

Probabilistic prediction of daily fire occurrence in the Mediterranean with readily available spatio-temporal data

Panagiota Papakosta,
Daniel Straub

The prediction of wildfire occurrence is an important component of fire management. We have developed probabilistic daily fire prediction models for a Mediterranean region of Europe (Cyprus) at the mesoscale, based on Poisson regression. The models use only readily available spatio-temporal data, which enables their use in an operational setting. Influencing factors included in the models are weather conditions, land cover and human presence. We found that the influence of weather conditions on fire danger in the studied area can be expressed through the FWI component of the Canadian Forest Fire Weather Index System. However, the prediction ability of FWI alone was limited. A model that additionally includes land cover types, population density and road density was found to provide significantly improved predictions. We validated the probabilistic prediction provided by the model with a test data set and illustrate it with maps for selected days.

Keywords: Fire Occurrence, Prediction, Canadian Forest Fire Weather Index, Poisson Regression

Introduction

Predicting the occurrence of wildfire incidents is an important component of fire management. Due to the uncertainties in the influencing factors, as well as to random effects in the fire process, such a prediction must necessarily be probabilistic. Various probabilistic models are proposed in the literature, including Poisson models (Cunningham & Martell 1973, Mandallaz & Ye 1997, Syphard et al. 2008), logistic regression (Vasconcelos et al. 2001, Preisler et al. 2004, Kalabokidis et al. 2007, Syphard et al. 2008, Chuvieco et al. 2009, Arndt et al. 2013), multiple regression (Sebastián-López et al. 2008, Oliveira et al. 2012), neural networks (Vasconcelos et al. 2001, Vasylakos et al. 2009) and Bayesian networks (Dlamini 2009). Recently, machine learning algorithms have been found to be well suited to modeling and predicting fire occurrences, due to their greater flexibility compared to classical regression analysis. In particular, the Maxent (maximum entropy) algorithm (Parisien & Moritz 2009) and methods based on decision tree learning,

such as the random forest algorithm (Masada et al. 2012, Oliveira et al. 2012), have been applied. The choice of the appropriate model depends on the influencing factors selected and their spatial and temporal resolution, as well as the purpose of the model prediction.

Past probabilistic models of fire occurrence use weather factors, anthropogenic factors or combinations thereof as explanatory variables (Plucinski 2012). The effect of climatic factors is often represented by components of the Canadian Forest Fire Weather Index System (CFFWIS - Martell et al. 1987, Wotton et al. 2003, Ager et al. 2014, Venäläinen et al. 2014). In these studies, the temporal resolution is daily and the spatial resolution is regional, ranging from 1 to several km² (except Venäläinen et al. 2014). Although CFFWIS was originally developed for Canadian climates and vegetation, it is commonly used for predicting fire occurrence in the Mediterranean (Viegas et al. 1999, Camia & Amatulli 2009, Venäläinen et al. 2014). Still, this necessitates that the CFFWIS indicators, in order to categorize

fire danger level (e.g., low, moderate, high), be adjusted to the specifics of the Mediterranean climates (Moriondo et al. 2006, Giannakopoulos et al. 2011, Dimitrakopoulos et al. 2011, Venäläinen et al. 2014).

Various studies have looked into the combined effect of weather and anthropogenic factors (Cardille et al. 2001, Pew & Larsen 2001, Amatulli et al. 2006, Kalabokidis et al. 2007, Syphard et al. 2008, Vilar et al. 2010, Padilla & Vega-García 2011, Miranda et al. 2012, Oliveira et al. 2012, Martínez-Fernández et al. 2013). The temporal resolution of these studies is seasonal or yearly, and thus the weather factors include mean, minimum and maximum temperatures, as well as cumulative precipitation. Common explanatory variables representing anthropogenic influences are population density, land use, distances to human-built infrastructures (Catry et al. 2010). However, many additional variables were studied, e.g., distance to campground (Chou & Chase 1993), holidays (Mandallaz & Ye 1997), ownership of housing (Cardille et al. 2001), proximity to urban areas and roads (Romero-Calcerrada et al. 2008, Ager et al. 2014), unemployment rate (Oliveira et al. 2012), rural exodus by means of population decrease (Martínez-Fernández et al. 2013), and hiking trail density (Arndt et al. 2013). The spatial resolution in these studies varies from cellular (1 km² grid – Pew & Larsen 2001) to regional.

The aim of this paper is the development of a daily probabilistic model for fire occurrences in Mediterranean climates, which includes both natural and anthropogenic factors. Such a daily predictive model with fine spatial resolution can eventually be helpful as a fire management tool. Here,

□ Engineering Risk Analysis Group, Technische Universität München, Theresienstr. 90, D-80333 München (Germany)

@ Panagiota Papakosta (patty.papakosta@gmail.com)

Received: Apr 24, 2015 - Accepted: Jul 07, 2016

Citation: Papakosta P, Straub D (2016). Probabilistic prediction of daily fire occurrence in the Mediterranean with readily available spatio-temporal data. *iForest* 10: 32-40. - doi: [10.3832/ifor1686-009](https://doi.org/10.3832/ifor1686-009) [online 2016-10-06]

Communicated by: Davide Ascoli

we show that the fire risk prediction at the mesoscale can be improved with readily available data on weather and anthropogenic factors, combined with a sound probabilistic model.

In the proposed model, the potential influence of weather conditions is represented by the Canadian Forest Fire Weather Index System (CFFWIS - Van Wagner 1987). The anthropogenic influence is represented through spatial variables such as land cover type and road density, which was found to be a relevant indicator of fire occurrence in Amatulli et al. (2006), Sypard et al. (2008) and Oliveira et al. (2012). The model is based on Poisson regression. Its results are daily maps of fire occurrence rates with 1 km² spatial resolution. The use of readily available data make the model easy to integrate into existing fire prediction systems. This can improve fire occurrence predictions due to the high spatio-temporal resolution (daily, 1 km²) of the proposed model and the incorporation of both weather and anthropogenic factors.

The model is applied to the island of Cyprus, where the model parameters are calibrated from observed fire events. Cyprus is part of the Eastern Mediterranean region, which is drier and warmer than the more commonly studied areas of Spain, Southern France and Northern Italy. The data is separated into a learning set and a validation set, which allows to investigate the predictive power of the proposed model. It is found that the best prediction can be achieved by combining the natural and anthropogenic factors. The main factors describing anthropogenic influences are found to be land cover, population density and road density.

Methodology

Canadian Forest Fire Weather Index System

The Canadian Forest Fire Weather Index System (CFFWIS - Van Wagner 1987) was first introduced across Canada in 1971 and is meanwhile adapted in several national fire danger estimation systems (Taylor & Alexander 2006, Camia & Amatulli 2009).

The input parameters required by CFFWIS are daily values of easily observed weather parameters (dry bulb temperature, wind speed, relative humidity and precipitation). CFFWIS consists of six components: three fuel moisture codes (FFMC: Fine Fuel Moisture Code; DMC: Duff Moisture Code; DC: Drought Code) and three fire behavior indices (ISI: Initial Spread Rate; BUI: Build Up Index; FWI: Fire Weather Index). A detailed description of CFFWIS is available in Van Wagner (1987) and Lawson & Armitage (2008).

Probabilistic model for predicting fire occurrence

Fig. 1 summarizes the proposed probabilistic model by means of a Bayesian Network (BN). In the BN, probabilistic dependence among the variables is represented graphically by means of arrows. This makes it convenient not only for graphical communication of the model but also for quantitative probabilistic modeling. For these reasons, BN are increasingly applied for risk assessment of natural hazards, e.g., for wildfire occurrence (Dlamini 2009), rockfall hazards (Straub 2005), avalanches (Grêt-Regamey & Straub 2006), tsunamis (Blaser et al. 2009) and earthquakes (Bayraktarli et al. 2005, Kuehn et al. 2011, Bensi et al. 2014). For a detailed introduction to BN, the reader is referred to Jensen & Nielsen (2007).

The BN in Fig. 1 models daily fire occurrence in a cell of 1 km², which is the spatial unit of this study. In the application presented in this paper, there is no difference between the BN model and the regression model. In fact, we use a regression approach to estimate the parameters of the BN model as explained later. However, when using the model for prediction, not all explanatory variables may be known with certainty. The BN allows modeling them as random variables, with a known distribution. As an example, the forecasted weather variables will be uncertain, which can be directly implemented in the BN.

In the presented study, data is available for all weather variables as well as all yellow variables. All these variables are con-

tinuous, with the exception of “land cover”, which has labeled states that are related to fuel type (e.g., forest, natural grasslands, olive groves, artificial surface, etc.). The orange variables are defined by the CFFWIS functions (see above). For given values of the weather variables, they are defined deterministically.

The fire occurrence rate λ , which is defined as the mean number of fires per day and km², is estimated from the data. In our model, it is a function of land cover, human population density, road density and FWI. The variable “fire occurrences” $N \in 0, 1, 2, \dots$ is the number of fires in one cell on one day. For a given daily fire occurrence rate λ , the number of fires follows a Poisson distribution, assuming independence among fire events for given occurrence rate. The conditional probability of observing n fires given λ is thus (eqn. 1):

$$Pr(N=n|\lambda) = \frac{(\lambda \alpha)^n}{n!} \exp(-\lambda \alpha)$$

where $n = 0, 1, 2, \dots, \lambda$ [Nr. Fires day⁻¹ km²] is the mean occurrence rate and $\alpha = 1$ km² is the area of the cell.

Observations of N are used to estimate λ based on eqn. 1, as described in the next section.

Poisson regression

The response variable is the number of fire occurrences N , which is a random variable described by the Poisson distribution with rate λ . This motivates the use of the generalized linear model of the Poisson regression for estimating λ (Mandallaz & Ye 1997). The rate λ is related to the explanatory variables $x = [x_1, \dots, x_k]$ by means of the link function (eqn. 2):

$$\log(\lambda) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k = \mathbf{x}^T \boldsymbol{\beta}$$

where $\boldsymbol{\beta} = [\beta_0, \dots, \beta_k]$ is the vector of regression coefficients. This link function ensures that λ is a non-negative real number. The mean occurrence rate is then given as (eqn. 3):

$$\lambda = \exp(\mathbf{x}^T \boldsymbol{\beta}) = \exp(\beta_0) \exp(\beta_1 x_1) \exp(\beta_2 x_2) \dots \exp(\beta_k x_k)$$

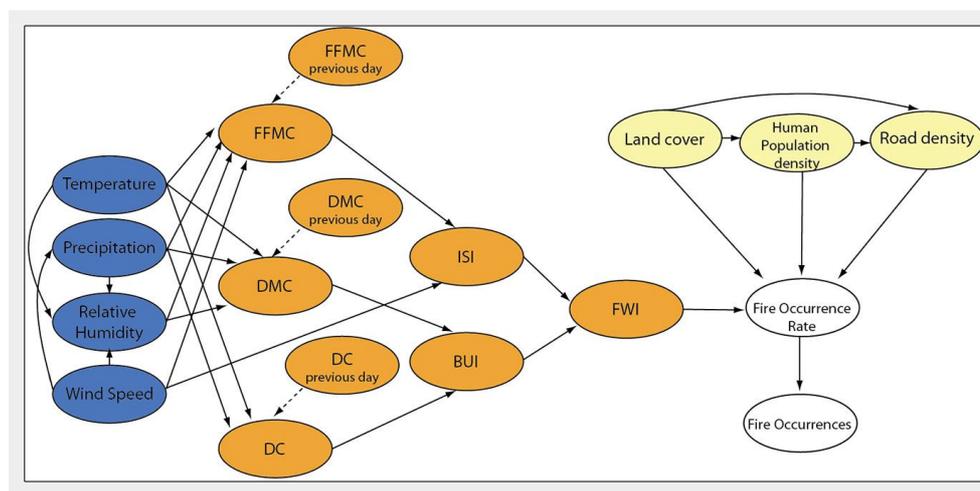


Fig. 1 - Bayesian Network for fire occurrence prediction. Blue nodes represent weather conditions; orange nodes are the components of the CFFWIS, which result in a FWI value; the variables in yellow represent the anthropogenic influence and the vegetation type; the variables in white are the predicted fire occurrence rate and the actual number of fires. The yellow variables change over space but are constant in time, whereas all other variables change both in time and space. Dashed arrows indicate a dependence on the value of the previous day.

Changing one of the explanatory variables from x_i to $x_i+\Delta x$, while keeping all other fixed, leads to a relative change in λ of (eqn. 4):

$$\left(\frac{\Delta \lambda}{\lambda}\right)_i = \left[\exp(\mathbf{x}^T \boldsymbol{\beta})\right]^{-1} \cdot \left[\exp(\beta_0)\exp(\beta_1 x_1) \dots \exp(\beta_i x_i + \Delta x) \dots \exp(\beta_k x_k) - \exp(\mathbf{x}^T \boldsymbol{\beta})\right] = \exp(\beta_i \Delta x) - 1$$

In the numerical investigations, several models are examined, which differ in the selection of the explanatory variables \mathbf{x} . These are selected from a set of variables describing land cover, human population density, road density and components of the CFFWIS. Land cover is a categorical variable; therefore, a separate binary variable x_i is defined for each of its categories. This variable takes value 1, if the land cover in this area belongs to this category, and value 0 otherwise.

Maximum likelihood estimation

Maximum likelihood estimation (MLE) is applied to determine the coefficients $\boldsymbol{\beta}$. For the Poisson regression model, the likelihood function follows from eqn. 1 as (eqn. 5):

$$L(\boldsymbol{\beta}|\mathbf{n}) = \prod_{i=1}^{m_d} \prod_{j=1}^{m_s} Pr(N_{ij} = n_{ij} | \lambda(\boldsymbol{\beta}, \mathbf{x}_{ij})) = \prod_{i=1}^{m_d} \prod_{j=1}^{m_s} \frac{\lambda(\boldsymbol{\beta}, \mathbf{x}_{ij})^{n_{ij}}}{n_{ij}!} \exp(-\lambda(\boldsymbol{\beta}, \mathbf{x}_{ij}))$$

where m_d is the number of days with observations and m_s is the number of spatial units with observations, n_{ij} is the number of fires observed on day i in the area j , \mathbf{x}_{ij} are

the values of the explanatory variables on day i in the area j .

The MLE is found as the value of $\boldsymbol{\beta}$ that maximizes $L(\boldsymbol{\beta}|\mathbf{n})$ (eqn. 6):

$$\boldsymbol{\beta}_{MLE} = \text{argmax } L(\boldsymbol{\beta}|\mathbf{n})$$

No analytical solution to this optimization problem exists. Numerical optimization must be applied. For this purpose, it is convenient to express the optimization problem in terms of the log-likelihood instead (eqn. 7):

$$\boldsymbol{\beta}_{MLE} = \text{argmax } \ln L(\boldsymbol{\beta}|\mathbf{n}) = \sum_{i=1}^{m_d} \sum_{j=1}^{m_s} n_{ij} \ln(\lambda(\boldsymbol{\beta}, \mathbf{x}_{ij})) - \ln(n_{ij}!) - \lambda(\boldsymbol{\beta}, \mathbf{x}_{ij})$$

In the numerical investigations, the simplex search method and the quasi-Newton method are used to solve eqn. 7, as implemented in the Matlab functions *fminsearch* and *fminunc*. The simplex search method is found to not converge in the models that include the categorical variable “land cover”.

Diagnostics

To compare different models, the Akaike Information Criterion (AIC) is employed (Akaike 1974). The AIC allows to compare models of different complexity. It is defined as (eqn. 8):

$$AIC = -2 \ln L(\boldsymbol{\beta}_{MLE}|\mathbf{n}) + 2(k+1)$$

where $\ln L(\boldsymbol{\beta}_{MLE}|\mathbf{n})$ is the maximum log-likelihood and $(k+1)$ is the number of coefficients β_i of the model. The first term in the

AIC accounts for the likelihood of the model, the second term punishes models with more parameters to avoid overfitting.

An additional comparison between models is performed with a validation data set \mathbf{n}_v , which is not used for estimating $\boldsymbol{\beta}_{MLE}$. The log-likelihood of $\boldsymbol{\beta}_{MLE}$ calculated with the validation data set \mathbf{n}_v , i.e., $\ln L(\boldsymbol{\beta}_{MLE}|\mathbf{n}_v)$, provides an additional indication of model accuracy.

Numerical investigations

Study area: Cyprus

We employ data from the Republic of Cyprus, which is selected due to its representative Eastern Mediterranean climate (short cool winters followed by long hot and dry summers), vegetation and fire history and data availability. The study area and the five weather stations used in the analysis are indicated in Fig. 2a. The natural areas on the island are mainly covered by coniferous forests (e.g., *Pinus brutia*), whereas the permanent cultivated areas are dominated by vineyards. The highest peak of the study area is Olympus mountain of the Troodos massif (1952 m a.s.l. - Fig. 2a). In the period 2001-2010, the mean annual number of fire occurrences in the study area was 215 and the mean annual burnt area was 29 km² (Joint Research Center/IES 2011). Data of fires suppressed by the state forest agency (Department of Forests of Cyprus) for the period 2006-2010 is shown in Fig. 2b. The dataset includes records of fires of all sizes, with 10% of recorded fires being less than 0.01 ha. The total number of recorded fires is 616, which

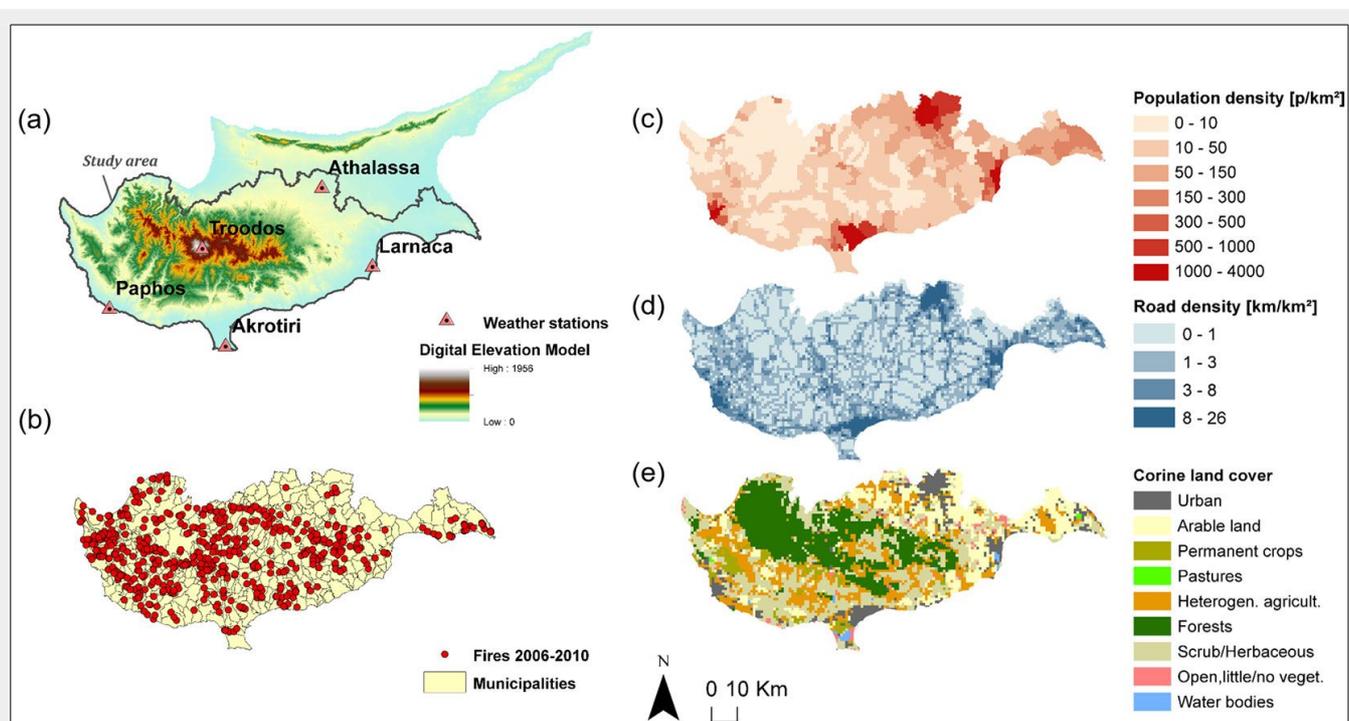
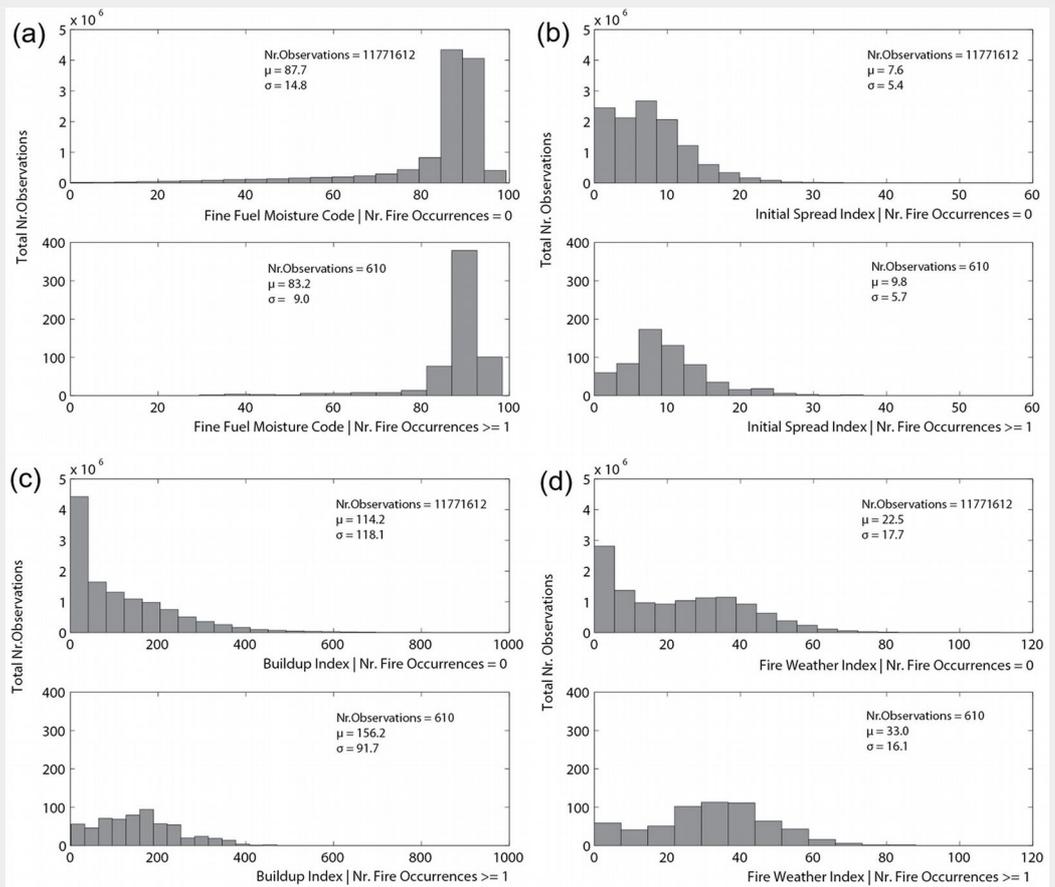


Fig. 2 - (a) ASTER Digital Elevation Model (m) showing the highest peak of the Troodos massif in white (1956 m a.s.l.) and the included five weather stations on Cyprus; (b) Municipality borders of the area of the numerical investigations and registered fire events during 2006-2010 (616 events); (c) population density; (d) road density; (e) land cover.

Fig. 3 - Histograms of (a) FFM, (b) ISI, (c) BUI and (d) FWI (2006-2010) conditional on fire occurrence. $CV = \sigma/\mu$ is the coefficient of variation.



corresponds to a mean annual number of fire occurrences of 123.

Data types and data sources

Both spatial and temporal explicit data are used in this study. Data are managed in a geodatabase processed with ArcGIS[®] 10.1 (ESRI, Redlands, CA, USA) and Python[®] 2.6.8 (Python Software Foundation, Wilmington, DE, USA) and are attached to a 1 km² grid covering the whole area of the case study (6447 grid cells). The population density in each grid cell (people km⁻²) is determined from the municipality census data (Fig. 2c). The road density (km km⁻²) is computed from the actual length of roads in each cell (Fig. 2d). The land cover type assigned to each cell is the one covering the largest area within that cell (Fig. 2e). According to Corine land cover (2006), forests and semi-natural areas together with agricultural areas cover the largest part of the study area. The land cover type “Pastures” is included into the “Urban-Wetland” land covers, since it covered only a small area of the case study (7 km²).

Weather data interpolation and CFFWIS calculation

Daily weather observations (extracted from 3 hr and 6 hr observations) are interpolated using Inverse Distance Weighting (IDW – Shepard 1968). Daily values of the CFFWIS components are then calculated for each grid cell based on the interpolated values.

Temperature is additionally adjusted to the altitude based on the normal lapse rate (0.65 °C/100m – Leemans & Cramer 1991). At each weather station i , the equivalent temperature at sea level is computed from the measured noon temperature T_i as $T_{o,i} = T_i + 0.0065 \cdot h_i$, where h_i is the altitude of the weather station in m. The IDW interpolation is performed using the $T_{o,i}$ values, resulting in a temperature value at sea level $T_{o,c}$ for each cell c . The daily noon value of temperature in each cell T_c is then computed as $T_c = T_{o,c} - 0.0065 \cdot h_c$. Here, h_c is the altitude at the center of cell c .

After the weather observations are interpolated, the daily FWI is calculated for each cell based on the formulation given in Van Wagner & Pickett (1985). The starting values of the fuel moisture codes for the first day (Jan 1) are the ones proposed in Lawson & Armitage (2008), *i.e.*, FFM = 85, DMC = 6, DC = 15. The starting values were reset every year.

Parameter estimation

After the data pre-processing, weather interpolation and FWI calculation, each of the 6447 grid cells is described by spatial information, noon daily weather conditions and FWI, and recorded fire events for the period 2006-2010 (11,772,222 records). Only the records of the period 2006-2009 (9,419,067 records) are used for parameter estimation.

Poisson regression with MLE is employed as described above. Various candidate

models for the fire occurrence rate λ were learnt with the data. All models are of the form given in eqn. 3 and differ only in the selection of parameters employed. From these models, five were selected and are presented in this paper.

Results

Preliminary data analysis

Preliminary analysis of the time series 2006-2010 is shown in Fig. 3 and in the Supplementary Material (Fig. S3, Fig. S4). As there are 616 recorded fires, the average occurrence rate of fires in this period is 5.5×10^{-5} fires d⁻¹ km⁻².

The results of Fig. 3 show that there is a statistically significant difference in the conditional means of ISI, BUI and FWI, which indicate their potential as explanatory variables in the regression model. However, by comparing the conditional distributions graphically, it is also clear that the components alone have only limited prediction ability. For example, fires occurred also on days and locations with FWI values close to zero.

Regression analysis

The investigated alternative candidate models included the components BUI, ISI and FWI of the CFFWIS. Maximum likelihood estimation (with respect to the learning set 2006-2009) results in the parameter values, which best explains the data for a given model. To compare the different

models, the AIC is applied, which corresponds to the maximum log likelihood value and combined with a term that punishes the use of additional model parameters to avoid overfitting (see eqn. 8). The model including both BUI and ISI (M_BUI_ISI) performed better than M_BUI and M_ISI, and all three models proved to perform worse than M_FWI. For this reason, FWI is selected to express fire weather conditions in the further investigated models.

In Tab. 1 models M1 to M5 are arranged according to their increasing number of explanatory variables, starting from M1 that includes FWI as the sole variable (M_FWI), to M5 with 12 variables, including FWI, road and population density and land cover types. As an example, the predicted rate of fires according to model M5 is (eqn. 9):

$$\lambda = \exp[-10.90 + 0.0329 \cdot FWI + 0.3217 \cdot Road\ density - 0.0234 \cdot (Road\ density)^2 - 0.0010 \cdot Pop\ density - 0.9681 \cdot Arable + 0.3235 \cdot Permanent - 0.0760 \cdot Heterogeneous + 0.1057 \cdot Forest + 0.0486 \cdot Shrubs - 0.1882 \cdot Open\ Spaces - 1.1863 \cdot Urban\ |Wet\ |Pastures]$$

In models M2, M3 and M5, road density as well as (road density)² are included as explanatory variables, to represent the non-linear effect of road density on the fire occurrence rate observed from Fig. S4c (in Supplementary Material). It is important to stress that road and population density are highly positively correlated, and are also dependent on land cover type.

Based on the learning data set, model M5 performs best, as it exhibits the lowest AIC, followed by M3 and M4. The estimated parameters of the explanatory variables FWI, road density and population density in all models M1-M5 are consistent. In models M4 and M5, the estimated parameters of the land cover types take slightly different values. They are higher in M4 due to the fact that in M5 the additional terms in the link function describing road and population density on average take a value slightly above zero.

It is also worthwhile noting that the variables describing road and population density in Model M5 are not independent of the land use type. Pearson's correlation coefficient *r* between population density and urban & wetlands land cover type is 0.48 and between road density and urban & wetlands is 0.59. Therefore, the variables population and road density in model M5 partly express the fact that fires are less likely in urban areas. In M4, where these variables are not present, this effect is fully described by β_{11} alone. Because of this dependence, there is also a significant correlation (*r* = 0.56) between population and

Tab. 1 - Selected models with explanatory variables and estimated parameters (2006-2009). (*): Permanent crops include olives, vineyards and fruits; (**): Urban-Wet-Past variable includes Urban areas, Wetlands and Pastures.

Explanatory variables	Param	Selected Models				
		M1	M2	M3	M4	M5
Intercept	β_0	-10.61	-10.95	-10.92	-10.90	-10.90
FWI	β_1	0.0278	0.0282	0.0302	0.0327	0.0329
Road density [km km ⁻²]	β_2	-	0.3236	0.3198	-	0.3217
(Road density) ² [km km ⁻²]	β_3	-	-0.0324	-0.0276	-	-0.0234
Population dens. [people km ⁻²]	β_4	-	-	-0.0018	-	-0.0010
Arable	β_5	-	-	-	-0.6501	-0.9681
Permanent*	β_6	-	-	-	0.8383	0.3235
Heterogeneous	β_7	-	-	-	0.4098	-0.0760
Forest	β_8	-	-	-	0.3497	0.1057
Shrub/Herbaceous	β_9	-	-	-	0.3279	0.0486
Open spaces	β_{10}	-	-	-	-0.1310	-0.1882
Urban-Wet-Past**	β_{11}	-	-	-	-0.9556	-1.1863
log-likelihood (2006-2009)	-	-5198.4	-5166.1	-5151.2	-5147.3	-5111.9
AIC (2006-2009)	-	10400.8	10340.2	10312.4	10312.6	10247.8

Tab. 2 - Relative change of occurrence rate $\Delta\lambda/\lambda$ with changing explanatory variables of model M5. For continuous variables FWI, population density and road density, the change of the variable is equal to one standard deviation σ . (*): $(\Delta\lambda/\lambda)_i = \exp[\Delta x(\beta_2 + 2\beta_3\mu_{RD} + \beta_3\Delta x)]$, with $\mu_{RD}=2.09$ being the mean value of road density.

Explanatory variables	Δx	$(\Delta\lambda/\lambda)_i$ eqn. 4
FWI	$\sigma = 17.7$	0.791
Population density	$\sigma = 316$	-0.271
Road density	$\sigma = 3.23$	0.614 *
(Road density) ²		
Arable	1	-0.620
Permanent	1	0.382
Heterogeneous	1	-0.073
Forest	1	0.111
Shrub/Herbaceous	1	0.050
Open spaces	1	-0.172
Urban-Wet-Past	1	-0.695

road density.

Eqn. 4 is used to compare the sensitivity of the studied models to changes in the explanatory variables. Tab. 2 shows the relative change of λ , as predicted by M5, when changing one explanatory variable and keeping all others fixed. For FWI, population density and road density, the change of the variable is equal to one standard deviation σ , whereas the land cover types change from 0 to 1. For higher FWI, λ

increases and for higher population density λ decreases. These results agree with results in Fig. S4 (Supplementary Material).

Prediction

Fire observations of the study area in 2010 are used to verify the predictive ability of the proposed models. The best model is the one that best discriminates the actual locations with fire occurrences from those without. This is described by the sum of

Tab. 3 - Predicted fire occurrence rate at the locations of fires shown in Fig. 4.

Day in 2010	Fire locations	Fire occurrence rate ($\times 10^{-5} d^{-1} km^{-2}$)				
		M1	M2	M3	M4	M5
Oct 8	a	7.2	7.1	7.8	9.1	9.4
	b	6.4	4.6	5.1	8.1	6.4
	c	6.0	4.3	4.8	7.5	5.9
	d	5.7	6.3	7.0	7.0	8.8
	e	5.4	4.0	4.4	6.6	5.3
Jun 26	a	3.9	5.3	5.8	4.7	5.9
	b	6.2	8.4	8.6	7.6	10.9
	c	5.1	8.0	8.0	6.0	10.7
2010	log-likelihood in study area	-1388.6	-1383.2	-1388.7	-1380.6	-1377.1

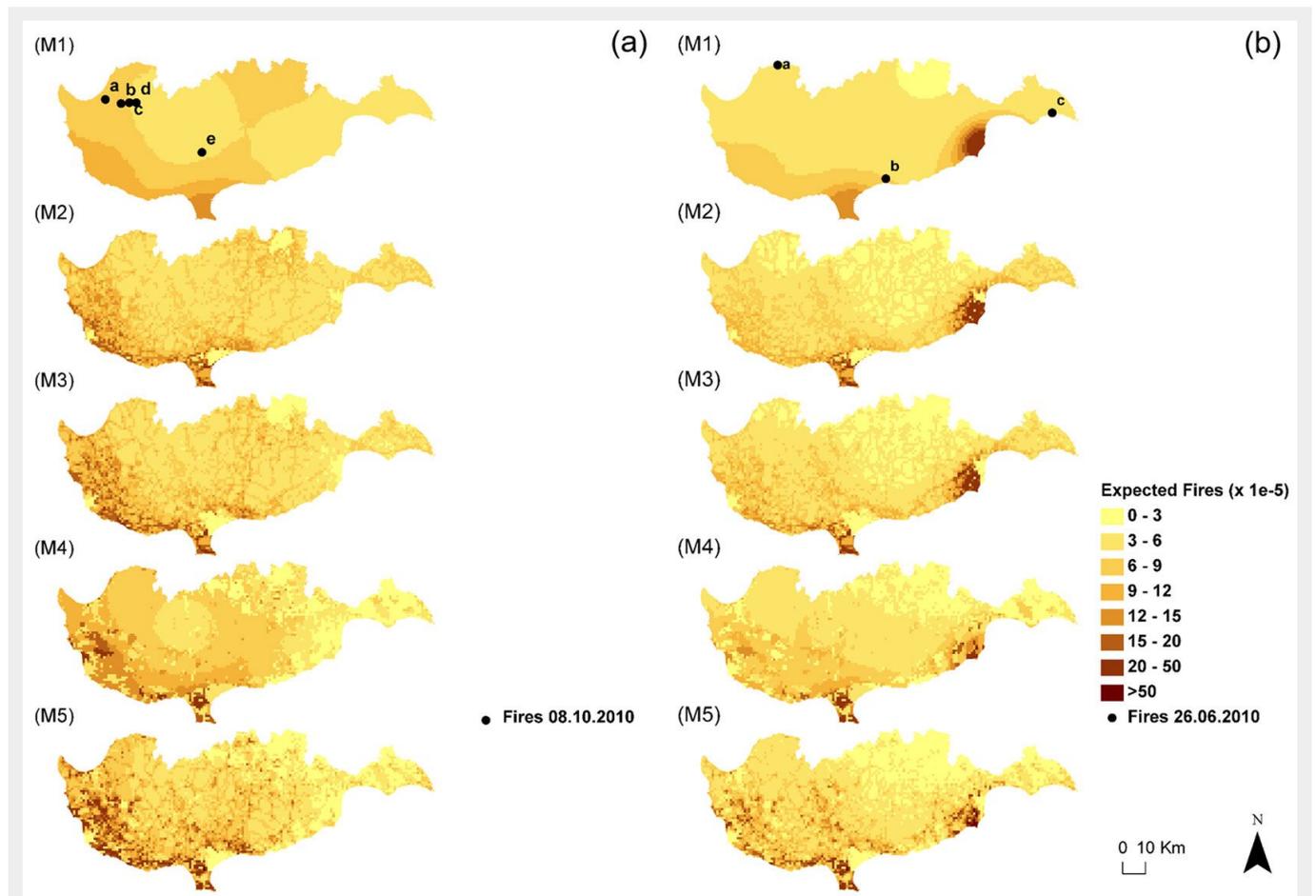


Fig. 4 - Expected occurrence rates of fires predicted by different regression models on (a) 8th October 2010 (day with maximum number of fires in 2010) and (b) on 26th June 2010 (day with second maximum number of fires and largest resulted burnt area ($3.4 \text{ km}^2 = 340 \text{ ha}$) in 2010). Black dots represent the registered fires on this day (a - e). The predictions are estimated by the models M1, M2, M3, M4, M5. Occurrence rate results are in the order of $1e-5$.

the likelihood values for all cells and all days of the prediction data set. Model M5 predicts the highest log-likelihood for the entire data of 2010 (Tab. 3), which indicates that this model is the best in predicting fire occurrence among all investigated models. Two days in 2010 with the highest num-

ber of fires are selected to investigate and demonstrate the prediction of the fire occurrence rate with the model (whose parameters were learnt by data for 2006-2009). Fig. 4a shows the expected number of fires as predicted by the models on October 8, 2010 – the day with the highest

number of fires in 2010 (5 fires) and for June 26, 2010. Urban centers are clearly visible in the maps as the areas with permanent low expected fires predicted by all models.

Models M4 and M5 generally predict higher occurrence rates than model M1, which includes only the influence of FWI (Tab. 3). However, it is reminded that to assess the predictive power of the model, it is not sufficient to focus on the prediction of fire occurrences. The prediction in all cells must be compared. To this end, one can compare the probability of the observed fire and no-fire events in the entire area on all days in 2010 as predicted by the models. This probability is equal to the likelihood of the final models computed with the 2010 data.

Since the likelihood is only a relative measure of prediction performance, additionally receiver operating characteristic (ROC) curves are computed for the dataset of 2010 and each model M1-M5 (Fig. 5). The ROC curves are computed by considering the binary variable, describing whether a fire occurred. The probability of one or more fires during one day is $1 - \exp(-\lambda)$. Model M5 has the biggest area under the ROC curve, i.e., it performs best among the

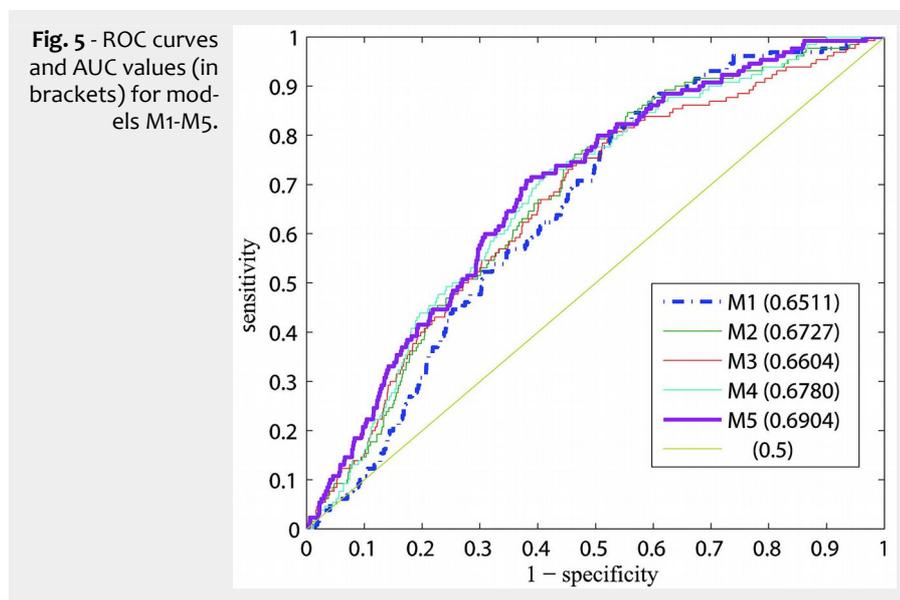


Fig. 5 - ROC curves and AUC values (in brackets) for models M1-M5.

other models, whereas model M1 has the lowest AUC and performs worse among the models. ROC curves are worse, and AUC values are lower, when they are computed for the fine spatial resolution employed here, as opposed to an analysis of a larger area. However, this is a mathematical artefact, but it is important to realize when comparing the values to other published studies. In larger areas, random effects are reduced, as follows from the law of large numbers. It is straightforward to compute ROC curves for the entire study area, since the fire occurrence rate is simply the sum of the occurrence rates in all cells. The AUC values computed for prediction in the entire study area in 2010 is 0.74. This AUC value is the same for all models, because in this computation, the spatial differentiation is lost.

Discussion

This study is a step towards an improved prediction of fire occurrence in the Mediterranean for fire management purposes. The selected probabilistic modeling approach provides a quantitative metric of the ability of different explanatory variables to predict daily fire occurrence. Of particular interest is the ability of the FWI, which was developed for Canada, to predict fire danger in the Mediterranean. As we found in this study, the FWI is a good indicator for fire danger also in the Eastern Mediterranean, even if its prediction ability is lower than in Canada and similar climates. In previous empirical studies, the components of the CFFWIS (FFMC, expressing fine fuel moisture; ISI, representing relative fire spread expected immediately after ignition; and BUI, expressing moisture content of heavier fuels) were found to be relevant indicators for predicting people-caused fire occurrence in Canada (Martell et al. 1987, Wotton et al. 2003). FWI was here chosen to express fire weather conditions as it proved to be more expressive than the intermediated components of the CFFWIS. A likely reason for this is that the studied fire events are the ones registered and suppressed by the forest fire department. It can be thus assumed that not all ignited fires are included in the data set. Since the included fire events are those that initiated a threat and suppression efforts had to be undertaken, the proposed model is potentially more relevant for fire management planning. Nevertheless, the observed FWI values in the study area are mostly in a limited range only (Fig. 3), which limits the ability of the FWI alone to discriminate days and locations with high fire danger from those with low fire danger. This indicates that there might be potential in adjusting the definition of the FWI to local conditions. It may also be investigated if selected weather parameters should be included as explanatory variables in addition to the FWI.

In agreement with previous studies (Cardille et al. 2001, Pew & Larsen 2001, Amat-

ulli et al. 2006, Kalabokidis et al. 2007, Syphard et al. 2008, Vilar et al. 2010, Padilla & Vega-García 2011, Miranda et al. 2012, Oliveira et al. 2012, Martínez-Fernández et al. 2013), we found that including anthropogenic factors as explanatory variables can significantly improve the prediction of fire occurrence. The comparison of different models showed that a model with land cover types, population and road density has a significantly better predictive ability than one based on FWI alone. Since such data is readily available, it is straightforward to include it in forecasting systems.

Further explanatory variables describing anthropogenic factors may be included in the analysis (see also Cardille et al. 2001, Oliveira et al. 2012, Martínez-Fernández et al. 2013). However, care should be taken not to introduce redundant variables. Already the three included explanatory variables (land cover type, population and road density) are partly redundant and inter-dependent, e.g., both population and road density are higher in urban areas. This dependency must be considered when transferring the model to other regions.

Due to the randomness of fire occurrence, there is a limitation to any prediction. This is evident in the results presented in this paper. Consider the predicted fire occurrence rate at locations and days where fires occurred, shown in Tab. 3: the rates predicted with the best models are approximately double the average rate of fires in the study area ($5.5 \cdot 10^{-5}$ day⁻¹ km⁻²). Therefore, while the developed models are able to identify days and locations with higher fire risks, they are not – and of course will not – be able to deterministically predict fire occurrences in advance. Nevertheless, the predictions can support the planning of preventive and mitigating measures. Importantly, they also improve the understanding of influential factors.

Conclusions

A probabilistic model was developed for predicting fire occurrences in the Mediterranean based on readily available data on weather conditions, human presence and land cover at the mesoscale. The model was learned with data from Cyprus. In agreement with existing forecasting systems, components of the CFFWIS are included to represent daily weather conditions. Among these components, FWI proved to express best the conditions favoring relevant fires. The final model including environmental and social factors was shown to provide improved predictions compared to a forecast based solely on FWI.

Acknowledgements

We gratefully acknowledge the support of Stefan Peters in data preprocessing and of Florian Klein in weather interpolation. We thank Areti Christodoulou from the Department of Forests of Cyprus for supplying the fire data and the comments of

Dr. Gavriil Xanthopoulos on the literature review. The comments of five anonymous reviewers to an earlier version are highly appreciated.

References

- Ager AA, Preisler HK, Arca B, Spano D, Salis M (2014). Wildfire risk estimation in the Mediterranean area. *Environmetrics* 25 (6): 384-396. - doi: [10.1002/env.2269](https://doi.org/10.1002/env.2269)
- Akaike H (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6): 716-723. - doi: [10.1109/TAC.1974.1100705](https://doi.org/10.1109/TAC.1974.1100705)
- Amatulli G, Rodrigues MJ, Trombetti M, Lovreglio R (2006). Assessing long-term fire risk at local scale by means of decision tree technique. *Journal of Geophysical Research: Biogeosciences* 111: G04S05. - doi: [10.1029/2005JG000133](https://doi.org/10.1029/2005JG000133)
- Arndt N, Vacik H, Koch V, Arpacı A, Gossow H (2013). Modeling human-caused forest fire ignition for assessing forest fire danger in Austria. *iForest - Biogeosciences and Forestry* 6 (5): 315-325. - doi: [10.3832/for0936-006](https://doi.org/10.3832/for0936-006)
- Bayraktarlı YY, Ulfkjær J, Yazgan U, Faber MH (2005). On the application of Bayesian probabilistic networks for earthquake risk management. In: *Proceedings of the "9th International Conference on Structural Safety and Reliability - ICOSSAR"* (Augusti et al. eds). Rome (Italy) 19-23 Jun 2005. Millpress, Rome, Italy, pp. 20-23. [online] URL: <http://www.merci.ethz.ch/Publications/bayota.pdf>
- Bensi M, Kiureghian AD, Straub D (2014). Framework for post-earthquake risk assessment and decision making for infrastructure systems. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 1 (1): 04014003. - doi: [10.1061/AJRUA6.0000810](https://doi.org/10.1061/AJRUA6.0000810)
- Blaser L, Ohrnberger M, Riggelsen C, Scherbaum F (2009). Bayesian belief network for tsunami warning decision support. In: *Proceedings of the 10th European Conference "ECSQARU 2009"*. Verona (Italy) 1-3 Jul 2009 (Sossai C, Chemello G eds). Springer, Berlin, Heidelberg, Germany, pp. 757-786.
- Camia A, Amatulli G (2009). Weather factors and fire danger in the Mediterranean. In: *"Earth Observation of Wildland Fires in Mediterranean Ecosystems"* (Chuvieco E eds). Springer, Berlin, Heidelberg, Germany, pp. 71-82. - doi: [10.1007/978-3-642-01754-4_6](https://doi.org/10.1007/978-3-642-01754-4_6)
- Cardille JA, Ventura SJ, Turner MG (2001). Environmental and social factors influencing wildfires in the Upper Midwest, United States. *Ecological Applications* 11 (1): 111-127. - doi: [10.1890/1051-0761\(2001\)011\[0111:EASFIW\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2001)011[0111:EASFIW]2.0.CO;2)
- Catry FX, Rego FC, Bação F, Moreira F (2010). Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire* 18 (8): 921-931. - doi: [10.1071/WF07123](https://doi.org/10.1071/WF07123)
- Chou YH, Chase RA (1993). Mapping probability of fire occurrence in San Jacinto Mountains, California, USA. *Environmental Management* 17 (1): 129-140. - doi: [10.1007/BF02393801](https://doi.org/10.1007/BF02393801)
- Chuvieco E, González I, Verdú F, Aguado I, Yerba M (2009). Prediction of fire occurrence from live fuel moisture content measurements in a Mediterranean ecosystem. *International Journal of Wildland Fire* 18 (4): 430-441. - doi: [10.1071](https://doi.org/10.1071)

/WFO8020

- Cunningham AA, Martell DL (1973). A stochastic model for the occurrence of man-caused forest fires. *Canadian Journal of Forest Research* 3: 282-287. - doi: [10.1139/x73-038](https://doi.org/10.1139/x73-038)
- Dimitrakopoulos AP, Bemmerzouk AM, Mitsopoulos ID (2011). Evaluation of the Canadian fire weather index system in an eastern Mediterranean environment. *Meteorological Applications* 18: 83-93. - doi: [10.1002/met.214](https://doi.org/10.1002/met.214)
- Dlamini WM (2009). A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland. *Environmental Modelling and Software* 25 (2): 199-208. - doi: [10.1016/j.envsoft.2009.08.002](https://doi.org/10.1016/j.envsoft.2009.08.002)
- Giannakopoulos C, Karali A, Roussos A, Hatzaki M, Xanthopoulos G, Kaoukis K (2011). Evaluating present and future fire risk in Greece. *Advances in Remote Sensing and GIS applications in Forest Fire Management. From local to global assessments. JRC Scientific and Technical Reports, European Commission, JRC-IES, Land Management and Natural Hazards Unit, Ispra, Italy*, pp. 181.
- Grêt-Regamey A, Straub D (2006). Spatially explicit avalanche risk assessment linking Bayesian networks to a GIS. *Natural hazards and Earth System Sciences* 6: 911-926. - doi: [10.5194/nhess-6-911-2006](https://doi.org/10.5194/nhess-6-911-2006)
- Jensen FV, Nielsen TD (2007). *Bayesian networks and decision graphs*. Springer, Series Information Science and Statistics, NY, USA, pp. 447. - doi: [10.1007/978-0-387-68282-2](https://doi.org/10.1007/978-0-387-68282-2)
- Joint Research Center/IES (2011). *Forest fires in Europe 2010*. Report no. 11, JRC Scientific and Technical Reports, European Commission, JRC-IES, Land Management and Natural Hazards Unit, Publications Office of the European Union, Luxembourg, pp. 92.
- Kalabokidis KD, Koutsias N, Konstantinidis P, Vasilakos C (2007). Multivariate analysis of landscape wildfire dynamics in a Mediterranean ecosystem of Greece. *Area* 39 (3) pp. 392-402. - doi: [10.1111/j.1475-4762.2007.00756.x](https://doi.org/10.1111/j.1475-4762.2007.00756.x)
- Kuehn NM, Riggelsen C, Scherbaum F (2011). Modeling the joint probability of earthquake, site, and ground-motion parameters using Bayesian Networks. *Bulletin of the Seismological Society of America* 101 (1): 235-249. - doi: [10.1785/0120100080](https://doi.org/10.1785/0120100080)
- Lawson BD, Armitage OB (2008). *Weather guide for the Canadian forest fire danger rating system*. Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre. Edmonton, Alberta, Canada, pp. 73. - [online] URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.372.7904&rep=rep1&type=pdf>
- Leemans R, Cramer WP (1991). The IIASA database for mean monthly values of temperature, precipitation and cloudiness on a global terrestrial grid. Report RR-91-18, International Institute for Applied Systems Analysis, Laxenburg, Austria, pp. 62. [online] URL: <http://epic.awi.de/29861/1/lee1991a.pdf>
- Martell DL, Otukol S, Stocks BJ (1987). A logistic model for predicting daily people-caused forest fire occurrence in Ontario. *Canadian Journal of Forest Research* 17 (5): 394-401. - doi: [10.1139/x87-068](https://doi.org/10.1139/x87-068)
- Martínez-Fernández J, Chuvieco E, Koutsias N (2013). Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression. *Natural Hazards and Earth System Science* 13 (2): 311-327. - doi: [10.5194/nhess-13-311-2013](https://doi.org/10.5194/nhess-13-311-2013)
- Miranda BR, Sturtevant BR, Stewart SI, Hammer RB (2012). Spatial and temporal drivers of wildfire occurrence in the context of rural development in northern Wisconsin, USA. *International Journal of Wildland Fire* 21 (2): 141-154. - doi: [10.1071/WF10133](https://doi.org/10.1071/WF10133)
- Mandallaz D, Ye R (1997). Prediction of forest fires with Poisson models. *Canadian Journal of Forest Research* 27: 1685-1694. - doi: [10.1139/x97-103](https://doi.org/10.1139/x97-103)
- Massada AB, Syphard DA, Stewart SI, Radeloff VC (2012). Wildfire ignition-distribution modeling: a comparative study in the Huron-Manistee national forest, Michigan, USA. *International Journal of Wildland Fire* 22 (2): 174-183. - doi: [10.1071/WF11178](https://doi.org/10.1071/WF11178)
- Moriondo M, Good P, Durao R, Bindi M, Giannakopoulos C, Corte-Real J (2006). Potential impact of climate change on fire risk in the Mediterranean area. *Climate Research* 31: 85-95. - doi: [10.3354/cr031085](https://doi.org/10.3354/cr031085)
- Oliveira S, Oehler F, San-Miguel-Ayanz J, Camia A, Pereira JMC (2012). Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and Random Forest. *Forest Ecology and Management* 275: 117-129. - doi: [10.1016/j.foreco.2012.03.003](https://doi.org/10.1016/j.foreco.2012.03.003)
- Padilla M, Vega-García C (2011). On the comparative importance of fire danger rating indices and their integration with spatial and temporal variables for predicting daily human-caused fire occurrences in Spain. *International Journal of Wildland Fire* 20 (1): 46-58. - doi: [10.1071/WF09139](https://doi.org/10.1071/WF09139)
- Parisien MA, Moritz MA (2009). Environmental controls on the distribution of wildfire at multiple spatial scales. *Ecological Monographs* 79 (1): 127-154. - doi: [10.1890/07-1289.1](https://doi.org/10.1890/07-1289.1)
- Pew KL, Larsen CPS (2001). GIS analysis of spatial and temporal patterns of human-caused wildfires in the temperate rain forest of Vancouver Island, Canada. *Forest Ecology and Management* 140 (1): 1-18. - doi: [10.1016/S0378-1127\(00\)00271-1](https://doi.org/10.1016/S0378-1127(00)00271-1)
- Plucinski MP (2012). A review of wildfire occurrence research. CSIRO Ecosystem Sciences and CSIRO climate Adaptation Flagship, Bushfires Cooperative Research Center (CRC), Canberra, Australia, pp. 2-25. [online] URL: http://www.bushfirecrc.com/sites/default/files/managed/research/attachment_g_fire_occurrence_literature_review_o.pdf
- Preisler HK, Brillinger DR, Burgan RE, Benoit JW (2004). Probability based models for estimation of wildfire risk. *International Journal of Wildland Fire* 13: 133-142. - doi: [10.1071/WF02061](https://doi.org/10.1071/WF02061)
- Romero-Calcerrada R, Novillo CJ, Millington JDA, Gomez-Jimenez I (2008). GIS analysis of spatial patterns of human-caused wildfire ignition risk in the SW of Madrid (Central Spain). *Landscape Ecology* 23 (3): 341-354. - doi: [10.1007/s10980-008-9190-2](https://doi.org/10.1007/s10980-008-9190-2)
- Sebastián-López A, Salvador-Civil R, Gonzalo-Jiménez J, SanMiguel-Ayanz J (2008). Integration of socio-economic variables for modeling long-term fire danger in Southern Europe. *European Journal of Forest Research* 127: 149-163. - doi: [10.1007/s10342-007-0191-5](https://doi.org/10.1007/s10342-007-0191-5)
- Shepard D (1968). A two-dimensional interpolation function for irregularly-spaced data. In: *Proceeding of the "23rd ACM National Conference"*. Las Vegas (NV, USA) 27-29 Aug 1968. ACM, New York, USA, pp. 517-524. [online] URL: <http://dl.acm.org/citation.cfm?id=810616>
- Straub D (2005). Natural hazards risk assessment using Bayesian networks. In: *Proceedings of the "9th International Conference on Structural Safety and Reliability - ICOSSAR"* (Augusti et al. eds). Rome (Italy) 19-23 Jun 2005. Millpress, Rome, Italy, pp. 2535-2542. [online] URL: http://www1.ethz.ch/ibk/emeritus/fa/people/straub/Presentation_Straub_ICOSSAR_2005_NH.pdf
- Syphard DA, Radeloff VC, Keuler NS, Taylor RS, Hawbaker TJ, Stewart SI, Clayton MK (2008). Predicting spatial patterns of fire on a southern California landscape. *International Journal of Wildland Fire* 17 (5): 602-613. - doi: [10.1071/WF07087](https://doi.org/10.1071/WF07087)
- Taylor RS, Alexander ME (2006). Science, technology, and human factors in fire danger rating: the Canadian experience. *International Journal of Wildland Fire* 15: 121-135. - doi: [10.1071/WF05021](https://doi.org/10.1071/WF05021)
- Van Wagner CE, Pickett TL (1985). Equations and FORTRAN program for the Canadian forest fire weather index system. *Forestry Technical Report 33*, Canadian Forestry Service, Government of Canada, Ottawa, Ontario, Canada, pp. 33. [online] URL: <http://www.cfs.nrcan.gc.ca/publications/?id=19973>
- Van Wagner CE (1987). Development and structure of the Canadian forest fire weather index system. *Forestry Technical Report 35*, Canadian Forestry Service, Ottawa, Ontario, Canada, pp. 44. - [online] URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.460.3231>
- Vasconcelos MJP, Silva S, Tomé M, Alvim M, Pereira JMC (2001). Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. *Photogrammetric Engineering and Remote Sensing* 67: 73-81.
- Vasilakos C, Kalabokidis K, Hatzopoulos J, Matsinos I (2009). Identifying wildland fire ignition factors through sensitivity analysis of a neural network. *Natural Hazards* 50 (1): 125-143. - doi: [10.1007/s11069-008-9326-3](https://doi.org/10.1007/s11069-008-9326-3)
- Venäläinen A, Korhonen N, Hyvärinen O, Koutsias N, Xystrakis F, Urbietta IR, Moreno JM (2014). Temporal variations and change in forest fire danger in Europe for 1960-2012. *Natural Hazards and Earth System Science* 14 (6): 1477-1490. - doi: [10.5194/nhess-14-1477-2014](https://doi.org/10.5194/nhess-14-1477-2014)
- Viegas DX, Bovio G, Ferreira A, Nosenzo A, Sol B (1999). Comparative study of various methods of fire danger evaluation in Southern Europe. *International Journal of Wildland Fire* 9 (4): 235-246. - doi: [10.1071/WF00015](https://doi.org/10.1071/WF00015)
- Vilar L, Woolford DG, Martell DL, Martín MP (2010). A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. *International Journal of Wildland Fire* 19 (3): 325-337. - doi: [10.1071/WF09030](https://doi.org/10.1071/WF09030)
- Wotton BM, Martell DL, Logan KA (2003). Climate change and people-caused forest fire occurrence in Ontario. *Climatic Change* 60 (3): 275-295. - doi: [10.1023/A:1026075919710](https://doi.org/10.1023/A:1026075919710)

Supplementary Material

Fig. S1 - Interpolated weather parameters on 8th October 2010. (a) wind speed (km h⁻¹), (b) temperature (°C), (c) relative humidity (%).

Fig. S2 - Estimated components of the Canadian Forest Fire Weather Index system on 8th October 2010. (a) Fine Fuel Moisture Code (FFMC), (b) Duff Moisture Code (DMC), (c) Drought Code (DC), (d) Initial Spread Index (ISI), (e) Buildup Index (BUI), (f) Fire Weather Index (FWI).

Fig. S3 - Daily values of Fire Weather Index (FWI), precipitation (mm) and noon dry-bulb temperature (°C) at Paphos weather station, and daily number of fire occurrences in the investigated area in 2006.

Fig. S4 - Observed mean occurrence rate (Nr.Fires day⁻¹ km²) conditional on (a) FWI, (b) Population density, (c) Road density and (d) Land cover types for 2006-2010.

Tab. S1 - Data types, resolution and sources.

Tab. S2 - Alternative models with explanatory variables and estimated parameters (2006-2009).

Link: Papakosta_1686@suppl001.pdf