Predicting above normal wildfire activity in southern Europe as a function of meteorological drought

L Gudmundsson1, F C Rego2, M Rocha2 and S I Seneviratne1

1 Institute for Atmospheric and Climate Science, ETH Zurich, Universitaetstrasse 16, 8092 Zurich, Switzerland
2 Centro de Ecologia Aplicada Baeta Neves, (CEABN/INBIO), Instituto Superior de Agronomia, Tapada da Ajuda, Lisboa, Portugal

E-mail: lukas.gudmundsson@env.ethz.ch

Received 11 April 2014, revised 13 July 2014
Accepted for publication 21 July 2014
Published 15 August 2014

Abstract
Wildfires are a recurrent feature of ecosystems in southern Europe, regularly causing large ecological and socio-economic damages. For efficient management of this hazard, long lead time forecasts could be valuable tools. Using logistic regression, we show that the probability of above normal summer wildfire activity in the 1985–2010 time period can be forecasted as a function of meteorological drought with significant predictability ($p < 0.05$) several months in advance. The results show that long lead time forecasts of this natural hazard are feasible in southern Europe, which could potentially aid decision-makers in the design of strategies for forest management.

Keywords: wildfire, drought, climate impact, forecast

1. Introduction

Although wildfires are recurrent natural phenomena and an important feature of many ecosystems (Bowman 2009, Moreira et al 2011), their occurrence is commonly perceived as a threat for human activities (Hardy 2005, Milad et al 2011, Thompson and Calkin 2011). Wildfires are complex phenomena that depend on a multitude of factors (Westerling et al 2006, Xiao and Zhuang 2007, Cruz and Alexander 2010, Krawchuk and Moritz 2010, Macias Fauria et al 2011, Moreira et al 2011). On the one hand, the availability of burnable biomass is a precondition for wildfire activity. On the other hand fuels will only ignite if their moisture content is sufficiently low. While fuel accumulation depends on both natural factors as well as on human interventions (Cruz and Alexander 2010, Bowman 2011, Moreira et al 2011), fuel moisture is governed by meteorological drivers (Westerling et al 2006, Macias Fauria et al 2011). Wildfire management therefore depends not only on knowledge of the amount of burnable biomass, but also on a robust quantification of changes in wildfire risk related to meteorological conditions (Thompson and Calkin 2011). This dependency of fire activity on meteorological conditions is often quantified using so called fire weather indices (Van Wagner 1987, Carvalho et al 2008, Ganatsas et al 2011) but also using statistical techniques that directly relate observed weather conditions to wildfire activity (Thompson and Calkin 2011).

The dependence of wildfires on meteorological conditions can be used to derive long lead time forecasts of fire activity. Such forecasts are either based on fire weather indices derived from weather forecast models (Roads et al 2010), or on statistical techniques that relate present meteorological conditions to fire activity in the future (Westerling et al 2002, Preisler et al 2004, Preisler and Westerling 2007, Preisler et al 2008).
Roads et al. 2010), their applicability in other fire prone regions is less clear. For example in southern Europe, investigations relating meteorological conditions to climate drivers have usually focused on relatively small areas (Pausas, 2004, Ganatsas et al. 2011, Pausas and Fernández-Muñoz 2012, Pausas and Paula 2012, de Vicente and Cre spo 2012, Papadopoulos et al. 2013, Turco, et al. 2013). While these investigations provide valuable tools for local management, the potential of large-scale and long lead time predictability of wildfire activity in southern Europe has not yet been demonstrated with scientific rigour.

To approach the question of predictability of wildfires, we rely on results showing that large scale droughts significantly affect continental scale wildfire activity (Xiao and Zhuang 2007). This, combined with the finding that drought impacts on Mediterranean fire regimes have recently increased (Pausas and Fernández-Muñoz 2012), suggests that wildfire activity in Southern Europe can be predicted using information on drought. In the following we present a pilot study demonstrating the potential for forecasting above normal wildfire activity as a function of meteorological drought several months in advance.

2. Data sources and processing

2.1. Meteorological drought in southern Europe

Meteorological drought for the 1985–2010 time window was quantified using the well established standardized precipitation index (SPI, McKee et al. 1993, Guttman 1999)), which has a long and proven record for drought characterization (e.g. Hayes et al. 2010, Mueller and Seneviratne 2012). SPI thus constitutes a comprehensive indicator for fuel moisture and fuel drying. For this study SPI is derived from the E-OBS dataset (version 9, Haylock et al. 2008), which provides reliable estimates of precipitation on a 0.5° grid in Europe. In a first step a backward moving average with accumulation time $\tau$ is applied to the time series of monthly precipitation. The accumulation time $\tau$ is also referred to as time scale in the literature. Here we consider time scales ranging from 1 to 12 months, where $\tau = 1$ corresponds to a precipitation anomaly accumulated over 1 month, and $\tau = 12$ corresponds to anomalies in precipitation accumulated over 12 months. In a second step the resulting series are standardized for each calendar month separately. For this, a gamma distribution is fitted to the data using maximum likelihood estimation and this distribution is subsequently used to transform the data to the standard normal distribution. The resulting SPI time series thus corresponds to accumulated precipitation anomalies over the past $\tau$ months expressed in units of the standard normal distribution. Positive values indicate above normal precipitation amounts, negative values indicate drought conditions. To emphasize large-scale phenomena and to reduce the effects of spatial heterogeneity, the gridded SPI series were spatially averaged to derive time series representing regional anomalies in southern Europe and two sub regions (the Iberian Peninsula and a region representing South Italy & Greece, figure 1).

![Figure 1](image_url) Mean wildfire occurrence in southern Europe for the 1985–2010 period. (a) Fraction of area burned in NUTS3 units in the region under investigation, including the Iberian Peninsula (red box) and the South Italy & Greece (blue box) sub-regions. The lower panels (b,c,d) show the mean observed total area burned (TAB) for each month.

2.2. Area burned by wildfires in southern Europe

The employed wildfire data (1985–2010) stem from the European Fire Database (EDF, European Commission 2009, 2011) that has been collected within the framework of the European Forest Fire Information System (EFFIS) and is held by the Joint Research Centre. The EFD contains monthly information on the Total Area Burned (TAB) at the level of NUTS3 units (Nomenclature of Units for Territorial Statistics), which correspond to local administrative units.

The monthly TAB series of the NUTS3 units were first aggregated to time series representing the total area burned in all of southern Europe as well as in the Iberian Peninsula and...
the South Italy & Greece sub regions (figure 1). These regional series are dominated by a pronounced seasonal cycle, which characterizes wildfire activity in southern Europe (figure 1). As wildfire activity is most pronounced throughout late summer and early autumn (figures 1(b)–(d)) we limit the analysis to July, August and September. To focus on departures of this mean pattern the regional series were converted to monthly anomaly series, by subtracting the mean annual cycle.

Figure 2(a) shows the relation between TAB anomalies and SPI$_1$ in southern Europe. Although a linear regression analysis indicates a significant dependence ($p < 0.01$), the resulting model has only low explanatory power ($R^2 = 0.13$). The large scatter in figure 2(a) thus shows, that many other factors such as fuel availability, fire suppression, or the stochastic nature of ignition also play a major role for determining regional anomalies in TAB. However, the slope of the regression is significantly smaller than zero ($p < 0.01$) showing that SPI$_1$ can indeed be used to derive information on wildfire activity in southern Europe. In addition, the regression line separates positive and negative anomalies quite elegantly. Hence the results not only suggest that SPI may not be sufficient to explain the magnitude of anomalies in regional TAB series, but also highlights that information on meteorological drought is sufficient to separate months with below- from months with above normal TAB. In the remainder of this study we focus on this property and therefore convert the anomaly series to binary series indicating above normal (positive) and below normal (zero and negative) wild fire activity in the respective regions.

3. Modelling above normal fire activity

3.1. Model setup

In the following we aim at modelling the monthly probability of above normal areas burned by wildfires as a function of meteorological drought, characterized by the SPI. For this we rely on logistic regression, an approach that is commonly used for modelling wild-fire risk (see e.g. Thompson and Calkin (2011) and references therein). For the reconstruction, $\pi$, the probability of above normal wildfire activity in July, August and September is modelled as function of SPI, such that:

$$\ln \left( \frac{\pi}{1 - \pi} \right) = a + b \times \text{SPI}_r,$$

where the left-hand side is known as the logit transformation. The model parameters $a$ and $b$ are estimated using standard regression techniques within the framework of Generalized Linear Models (GLM) (Venables and Ripley 2002, Zuur et al 2009). SPI accumulation times, $\tau$, ranging from one to 12 months are assessed. Note that only one model is estimated for each (sub) region (figure 1) that is applicable for all three months under investigation. Figure 2(b) shows a logistic regression example, highlighting the models ability to predict the probability of above and below-normal wild fire activity as a function of SPI. This illustrates that the likelihood of above normal wildfire activity increases as meteorological droughts get more extreme and thus highlights the importance of fuel dryness for wildfire forecasting.

The model setup for forecasting the probability of above normal area burned by wildfires is identical, except that SPI$_{\text{res}}$ of the preceeding one, two, or more months is used. For

![Figure 2](image-url)
instance, to predict the probability of above normal fire occurrence in July with one month lead time SPI\textsubscript{τ} for June is used. For the same prediction with a two month lead time SPI\textsubscript{τ} of May is used.

3.2. Model selection and validation

An important property of probabilistic models is that they do not predict the occurrence of events (here above normal area burned by wildfires) directly but only the chance that an event occurs. Consequently users are faced with the challenge of deciding at which predicted probability warnings are issued, hereafter referred to as the threshold probability. The nature of probabilistic predictions also implies that the issued warnings capture only a fraction of the observed events, and that some events occur without warning. This fraction of events for which a warning is issued is referred to as ‘hit rate’ and increases with decreasing threshold probability. On the other hand, decreasing the threshold probability also implies that the ‘false alarm rate’ i.e. the fraction of issued warnings for which no events occur increases. Depending on the objective, the acceptable hit- and false alarm rates can vary considerably and consequently there cannot be a general recommendation for the threshold probability. To get a comprehensive overview on wildfire predictability in a decision making context we therefore evaluate the hit rate and the false alarm rate for a number of threshold probabilities. The resulting diagram which plots the hit rate as a function of the false alarm rate is referred to as the Relative Operating Characteristics or Receiver Operating Characteristics (ROC) curve (Rego and Machado 1993, Mason and Graham 2002, Wilks 2011) (see figure 2(c)). If the ROC curve is above the identity line the probabilistic forecast is better than random guessing. This can formally be assessed by estimating the area under the ROC curve (later referred to as ROC-Area) which takes values larger than 0.5 in this case and equals 1.0 for the perfect forecast. Interestingly it can be shown that the ROC-Area ‘defines the probability that the forecast probability issued for when an event occurs is greater than for when there is no event’ (Mason and Graham 2002). In this study ROC curves are estimated at maximum resolution, i.e. that each predicted probability is used once as threshold probability. This allows for a straightforward estimation of ROC-Area (Mason and Graham 2002). In addition, the significance of ROC-Area is tested using a Mann-Whitney U-test (Mason and Graham 2002). The ROC curve can further be used to find the threshold probability, \( \theta \), at which the difference between the hit rate and the false alarm rate is maximal (see figure 2(c)). This ‘optimal’ threshold maximizes the hit rate while minimizing the false alarm rate of the model.

To evaluate predictability independently from the data used for model identification (fitting) we derive the ROC statistics using leave-one-out cross-validation (LOO-CV) (e.g. Hastie et al. 2001). For LOO-CV one entry is removed from the data and the statistical model is fitted to the remaining data. The resulting model is subsequently used to predict the value of the entry that has been left out. This procedure is repeated until each entry has been left out once. The ROC statistics are computed from the resulting set of LOO-CV predictions, which is independent from the data.

![Figure 3](https://example.com/figure3.png)
used for fitting the model and consequently allows for an unbiased assessment of model performance.

4. Results and discussion

Figure 3 shows the ROC-Area for models predicting the probability of above normal burned area as a function of SPI with different accumulation times and lead times up to 12 months in the three regions under investigation. The best significant models for each region and each lead time are displayed in figure 4 and the corresponding model parameters are shown in table 1. The ROC curves of the selected models, including additional information on model performance are shown in figure 5.

Overall there are significant links between above normal area burned by wildfire and SPI in southern Europe, indicating that information on drought conditions is suitable predictor for wildfire activity. The ROC-Area for reconstruction (lead time: 0 months) is significant for most SPI accumulation times ($r$), indicating a robust relation between

| Lead time | $a$         | $b$         | $r$         |
|-----------|-------------|-------------|-------------|
| Southern Europe |             |             |             |
| 0         | $-0.24(\pm0.26)$ | $-2.43(\pm0.69)$ | 2           |
| 1         | $-0.18(\pm0.25)$ | $-2.03(\pm0.69)$ | 1           |
| 2         | $-0.18(\pm0.25)$ | $-2.27(\pm0.78)$ | 1           |
| 3         | $-0.18(\pm0.25)$ | $-1.65(\pm0.65)$ | 2           |
| 4         | $-0.64(\pm0.27)$ | $-2.57(\pm0.73)$ | 5           |
| 5         | $-0.62(\pm0.26)$ | $-2.28(\pm0.70)$ | 4           |
| 6         | $-0.57(\pm0.25)$ | $-1.68(\pm0.65)$ | 3           |
| 11        | $-1.03(\pm0.28)$ | $+1.22(\pm0.60)$ | 1           |

Table 1. Parameters of equation (1) for the best significant models (see figures 3 and 4). Values in brackets are the standard errors of the parameter estimates.
above normal burned area and meteorological drought. The best performing model is found for $\tau = 2$. For forecasts with one month lead time, the ROC-Area decreases slightly and the best performing model is found for $\tau = 1$. For two and more months lead time no significant model is found.

For the Iberian Peninsula, SPI with longer accumulation times have larger effects on the probability of above normal area burned by wildfires. For reconstruction, the best significant model if found for $\tau = 5$, emphasizing the importance of longlasting drought conditions for wildfire dynamics in this region. This is also reflected in the long SPI accumulation time ($\tau = 4$) for forecasts with one month lead time. Further, the dependence on prolonged dry episodes is a likely cause for the significant predictability of above normal wildfires in this region up to two months in advance.

In South Italy & Greece the probability of above normal wildfire activity is only significantly related to SPI values with short accumulation times. For reconstruction, the best significant model is found for $\tau = 2$ and models based on longer SPI accumulations are not significant. Also for this region one month ahead predictions of the probability of above normal burned areas have significant predictability ($\tau = 1$). Interestingly the logistic regression models for two months lead time in this region exhibit a significant inverse response. In other words, the probability of above normal area burned by wildfires increases if SPI indicates wet conditions two months before the event. This result suggests that a surplus of water in the growing season can cause a larger buildup of burnable biomass, which in turn is a precondition for a large spatial extent of wildfires. Finally the model for eleven months lead time suggests a significant inverse response. This might be related to the fact that low wildfire activity in wet years can lead to a larger buildup of fuel, triggering above normal burned areas in the following year. However, as the interpretation of this forecast with an extremely long lead time is not straightforward we limit further discussions to lead times not longer than two months. While the latter findings are not directly linked to the drought related process of fuel drying, they clearly show that significant predictability of regional wildfire activity can be yielded from antecedent meteorological conditions.

Figure 5. ROC curves of the best significant logistic regression models, predicting above normal area burned in each region and at different lead time (selected in figure 3). The individual graphical elements are described in figure 2(c).
The significant relation of above normal area burned by wildfire and SPI can consequently be used to predict the temporal evolution of the wildfire activity in the near future. Figure 6 exemplifies this, showing predictions of above normal wildfire probabilities in southern Europe with a lead time of one month. Although the temporal evolution of the predicted probabilities and the occurrence of above normal wildfire activity do not match perfectly, the ROC statistics (figure 3(b) and figure 5) clearly show that decisions based on such a forecast are significantly superior to random guessing.

5. Summary and conclusions

The presented analysis shows that the probability of above normal wildfire activity in large geo-climatic regions is significantly related to meteorological drought, despite the fact that other factors such as fuel availability or fire suppression also play an important role for wildfire dynamics. We have shown that this relation can not only be used for reconstructing above normal wildfire occurrence in southern Europe, but also to forecast the probability of such occurrences up to two months in advance. Nevertheless it is important to note that these results can only serve as a proof of principle, demonstrating the potential for forecasting wildfire occurrences in southern Europe. Limitations are e.g. the sub-continental perspective, preventing local interpretation and the fact that the fuel availability was not taken into account. Especially the pronounced differences between the Iberian Peninsula and the South Italy & Greece sub regions highlight the fact that different processes contribute to the predictability of wildfires at different locations. Therefore we anticipate that more sophisticated approaches, explicitly considering e.g. the spatial distribution and the buildup of burnable biomass may further improve the seasonal predictability of area burned by wildfires in the considered region. Despite these limitations, the results suggest that predicting the probability of above normal wildfire activity in Southern Europe on the basis of preceding drought conditions is feasible. Such predictions may in turn support regional authorities concerned with wildfire management, in their decision making process.

Acknowledgments

This research contributes to the European Union (FP7) funded project DROUGHT-R&SPI (www.eu-drought.org/, contract no. 282769). We thank the Joint Research Centre of the European Commission, especially Dr Andrea Camia and Dr Jesus San-Miguel, for providing the fire data for the DROUGHT-R&SPI project. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (www.ecad.eu).

References

Bowman D M J S et al 2011 J. Biogeogr. 38 2223–36
Bowman D M J S et al 2009 Science 324 481–4
Carvalho A, Flannigan M D, Logan K, Miranda A I and Borrego C 2008 Int. J. Wildland Fire 17 328–38
Cruz M G and Alexander M E 2010 Int. J. Wildland Fire 19 377–98
de Vicente J and Crespo F 2012 Int. J. Wildland Fire 21 1030–41
European Commission 2009 Forest fires in Europe 2008 JRC Scientific and Technical Report 9 Joint Research Centre
European Commission 2011 Forest fires in Europe 2010 JRC Scientific and Technical Report 11 Joint Research Centre
GANATAS P, Antonis M and Mariantti T 2011 Agric. Forest Meteorol. 151 241–50
Guttman N B 1999 J. Am. Water Resour. Assoc. 35 311–22
Hardy C C 2005 Forest Ecol. Manage. 211 73–82
Hastie T, Tibshirani R and Friedman J H 2001 The Elements of Statistical Learning (Berlin: Springer)
Hayes M, Svoboda M, Wall N and Wihalm M 2010 Bull. Am. Meteorol. Soc. 92 485–8
Haylock M R, Hofstra N, Klein Tank A M G, Klok E J, Jones P D and New M 2008 J. Geophys. Res. 113 D20119
Keywood M, Kanakidou M, Stohl A, Dentener F, Grassi G, Meyer C P, Torseth K, Edwards D, Thompson A M, Lohmann U and Burrows J 2013 Crit. Rev. Environ. Sci. Technol. 43 40–83
Krawchuk M A and Moritz M A 2010 Ecology 92 121–32
Macias Fauria M, Michaletz S T and Johnson E A 2011 Wiley Interdiscip. Rev.: Clim. Change 2 99–112
Mason S J and Graham N E 2002 J. R. Meteorol. Soc. 128 2145–66
McKee T, Doesken N and Kleist J 1993 5th Conf. on Applied Climatology (Anaheim CA) pp 179–84
Milad M, Schaich H, Bürgi M and Konold W 2011 Forest Ecol. Manage. 261 829–43
Moreira F et al 2011 J. Environ. Manage. 92 389–402
Mueller B and Seneviratne S I 2012 Proc. Natl. Acad. Sci. 109 12398–403
Papadopoulos A, Paschalidou A, Kassomenos P and McGregor G 2013 Theor. Appl. Climatol. 112 113–26
Pausas J 2004 Clim. Change 63 337–50
Pausas J G and Fernández-Muñoz S 2012 Clim. Change 11 215–26
Pausas J G and Paula S 2012 Glob. Ecol. Biogeogr. 21 1074–82
Preisler H, Brillinger D, Burgan R and Benoit J 2004 Int. J. Wildland Fire 13 133–42
Preisler H K, Chen S, Fujioka F, Benoit J W and Westerling A L 2008 Int. J. Wildland Fire 17 305–16
Preisler H K and Westerling A L 2007 J. Appl. Meteorol. Climatol. 46 1020–30
Rego F C and Machado C A 1993 12th conf. on Fire and Forest Meteorology 544–51
Roads J, Tripp P, Juang H, Wang J, Fujioka F and Chen S 2010 Int. J. Wildland Fire 19 399–414
Thompson M P and Calkin D E 2011 *J. Environ. Manag.* 92 1895–909

Turco M, Llasat M, Hardenberg J and Provenzale A 2013 *Clim. Change* 116 665–78

Turco M, Llasat M C, Tudela A, Castro X and Provenzale A 2013 *Nat. Hazards Earth Syst. Sci.* 13 649–52

Van Wagner C 1987 Development and structure of the canadian forest fire weather index system *Forestry Technical Report 35 Canadian Forestry Service Ottawa*

Venables W N and Ripley B 2002 *Modern Applied Statistics with S* (Berlin: Springer)

Westerling A L, Gershunov A, Cayan D R and Barnett T P 2002 *Int. J. Wildland Fire* 11 257–66

Westerling A L, Hidalgo H G, Cayan D R and Swetnam T W 2006 *Science* 313 940–3

Wilks D S 2011 Statistical Methods in the Atmospheric Sciences 3rd edn *International Geophysics Series* vol 100 (New York: Academic Press)

Xiao J and Zhuang Q 2007 *Environ. Res. Lett.* 2 044003

Zuur A, Ieno E, Walker N, Saveliev A and Smith G 2009 Mixed effects models and extensions in ecology with R *Statistics for Biology and Health* (New York: Springer)