautard et al. (2007) and Stéfanon et al. (2012b) showed that summer heatwave occurrences were also impacted by rainfall deficit in previous months.

Our first attempt at linking fire and weather data used regression techniques to forecast the conditional mean. This approach failed, with maximum  $R^2$  of 0.10 and 0.05 for



**Figure 1.** Evolution of 5th (blue), 25th (green), 50th (red), 75th (cyan) and 95th (purple) quantiles of BA (data set  $BA_M$  and  $BA_E$ ) and FRP (FRP data set) with  $\Delta T_2$ ,  $WS_{10}$  and  $\Delta N_{\text{precip}}$ . Top row corresponds to  $BA_M$ , middle to  $BA_E$  and bottom to FRP. The red shaded area corresponds to 90 % confidence intervals for the 95th quantile.

the FRP and BA data sets respectively using artificial neural networks. We therefore chose to focus our analysis on the variability of the distributions of BA and FRP with respect to weather, and at first on the variations of the quantiles of these distributions. Figure 1 shows the variations of the 5th, 25th, 50th, 75th and 95th quantiles of BA and FRP for data sets  $BA_M$ ,  $BA_E$  and FRP with respect to the selected covariates. The methodology consists in splitting the data sets into seven subsets containing an equal number of points. This allows comparable uncertainties for each subset. The number of bins was chosen as a trade-off between the smoothness of the curve and the significance of the curve fluctuations. These statistics were bootstrapped 1000 times, allowing an accurate estimation of each quantile and of the associated confidence intervals. First, we can see that these variations depend heavily on the selected quantile. In particular the 5th quantile seems roughly constant whereas the 95th is more variable. BA and FRP show strong responses to  $\Delta T_2$ , with general growth of fire size and radiative power. For the  $BA_E$  and FRP data sets, BA and FRP are growing functions of  $WS_{10}$ . This is not seen for the  $BA_M$  data set. However Hernandez et al. (2015b) show that by conditioning on  $\Delta T_2$  significant variations of BA and FRP can be observed at the 70 and 90 % confidence levels respectively. We observe that BA and FRP decrease with increasing  $\Delta N_{\text{precip}}$ . In the following we use  $\Delta T_2$ ,  $WS_{10}$  and  $\Delta N_{\text{precip}}$  to reconstruct the conditional distribution functions of BA and FRP.

### 3 BA and FRP distributions

Figure 1 shows that the variability of BA and FRP is very high, and a proper way to build a risk metric would be to compute probabilities of large fire size or large intensity using these variations. A way of doing so would be to model the conditional distributions of BA and FRP with respect to weather. To achieve this goal we want to find a parametric distribution which fits these variables well. In this section we proceed to this task independently of the weather covariates in order to provide good models for the distributions of BA and FRP. The meteorological covariates will be reintegrated at the beginning of Sect. 4.

As BA and FRP have very skewed distributions it becomes easier to study their logarithm. We therefore from this point onward only discuss the modelling of  $\log_{10}(BA)$  and  $\log_{10}(FRP)$ . We also subtract a threshold to each variable ( $\log_{10}(25)$  for the BA data sets and  $\log_{10}(4)$  for the FRP data set), so as the data starts approximately at 0 and is always non-negative.

The parametric forms that are tested for the distributions of the transformed fire variables are the following:

- the Exponential distribution,

$$f(x; \beta) = \beta e^{-\beta x}, \quad x \in \mathbf{R}^+;$$

- the Normal distribution,

$$f(x; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}, \quad x \in \mathbf{R};$$

**Table 1.** AD2R values for all different distributions and for all data sets. The AD2R values for the chosen distributions are in bold.

Criterion	Data set	Normal	Exponential	Cauchy	Gamma	Logistic	Log-Normal	GEV
AD2R	BA <sub>M</sub>	190	370	394	20.0	58.3	103	<b>20.3</b>
	BA <sub>E</sub>	51.4	648	355	23.6	16.5	83.5	<b>3.45</b>
	FRP	674	4951	1626	<b>15.2</b>	121	124	17.8

– the Cauchy distribution,

$$f(x; x_0, a) = \frac{1}{\pi a \left( 1 + \left( \frac{x - x_0}{a} \right)^2 \right)}, x \in \mathbf{R};$$

– the Gamma distribution,

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x \in \mathbf{R}^+;$$

– the Logistic distribution,

$$f(x; \mu, s) = \frac{e^{-\frac{x-\mu}{s}}}{s \left( 1 + e^{-\frac{x-\mu}{s}} \right)}, x \in \mathbf{R};$$

– the Log-Normal distribution,

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, x \in \mathbf{R}^+;$$

– the Generalized extreme value (GEV) distribution,

$$f(x; \mu, \sigma, \xi) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi-1} e^{-\left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi}},$$

$x \in \mathbf{R}, x \geq \mu - \sigma/\xi, \xi > 0.$

Here  $f$  denotes the corresponding probability density function.

If  $Y$  is a random variable, the truncated exponential distribution for  $\log Y$  correspond to the truncated pareto distribution for  $Y$ . As the Truncated Pareto distribution was shown alongside with the Tapered Pareto distribution to be a good fit for the distribution of BA (Schoenberg et al., 2003), we included the exponential distribution in our possible forms for  $\log_{10}(\text{BA}/25)$  and  $\log_{10}(\text{FRP}/4)$ .

We fitted all these distributions for each data set (BA<sub>M</sub>, BA<sub>E</sub> and FRP) using the minimization of the AD2R goodness-of-fit criterion (Anderson and Darling, 1954) as fit-

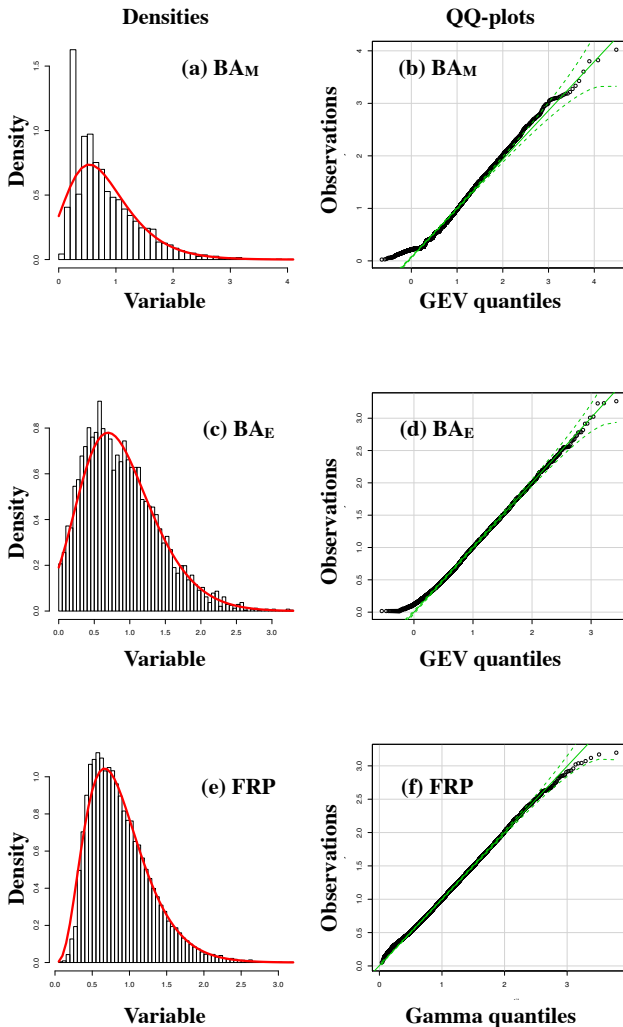
ting method. The AD2R criterion is defined as follows:

$$\text{AD2R}(F) = \int \left( \hat{F}_n(x) - F(x) \right)^2 \Psi(x) dx,$$

with  $\Psi(x) = (1 - F(x))^{-2}$  (1)

with  $\hat{F}_n$  being the empirical, step-wise cumulative density function of the data to fit and  $F$  the cumulative density function for which the AD2R criterion is calculated. The choice of the function  $\Psi$  gives more weight to the quality of the fit for the right tail of the distribution. If  $F(x)$  and  $\hat{F}_n(x)$  were to have different asymptotic behaviours for large values of  $x$  the AD2R criterion would be very large. The minimization of the AD2R criterion then has the theoretical advantage of making a better fitting of the distribution for larger values of the selected variable. All the AD2R values found for each distribution and data set are available in Table 1. Computations were done in R (R Core Team, 2013) using the “fitdistplus” package (Delignette-Muller and Dutang, 2015). We see that for the BA data sets there are two distributions selected, Gamma and GEV. We will continue using only the GEV distribution since the difference seen for the BA<sub>M</sub> data set between these two distribution is very small (AD2R values of 20.3 for the GEV distribution and 20.0 for the Gamma distribution), whereas for the BA<sub>E</sub> data set the difference is much larger (AD2R values of 3.45 for the GEV distribution and 23.6 for the Gamma distribution). For FRP the Gamma distribution is selected. Surprisingly the Exponential distribution fits the BA data sets poorly. This could be due to the absence of the <25 ha wildfires in our BA<sub>M</sub> and BA<sub>E</sub> data sets, whereas they are taken into account in Schoenberg et al. (2003).

Figure 2 shows the normalized histograms and modelled densities of BA and FRP with accompanying QQ-plots for all considered data sets. The QQ-plots were computed using the car package (Fox and Weisberg, 2011). For values of BA smaller than 40 ha, the QQ-plots depart from the 95 %-level confidence intervals. Conversely, the QQ-plots are within the confidence intervals for larger values. The distribution fits better the BA<sub>E</sub> data set than the BA<sub>M</sub>. It may be due to the methodology of construction of this data set, which considers burned only the fraction of the burning MCD64A1 pixels of surface 25 ha covered by vegetation. A preference for multiples of 25 ha arises and it is detrimental for the accuracy on the distribution tails of BA, and especially the lower percentiles. However, the fit is still accurate enough for our purpose. As only the largest wildfires are controlled by the



**Figure 2.** Normalized histograms, modelled densities (a, c, e) and QQ-plots (b, d, f) for the GEV and Gamma distributions for the BA and FRP data sets respectively. The fitting method used is the AD2R criterion minimization. On the densities panels the normalized histograms are in black and the modelled distribution in red. The dashed green lines on the QQ-plots are the 95 % confidence envelopes.

weather conditions (Hernandez et al., 2015a), having an accurate fit of the high values of BA and FRP is enough for our modelling framework. Caution should therefore be taken when trying to interpret these distributions for low values of BA. For FRP, the QQ-plot remains within the 95 %-level confidence intervals for all values. Besides the AD2R criterion, Fig. 2 shows that the GEV and Gamma models fit the data accurately and can be considered suited for our model. In the following, we will take the strong hypothesis that the observations coming from the BA and FRP data sets have respectively GEV and Gamma distributions conditionally to the weather. This hypothesis was tested on large subsets of the data sets corresponding to particularly favourable or un-

favourable weather conditions. We take as favourable conditions  $\Delta T_2 \geq 5 \text{ K}$  and  $WS_{10} \geq 6 \text{ m s}^{-1}$  and as unfavourable conditions  $\Delta T_2 \leq 0$  and  $WS_{10} \leq 3 \text{ m s}^{-1}$ . We find that the hypothesis holds well for the  $BA_E$  and FRP data sets, but that there are more discrepancies with the  $BA_M$  data set, which is coherent with the deviations seen in Fig. 2. This hypothesis is used to obtain the conditional distribution of BA and FRP with respect to  $\Delta T_2$ ,  $WS_{10}$  and  $\Delta N_{\text{precip}}$ .

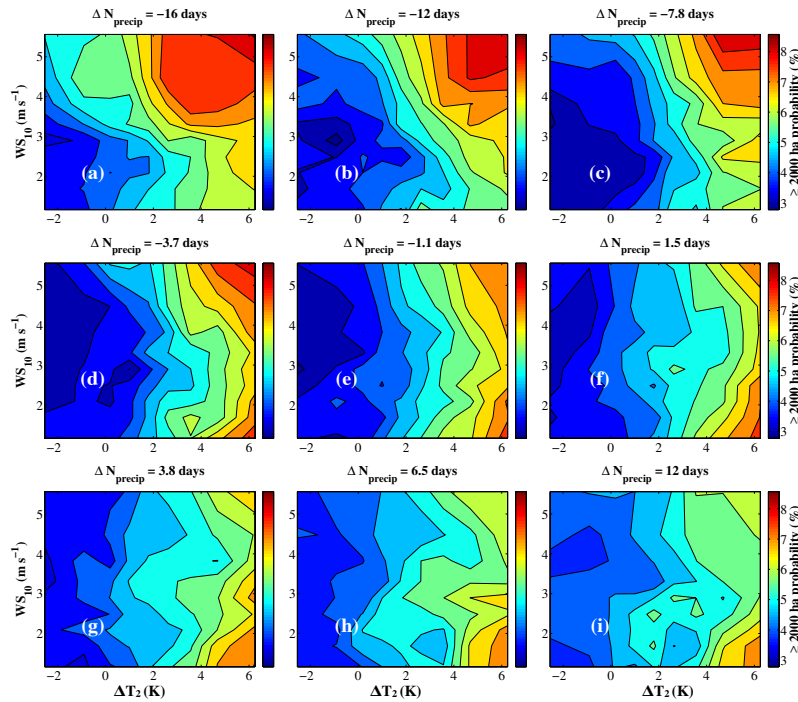
## 4 Fire risk assessment using meteorological covariates

### 4.1 Methodology

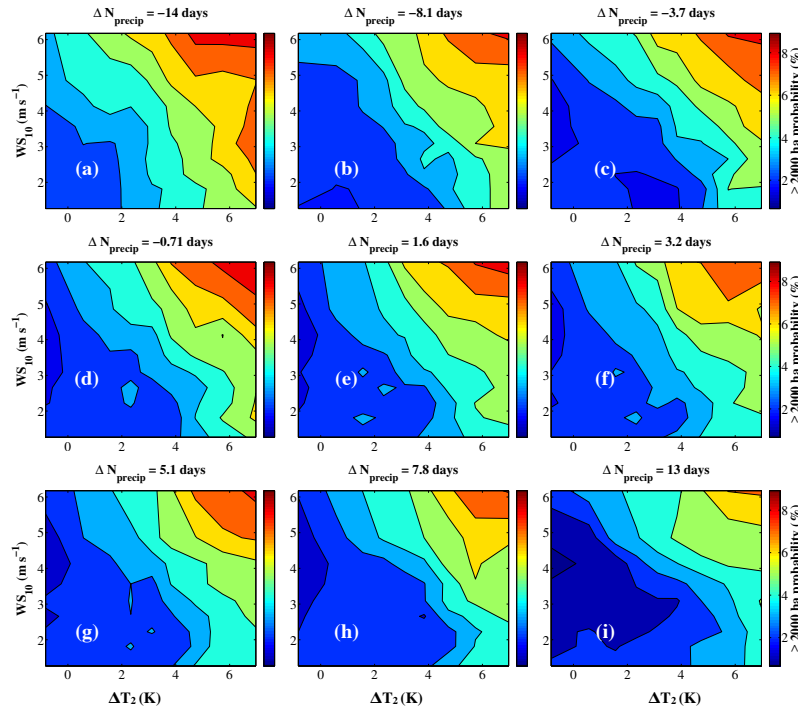
The general framework of our methodology is the parametric estimation of the conditional probability density function of BA or FRP with respect to  $\Delta T_2$ ,  $WS_{10}$  and  $\Delta N_{\text{precip}}$ . In other words we seek  $f_{Y|\mathbf{X}}(y) = f_Y(y|\mathbf{X} = \mathbf{x})$  with  $y$  the fire variable,  $\mathbf{X}$  the meteorological covariates and  $\mathbf{x}$  a specific value taken by the covariates. We made the hypothesis in the previous section that  $f_{\log_{10}(\text{BA}/25)} \sim \text{GEV}(\mu, \sigma, \xi)$  and  $f_{\log_{10}(\text{FRP}/4)} \sim \text{Gamma}(\alpha, \beta)$  for all subsets of our data sets. Therefore to approximate the values of the parameters of these distributions we need to compute the distribution of  $y$  near the point  $\mathbf{X} = \mathbf{x}$ . To do so we choose to retain the 10 % of our data sets nearest of the point  $\mathbf{X} = \mathbf{x}$  and to estimate the parameters of the distribution by minimizing the AD2R criterion. The fraction of nearest neighbours was chosen to be sufficient to estimate a distribution function. The calculation of these nearest neighbours was done in R using the FNN package (Beygelzimer et al., 2013). It must be noted that due to the curse of dimensionality taking a larger number of covariates would lead to a very large inaccuracy on  $\mathbf{x}$  (Hastie et al., 2009, pp. 22–23). In order to tackle this issue we select only three covariates for our density estimation. The choice of these covariates was done using Fig. 1. We wish to retain covariates that cover a broad range of temporal variability and for which BA and FRP exhibit strong significant variability. We therefore choose to take  $\mathbf{X} = (\Delta T_2, WS_{10}, \Delta N_{\text{precip}})$  for all data sets. For computation purposes we choose not to estimate  $f_{Y|\mathbf{X}}$  at each possible value of  $\mathbf{x}$ . Instead we take the values of  $\mathbf{x}$  corresponding to the 1st to 9th deciles of each of its components. This makes  $9^3 = 729$  values of  $\mathbf{x}$  for which each conditional distribution parameters are estimated. In order to obtain asymptotic confidence intervals for our estimates of the conditional distribution parameters and of the probability of large or intense events we perform 500 bootstrap estimations of these parameters using the determined nearest neighbours. Bootstrap estimation was done using the bootstrap R package (Leisch and Tibshirani, 2014).

### 4.2 Results

Figures 3, 4 and 5 show the estimated probability contours of particularly large or intense fire events computed from our method. These events are defined by the wildfire ex-

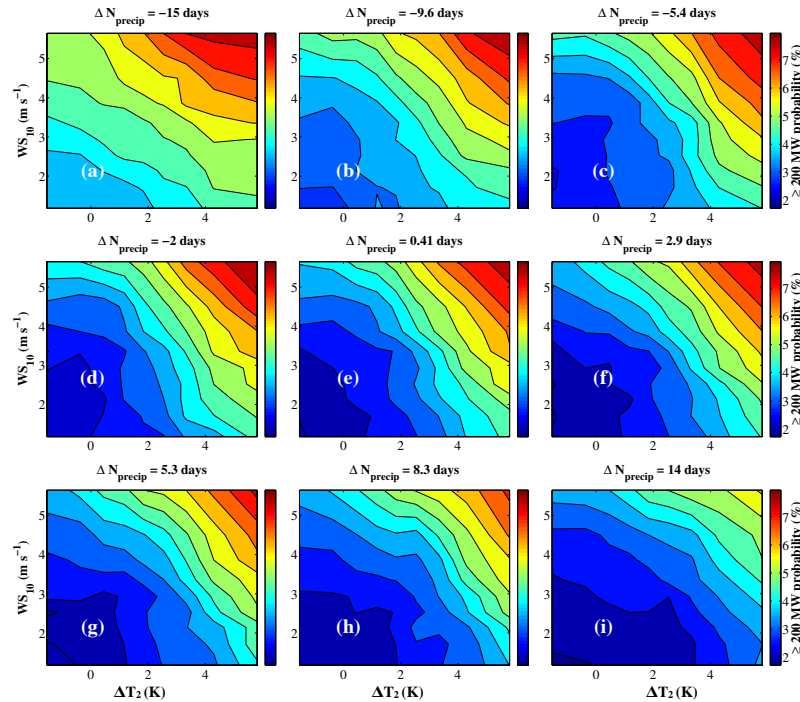


**Figure 3.** Estimated probabilities of fire size (BA) exceeding the 2000 ha threshold ( $BA_M$  data set). The  $x$  axis is the 2 m air temperature anomaly ( $\Delta T_2$ ), the  $y$  axis the 10 m wind speed ( $WS_{10}$ ) and each panel stands for values of January–June precipitation days anomaly ( $\Delta N_{precip}$ ) centred on the given value on the panel titles.



**Figure 4.** Estimated probabilities of fire size (BA) exceeding the 2000 ha threshold ( $BA_E$  data set). The  $x$  axis is the 2 m air temperature anomaly ( $\Delta T_2$ ), the  $y$  axis the 10 m wind speed ( $WS_{10}$ ) and each panel stands for values of January–June precipitation days anomaly ( $\Delta N_{precip}$ ) centred on the given value on the panel titles.





**Figure 5.** Estimated probabilities of fire intensity (FRP) exceeding the 200 ha threshold. The  $x$  axis is the 2 m air temperature anomaly ( $\Delta T_2$ ), the  $y$  axis the 10 m wind speed ( $WS_{10}$ ) and each panel stands for values of January–June precipitation days anomaly ( $\Delta N_{\text{precip}}$ ) centred on the given value on the panel titles.

**Table 2.** Mean values of the standard deviations calculated from the nearest neighbours search ( $\Delta T_2$ ,  $WS_{10}$  and  $\Delta N_{\text{precip}}$ ).

Data set	$\Delta T_2$ (K)	$WS_{10}$ ( $\text{m s}^{-1}$ )	$\Delta N_{\text{precip}}$ (days)
$BA_M$	1.41	0.74	4.4
$BA_E$	1.22	0.83	3.9
FRP	1.17	0.74	4.4

ceeding the 2000 ha or 200 MW thresholds in BA or FRP respectively. These thresholds correspond approximately to the 95th quantiles of each variable. The values of each class of  $\Delta N_{\text{precip}}$  corresponds to the mean of the  $\Delta N_{\text{precip}}$  of each decile. Each panel displays the mean distribution of the corresponding  $\Delta N_{\text{precip}}$  class. The uncertainty of the distribution can be inferred from Table 2 which displays the average standard deviation of each covariate. The probability of large BA occurring is a growing function of  $\Delta T_2$  (Figs. 3 and 4). The two modes of higher BA commented and analysed in Hernandez et al. (2015b) are visible in Fig. 3. There is a clear significant increase in large BA probabilities with increasing  $\Delta T_2$  and  $WS_{10}$  for low values of  $\Delta N_{\text{precip}}$ . The role of  $WS_{10}$  is significantly damped when  $\Delta N_{\text{precip}}$  rises (wetter January–June time period) and  $\Delta T_2$  becomes the main driving factor for the  $BA_M$  data set (Fig. 3). Accounting for the confidence intervals of the estimated probabilities (not shown)

shows that  $WS_{10}$  has no explanatory value in the pattern of the probability at the 90 % confidence level. The variations between the minima and maxima of the estimated probabilities are significant at the 90 % confidence level. However the two modes are hard to distinguish statistically because of the low number of points in our BA data sets (5821 for  $BA_M$  and 4840 for  $BA_E$ ). The difference of results between the  $BA_M$  and  $BA_E$  probabilities is due to the  $BA_E$  data set spanning over the 2006–2012 time period, therefore missing the 2003 and 2005 megafire events which are present in the  $BA_M$  data set (San-Miguel-Ayanz et al., 2013). Regarding fire intensity, FRP is a growing function of  $WS_{10}$ ,  $\Delta T_2$  and a decreasing function of  $\Delta N_{\text{precip}}$ , which is significant at the 90 % confidence level. The variability linked to  $\Delta T_2$  and  $WS_{10}$  is discussed in Hernandez et al. (2015a) and found back on this figure. Because we use a meteorological covariate depending on past weather ( $\Delta N_{\text{precip}}$ ), a seasonal preconditioning of high fire risk can be assessed. When a drought occurs in the past months ( $\Delta N_{\text{precip}} \leq -7$  days) the highest probabilities of large BA can be found for high values of both  $\Delta T_2$  and  $WS_{10}$  (Fig. 3). For higher values of the past months precipitation anomaly ( $\Delta N_{\text{precip}} \geq 7$  days), the highest risk corresponds to heatwaves, with high  $\Delta T_2$  and low  $WS_{10}$ . This difference could be exploited to adapt fire mitigation strategies and take into account seasonal weather information. The absence of the 2003 and 2005 megafire events (San-Miguel-Ayanz et al., 2013) limits the number of observations used

to derive the parameters of the distributions, therefore explaining the absence of significant discrimination between situations of spring drought and the others in the  $BA_E$  data (Fig. 4).

Let us illustrate the information provided by our method by focusing on the 2003 megafire event in Portugal. We take the largest wildfire event of the  $BA_M$  data set (262 520 ha BA, 731 MW FRP). It is recorded at  $[-7.65^\circ \text{ E}, 40^\circ \text{ N}]$  and the considered weather is that of the  $[-7.50^\circ \text{ E}, 39.75^\circ \text{ N}]$  ERA-Interim grid point. Figure 6 shows the time evolution of the probability of large BA and FRP with the corresponding 90 % confidence intervals. Two black lines show the beginning and the end of the fire event. During the wildfire the probability of large BA peaks to 7 %, whereas it stays at about 3 % ( $BA_M$ ) or 2 % ( $BA_E$ ) the rest of the time. The probability of large FRP behaves the same way, going from 3 % to more than 6 %. The variations of these estimated probabilities are significant at the 90 % confidence level. The “background” probability refers to the background fire risk of large or intense fire events during summer. We also see a secondary peak before the fire event, even though no fire occurred. Our method can be used to identify time periods when fire risk is especially high. When a fire occurs during one of these “extreme” periods, the fire event has high odds of being catastrophic.

Regarding the uncertainties of the method the mean standard deviation of the meteorological covariates have been calculated (Table 2). They stem from our nearest neighbours approach. The uncertainties on the meteorological features are fairly small and, with the exception of  $\Delta N_{\text{precip}}$ , fall within measurement error. Figure 7 shows the normalized histograms of the estimated probabilities and of the confidence intervals lengths for all July–August time periods everywhere a fire is detected. We also quantify the mean and standard deviation of the “background” and “extreme” fire risk regimes. To do this, the densities of the estimated probabilities are fitted with a mixture of two Gaussians, representing the “background” and “extreme” fire risk regimes. The model can be written as follows:

$$\begin{aligned} \text{PDF}(x; \mu_1, \sigma_1, \mu_2, \sigma_2, \alpha) &= \frac{\alpha}{\sigma_1 \sqrt{2\pi}} \\ &\times \exp\left\{-\frac{(x - \mu_1)^2}{2\sigma_1^2}\right\} + \frac{1 - \alpha}{\sigma_2 \sqrt{2\pi}} \\ &\times \exp\left\{-\frac{(x - \mu_2)^2}{2\sigma_2^2}\right\}. \end{aligned} \quad (2)$$

For  $BA_E$  data set, the distinction between “background” and “extreme” is more difficult than for  $BA_M$  due to the absence of major megafires in the data set (2003 and 2005). Otherwise, the mean probability that a fire exceeds 2000 ha is around 4 % for “background” summer fire risk conditions with a standard deviation of 0.5 % and increases to 5 % in extreme weather conditions favourable to larger fires. A similar behaviour is found for fire intensity with an even more

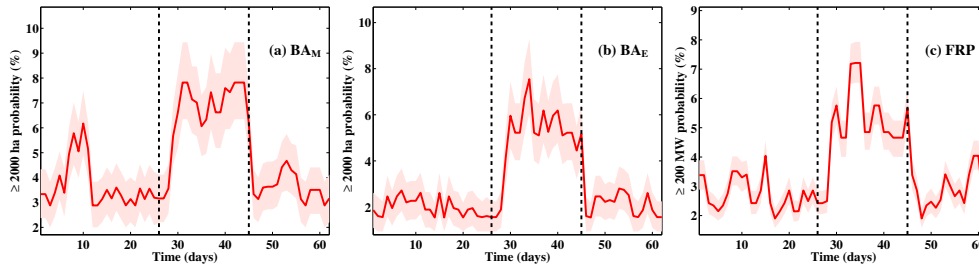
distinguishable two-mode distribution. The mean probability that a fire exceeds 200 MW is around 2.4 % for “background” summer fire risk conditions with a standard deviation of 0.3 % and increases to 3.6 % in extreme weather conditions favourable to intense fires. The 90 %-level confidence interval lengths remain large for the BA data sets, with typical values of 2 and 1.8 % for the  $BA_M$  and  $BA_E$  data sets respectively. For the FRP data set these lengths are smaller because of the larger number of data points, with a mean value of approximately 1 %.

## 5 Conclusions

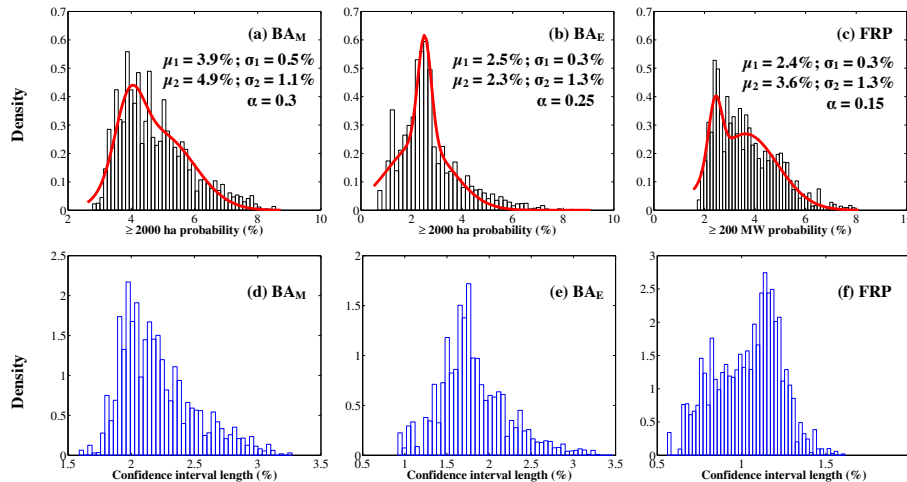
Statistical modelling of burnt area (BA) and fire radiative power (FRP) was investigated in this article. Using maximum goodness-of-fit techniques the density functions of  $\log_{10}(\text{BA})$  and  $\log_{10}(\text{FRP})$  were found to be well represented by GEV and Gamma distributions respectively. Using the hypothesis that this result holds for the conditional distribution of the fire variables with respect to meteorological covariates, a methodology for its estimation with three weather parameters was designed. Surface wind speed, 2 m air temperature anomaly and rainfall occurrence anomaly in January–June were selected to fit BA and FRP. The statistical model proved to be efficient in associating large fire risk with previous fire events, and so with rather low uncertainties. Such a model would be useful for the design of a data-driven wildfire alert system in the Mediterranean Basin taking into account seasonal trends and weather forecasts.

- Our model allows to discriminate accurately jumps between “background” summer fire risk regime and an “extreme” additional fire risk regime, corresponding to higher probability of occurrence of larger fire size or intensity associated with specific weather conditions;
- our model provides information for both the fire size and the fire intensity;
- our model provides an estimation of the probability of risk to exceed given values of fire size and fire intensity each time meteorological forcing data are available, that is typically on an hourly to 6-hourly basis;
- our model includes enhanced fire risk preconditioning by precipitation occurrence anomaly during the preceding months.

However, this work must be seen as a first step towards fire risk forecasting and a thorough analysis is required to assess the model performance in forecast mode. In this study we use parametric distributions as they provide a simple framework to model fire risk with a limited number of coefficients, which can be of interest for the implementation of a fire risk forecast system. Non-parametric estimations of the conditional distributions of the fire variables with respect to the



**Figure 6.** Probabilities of observing a  $\geq 2000$  ha wildfire calculated from the  $BA_M$  (a) and  $BA_E$  (b) data sets and probabilities of observing a  $\geq 200$  MW wildfire calculated from the FRP data set (c) as a function of time for the 2003 July–August period nearest the largest wildfire occurring in Portugal this season. Black dashed lines show the beginning and the end of the wildfire event. In light shaded red are the 90 % confidence intervals.



**Figure 7.** Normalized histograms of the estimated probabilities (black), PDFs of the mixture model (red) and normalized histograms of the 90 %-level confidence intervals lengths (blue). (a, d) are for  $BA_M$ , (b, e) for  $BA_E$  and (c, f) for FRP. The data set is made of each July–August time period everywhere a fire is detected. The parameters of the gaussian mixture model (Eq. 2) are displayed on each panel of the top row.

meteorological covariates could be performed (Brunel et al., 2010). This would lead to longer computation times but probably more accurate estimations of the conditional distributions and associated probabilities. More complete data sets for BA would allow a better estimation of the conditional distribution for this particular variable and help further reduce the uncertainties. Finally, in order to improve fire risk forecasting meteorological driving factors of fire size and intensity can be used to reconstruct a conditional distribution function of either variable. Such a method provides much more information than commonly used fire risk indicators (e.g. the Canadian Fire Weather Index) as one gets the distribution of all possible fire sizes and intensities given the meteorological covariates rather than an estimation of the fire intensity alone. The method also allows a multi-timescale analysis of the fire risk level as it accounts for preconditioning build-up by past months drought and the instantaneous wind speed and temperature anomaly with respect to the daily climatology. It thus produces a contrasted day-to-day probabil-

ity of large fire size and intensity which can be combined into a single fire risk indicator.

**Acknowledgements.** This work contributes to the HyMeX program (HYdrological cycle in The Mediterranean EXperiment – Drobinski et al. (2014)) through INSU-MISTRALS support and the GEWEX hydroclimate panel of the World Climate Research Program (WCRP). Data were provided by the European Forest Fire Information System – EFFIS (<http://effis.jrc.ec.europa.eu>) of the European Commission Joint Research Center.

The topical editor V. Kotroni thanks V. Moron and one anonymous referee for help in evaluating this paper.

## References

- Alonso-Betanzos, A., Fontenla-Romero, O., Guijarro-Berdiñas, B., Hernández-Pereira, E., Inmaculada Paz Andrade, M., Jiménez, E., Luis Legido Soto, J., and Carballas, T.: An intelligent system for forest fire risk prediction and fire fighting management in Galicia, *Expert Syst. Appl.*, 25, 545–554, 2003.

- Amatulli, G., Camia, A., and San-Miguel-Ayanz, J.: Estimating future burned areas under changing climate in the EU-Mediterranean countries, *Sci. Total Environ.*, 450, 209–222, 2013.
- Anderson, T. W. and Darling, D. A.: A test of goodness of fit, *J. Am. Stat. Assoc.*, 49, 765–769, 1954.
- Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., and Li, S.: FNN: Fast Nearest Neighbor Search Algorithms and Applications, r package version 1.1, available at: <http://CRAN.R-project.org/package=FNN> (last access: 8 December 2015), 2013.
- Brunel, E., Comte, F., and Lacour, C.: Minimax estimation of the conditional cumulative distribution function, *Sankhya Ser. A*, 72, 293–330, 2010.
- Cardil, A., Salis, M., Spano, D., Delogu, G., and Molina Terrén, D.: Large wildland fires and extreme temperatures in Sardinia (Italy), *iForest-Biogeosciences and Forestry*, 7, 162–169, 2014.
- Cortez, P. and Morais, A. D. J. R.: A data mining approach to predict forest fires using meteorological data, in: Proceedings of the 13th Portuguese Conference on Artificial Intelligence, 3–7 December 2007, Guimarães, Portugal, 512–523, 2007.
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, 2011.
- Delignette-Muller, M. L. and Dutang, C.: fitdistrplus: An R Package for Fitting Distributions, *J. Stat. Softw.*, 64, 1–34, <http://www.jstatsoft.org/v64/i04/>, 2015.
- Dimitrakopoulos, A. P., Vlahou, M., Anagnostopoulou, C. G., and Mitsopoulos, I.: Impact of drought on wildland fires in Greece: implications of climatic change?, *Climatic Change*, 109, 331–347, 2011.
- Drobinski, P., Ducrocq, V., Alpert, P., Anagnostou, E., Béranger, K., Borga, M., Braud, I., Chanzy, A., Davolio, S., Delrieu, G., Estournel, C., Filali Boubrahmi, N., Font, J., Grubisic, V., Gualdi, S., Homar, V., Ivancan-Picek, B., Kottmeier, C., Kotroni, V., Lagouvardos, K., Lionello, P., Llasat, M. C., Ludwig, W., Lutoff, C., Mariotti, A., Richard, E., Romero, R., Rotunno, R., Roussot, O., Ruin, I., Somot, S., Taupier-Letage, I., Tintore, J., Uijlenhoet, R., and Wernli, H.: HyMeX, a 10-year multidisciplinary program on the Mediterranean water cycle, *B. Am. Meteorol. Soc.*, 95, 1063–1082, 2014.
- European Commission: Forest Fires in Europe 2009, Office for Official Publication of the European Communities, Luxembourg, EUR 24502 EN, 81, 2010.
- Fox, J. and Weisberg, S.: An R Companion to Applied Regression, Sage, Thousand Oaks CA, second edn., <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>, 2011.
- Ganteaume, A., Camia, A., Jappiot, M., San-Miguel-Ayanz, J., Long-Fournel, M., and Lampin, C.: A review of the main driving factors of forest fire ignition over Europe, *Environ. Manage.*, 51, 651–662, 2013.
- Giglio, L., Randerson, J. T., van der Werf, G. R., Kasibhatla, P. S., Collatz, G. J., Morton, D. C., and DeFries, R. S.: Assessing variability and long-term trends in burned area by merging multiple satellite fire products, *Biogeosciences*, 7, 1171–1186, doi:10.5194/bg-7-1171-2010, 2010.
- Haines, D. A., Main, W. A., Frost, J. S., and Simard, A. J.: Fire-danger rating and wildfire occurrence in the northeastern United States, *Forest Sci.*, 29, 679–696, 1983.
- Hastie, T., Tibshirani, R., and Friedman, J.: The elements of statistical learning, vol. 2, Springer, New York, USA, 2009.
- Hernandez, C., Drobinski, P., and Turquety, S.: How much does weather control fire size and intensity in the Mediterranean region?, *Ann. Geophys.*, 33, 931–939, doi:10.5194/angeo-33-931-2015, 2015a.
- Hernandez, C., Drobinski, P., Turquety, S., and Dupuy, J.-L.: Size of wildfires in the Euro-Mediterranean region: observations and theoretical analysis, *Nat. Hazards Earth Syst. Sci.*, 15, 1331–1341, doi:10.5194/nhess-15-1331-2015, 2015b.
- Leisch, F. and Tibshirani, R.: bootstrap: Functions for the Book “An Introduction to the Bootstrap”, r package version 2014.4, available at: <http://CRAN.R-project.org/package=bootstrap> (last access: 8 December 2015), 2014.
- Linn, R., Reisner, J., Colman, J. J., and Winterkamp, J.: Studying wildfire behavior using FIRETEC, *Int. J. Wildland Fire*, 11, 233–246, 2002.
- Moriondo, M., Good, P., Durao, R., Bindi, M., Giannakopoulos, C., and Corte-Real, J.: Potential impact of climate change on fire risk in the Mediterranean area, *Clim. Res.*, 31, 85–95, 2006.
- Pereira, M. G., Trigo, R. M., da Camara, C. C., Pereira, J. M., and Leite, S. M.: Synoptic patterns associated with large summer forest fires in Portugal, *Agr. Forest Meteorol.*, 129, 11–25, 2005.
- Preisler, H. K., Brillinger, D. R., Burgan, R. E., and Benoit, J.: Probability based models for estimation of wildfire risk, *Int. J. Wildland Fire*, 13, 133–142, 2004.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, available at: <http://www.R-project.org/> (last access: 8 December 2015), 2013.
- Rothermel, R. C.: A mathematical model for predicting fire spread in wildland fuels, Res. Pap. INT-115. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Publication of U.S. Government, 40 pp., 1972.
- Russo, L., Russo, P., Vakalisc, D., and Siettos, C.: Detecting Weak Points of Wildland Fire Spread: A Cellular Automata Model Risk Assessment Simulation Approach, *Chem. Eng.-New York*, 36, 253–258, 2014.
- Sakr, G. E., Elhadj, I. H., and Mitri, G.: Efficient forest fire occurrence prediction for developing countries using two weather parameters, *Eng. Appl. Artif. Intel.*, 24, 888–894, 2011.
- San-Miguel-Ayanz, J., Moreno, J. M., and Camia, A.: Analysis of large fires in European Mediterranean landscapes: lessons learned and perspectives, *Forest Ecol. Manage.*, 294, 11–22, 2013.
- Schoenberg, F. P., Peng, R., and Woods, J.: On the distribution of wildfire sizes, *Environmetrics*, 14, 583–592, 2003.
- Sedano, F., Kempeneers, P., San Miguel, J., Strobl, P., and Vogt, P.: Towards a pan-European burnt scar mapping methodology based on single date medium resolution optical remote sensing data, *Int. J. Appl. Earth Obs.*, 20, 52–59, 2013.
- Sousa, P. M., Trigo, R. M., Pereira, M. G., Bedia, J., and Gutiérrez, J. M.: Different approaches to model future burnt area in the Iberian Peninsula, *Agr. Forest Meteorol.*, 202, 11–25, 2015.

- Stéfanon, M., Drobinski, P., D'Andrea, F., and de Noblet-Ducoudré, N.: Effects of Interactive Vegetation Phenology on the 2003 Summer Heat Waves, *J. Geophys. Res.*, 117, D24103, doi:10.1029/2012JD018187, 2012a.
- Stéfanon, M., D'Andrea, F., and Drobinski, P.: Heatwave classification over Europe and the Mediterranean region, *Env. Res. Lett.*, 7, 014023, doi:10.1088/1748-9326/7/1/014023, 2012b.
- Trigo, R. M., Pereira, J., Pereira, M. G., Mota, B., Calado, T. J., Dacamara, C. C., and Santo, F. E.: Atmospheric conditions associated with the exceptional fire season of 2003 in Portugal, *Int. J. Climatol.*, 26, 1741–1757, 2006.
- Turquety, S., Menut, L., Bessagnet, B., Anav, A., Viovy, N., Maignan, F., and Wooster, M.: APIFLAME v1.0: high-resolution fire emission model and application to the Euro-Mediterranean region, *Geosci. Model Dev.*, 7, 587–612, doi:10.5194/gmd-7-587-2014, 2014.
- Van Wagner, C. E.: Structure of the Canadian forest fire weather index, *Can. For. Serv. Publ.*, 1333, 44 pp., 1974.
- Van Wagner, C. E.: Development and structure of the Canadian forest fire weather index system, Canadian Forestry Service, Headquarters, Ottawa, Canada, Forestry Technical Report, vol. 35, 35 pp., 1987.
- Van Wagner, C. E. and Pickett, T. L.: Equations and FORTRAN program for the Canadian forest fire weather index system, Canadian Forestry Service, Petawawa National Forestry Institute, Chalk River, Ontario, Canada, Forestry Technical Report, vol. 33, 18 pp., 1985.
- Vautard, R., Yiou, P., D'andrea, F., De Noblet, N., Viovy, N., Casou, C., Polcher, J., Ciais, P., Kageyama, M., and Fan, Y.: Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit, *Geophys. Res. Lett.*, 34, L07711, doi:10.1029/2006GL028001, 2007.
- Zampieri, M., D'Andrea, F., Vautard, R., Ciais, P., de Noblet-Ducoudré, N., and Yiou, P.: Hot European summers and the role of soil moisture in the propagation of Mediterranean drought, *J. Climate*, 22, 4747–4758, 2009.