University of Tartu

Faculty of Science and Technology

Institute of Ecology and Earth Sciences

Department of Geography

Master's thesis in Geoinformatics for Urbanised Society (30 ECTS)

Mapping Wildfire Susceptibility of Sardinia Island, Italy

Behzad Valipour Shokouhi

Supervisor: PhD Evelyn Uuemaa

Approved for defence:

Supervisor:

Head of Department:

Tartu 2020

Magistritöö õppekaval Linnastunud ühiskonna geoinformaatika: Metsatulekahjude tuleohu kaardistamine Sardiinia saarel Itaalias

Sisututvustus
Metsatulekahjude tõttu on kogu maailmas kaotatud suur osa metsi, põllumajandusmaid, looduslikke alasid ja inimeste elusid. Igal aastal registreeritakse Lõuna-Euroopas põlengute hooajal rohkem kui 30 000 tulekahju. Seega tuleb tulekahjude puhkemise ja nende tekketingimuste vältimise nimel pingutusi teha. Käesolevas uuringus analüüsiti tulekahjude ruumilisi mustreid, kasutades kaht masinõppe meetodit: logistilist regressiooni ja otsustusmetsa (Random Forest). Mudelite koostamisel kasutati viimase kaheksa aasta (2010‒2018) andmeid toimunud tulekahjude kohta ning kuutteist sõltumatut muutujat (maapinna kalle, aspekt, kõrgus merepinnast, topograafiline asukoha indeks (TPI), topograafiline kareduse indeks (TRI), kuu keskmine temperatuur ja sademete hulk, tuule kiirus, tuule suund, mullaniiskus, maakasutus, rahvaarv piirkonnas, õhuniiskus ning (põlengu)ala kaugus teedest, asulatest ja jõgedest). Mudelite tulemuslikkuse optimeerimiseks rakendati kaheastmelist parameetrite valiku meetodit. Mudelite tulemuslikkust hinnati statistiliste indeksipõhiste hindajate ning (*ROC-AUC Receiver Operating Characteristic curve*) graafiku abil abil. Tulemused näitasid, et otsustusmetsa täpsus oli veidi parem (73%) kui logistiline regressioon (64%)

Võtmesõnad: tuleoht, ruumiline prognoosimine, masinõpe, otsustusmets, logistiline regressioon.
CERCS kood:

Master Thesis in Geoinformatics for Urbanized Society: Mapping Wildfire Susceptibility of Sardinia Island, Italy

Abstract
Forests, agricultural lands, natural areas and lives of the citizens all around the world are lost and endangered by wildfire. Annually, more than 30,000 fires are recorded in Southern Europe during the fire seasons. Therefore, efforts are required to avoid the outbreak of wildfires and to decrease the condition of wildfire occurrence. This study analyzed the spatial patterns of the fires using two machine learning approaches: Logistic Regression and Random Forest. Historical data about fires for the last eight years (2010‒2018) with sixteen independent variables (slope degree, aspect, altitude, topographic position index (TPI), topographic roughness index (TRI), mean monthly temperature and rainfall, wind speed, wind direction, soil moisture, land use, population, humidity and proximity to roads, settlement and rivers) were used to train the models. To optimize the performance of the models, a two-step feature selection method including multi-collinearity and Gain Ratio analysis for random forest and Wald test for logistic regression were applied. The performance of the models was estimated using statistical index-based evaluators and Receiver Operating Characteristic curve (ROC-AUC). The results were illustrated that the performance of Random Forest with 73% overall accuracy is better than Logistic Regression with 64% overall accuracy.

Keywords: Fire risk, Spatial prediction, Machine Learning, Random Forest, Logistic Regression
CERCS Code:

Contents

Introduction

One of the man-made and natural disasters which leads to the death of people, damage to housing areas and finally destruction of natural surroundings is wildfire. Wildfire has a negative influence on the agricultural systems by destroying not only vegetation but also organic matter in the soil which is essential for the fertility of the soil (Suryabhagavan et al., 2016). On the other hand, wildfires lead to soil erosion due to fire effects on the growth of grass, herbs, and shrubs (Kandya et al., 1998; Suryabhagavan et al., 2016). It can also disturb the balance of soil chemicals, water levels, and soil pH. Various factors can increase the wildfire risk, for instance, during the El Niño years of 1997–1998 and the 2005–2006 dry seasons in the Brazilian Amazon, 1000 km^2 of forest ruined (Nasiri, 2013). Heat wave (HW) of summer 2003 in the Europe caused over 25,000 fires in the south (Portugal, Spain, Italy, France), the north and central countries (Austria, Finland, Denmark and Ireland) and more than 650,000 ha of forest and other environmental areas were burned (Parente et al., 2018). In the Mediterranean basin, every year a large number of wildfires are recorded and thousands of hectares of natural environment are lost (Ager et al., 2014; Mouillot et al., 2005). Based on the European Forest Fire Information System during the last 8 years, 700,000 ha of the land were burned in the European countries (EFFIS). Therefore, it is crucial for planers and organization in charge of protection and management of natural habitats to have access to accurate and up to date information related to wildfire susceptibility to define develop mitigation measures.

Regarding the evaluation of fire susceptibility and forecasting potential future fires, Pastor et al. (2003), proposed three categories for classification models including physics-based models, semi-empirical, and empirical. Physics-based models are based on the fluid dynamics, combustion, and heat transfer but it depends on the area and the data with high accuracy (Hong et al., 2018). The semi-empirical models are based on the principle of energy conservation which needs numerous laboratory works in order to name the parameters (Hong et al., 2018). Empirical models comprise the analysis of a wildfire dataset and related variables on the basis of statistical or data mining methods (Chen et al., 2017; Hong et al., 2018; Mell et al., 2007). In this technique, statistical correlations are considered among the available parameters which are known or can have an impact on the happening of wildfires to produce time-based prediction or to define the high-risk areas (Hong et al., 2018). Empirical models enable to determine the spatial distribution of the phenomenon and prediction of the future changes in the dependent variable and the possible influencing factors (Hong et al., 2018; Pourtaghi et al., 2016). Empirical methods which have been utilized in several studies are including Logistic Regression (Vacik et al., 2013), fuzzy logic (H. reza Pourghasemi et al., 2016), decision tree (Camp et al., 1997), Random Forests (RF) (Oliveira et al., 2012a), Kernel Logistic Regression (Tien Bui et al., 2016), artificial neural networks (ANN) (Satir et al., 2016). Although many studies have been carried out regarding wildfire susceptibility and various developed models have been utilized, it is not still obvious that which method is most appropriate to assess the wildfire risk nor are clear all the factors.

Sardinia was chosen as study area because this island encounters hundreds of wildfires every year. Fires have strong impacts on the environment and considerable damage to properties, along with loss of human lives. Thus, the ongoing progress in causes of fire along with the development in technologies and modelling methods and awareness of public and politicians about the fire risks on the scientific bases is crucial in order to reduce the risk associated with fires. Various studies have been carried out in Sardinia concerning wildfire risk assessment. For instance, Ager et al. (2014) utilized nonparametric statistical models to determine burn probabilities based on the weather and human factors on the fire events and size in Sardinia, Italy, and Corsica, France. The produced map can be utilized to (1) assign extinction resources to specific region during the fire season; (2) improvement of the current fire indices; (3) simulation the wildfire to understand the relationship between land use and climate change and burn probability and intensity. Fiorucci et al., (2008) used the fuel moisture model and the fuel load model for the risk distribution evaluation and generated wildland fire danger rating system, called RISICO to estimate the fire risk. Salis et al., (2014) applied large-scale wildfire exposure factors assessment to map burn probability and fire intensity based on the key factors including weather, fuel, topography, and spatial ignition patterns.

The aim of current study is firstly, to explore the spatiotemporal patterns of wildfire in Sardinia to investigate the important factors contributing to the fire occurrence with consideration of the soil moisture as plant physiological factor and indicator of drought. Secondly, to test one simple correlation method (Logistic Regression) and one more sophisticated machine learning method (Random Forest) to model the wildfire risk. Thirdly, to produce probability map of the fire occurrence using the most related variables.

1. Theory

1.1 Wildfire Risk Management

Influence of the wildfires on human life, ecosystem, and other precious sources leads to the global attempt to consider factors contributing to wildfire incident and behavior (Thompson & Calkin, 2011). The existence of uncertainties in the wildfire management is due to different fire behavior, lack of precise data, different ecological response to fire, different reaction of fires to management treatments (suppression, fuel reduction, etc.) and finally limitation in understanding priority of resources at risk (Thompson & Calkin, 2011). Uncertainty refers to a lack of information, and it is possible to express it quantitively or as in terms of risk management (Thompson & Calkin, 2011). It is required to understand the nature of uncertainties and figure out the suitable time to add it to quantitative risk management frameworks and vice versa. In addition, it is also necessary to identify the supplementary information to decrease the effects of uncertainty (Thompson & Calkin, 2011).

Risk assessment is a decision support tool which integrates information about the probability and response of sources to risk factors to help decision-making (Sikder et al., 2006; Thompson & Calkin, 2011). Researchers and scientist in management are developing risk assessment used for strategic and operational decision making as support tools (Thompson & Calkin, 2011). Lack of access to fire risk assessments makes decision and management inefficient (Bar Massada et al., 2009; Thompson & Calkin, 2011). Studies proved that it is important to risk-informed and decision-making to know how risk in different management situation can influence various resource changes (O'Laughlin, 2005; Roloff et al., 2005; Thompson & Calkin, 2011). Risk assessments models can assist to forest management decision-making in the active and pre-fire period (González et al., 2006; Thompson & Calkin, 2011). Wildfire management plans are in different scales from incident-specific to regional/national assessment with various applications including fire prevention, fire detection, development, and initial attack dispatch, large fire management, and strategic planning and fuel management (Thompson & Calkin, 2011). The purpose of fire risk management is to analyze both exposure and effects, and finally to produce suitable management response to reduce exposure and mitigation of various impacts (Fairbrother et al., 2005; Finney, 2005; Thompson & Calkin, 2011). Obvious definition of management aims, and understanding the relevant management of priorities play an important role in the effectiveness of wildfire risk management (Thompson & Calkin, 2011).

1.2 Wildfire Risk Factors

Each year from 273 to 567 Mha, with average of 383 Mha of different parts of the world are burning (Schultz et al., 2008). Flannigan et al. (2005) classified the factors influencing the fire activities in four categories: fuels, climate-weather, ignition agents and people. Topography is one of the important aspects of the ignition agents which shows where fires start and spread (Torres et al., 2018).

Fuel amount, type, continuity, structure and moisture content are significant components for fire to happen and spread (Linn et al., 2010; Paton, 2014). For instance, there should be at least 30% of fuel for fire ignition and spread. In other words, fuel continuity is required for fire (Paton, 2014).

This factor is important in most drier regions all around the world where precipitation before fire seasons favors the growth of the vegetation which in turn will be burnt by a fire (Meyn et al., 2007; Paton, 2014; Swetnam et al., 1998). Fuel structure is also an important factor so that understory trees and shrubs in a forest can start to burn faster than tree crowns. In addition, they cause faster movement of the fire to other parts. Fuel moisture is a very important factor which has an effect on both fire occurrence and behavior (Paton, 2014). Obviously, greater heat is required to dry the fuel with high moisture which decreases the fuel consumption rate (Fiorucci et al., 2008). However, fuel with lower moisture can increase the facility of ignition, spread rate and intensity (Catchpole et al., 2001; Fiorucci et al., 2008). Studies show that as the density of the forest increases, the air movement decreases so that evaporation and fuel moisture rates go down (Cawson et al., 2017). Song et al., (2017) discovered that as Tree Cover Density (TCD) is high, fire can spread all around the landscape quickly.

Macroclimatic conditions including temperature, precipitation, wind, and atmospheric moisture are influencing both fuel and ignitions (Cawson et al., 2017; Paton, 2014). Fuel moisture which can be a key factor of the fuel and the amount of the vegetation in the area both depend on the weather and climate (Paton, 2014). Additionally, one of the two main factors of fire occurrence is lightning specified by meteorological conditions (Paton, 2014). Compared to other climatic variables, temperature is the most important factor. Studies have proved that high temperature can start and increase fire activity (Flannigan et al., 2005; Gillett et al., 2004; Parisien et al., 2011; Paton, 2014). There are three important reasons why high temperature can cause wildfire: (1) evaporation rate increases in the warmer temperatures which decrease the amount of the water in the ground and dead fuel moisture if there is not enough precipitation (Paton, 2014); (2). Lightning can cause ignition in warmer temperature (Colin Price et al, 1994); (3) The length of the fire season can be extended as the temperature is warmer (Westerling et al., 2006; Wotton et al., 1993). In general, climate change can increase the frequency and severity of wildfire (Paton, 2014).Gillett et al. (2004) mention that wildfires in Canada between 1974 and 2014 have taken place due to the temperature increase caused by the human.

Humans have also influence on these factors directly or indirectly (Knorr et al., 2014). For instance, grazing can change the amount of fuel and shape of the lands (ARCHIBALD et al., 2009), decrease fires via roads, defrayals and suburban structures (Syphard et al., 2007). Archibald et al. (2009) used the satellite imagery and found out that in Southern parts of Africa the number of burnt areas goes down when the population is above 20 people per km^2. In another study, Lehsten et al. (2010) reached the same result so that they noticed that for the whole African continent the frequency of the fires decrease as population density goes up. Pausas et all, (2012) addressed that 80% decrease in the rural population density led to the increment of wildfire occurrence and (Moreira et al., 2001) revealed that high grazing activity can decrease the fire activity. The consequences of urbanization increase the human-caused wildfires in the different part of the world so that in the southern and eastern United States 66% of ignition were made by individuals (Grala et al., 2017).

In comparison to weather information, the topography is static apart from the human-made changes. For instance, a wildfire burns rapidly as the slope is increasing so that it makes unburned fuels more flammable (Figure 1). In addition, rapid movement of the wind in slopes which can fast distribution of the wildfire. Aspect is another important factor so that it shows how much radiated heat receives from the sun. Studies proved that south- and west-facing slopes have drier vegetations than north- and east-facing of slopes in northern hemisphere (Ebel, 2012) (Figure 1).

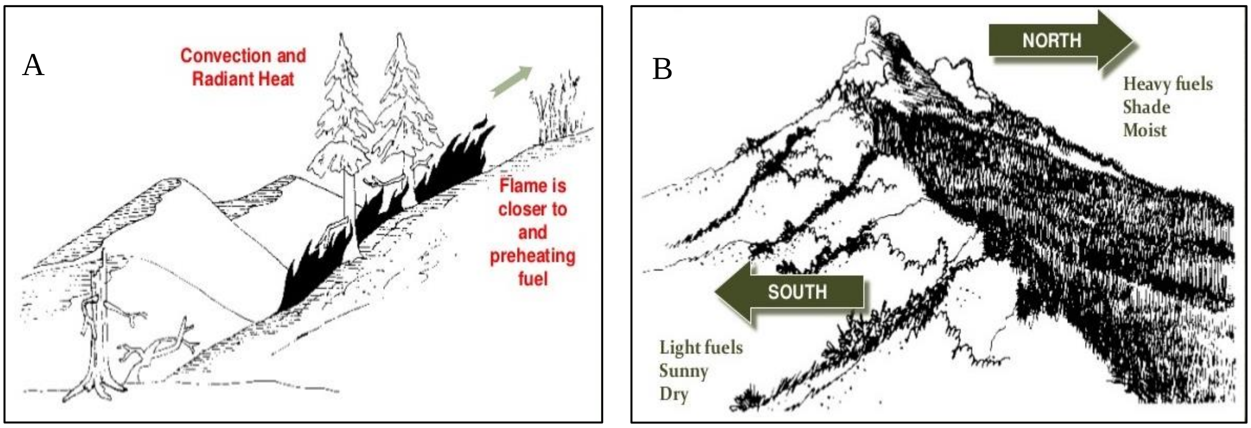

Figure 1. A) Influence of the slope on the wildfire B) Influence if the aspect on the wildfire Source: Introduction to Wildland Fire Behavior, © National Wildfire Coordinating Group

1.3 Fire risk analysis

The “wildfire risk” illustrates the danger of the fire beginning and spreading. It also indicates the amount of the damage on environmental and human resources. The term “wildfire danger” refers to factors have influence on the inception, spread and resistance to control. That is, “wildfire danger” is part of the “wildfire risk” (Martín et al., 2019). Wildfire susceptibility estimation assists to understand the fundamental patterns of the wildfire danger ignition and its causative factors in the complex landscapes (Jaafari et al., 2018). Wildfire ignition modelling has been developed over time and these days, most models combines geographic information with remote sensing data to produce wildfire susceptibility (Chowdhury et al., 2013; Martín et al., 2019). Chuvieco et al., (2014) implement GIS and remote sensing techniques to estimate the daily wildfire risk maps for mainland Spain. Nhongo et al., (2019) used Logistic Regression for wildfire occurrence model. This method is one of the statistical approaches mostly used, both for prediction of fire risk and to specify the causes of fire, at the global scale. In the more recent studies, researchers applied machine learning tools to map the wildfire susceptibility due to complexity of wildland fires and the abundance of causative factors (Jaafari et al., 2018). As an example, Gigović et al., (2019) compare the Random Forest and support vector algorithms to evaluate forest fire susceptibility maps. In general, the evolution of the machine learning method is better than statistical ones. However, as Bui et al.,(2016) mentioned there is no accurate approach to model and predict wildfires on a regional scale due to complexity of the relationships between the ignition factors.

1.3.1 Logistic Regression

In Logistic Regression, the independent variables are binary such as presence or absence of the fire (Bisquert et al., 2012). The a Logistic Regression enables to model the conditional probability of the fire occurring from the dependent variables (Bisquert et al., 2011). Logistic regression can also be used to estimate the importance a of each variable. The Wald test (also called the Wald Chi-Squared Test) is used to determine whether the explanatory variables are significant or not (Martínez et al., 2009). The application of this approach is in the more complex models such as an ANN, in order to decide which variables or combination of variables should be used (Bisquert et al., 2012). Logistic Regression has been widely used to map the fire ignition probability (Chang et al., 2013; Lozano et al., 2007; Martell et al., 1987; Nhongo et al., 2019; Padilla et al., 2011; Preisler et al., 2011; Vilar del Hoyo et al., 2011) since the combination of categorical and continuous variables; plus, non-normally distributed variables can be utilized (Catry et al., 2009; Chang et al., 2013). The fundamental function of the Logistic Regression is as follow:

$$f(Z_i) = \frac{1}{1+ e^{Z_i}} \qquad (1)$$

where the $f(Z_i)$ is the estimation of probability of fire occurrence, and Z_i is linear function of the explanatory variables for each observation which based on the adjustment of maximum likelihood; B0 is intercept, B_i coefficient of the linear regression and X_i is the value of the independent variables:

$$Z_i = B_0 + B_1X_1 + B_2X_2 + \cdots + B_iX_i \qquad (2)$$

1.3.2 Random Forest

One of the simplest machine learning techniques is Random Forest (RF). Random Forest was introduced by (Ho et al., 1994). RF is based on the classification and regression tree (CART). CART has been used in different studies for the fire risk prediction (Amatulli et al., 2006; Lozano et al., 2008; McKenzie et al., 2000; Oliveira et al., 2012a). Combination of many trees forms the RF so that each tree is created by bootstrap samples. A third of the total sample put aside for validation (the out-of-bag predictions -OOB). Random subset of the predictors is used to specify each tree. The average of the outcomes of all trees considers as the final result (Breiman, 2001; Cutler et al., 2007; Ghorbanzadeh et al., 2019). In the wildfire Susceptibility prediction RF is one of the most usable non-parametric ensemble learning methods and has proved to have high accuracy in modelling the complex relationships between variables studies (Ghorbanzadeh et al., 2019; Jaafari et al., 2019; Rihan et al., 2019). Although there are some advantages for this method, it has also some disadvantages. For instance, so called "black box" issue since examination of each tree separately is not possible (Prasad et al., 2006). Regression coefficients and confidence intervals also are not calculated in RF (Cutler et al., 2007). However, variable importance measures can be calculated and therefore it is possible to identify each factor's contribution to the model (Grömping, 2009). Two important parameters are required by RF to run including the square root of the number of factors (m_{try}) and the number of trees to run the model (n_{tree}). These parameters can decrease the generalization error (Breiman, 2001; Gigović et al., 2019). The Gain Index is used by RF to determine the optimal subset of variables (purest subset). This method is useful to only select variables able to represent the entire dataset (Jaafari et al., 2018). It is calculated as follow:

$$\text{Entropy} = -\sum_{i=0}^{k} P\,(c_i) log_2(P(c_i)) \quad (3)$$

$$\text{GainRatio(f, c)} = \frac{Gain(f,c)}{Splitinfo(f,c)} \quad (4)$$

In this equation, f is representative of the training dataset, c is class attribute, k is the number of independent features, C represents entropy corresponds to the uncertainty about the value of the class attribute, P (ci) is the probability that C = ci, and SplitInfo indicates the information provided by dividing attribute C of the training data set f into m subsets (Jaafari et al., 2018). SplitInfo is calculated as follow:

$$\text{Splitinfo(f, c)} = -\sum_{j=1}^{m} \frac{|f_j|}{|f|} log_2 \frac{|f_j|}{|f|} \quad (5)$$

1.4 Uncertainty in wildfire risk management

Thompson et al. (2010) adopt the uncertainty typology by Ascough et al. (2008) for wildfire risk management. According to this all uncertainties can be classified into four main groups including linguistic uncertainty, variability uncertainty, knowledge uncertainty, and decision uncertainty. Linguistic uncertainty is contextual dependency, vagueness of words that can make it difficult to explain the results so that there is more than one valid way to understand and explain the management system (Brugnach et al., 2011; Thompson & Calkin, 2011). The term “risk” can be used in different meanings in wildfire risk management (Emilio Chuvieco et al., 2010; Schmoldt, 2001). For instance, risk is defined as the probability of the fire event by Hardy, (2005), while (Finney, 2005) defined as probabilistic expectation of net resource values change in response to fire.

Variability uncertainty is defined as inner variability which shows itself in natural systems (Thompson & Calkin, 2011). The examples of variability uncertainty can be the frequency and spatial pattern of ignitions, the variability of weather conditions etc. In several studies, probabilistic methods were used to dealing with variability (Bar Massada et al., 2009; Carmel et al., 2009; Krougly et al., 2009; Podur et al., 2010; Thompson & Calkin, 2011).

Knowledge uncertainty defines as the limitation of the knowledge and scientific understanding (Thompson & Calkin, 2011). This kind of uncertainty is related to how we conceptualize the natural processes or how the choose to model them (Thompson & Calkin, 2011). Cruz et al. (2010) stressed that there are knowledge gaps and errors in modeling system of crown fire behavior. Non-probabilistic methods are common to manage knowledge uncertainty (Thompson & Calkin, 2011). In this case, information from the experts is the best source for judgment (Borchers, 2005). This information can be utilized in several ways including knowledge-based systems, hierarchical multi-attribute models, logic models, fuzzy set theory, and hybrids thereof (Hessburg et al., 2007; Macgregor et al., 2008; Thompson & Calkin, 2011; Vadrevu et al., 2010).

Another kind of uncertainty which is related to imperfect information about social cost/benefit analysis is decision uncertainty (Thompson & Calkin, 2011). Lack of information about social preference/values causes limitation regarding the management of social welfare. The value measurement method is utilized for handling this type of uncertainty such as Analytic Hierarchy Process, utility theory, outranking models, and social choice theory (Diaz-Balteiro et al., 2008; Mendoza et al., 2006; Thompson & Calkin, 2011). Dealing with decision uncertainty is not easy because of different stakeholders have various perspectives, perception, and aims which can be changed during the time or as new information releases (Thompson & Calkin, 2011). Decision uncertainty is complicated because people often are not sure about their own preferences (Brown et al., 2008; Rieskamp et al., 2006; Thompson & Calkin, 2011). Table 1 gives overview of different uncertainties related to wildfire management and how to mitigate the uncertainties.

Table 1. Uncertainties related to in wildfire management. Adopted from (Thompson, Calkin, et al., 2011)

Wildland fire context	Uncertainty type	Methodology
Fire occurrence	Variability	Probability-based
Fire behavior	Variability; knowledge	Probability-based; Expert system
Accounting for role of climate change	Variability; knowledge	Probability-based; Expert system
Interaction of fire with other disturbance	Variability; knowledge	Probability-based; Expert system
Temporal vegetation & fuel dynamics	Variability; knowledge	Probability-based; Expert system
Ecological response to fire	Knowledge	Expert system
Efficacy of management treatments	Knowledge	Expert system
Valuation of non-market resources	Decision	Value measurement

2. Data and methodology

2.1 Study area

Sardinia island with the area of approximately 24,100 square kilometers is the second largest island in the western Mediterranean Sea located between 38° 51' and 41° 18' latitude north and 8° 8' and 9° 50' east longitude (Figure 2). Forest including *Quercus ilex L.*, *Quercus suber L.*, and *Quercus pubescens Willd* covers the most part of the island (Shokouhi et al., 2019). At higher elevations, the oak formation merges with *Castanea sativa Mill.* and *Ilex aquifolium L.* The coniferous stands (*Pinus spp.*) are restricted to 3%. Mediterranean maquis and garrigue cover 28% of the island (Ager et al., 2014). The topography of the island includes small hills and mountains. Highest elevation is 1850 m a.s.l. and located in the middle of the island. The climate of the island includes mild and rainy winters; dry and hot summers (Ager et al., 2014).Summers (From May until September) are threatened by the shortage of water and wildfire (Ager et al., 2014). The average temperature fluctuates between 12ºC and 17ºC, and precipitation is between 500 mm in the south to 1300 mm in the highest area (Ager et al., 2014). According to information collected from burnt areas between 1971 and 2015, 2,931 fires per year has happened in the study area burned about

18,800 ha (Cardil et al., 2017). These wildfires influenced various part of the island and damaged agriculture areas and urban infrastructures.

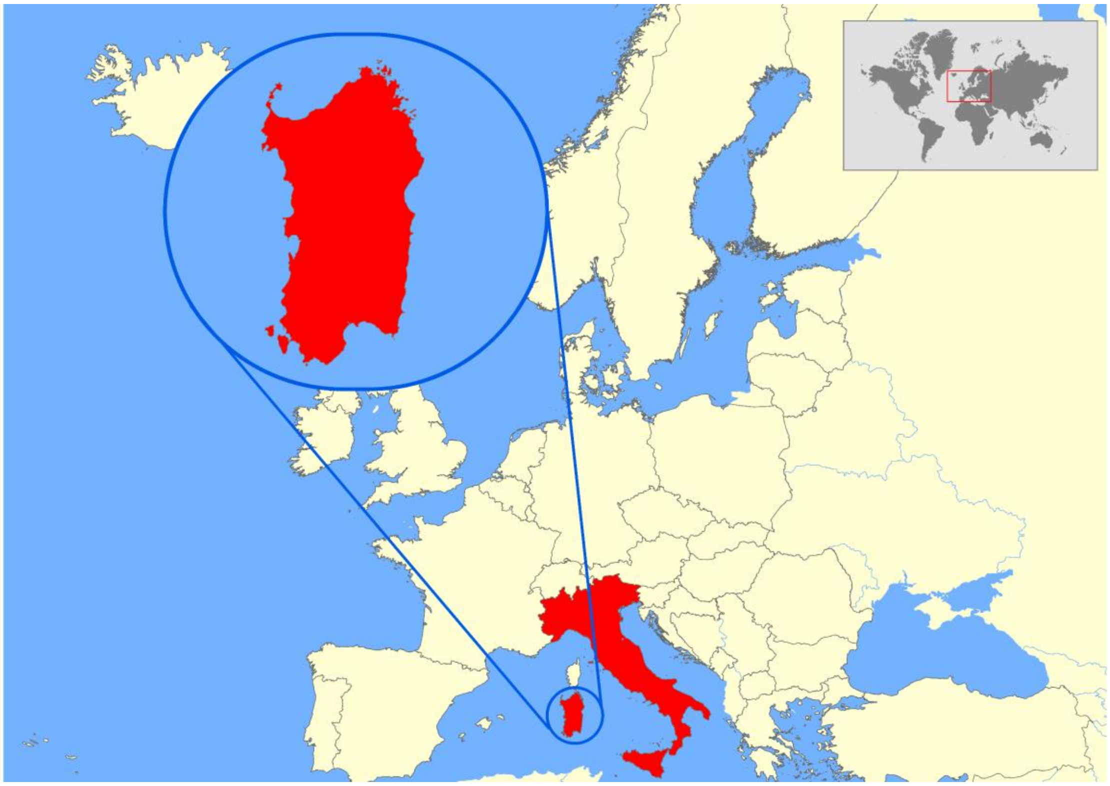

Figure 2. Study area

2.2 Data

2.2.1 Dependent variable - wildfire occurrence

Fire data was obtained from NASA Fire Information for Resource Management System (https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms). Active fire sources are collected by the MODIS sensor on board the Aqua and Terra platforms with spatial resolution of 1km. Each location of the fire collected by MODIS sensor represents the centre of the 1×1 km^2 pixels determined by the algorithm as one or more fires in surrounded area. In some cases, product can underestimates the fire occurrence including short duration fires, cloud coverage and heavy smoke (Anderson et al., 2015). Overestimation can also happen due to high temperature of the feature such as sandy soil, rock and etc. (Nhongo et al., 2019). Therefore, to avoid the commission error, only fire pixels with high confidence was used (> 70% confidence). 1041 fires were extracted from 2010 to 2018 using the FIRMS database. Figure 3-a illustrates the number of the wildfires in each month, and Figure 3-b shows the location of them in the study area during the study period. All the fires occurred between May and October. July has been experienced the large number of the wildfires (Figure 3-a). Since in some areas fires were happend in the several times in different years, wildfire data was checked to aviod the double fires. In the end, there was 944 burned pixels. Binary classification method is utilized to model the wildfire susceptibility. With this concept, 944 grid cells were classified as fires and were assigned the value of “1”, and the same number of grid cells without any evidence of wildfir were sampled for non-fire and coded as “0”.

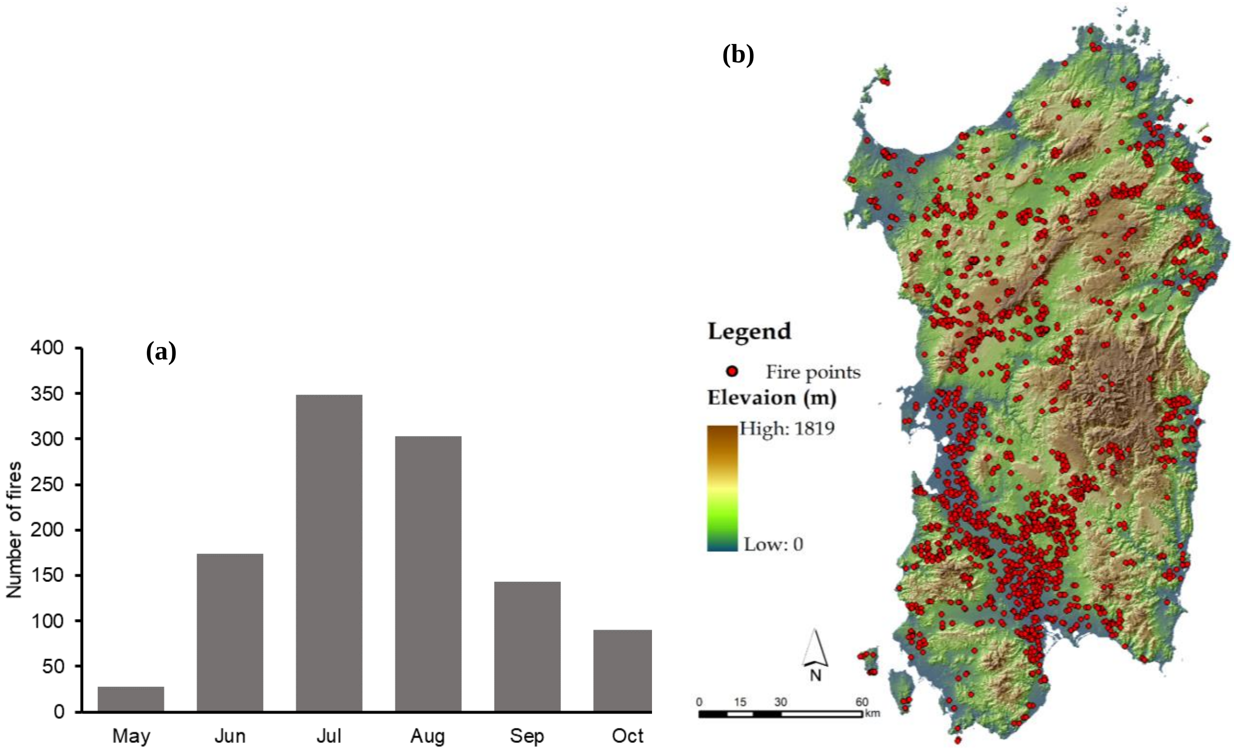

Figure 3. (a) Fire number in Sardinia for the period 2010 – 2018; (b) Location of the wildfires in the study area during the last 8 years

2.2.2 Explanatory variables

Several factors can play important role in wildfire risk (Adab et al., 2013; Catry et al., 2009; F. Chen et al., 2015; W. Chen et al., 2017; Jaafari et al., 2017, 2018; Mhawej et al., 2015; Oliveira et al., 2012a; Pourtaghi et al., 2016; Tien Bui et al., 2016). Based on the previous studies and availability of the data for the study area; as well as, the traits of the region, 16 factors were adopted for modeling the wildfire susceptibility (Table 2)

Table 2. The explanatory environmental variables used for modelling.

Variable type	Variable name	Data source	Resolution
Topography	Elevation	(EU-DEM v1.1)	25 m
	Slope (Gradient)		25 m
	Slope direction (Aspect)		25 m
	Topographic position index (TPI)		25 m
	Topographic roughness index (TRI)		25 m
Land cover	Fuel/Vegetation Map	(CORINE Land Cover)	25 m
Infrastructure	Proximity to road	(Sardinian Geo-Portal)	1:250 000
	Proximity to river	(Pan-European, Water & Wetness)	2.5 m
	Proximity to settlement	(European Settlement Map)	10 m
Climatic	Mean monthly temperature	(UERRA)	5 km
	Cumulative monthly precipitation		5 km
	Average monthly relative humidity		5 km
	Average wind direction		5 km
	Average wind speed		5 km
Geology	Average monthly soil moisture	(UERRA)	5 km
Demographic	Average population density	(ORNL's LandScan)	1km

Land cover data

Previous studies have showed the association between land cover and fire occurrence (Martínez et al., 2009; Oliveira et al., 2012a). Land cover data was extracted from CORINE land cover dataset (CLC, EEA, 2018). This data includes 44 land cover classes with Minimum Mapping Unit (MMU) of 25 ha of areal phenomena. Updates have been produced in 2000, 2006, 2012, and 2018. In this study, the latest version with 25 m resolution was selected and most important classes based on the previous studies in the area (Bajocco et al., 2008; Guglietta et al., 2015) were extracted from them (Figure 4).

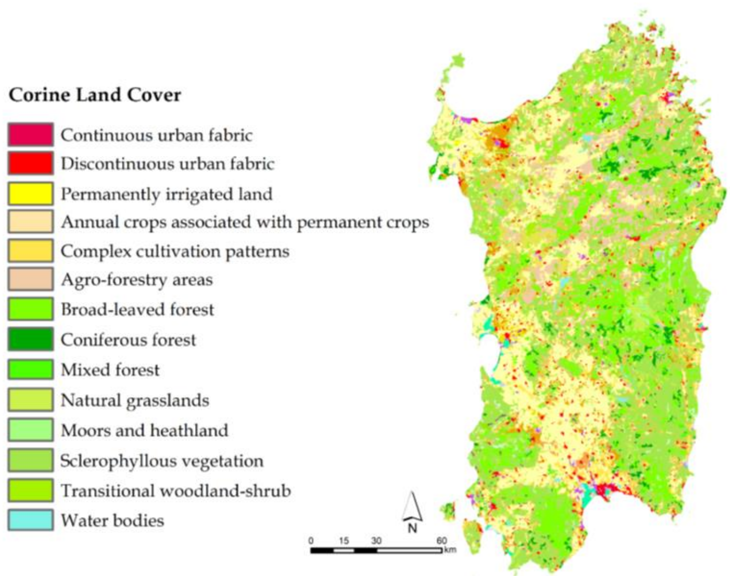

Figure 4. Land cover of Sardinia

Topography data

Vegetation distribution and productivity, moisture gradients and energy and water balances are under the influence of the topographic variables. On the other hand, the mentioned factors can also affect the wildfire (Alexandre et al., 2016). Topographic data was derived from EU-DEM v1.1 (García et al., 2015) which is a digital elevation model (DEM) with 25m horizontal resolution and provided by Copernicus land monitoring service . This product is based on the SRTM and ASTER GDEM data with 2.9 meters RMSE as overall vertical accuracy. ArcGIS and SAGA GIS were used to create topographic position index (TPI), topographic roughness index (TRI), elevation, slope, aspect (Figure 5).

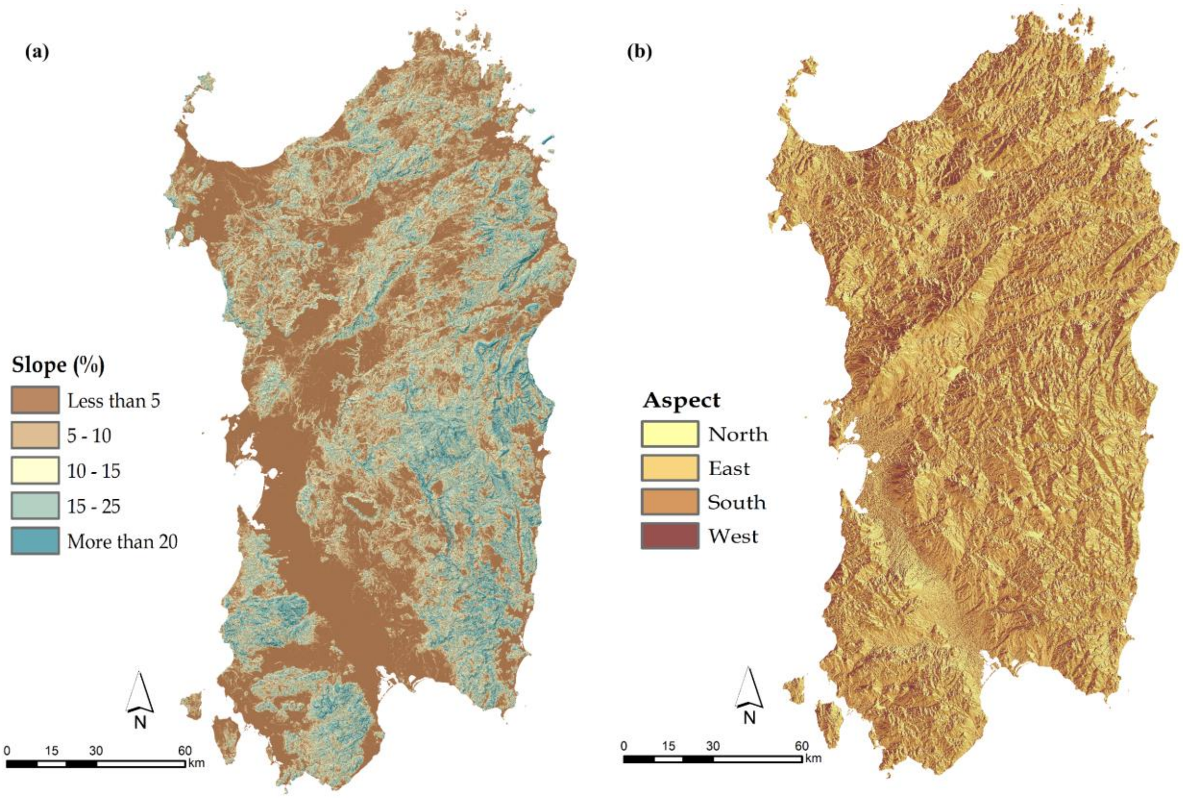

Figure 5. Topographical variables for wildfire susceptibility: (a) slope; (b) aspect

TPI measures the difference between the elevation at central point and mean elevation at the predefined neighborhood (Wilson et al. , 2000). The result is called relative topographic position of the central point (De Reu et al., 2013). TPI is calculated as follows:

$$TPI = Z_0 - \overline{Z} \tag{6}$$

$$\overline{Z} = \frac{1}{n_R} \sum_{i \in R} Z_i \tag{7}$$

where (Z_0) is elevation at the central point, and $(\overline{Z})$ is the average elevation in the predetermined radius (R). TPI includes positive and negative values showing whether the position is lower or higher than the average. (Figure 6.b)

TRI (Figure 6.a) indicates the changes of the land surface in the particular area (Oliveira et al., 2012). TRI is calculated as follows:

$$TRI = Y[\sum(x_{ij} - x_{00})^2]^{1/2} \qquad (8)$$

where x_{ij} is the elevation of each neighbor cell to cell (0,0). Zero value shows the flat areas, while steep ridges indicates with positive values (Kmoch et al., 2019)

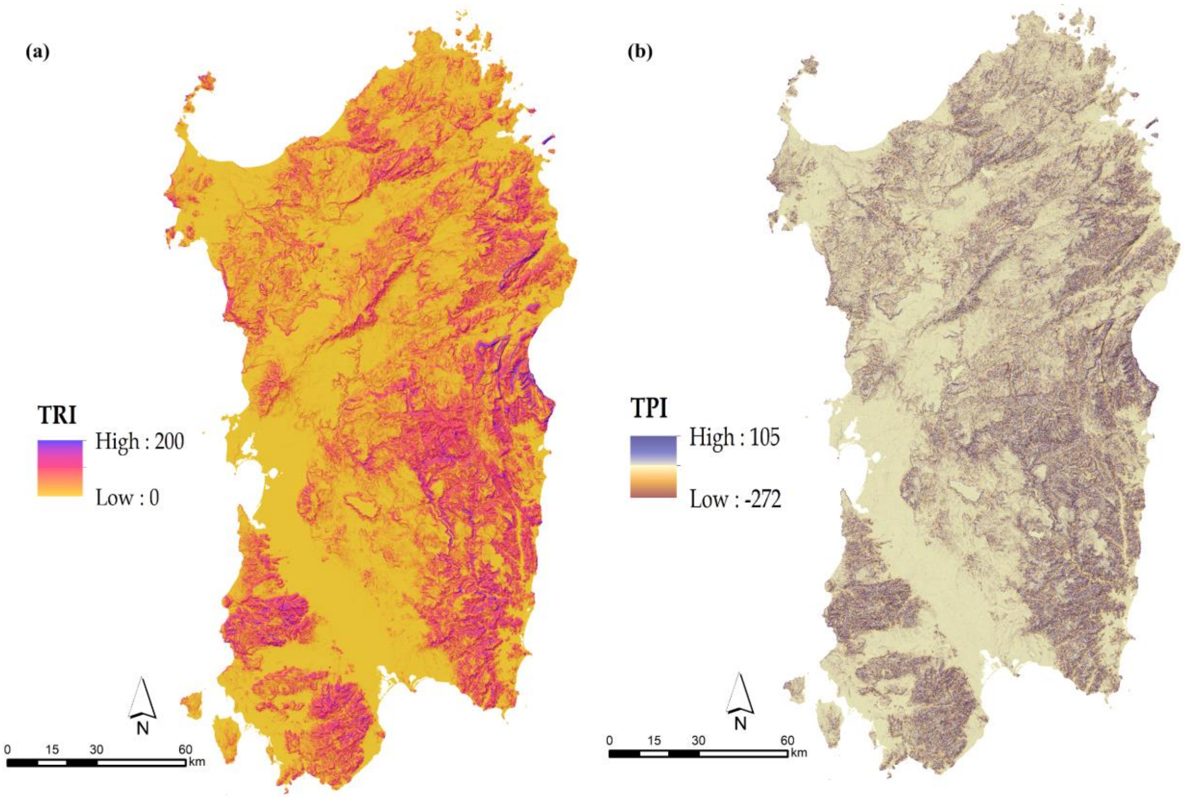

Figure 6. Topographical variables for wildfire susceptibility: (a) TRI; (b) TPI

Population density and proximity to roads

Studies have revealed that majority of the fires in the European Union are human-caused (Oliveira et al., 2012). Thus, population density can be one of the important factors of the fire occurrence. ORNL's LandScan global population distribution data with 1 km resolution (Rose et al., 2018) was used for population density. For the simplicity, population densities from 2010 to 2018 were averaged (Figure 8.a). We also used the available road map of the study area from Sardinian Geo-Portal (http://www.sardegnageoportale.it/) and European settlement map (Ag et al., 2017) for calculation of the proximity to the roads and settlements as one of the human activities which can affect fire risk . Proximity to the roads (Figure 7.a) and settlements were calculated by using ArcMap 10.5.

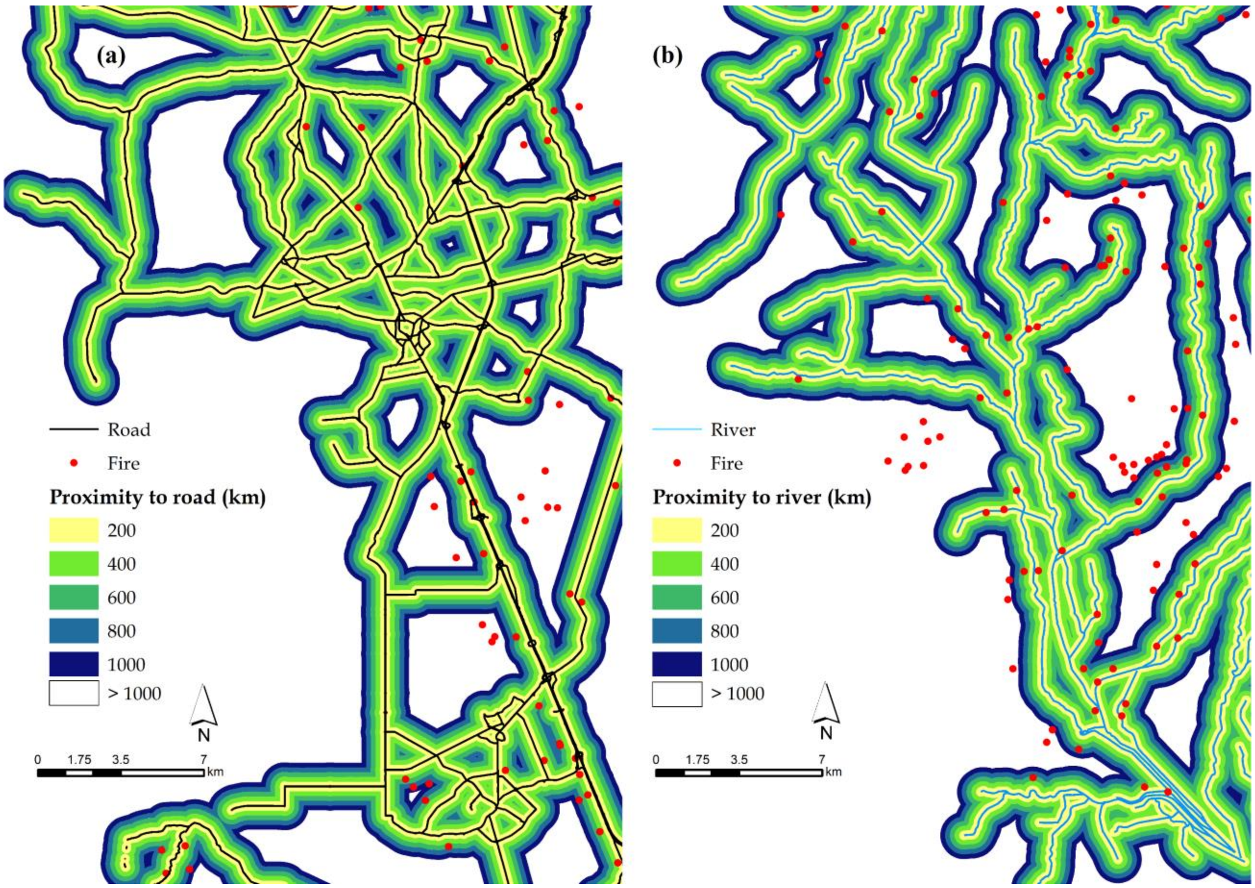

Figure 7. (a) Proximity to the roads; (b) Average population density between 2010 - 2018

Proximity to the rivers

Moisture is one of the factors which influence fire ignition. Water sources are especially in warm weather can produce a large amount of the moisture due to evaporation. Therefore, proximity to water sources can be important in wildfire risk analysis. For this purpose, EU-Hydro River Network (Gallaun et al., 2019) provided by Copernicus Land Cover Service with 2.5 m resolution was utilized (Figure 7.b).

Soil moisture

Soil moisture can be used as indicator of drought (Laguardia et al., 2008; Oliveira et al., 2012). The soil moisture influences the moisture content of the live fuels, and thus, it determines the required heat for plants' ignition (Bartsch et al., 2009; Chuvieco et al., 2004; Oliveira et al., 2012). In this study, the soil moisture dataset for the study period (2010 - 2018) used, provided by Copernicus Climate Data Store (Schimanke et al., 2019) (Figure 8.b).

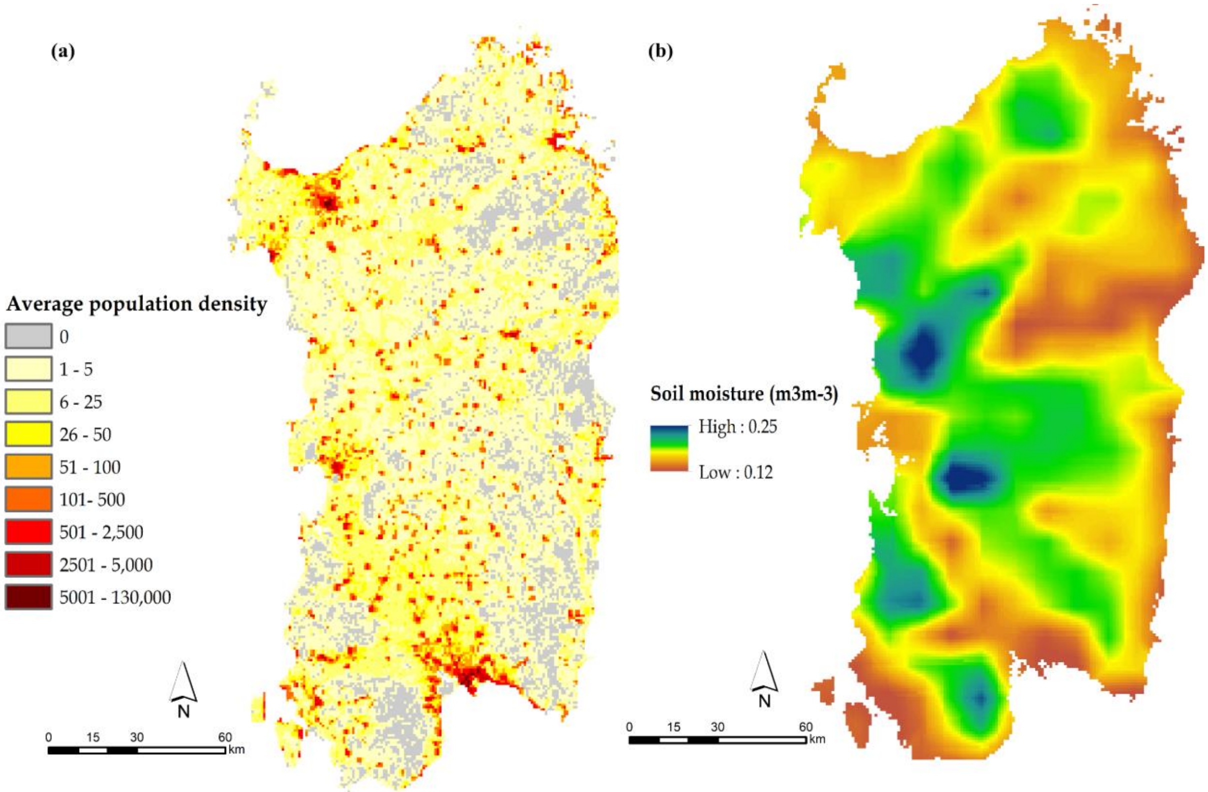

Figure 8. (a) proximity to rivers; (b) amount of the soil moisture

Climate

Weather conditions can have an influence on the amount of the moisture and fuel accumulation which in turn affects the probability of the fire occurrence (Oliveira et al., 2012; Syphard et al., 2008). Therefore, in this study, UERRA (Uncertainties in Ensembles of Regional Re-Analyses) climate data from 2010 to 2018 were used from Copernicus Climate Data Store (Schimanke et al., 2019). The UERRA includes European regional meteorological reanalyses of Essential Climate Variables (ECVs) for several decades. The data is available from 1961 to present with two different spatial resolutions including 11km for the UERRA-HARMONIE system and 5.5km for the MESCAN-SURFEX system. In the current study, 5.5 km resolution was utilized (Figure 9).

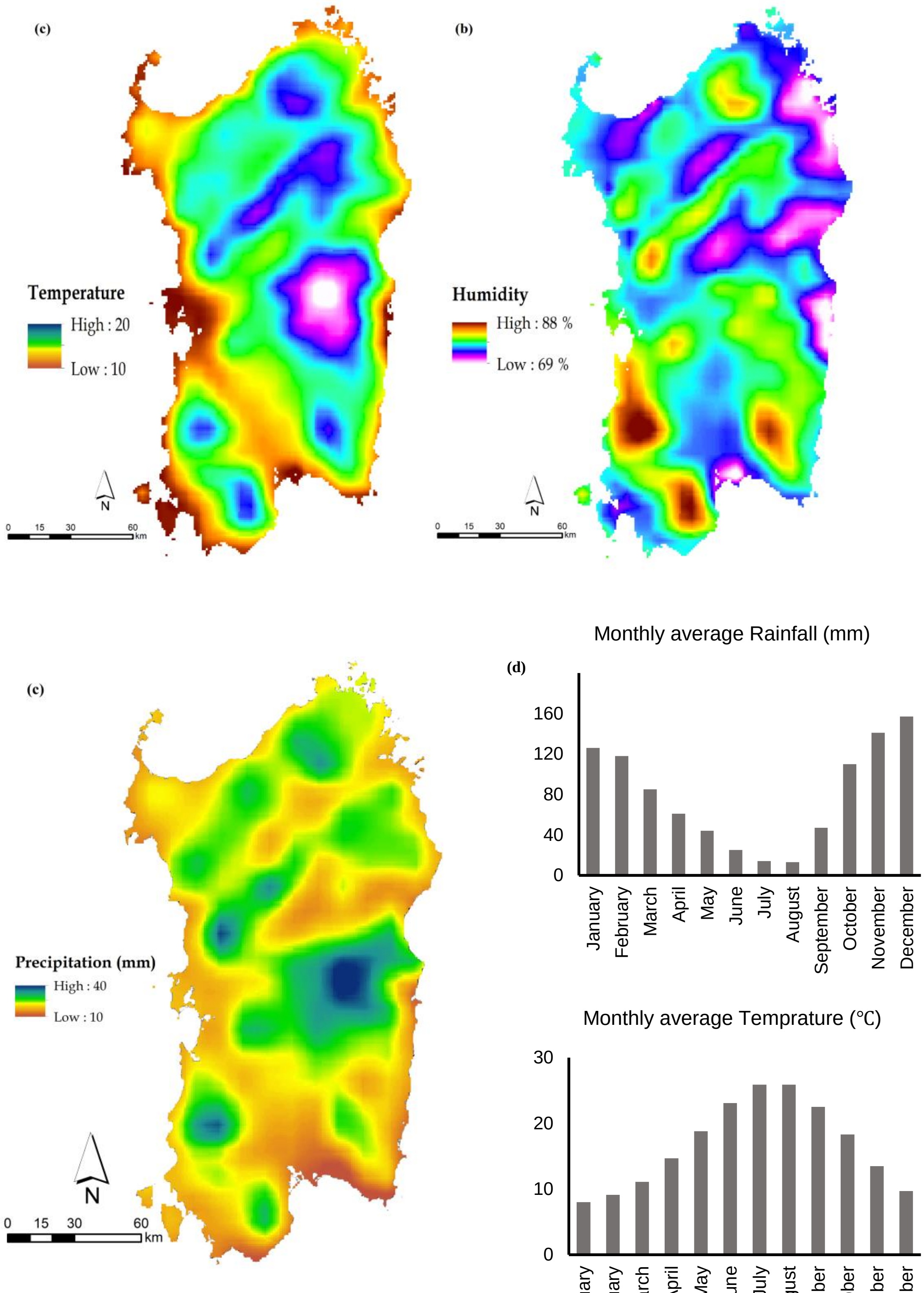

Figure 9. Climatic variables (a) spatial variability of average temperature (℃) in the study area in 10 years (b) spatial variability of average humidity (c) spatial variability of rainfall (d) monthly average temperature and rainfall

To avoid occurrence of correlation between independent variables in the model (Gigović et al., 2019), multicollinearity analysis was performed. Collinearity can decrease the accuracy of the estimation (Dormann et al., 2013). The variance inflation factors (VIF) (Liao et al., 2012) and tolerance were used to determine the multicollinearity among the wildfire independent variables. VIF > 5 and tolerance < 0.2 indicates multicollinearity (Jaafari et al., 2018). "USDM" package (Naimi et al., 2014) in R was utilized for this purpose. Table 3 shows the result of the analysis, and elevation was removed from the further analysis because of very high collinearity.

Table 3. The result of multi-collinearity tests

Variable	**Collinearity statistics**	
	Tolerance	**VIF**
Temperature	0.29	3.41
Precipitation	0.23	4.38
Wind direction	0.65	1.53
Wind speed	0.70	1.44
Humidity	0.43	2.30
Elevation	**0.11**	**9.32**
TRI	0.75	1.33
Slope	0.27	3.65
Aspect	0.90	1.11
TPI	0.92	1.08
LULC	0.88	1.13
Proximity to settlement	0.74	1.36
Proximity to road	0.79	1.26
Proximity to river	0.77	1.29
Soil moisture	0.24	4.08
Mean population density	0.88	1.14

2.3 Analysis

To determine which variables are influencing the model, two separate methods were utilized for each model. Wald Chi-Squared Test was calculated to determine the significant variables for Logistic Regression. Gain ratio with 10-fold cross validation technique, to select the much more optimal variables which can be contributed in the wildfire occurrence in our study area for Random Forest.

2.3.1 Modelling

For model assessment, the fire and non-fire data was randomly split into training (70%) (1321Grid cells) and test (30%) (567 grid cells) sets. For modelling, two methods were used: Logistic Regression and Random Forest. All the code generated for this study is available in GitHub (https://github.com/behzad89/Thesis).

Logistic Regression

Logistic regression is a flexible approach compared to other ones since both continues and categorical variables can be used (Nhongo et al., 2019). The model was made using the "stats v3.6.2" package in R-3.6.2 software (R Core Team, 2019)

Random Forest

Random Forest model has illustrated potentiality to apply in the assessment of fire-related phenomena (Oliveira et al., 2012a). In the first step Gain ratio method was applied for determination of the much optimal variables using "FSelectorRcpp" package in R software (Zygmunt et al., 2019). The variables with "0" gain ratio (average merit) were removed from the next modelling process. In the last step, the model was trained using the important factors using "randomForest" package (Breiman, 2001) in R-3.6.2.

2.3.2 Accuracy assessment

The following methods were applied to assess the performance of the both models:

Statistical index-based measures

Statistical index-based measure is an approach to measure the performance of the model including accuracy, sensitivity, specificity, precision, F-measure, and the Kappa index. Accuracy indicates the fraction of correct classified pixels. Sensitivity is the proportion of true positives which are correctly classified as positives by the classifier. Specificity determines the ability of a classifier to determine the negative results i.e. the proportion of correctly specified non-fire pixels out of all pixels correctly classified as non-fire plus which incorrectly determined as fire pixels. Precision is defined as the fraction of the pixels which are truly positive among all the pixels which model predicted positive. The F-measure is defined as the harmonic mean of precision and recall. These measures were calculated using confusion matrices (Table 4).

$$Accuracy = \frac{TP+TN}{TP+ TN+ FP+ FN} \qquad (9)$$

$$Sensitivity = \frac{TP}{TP+ FN} \qquad (10)$$

$$Specificity = \frac{TN}{TN+ FP} \qquad (11)$$

$$Precision = \frac{TP}{TP+ FP} \qquad (12)$$

$$F - measure = \frac{2TP}{2TP+FP+FN} \qquad (13)$$

In the above-mentioned equations TN (true negative) and TP (true positive) represent the number of pixels which classified as fire and FP (false positive) and FN (false negative) indicate the number of the pixels assigned incorrectly (Jaafari et al., 2018; Pham et al., 2017).

Table 4. Classification table (Confusion matrix)

		Predicted	
		No-fire (0)	Fire (1)
Actual	No-fire (0)	TN	FP
	Fire (1)	FN	TP

Receiver operating characteristic curve (ROC curve)

The common method to evaluation of the prediction models is ROC. It represents the global performance of the classifiers (Jaafari et al., 2015). It graphically illustrates the trade-offs at each cut-off for diagnostic test; i.e. trade-offs between the 1-specifity and sensitivity (Fan et al., 2006; Jaafari et al., 2018). The X axis of the plot is representative of the 1–specificity, and the Y axis is representative of the sensitivity.

$$X = 1 - \text{specificity} = 1 - \left(\frac{TN}{TN+FP}\right) \qquad (14)$$

$$Y = Sensitivity = \left(\frac{TP}{TP+FN}\right) \qquad (15)$$

The quantitative measurement of the ROC curve is based on the area under curve (AUC). The maximum value for the AUC is 1 which indicates theoretically high performance of the model with 100% sensitive (no false positives; all non-fire pixels properly classified) and 100% specific (no false negatives; all fire pixels properly classified). The interpretation of the AUC value can be done at different scales. However, in general AUC values below 0.6 indicate a poor, 0.6 – 0.7 a normal, 0.8 – 0.9 a good, and greater than 0.9 an excellent model performance (Fan et al., 2006; Hosmer Jr et al., 2013; Jaafari et al., 2018). The ROC curve was calculated for both trained models.

2.3.3 Development of the wildfire probability map

The trained and validated models were used to create the wildfire probability. Given that the Logistic Regression cannot produce the probability map automatically while Random Forest is able to generate. Two different approaches were used. For the Logistic Regression, all the independent variables were imported to the ArcGIS 10.6 environment and the probability of the fire risk was produced based on equation (2). For Random Forest, the probability of the wildfire of each pixel was computed in R-3.6.2. The generated maps were classified in five classes including very low, low, moderate, high and very high. To test the reliability of the generated maps, frequencies of whole five classes were calculated. It is expected that fire frequency increase from low to very high-risk levels (Jaafari et al., 2018). Figure 10 displays the flowchart of the overall process of the study.

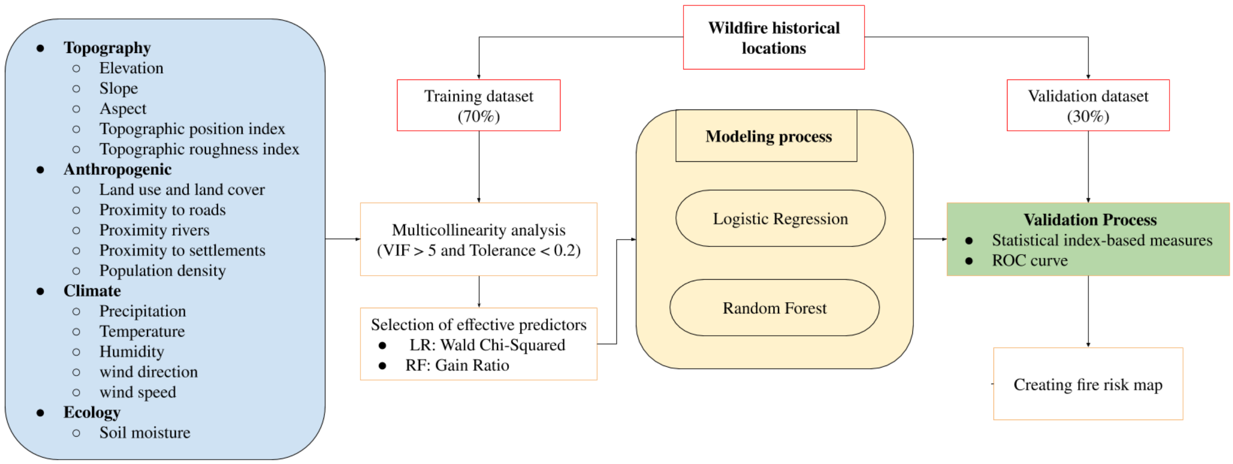

Figure 10. The general workflow of the study

3. Results and Discussion

3.1 Frequency of wildfire

The frequency of the fire occurrence in relation to environmental variables is illustrated in Figure 11. Since steep slope areas have lower fertility and water capacity, and high elevation areas have high transportation costs of food, croplands are located in the lower elevation and slope (Li et al., 2015). In our study, the majority of the fires happened at the low elevation and on gentle slopes and it decreases regularly. On the other hand, consideration of the frequency of the fires in different land uses illustrates that the number of fires in the non-irrigated arable land is more than others. It indicates the relationships between agricultural areas, terrain conditions and fire frequency.

Aspect indicates the amount of energy received from the sun, and it has a close relationship with wind direction (Cawson et al., 2017; Ebel, 2012). In Sardinia, it can be seen the major number of fires happened in the south-western facing slopes. This is, this slope face includes more drier vegetations causing the fires. Besides, the frequency of the fire increases in the south-west wind direction indicating how wind reducing fuel moisture by increasing evaporation in those areas.

Air temperature has a direct influence on the fire ignition since it provides the heat requirements for ignition (Gillett et al., 2004). Additionally, the type of surface and land cover can also affect the temperature. In our study area, most of the fires happened at high temperatures with directly or indirectly being supported by the surfaces and land covers which raise the fuels to their ignition temperature.

Based on the Italian Meteorology Department the annual precipitation on the island is around 500 mm, while the graph shows that all the fires have happened during the low precipitation season. However, the amount of the soil moisture and humidity where the most fire occurred is 75% - 80% and 0.15 – 0.18 (m^3 m^3), respectively (reasonable due to the non-irrigated arable land), but the other climatic and topographical factors can decrease the amount of them considerably during fire seasons. Nhongo et al.(2019) indicate that high temperatures decrease the moisture, and increase the susceptibility of fire.

Land cover and land use (LCLU) are important factors regarding the strong relation between vegetation phenology and spatiotemporal wildfire distribution (De Angelis et al., 2012). Therefore, the determination of them can help a lot to find the area with high fire risk. Studies show that shrublands can start burning fast compared to others in fire seasons in Mediterranean areas (Oliveira et al., 2012). The current study presents that approximately 60% of fires take place on arable lands and and almost 20% on shrublands (Sclerophyll vegetations) (Figure 11). Bacciu et al., (2011) revealed that in 2007 and 2009 this value was 48% and 51% for arable lands, respectively. The plausible explanations for this increase are first, use of fire to make new pastures, removing crop residues, and reducing fuel load (Montiel et al., 2010; Salis et al., 2014) which can increase the fire ignitions. Secondly, decrease in land abandonment process in rural areas which ends up with reduction in shrublands which are responsible for the generation of high-intensity fires (Ager et al., 2014; Pausas, 2004; Salis et al., 2014).

Moisture in the atmosphere is in the form of water vapor, and the amount of moisture that is presented in the atmosphere affects the amount of moisture in the fuel. With increased distance from the river the air humidity is also lower because of smaller evapotranspiration. That is, the fuel can dry as temperature rises. The lower the relative humidity, the more easily a fire will begin and burn. The proximity to rivers shows the number of fires and distance increases together.

Studies in the region have shown that more than 90% of the fire ignition is because of human activities (Ager et al., 2014). However, surprisingly the results of the current study showed that the most of fires happened in the low populated areas. However, considering the distance to roads and settlements reveals that fires were close to them. On the other hand, the number of fires in the various landscapes shows that they are all accessible with people (e.g. farmers). Therefore, it can be deduced that the human component can affect ignition probabilities in the region.

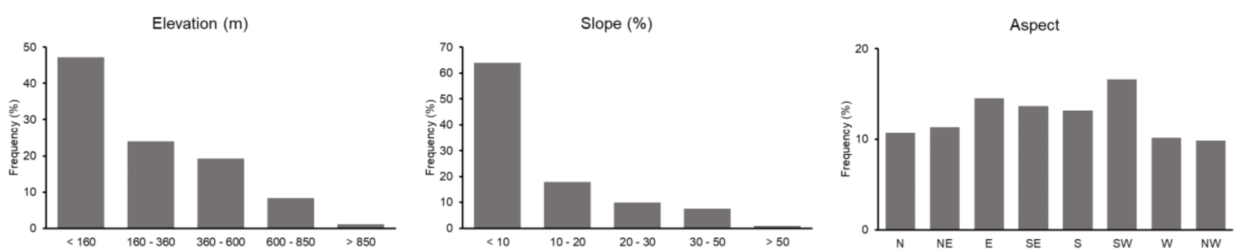

Figure 11. Frequency of fired pixels in relation to different environmental variables. (Continue)

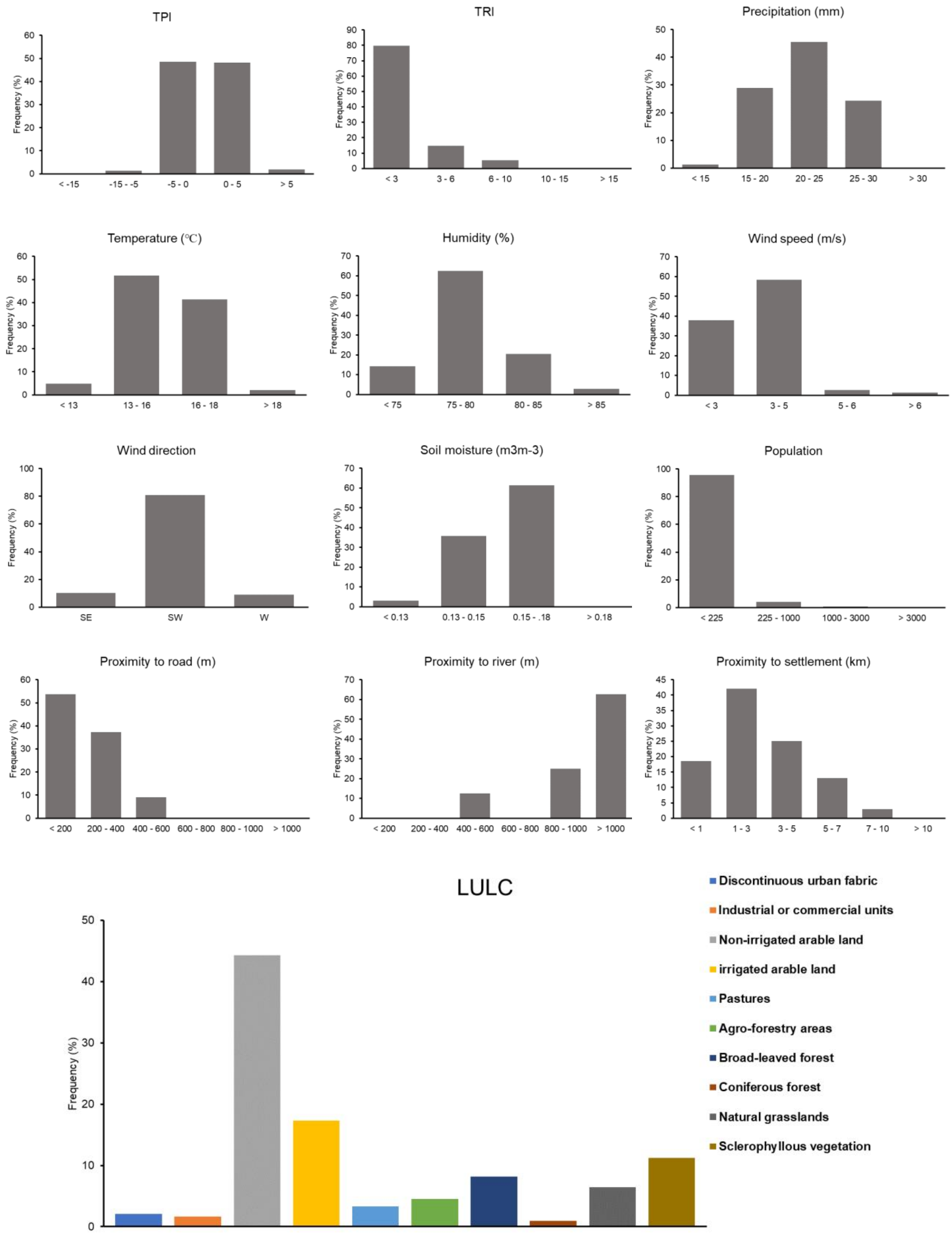

Figure 11. Frequency of fired pixels in relation to different environmental variables.

3.2 Modelling results

Logistic Regression

After several interactions with different variables, the final model was selected with the most correlated factors with wildfire including soil moisture, temperature, wind direction, humidity, population density, proximity to road and slope. Selected variables have a significant relationship with the wildfire probability ($p < 0.05$). The other factors were excluded. Table 5 represents the significance of independent variables and their related coefficients. According to the result, there is a negative relationship between the wildfire occurrence and humidity, slope, population density, proximity to road and proximity to settlement.

Table 5. Results of binary Logistic Regression model

Variable	Coefficients	S.E.	Wald	df	Sig.
Soil moisture	19.340	3.829	25.511	1	0.000
Temperature	0.231	0.041	31.701	1	0.000
Wind Direction	0.023	0.003	47.568	1	0.000
Humidity	-0.058	0.025	5.482	1	0.019
TRI	-0.006	0.001	20.774	1	0.000
Population	-0.001	0.000	11.146	1	0.001
Proximity to road	-0.0002	0.000	16.473	1	0.000
Proximity to settlement	-0.00015	0.000	29.377	1	0.000
Constant	-5.685	1.718	10.950	1	0.001

Random forest

The possibility of checking the variable importance (contribution of each variable) is the main advantage of the Random Forest. The result of the feature importance analysis using Gain ratio method (Figure 12) shows that soil moisture is an important factor, followed by slope, TRI, proximity to road and wind direction.

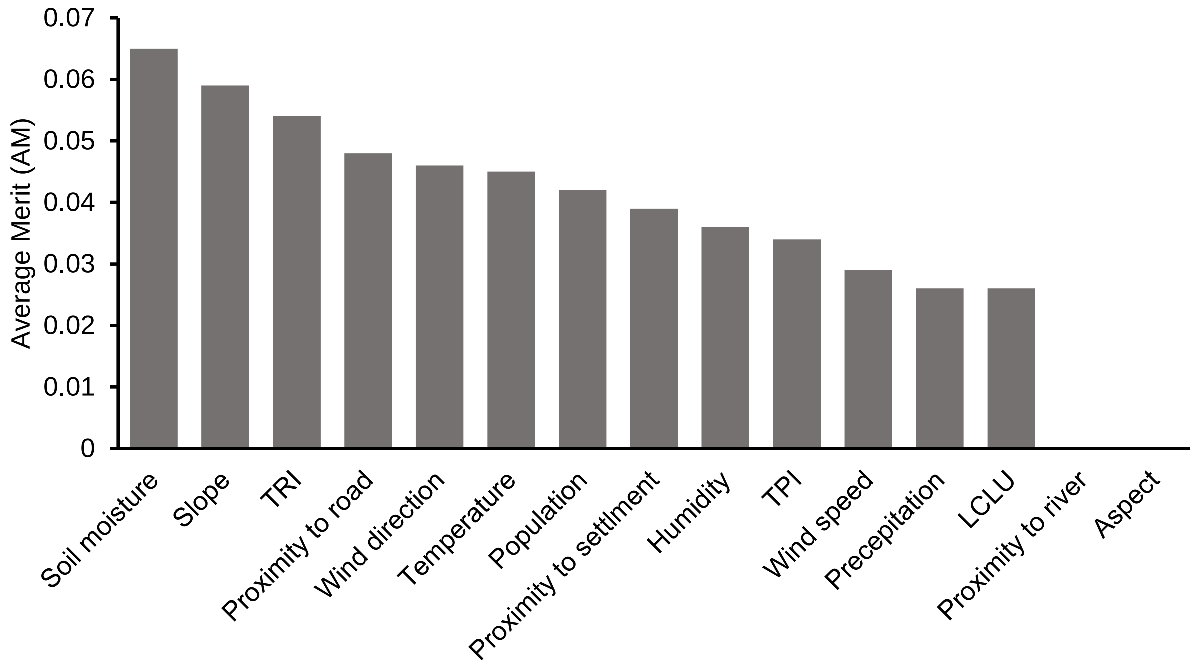

Figure 12. The relative contribution of variables to the models based on the Gain Ratio method.

Development of the model for determination of the wildfire risk is not simple due to frequent changes in the environmental factors of landscape including geology, topography, climatology etc. Although there are several methods proposed by researchers to predict natural disasters, there are still delay evaluating the capability of these methods regarding the wildfire prediction (Jaafari et al., 2018). In this study, several potential environmental variables as predictors in wildfire ignition were analysed to develop two different models: Random Forest and Logistic Regression. Feature importance analysis employed to determine the most relevant factors in the fire risk modelling. It was found out that soil moisture is one of the relevant factors among selected ones. Studies proved that the changes in the moisture content of vegetations increase the risk of the fire occurrence in the area (Oliveira et al., 2012a). Analysing amount of the rainfall in summer and winter seasons also proved that soil moisture is the major factor in fire risk in the Mediterranean region (Scoccimarro et al., 2011). In the current study, the positive relationship of soil moisture with the fire occurrence indicates that a fire can happen when the soil surface layer is drier because soil moisture directly affects the moisture of the fuels.

Climate variability affects the occurrence of large wildfires in different time scales via the existence and flammability of the fuels (Bodini et al. , 2012). In Sardinia, the major of the fire phenomenon happens during the summer. Masala et al., (2008) discovered that both fire numbers and burned areas are highly correlated with rainfall and temperature in Italy. The positive relationship exists between temperature and fire ignition which suggests that the risk of the wildfire is high in the higher temperature. In this study, It was discovered that the most frequent wind directions were S, SE and SW and has positive relationship with fire occurrence. Fois et al., (2012) verified that wind direction combined with slope are major factors in increase of the fire intensity in the island. Among climatological variables, relative humidity is one of the important factors with negative relationship with fire susceptibility in Sardinia. It means high air relative humidity decrease the possibility of fires. Humidity above 60% can avoid burning of vegetation, and optimal condition for beginning of fire is between 30% and 40% (Nhongo et al., 2019).

Topography influences vegetation structure, ignitable moisture and air humidity. Among the topographical factors, slope and TRI are significant and negative effect on fire occurrence. It looks logical since this study discovered that agricultural lands located in low elevation are more prone to fire. It indicates the prevalence of human-induced fires. The negative relationship between the anthropogenic factors and fire occurrence shows that human activities are the cause of the wildfire in the region. The result shows that ignition probability increase as the distance to roads reduces. According to Ager et al., (2014) the distance to road has a strong effect on fire probability in the fire season in Sardinia island. The reason is due to crop cultivation, or agro-pastoral field burning which is so common in Sardinia. Understanding interactions between fire and distance to road can also help to find convenient location for arsonists and suppression resources. The areas near human settlements are also more vulnerable to fires. It is because of illegal development of houses or infrastructure and fires produced by human (Nhongo et al., 2019). Previous studies confirmed that 15.7% of fires occurred in the rural-urban interface in Sardinia (Sirca et al., 2017).

Table 6. Validation of trained models

Measures (%)	Training dataset		Validation dataset	
	LR	RF	LR	RF
Accuracy	66.6	73	64.35	72.6
Sensitivity	70	73.4	66.8	73
Specificity	63.3	72.4	61.8	72
Precision	65.7	72.8	64.7	72.6
F-measure	67.8	73.1	65.7	72.8

The performance of both generated models was analysed based on the classification tables for training and test datasets (Table 6). From the analysis different statistical index-based performance measures for both datasets, we could find that Random Forest provides a training accuracy of 73% and testing accuracy 72.6%, while the result for Logistic Regression is lower with 66.6% and 64.35%, respectively. The accuracy of both models is listed in the Table 6. The Random Forest classified 73% of testing fired pixels into the fire class (sensitivity), and 72% of the non-fire pixels into the non-fire class (Specificity). Precision for Random Forest is 72.6% interpreted as the percentage of the fire pixels correctly classified into the fire pixels. The ROC curve analysis (Figure 13) is for both models are acceptable with 77% for Random Forest and 72% for LR. ROC curve analysis revealed that the performance of Random Forest slightly better than Logistic Regression, but the difference is not significant.

The performance of the machine learning approaches heavily depends on the selected algorithm to implement. For example, if the dependant variable (fire in this case) is not linearly separable in n-dimensional space, then it is required to use a more complex (non-parametric) model to achieve higher prediction performance. It should be also considered that complex models like Random Forest can lead to overfitting if not properly tuned (Kaitlin et al., 2018). Overfitting is defined as high performance of the model on a training set, while does not generalize well on unseen data. On the other hand, less complex (parametric based) models like Logistic Regression has better performance in the variables that are linearly separable. However, these models cannot learn sufficiently from the patterns in the data for accurate prediction resulting in underfitting (Guo et al., 2016). In the current study, given that the complexity and randomness of the datasets is high, Random Forest performed better than Logistic regression. In addition, since the Random Forest analyses the subset of the training data with bagging and subset of features which decreases the influence of highly variable data and the spatial outliers, therefore, it provides more accurate and reliable result (Hong et al., 2018)

One important issue which can decrease the performance of the both models is spatial autocorrelation. Autocorrelation addresses that two locations close to each other are not independent (Dormann et al., 2007). Spatial autocorrelation is not often considered by fire studies, but this can cause type I error and lead to wrong estimation of parameters and important misinterpretation (Portier et al., 2018). A type I error is a sort of fault that happens during the hypothesis testing process when the null hypothesis is true and it was rejected (i.e. noise causes false positives, while there is no correct effect). Studies confirmed the considerable improvement in the fire modelling with accounting for spatial autocorrelation. Furthermore, it can reduce some uncertainties regarding the contribution of various factors relative to each other (Portier et al., 2018).

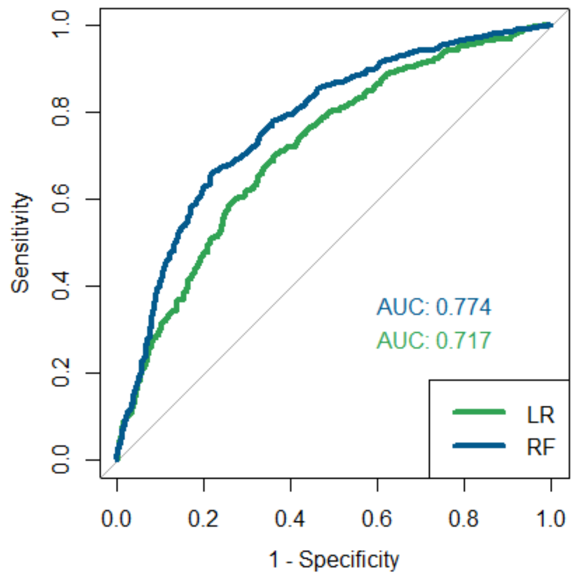

Figure 13. The ROC-AUC for the validation dataset

3.3 Wildfire probability maps

Probability maps were generated using the trained best models (Figure 14). Degree of fire risk susceptibility in the study area was determined in five categories including very low (less than 0.23), low (0.23 – 0.38), moderate (0.38 – 0.50), high (0.50 – 0.75) and very high (greater than 0.75). Based on the applied models the wildfire occurrence rankings are different. Frequency of the fires in the separate classes (Figure 15) explains consistently increase in the number of the fires from zone of the very low to very high. However, Random Forest shows the higher values than Logistic Regression in very high zones with proximately 70% frequency, but the frequency for high and moderate zones in Logistic Regression is larger than Random Forest with almost 30% and 20%, respectively. In each map, the highest frequency belongs to very high susceptibility level followed by other four classes (high, moderate, low and very low), respectively.

Overall, the Frequency of susceptibility levels indicates the capability of the both models for properly determination of the different fire risk levels in the region. Produced probability maps for both models represent that the west and the south-west regions of the island are more prone to fire. They also present that the coastal areas are more threated by fire. In the last decade, the number of the tourists has been increased during the late spring and the summer season in the coastal areas which leads to dramatically increase of the fire risk due to weather and fuel conditions and increasing of population due to tourism (Sirca et al., 2017). Comparison of the our finding with the previous study on the island (Ager et al., 2014) revealed that accuracy achieved by Random Forest and Logistic Regression is reliable. This comparison also explains that during the study period the wildfire spatial-temporal pattern on the island has not been changed considerably.

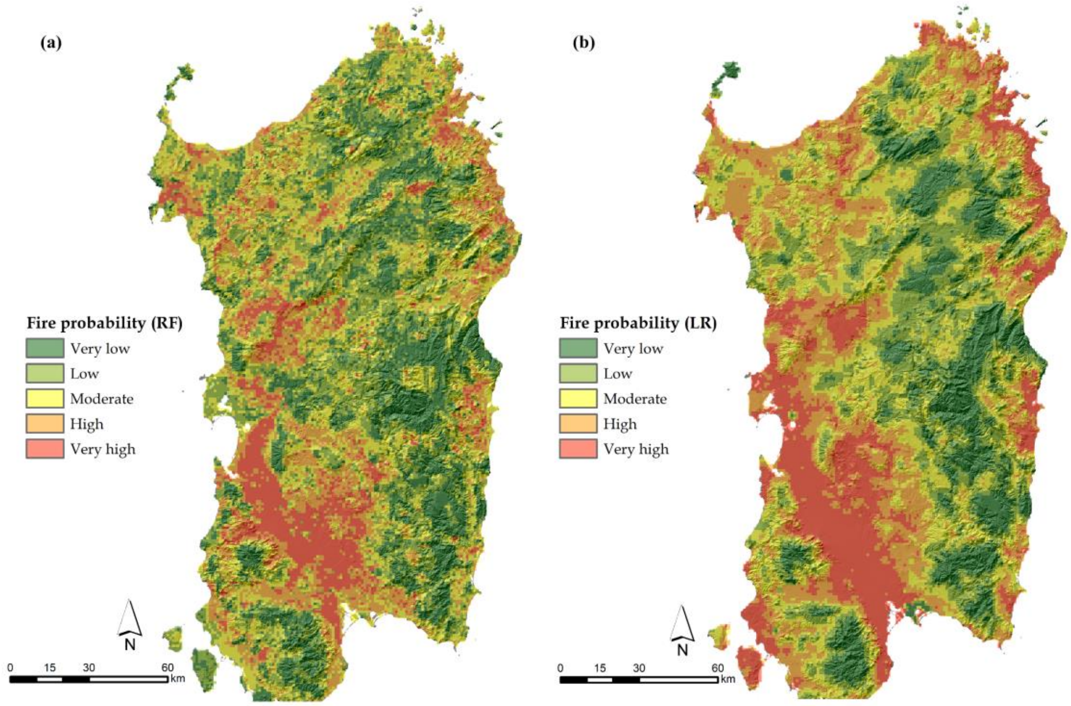

Figure 14. Maps of the probability of fire occurrence for (a) Random Forest and (b) Linear Regression models.

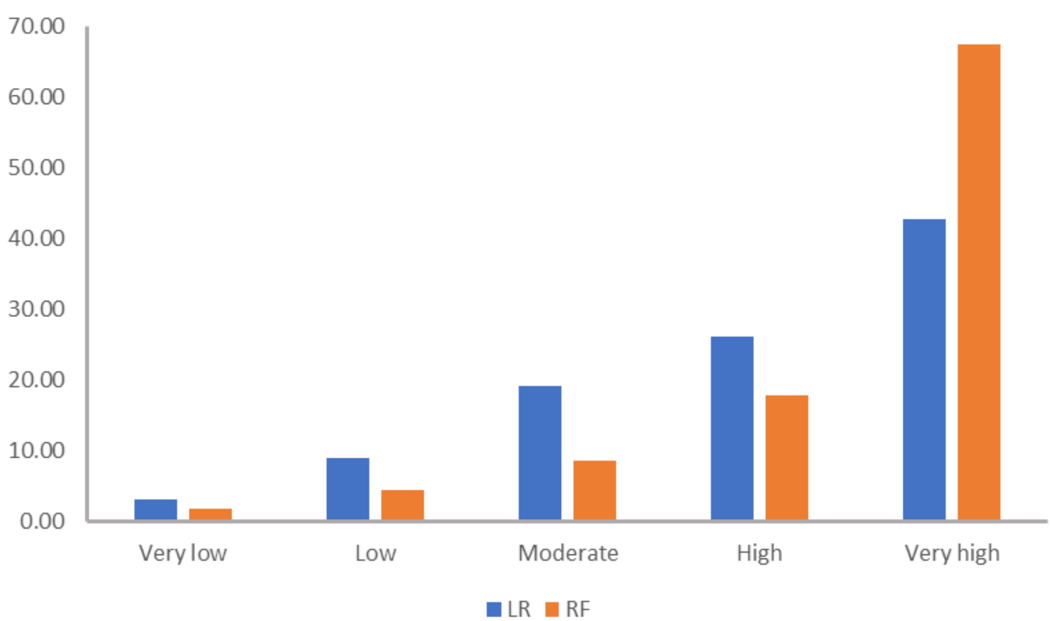

Figure 15. Frequency of susceptibility levels in two wildfire susceptibility maps.

Conclusion

Previous studies proved that fires in Sardinia has complicated relation with various factors including weather conditions, vegetation, and topography and human activities (Ager et al., 2014). These days, changes in the climate and land use on one hand, and significant increase in the interference of the human on the environmental ecosystem on the other hand, requirement for the sustainable wildfire management has become important. For the long-term wildfire management plans, the reliable prediction models are needed to accelerate decision-making process with provision of the reliable information. In this study, topography, climatology, anthropogenic and geology related variables to wildfire was examined to give the wide horizon to wildfire management system regarding the allocation of the fire prevention resources. Moreover, using the data-processing techniques, statistical and machine learning approaches, the prediction of wildfire Susceptibility in the large area became feasible. This can allow managers and decision makers to see more details and perspectives to understand contribution of the different factors in the probability of wildfire occurrence in the different location and various level of the fire susceptibility over the landscape. As a result, it was presented that climatic and anthropomorphic variables have more influence on the fire occurrence in Sardinia. Random Forest shows more robust result compared to LR with 73 percent accuracy. With AUC values over 70%, the performance of the both classifiers are normal. The generated map illustrated that the west, the south-west and coastal regions of the island are more prone to fire.

Although the current study focused on different factors contributed to fire ignition, it should be remembered that there are other factors having a close relationship with fire such as soil type, topographic wetness index (TWI), Potential solar radiation, Recreation area etc. To achieve the best understanding of fire risk, it is important to consider other important explanatory variables. The quality of the data and sources involved also play an important role in reliability of the wildfire risk prediction. It is crucial to utilize the accurate wildfire records and important explanatory variables.

For further study, it is suggested to consider spatial autocorrelation to better understanding of the relationship between the factors. It is also suggested to extend the current study in the area to test more recent prediction techniques in GeoAI including deep learning which can sufficiently model the wildfire susceptibility across the island (Q. T. Bui, 2019). Comparison of the different data mining models could assist to adequately customize the models to achieve the high quality in different conditions and environmental settings.

Magistritöö õppekaval Linnastunud ühiskonna geoinformaatika: Metsatulekahjude tuleohu kaardistamine Sardiinia saarel Itaalias

Behzad Valipour Shokouhi

Kokkuvõte

Metsatulekahjud võivad põhjustada rasket majanduslikku, sotsiaalset ja keskkonnakahju. Metsatulekahjud on osa ökosüsteemi protsessidest ning mõjutavad fauna ja floora taastumist, arengut ja ruumilist levikut (Nhongo *et al.*, 2019). Euroopa Metsatulekahjude Infosüsteemi andmetel on enam kui 95% metsatulekahjudest põhjustanud inimeste poolt ning Vahemere maades on see protsent isegi suurem. Tulekahjude haldamiskavades mängib olulist rolli metsatulekahjude esinemistõenäosus. Sellega arvestamine aitab vähendada tulekahju põhjustatud kahjusid (Leuenberger *et al*., 2018). Metsatulekahjudele vastuvõtlike alade kaardistamiseks on vajalik tulekahju tekkimist mõjutavate tegurite kindlakstegemine. Enamik neist teguritest on inimtekkelised (või inimtegevusest põhjustatud), nt kaugus teedest, kuid seotud ka keskkonnateguritega (nt kliima, topograafia). Selles kontekstis saab mudeli ettevalmistamisel peamiste muutujate kombineerimiseks kasutada geograafilisi infosüsteeme (GIS). Kõige usaldusväärsemate mudelite saavutamiseks on siiski vaja statistilist analüüsi, et täpsustada tulekahju põhjustavate erinevate muutujate olulisust. Itaalias on igal aastal sadu metsatulekahjusid ning seetõttu valiti uuringualaks Itaalia, täpsemalt Sardiinia. Käesoleva töö eesmärgid olid: (1) uurida Sardiinia metsatulekahjude ajalis-ruumilisi mustreid, et teha kindlaks tulekahjude puhkemise olulised tegurid; (2) koostada metsatulekahjude esinemistõenäosuse kaart, mis põhineb kahel erineval meetodil, logistilisel regressioonil ja otsustusmetsal, kasutades kõige olulisemaid muutujaid; (3) pakkuda piirkonna tuleohte käsitlevat teavet tulekahjude haldamiseks. Selles uuringus kasutati kuutteist muutujat (maapinna kalle, aspekt, kõrgus merepinnast, topograafilise asukoha indeks (TPI), topograafiline kareduse indeks (TRI), kuu keskmine temperatuur ja sademete hulk, tuule kiirus, tuule suund, mullaniiskus, maakasutusviis, rahvaarv piirkonnas, õhuniiskus ning (põlengu)ala kaugus teedest, asulatest ja jõgedest), mis valiti piirkonnas varem tehtud uuringute põhjal. Seejärel kasutati kõige olulisemate tegurite valimiseks kaheastmelisi statistilisi mudeleid (multikollineaarsuse analüüsi, juurdekasvu suhet ja Waldi hii-ruut-testi). Muutujate suhteline mõju näitas, et uuringupiirkonnas olid klimaatilised ja sotsiaalmajanduslikud tegurid tuleohu osas olulisemad kui topograafilised muutujad. Valitud muutujaid kasutati sisendina kahele masinõppe meetodile: logistilisele regressioonile ja otsustusmetsale. Tulemused näitasid juhusliku metsa meetodi paremat toimimist (üldine täpsus 73%) võrreldes logistilise regressiooniga (üldine täpsus 67%). Tuleohu tõenäosuskaardid loodi viies klassis: väga väike (tõenäosus tuleohuks alla 0,23), väike (0,23–0,38), mõõdukas (0,38–0,50), suur (0,50–0,75) ja väga suur (rohkem kui 0,75). Loodud kaartide paikapidavust katsetati viie klassi esinemissageduste arvutamise teel. Kaardid näitavad, et tuleohtlikumad on saare lääne-, edela- ja rannikupiirkonnad. Üldiselt võib tulekahjude puhkemise ning keskkonna- ja inimtegurite vastastikmõju mõistmine toetada tulekahjude haldamissüsteemi, aidates leida kohti, kus on suurem tõenäosus tulekahju tekkeks ja suunata sinna piirkonda rohkem tuletõrje ressursse. Lisaks saab

selliseid nüüdisaegseid vahendeid kasutada metsakaitsekavade toetamiseks ning metsatulekahjude negatiivse mõju vähendamiseks inimkeskkonnale.

Acknowledgment

I would like to express my sincere gratitude to my supervisor Dr. Evelyn Uuemaa for the continuous support of my M.Sc. study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my M.Sc. study.

Last but not the least, I must express my very profound gratitude to my family: my parents and my brother, my friends: Mehrdad Eslami and Saleh Rahimlou, my colleagues at CollectiveCrunch Oy. They kept me going on and this work would not have been possible without their input.

References

A. Bodini, E. Entrade, QA Cossu, P. Fiorucci, and G. B. (2012). Analysis of climatic conditions influencing wildfire static risk in Sardinia and Liguria (Italy). *International Conference on Fire Behavior and Risk Modeling*.

Adab, H., Kanniah, K. D., & Solaimani, K. (2013). Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. *Natural Hazards*, *65*(3), 1723–1743. https://doi.org/10.1007/s11069-012-0450-8

Ager, A. A., Preisler, H. K., Arca, B., Spano, D., & Salis, M. (2014). Wild fi re risk estimation in the Mediterranean area. *Environmetrics*, (September 2013). https://doi.org/10.1002/env.2269

Alexandre, P. M., Stewart, S. I., Mockrin, M. H., Keuler, N. S., Syphard, A. D., Bar-Massada, A., … Radeloff, V. C. (2016). The relative impacts of vegetation, topography and spatial arrangement on building loss to wildfires in case studies of California and Colorado. *Landscape Ecology*, *31*(2), 415–430. https://doi.org/10.1007/s10980-015-0257-6

Amatulli, G., Rodrigues, M. J., 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*(G4). https://doi.org/10.1029/2005JG000133

Anderson, L. O., Aragão, L. E. O. C., Gloor, M., Arai, E., Adami, M., Saatchi, S. S., … Duarte, V. (2015). Disentangling the contribution of multiple land covers to fire-mediated carbon emissions in Amazonia during the 2010 drought. *Global Biogeochemical Cycles*, *29*(10), 1739–1753. https://doi.org/10.1002/2014GB005008

Archibald, S., Roy, D. P., van Wilgen, B. W., & Scholes, R. J. (2009). What limits fire? An examination of drivers of burnt area in Southern Africa. *Global Change Biology*, 15(3), 613-630. https://doi.org/10.1111/j.1365-2486.2008.01754.x

Arndt, N., Vacik, H., Koch, V., Arpaci, A., & Gossow, H. (2013). Modeling human-caused forest fire ignition for assessing forest fire danger in Austria. *iForest-Biogeosciences and Forestry*, *6*(6), 315. https://doi.org/10.3832/ifor0936-006

Ascough Ii, J. C., Maier, H. R., Ravalico, J. K., & Strudley, M. W. (2008). Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. *Ecological modelling*, *219*(3-4), 383-399. https://doi.org/https://doi.org/10.1016/j.ecolmodel.2008.07.015

Bajocco, S., & Ricotta, C. (2008). Evidence of selective burning in Sardinia (Italy): Which land-cover classes do wildfires prefer? *Landscape Ecology*, *23*(2), 241–248. https://doi.org/10.1007/s10980-007-9176-5

Bar Massada, A., Radeloff, V. C., Stewart, S. I., & Hawbaker, T. J. (2009). Wildfire risk in the wildland–urban interface: A simulation study in northwestern Wisconsin. *Forest Ecology and Management*, *258*(9), 1990–1999. https://doi.org/https://doi.org/10.1016/j.foreco.2009.07.051

Bartsch, A., Balzter, H., & George, C. (2009). The influence of regional surface soil moisture anomalies on forest fires in Siberia observed from satellites. *Environmental Research Letters*, *4*(4). https://doi.org/10.1088/1748-9326/4/4/045021

Bisquert, M., Caselles, E., Snchez, J. M., & Caselles, V. (2012). Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. *International Journal of Wildland Fire*, *21*(8), 1025–1029. https://doi.org/10.1071/WF11105

Bisquert, M., Sánchez-Tomás, J., & Caselles, V. (2011). Fire danger estimation from MODIS Enhanced Vegetation Index data. Application to Galicia region (Northwest Spain). *International Journal of Wildland Fire*, *20*, 465–473. https://doi.org/10.1071/WF10002

Borchers, J. G. (2005). Accepting uncertainty, assessing risk: Decision quality in managing wildfire, forest resource values, and new technology. *Forest Ecology and Management*, *211*(1), 36–46. https://doi.org/https://doi.org/10.1016/j.foreco.2005.01.025

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brown, T. C., Kingsley, D., Peterson, G. L., Flores, N. E., Clarke, A., & Birjulin, A. (2008). Reliability of individual valuations of public and private goods: Choice consistency, response time, and preference refinement. *Journal of Public Economics*, *92*(7), 1595–1606. https://doi.org/https://doi.org/10.1016/j.jpubeco.2008.01.004

Brugnach, M., Dewulf, A., Henriksen, H. J., & van der Keur, P. (2011). More is not always better: Coping with ambiguity in natural resources management. *Journal of Environmental Management*, *92*(1), 78–84. https://doi.org/https://doi.org/10.1016/j.jenvman.2010.08.029

Bui, D. T., Le, K. T. T., Nguyen, V. C., Le, H. D., & Revhaug, I. (2016). Tropical forest fire susceptibility mapping at the Cat Ba National Park area, Hai Phong City, Vietnam, using GIS-based Kernel logistic regression. *Remote Sensing*, *8*(4), 1–15. https://doi.org/10.3390/rs8040347

Bui, Q. T. (2019). Metaheuristic algorithms in optimizing neural network: a comparative study for forest fire susceptibility mapping in Dak Nong, Vietnam. *Geomatics, Natural Hazards and Risk*, *10*(1), 136–150. https://doi.org/10.1080/19475705.2018.1509902

C Fois, F Casula, M Salis, C Dessy, S Falchi, S Mavuli, A Piga, S Sanna, D. S. (2012). Fire behavior analysis of the Anela wildfire (Sardinia, 1945). *Modeling Fire Behavior and Risk*.

Camp, A., Oliver, C., Hessburg, P., & Everett, R. (1997). Predictinglate-successional fire refugia pre-dating European settlement in the Wenatchee mountains. *Forest Ecology and Management*, *95*(1), 63–77. https://doi.org/10.1016/S0378-1127(97)00006-6

Cardil, A., Delogu, G. M., & Molina-Terrén, D. M. (2017). Fatalidades em incêndios florestais de 1945 a 2015 na Sardenha (Itália). *Cerne*, *23*(2), 175–184. https://doi.org/10.1590/01047760201723022266

Carmel, Y., Paz, S., Jahashan, F., & Shoshany, M. (2009). Assessing fire risk using Monte Carlo simulations of fire spread. *Forest Ecology and Management*, *257*(1), 370–377. https://doi.org/https://doi.org/10.1016/j.foreco.2008.09.039

Catchpole, E. A., Catchpole, W. R., N.R.Viney, McCaw, W. L., & Marsden-Smedley, J. B. (2001). Estimating fuel response time and predicting fuel moisture content from field data. *International Journal of Wildland Fire*, *10*(2), 215–222. Retrieved from https://doi.org/10.1071/WF01011

Catry, F. X., Rego, F. C., Bação, F. L., & Moreira, F. (2009). Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire*, *18*(8), 921–931. Retrieved from https://doi.org/10.1071/WF07123

Cawson, J. G., Duff, T. J., Tolhurst, K. G., Baillie, C. C., & Penman, T. D. (2017). Fuel moisture in Mountain Ash forests with contrasting fire histories. *Forest Ecology and Management*, *400*, 568–577. https://doi.org/10.1016/j.foreco.2017.06.046

Chang, Y., Zhu, Z., Bu, R., Chen, H., Feng, Y., Li, Y., … Wang, Z. (2013). Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China. *Landscape Ecology*, *28*(10), 1989–2004. https://doi.org/10.1007/s10980-013-9935-4

Chen, F., Du, Y., Niu, S., & Zhao, J. (2015). Modeling forest lightning fire occurrence in the Daxinganling Mountains of Northeastern China with MAXENT. *Forests*, *6*(5), 1422–1438. https://doi.org/10.3390/f6051422

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., … Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *CATENA*, *151*, 147–160. https://doi.org/10.1016/J.CATENA.2016.11.032

Chowdhury, E. H., & Hassan, Q. K. (2013). Use of remote sensing-derived variables in developing a forest fire danger forecasting system. *Natural Hazards*, *67*(2), 321–334. https://doi.org/10.1007/s11069-013-0564-7

Chuvieco, E, Aguado, I., Jurdao, S., Pettinari, M. L., Yebra, M., Salas, J., … Martínez-Vega, F. J. (2014). Integrating geospatial information into fire risk assessment. *International Journal of Wildland Fire*, *23*(5), 606–619. Retrieved from https://doi.org/10.1071/WF12052

Chuvieco, Emilio, Aguado, I., & Dimitrakopoulos, A. P. (2004). Conversion of fuel moisture content values to ignition potential for integrated fire danger assessment. *Canadian Journal of Forest Research*, *34*(11), 2284–2293. https://doi.org/10.1139/X04-101

Chuvieco, Emilio, Aguado, I., Yebra, M., Nieto, H., Salas, J., Martín, M. P., … Zamora, R. (2010). Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecological Modelling*, *221*(1), 46–58. https://doi.org/https://doi.org/10.1016/j.ecolmodel.2008.11.017

Colin Price David Rind. (1994). Possible implications of global climate change on global lightning distributions and frequencies. *Journal of Geophysical Research: Atmospheres*, *99*(D5), 10823–10831. https://doi.org/10.1029/94JD00019

Copernicus Land Monitoring Service. CORINE Land Cover. (2018). Retrieved from https://land.copernicus.eu/pan-european/corine-land-cover

Cruz, M. G., & Alexander, M. E. (2010). Assessing crown fire potential in coniferous forests of western North America: a critique of current approaches and recent simulation studies.

International Journal of Wildland Fire, *19*(4), 377–398. Retrieved from https://doi.org/10.1071/WF08132

Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). RANDOM FORESTS FOR CLASSIFICATION IN ECOLOGY. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1

De Angelis, A., Bajocco, S., & Ricotta, C. (2012). Phenological variability drives the distribution of wildfires in Sardinia. *Landscape Ecology*, *27*(10), 1535–1545. https://doi.org/10.1007/s10980-012-9808-2

De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., … Crombé, P. (2013). Application of the topographic position index to heterogeneous landscapes. *Geomorphology*, *186*, 39–49. https://doi.org/10.1016/j.geomorph.2012.12.015

Diaz-Balteiro, L., & Romero, C. (2008). Making forestry decisions with multiple criteria: A review and an assessment. *Forest Ecology and Management*, *255*(8), 3222–3241. https://doi.org/https://doi.org/10.1016/j.foreco.2008.01.038

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., … Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 027–046. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dupuy, J.-L. (1997). *Mieux comprendre et prédire la propagation des feux de fôrets : expérimentation, test et proposition de modèles*. Retrieved from http://www.theses.fr/1997LYO10080

Ebel, B. (2012). *Impacts of Wildfire and Slope Aspect on Soil Temperature in a Mountainous Environment*. *Vadose Zone Journal* (Vol. 11). https://doi.org/10.2136/vzj2012.0017

EFFIS. COPERNICUS Emergency Management Service. Retrieved from http://effis.jrc.ec.europa.eu/about-effis/technical-background/european-fire-database/

European Settlement Map. (2017). Retrieved from https://land.copernicus.eu/pan-european/GHSL/european-settlement-map

F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., … Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, *30*(5), 609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

Fairbrother, A., & Turnley, J. G. (2005). Predicting risks of uncharacteristic wildfires: Application of the risk assessment process. *Forest Ecology and Management*, *211*(1), 28–35. https://doi.org/https://doi.org/10.1016/j.foreco.2005.01.026

Fan, J., Upadhye, S., & Worster, A. (2006). Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Emergency Medicine*, *8*(1), 19–20. https://doi.org/DOI: 10.1017/S1481803500013336

Fernández-muñoz, J. G. P. S. (2012). Fire regime changes in the Western Mediterranean Basin : from fuel-limited to drought-driven fire regime, 215–216. https://doi.org/10.1007/s10584-011-0060-6

Finney, M. A. (2005). The challenge of quantitative risk analysis for wildland fire. *Forest Ecology and Management*, *211*(1), 97–108. https://doi.org/https://doi.org/10.1016/j.foreco.2005.02.010

Fiorucci, P., Gaetani, F., & Minciardi, R. (2008). Development and application of a system for dynamic wildfire risk assessment in Italy. *Environmental Modelling & Software*, *23*(6), 690–702. https://doi.org/10.1016/J.ENVSOFT.2007.05.008

Flannigan, M. D., Logan, K. A., Amiro, B. D., Skinner, W. R., & Stocks, B. J. (2005). Future Area Burned in Canada. *Climatic Change*, *72*(1), 1–16. https://doi.org/10.1007/s10584-005-5935-y

García, J. C., Antonio, J., & Garzón, A. (2015). EU-DEM Upgrade Documentation EEA User Manual EU-DEM Upgrade, (1/2), 12. Retrieved from (https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1

Ghorbanzadeh, O., Valizadeh Kamran, K., Blaschke, T., Aryal, J., Naboureh, A., Einali, J., & Bian, J. (2019). Spatial Prediction of Wildfire Susceptibility Using Field Survey GPS Data and Machine Learning Approaches. *Fire*, *2*(3), 43. https://doi.org/10.3390/fire2030043

Gigović, L., Pourghasemi, H. R., Drobnjak, S., & Bai, S. (2019). Testing a New Ensemble Model Based on SVM and Random Forest in Forest Fire Susceptibility Assessment and Its Mapping in Serbia's Tara National Park. *Forests*, *10*(5), 408. https://doi.org/10.3390/f10050408

Gillett, N. P., Weaver, A. J., Zwiers, F. W., & Flannigan, M. D. (2004). Detecting the effect of climate change on Canadian forest fires. *Geophysical Research Letters*, *31*(18). https://doi.org/10.1029/2004GL020876

González, J. R., Palahí, M., Trasobares, A., & Pukkala, T. (2006). A fire probability model for forest stands in Catalonia (north-east Spain). *Annals of Forest Science*, *63*(2), 169–176. https://doi.org/10.1051/forest:2005109

Grala, K., Grala, R. K., Hussain, A., Cooke, W. H., & Varner, J. M. (2017). Impact of human factors on wildfire occurrence in Mississippi, United States. *Forest Policy and Economics*, *81*(May), 38–47. https://doi.org/10.1016/j.forpol.2017.04.011

Grömping, U. (2009). Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *The American Statistician*, *63*(4), 308–319. https://doi.org/10.1198/tast.2009.08199

Guglietta, D., Migliozzi, A., & Ricotta, C. (2015). A Multivariate Approach for Mapping Fire Ignition Risk: The Example of the National Park of Cilento (Southern Italy). *Environmental Management*, *56*(1), 157–164. https://doi.org/10.1007/s00267-015-0494-0

Guo, F., Wang, G., Su, Z., Liang, H., Wang, W., Lin, F., & Liu, A. (2016). What drives forest fire in Fujian, China? Evidence from logistic regression and Random Forests. *International Journal of Wildland Fire*, *25*(5), 505–519. https://doi.org/10.1071/WF15121

Hardy, C. C. (2005). Wildland fire hazard and risk: Problems, definitions, and context. *Forest Ecology and Management*, *211*(1), 73–82. https://doi.org/https://doi.org/10.1016/j.foreco.2005.01.029

Heinz Gallaun, Kerstin Dohr, Martin Puhm, André Stumpf, J. H. (2019). Copernicus land monitoring service reference data: eu-hydro river net user guide 1.0 Retrieved from https://land.copernicus.eu/imagery-in-situ/eu-hydro

Hessburg, P. F., Reynolds, K. M., Keane, R. E., James, K. M., & Salter, R. B. (2007). Evaluating wildland fire danger and prioritizing vegetation and fuels treatments. *Forest Ecology and Management*, *247*(1), 1–17. https://doi.org/https://doi.org/10.1016/j.foreco.2007.03.068

Ho, Tin Kam, Jonathan J. Hull, and S. N. S. (1994). Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis & Machine Intelligence*.

Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A. X., & Xu, C. (2018). Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China. *Science of the Total Environment*, *630*, 1044–1056. https://doi.org/10.1016/j.scitotenv.2018.02.278

Hosmer Jr, Stanley Lemeshow, and R. S. (2013). *Applied logistic regression*.

Jaafari, A., Gholami, D. M., & Zenner, E. K. (2017). A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran. *Ecological Informatics*, *39*, 32–44. https://doi.org/10.1016/J.ECOINF.2017.03.003

Jaafari, A., Najafi, A., Rezaeian, J., Sattarian, A., & Ghajar, I. (2015). Planning road networks in landslide-prone areas: A case study from the northern forests of Iran. *Land Use Policy*, *47*, 198–208. https://doi.org/https://doi.org/10.1016/j.landusepol.2015.04.010

Jaafari, A., & Pourghasemi, H. R. (2019). 28 - Factors Influencing Regional-Scale Wildfire Probability in Iran: An Application of Random Forest and Support Vector Machine. In H. R. Pourghasemi & C. B. T.-S. M. in G. I. S. and R. for E. and E. S. Gokceoglu (Eds.) (pp. 607–619). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-815226-3.00028-4

Jaafari, A., Zenner, E. K., & Pham, B. T. (2018). Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. *Ecological Informatics*, *43*(June 2017), 200–211. https://doi.org/10.1016/j.ecoinf.2017.12.006

Kirasich, K., Smith, T., & Sadler, B. (2018). Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets. *SMU Data Science Review*, *1*(3), 9. https://scholar.smu.edu/datasciencereview.

Kandya, A. K., Kimothi, M. M., Jadhav, R. N., & Agarwal, J. P. (1998). Application of Geographic Information System in Identification of'Fire-Prone'Areas-a Feasibility Study in Parts of Junagadh (Gujarat). *Indian forester*, *124*(7), 531-536.

Kmoch, A., Kanal, A., Astover, A., Kull, A., Virro, H., Helm, A., … Uuemaa, E. (2019). EstSoil-EH v1.0: An eco-hydrological modelling parameters dataset derived from the Soil Map of Estonia (data deposit), (October), 1–29. https://doi.org/10.5281/ZENODO.3473290

Knorr, W., Kaminski, T., Arneth, A., & Weber, U. (2014). Impact of human population density on fire frequency at the global scale, 1085–1102. https://doi.org/10.5194/bg-11-1085-2014

Krougly, Z. L., Creed, I. F., & Stanford, D. A. (2009). A stochastic model for generating disturbance patterns within landscapes. *Computers & Geosciences*, *35*(7), 1451–1459. https://doi.org/https://doi.org/10.1016/j.cageo.2008.05.010

Laguardia, G., & Niemeyer, S. (2008). On the comparison between the LISFLOOD modelled and the ERS/SCAT derived soil moisture estimates. *Hydrology and Earth System Sciences*, *12*(6), 1339–1351. https://doi.org/10.5194/hess-12-1339-2008

Lehsten, V., Harmand, P., Palumbo, I., & Arneth, A. (2010). Modelling burned area in Africa, 3199–3214. https://doi.org/10.5194/bg-7-3199-2010

Leuenberger, M., Parente, J., Tonini, M., Pereira, M. G., & Kanevski, M. (2018). Environmental Modelling & Software Wild fi re susceptibility mapping : Deterministic vs . stochastic approaches. *Environmental Modelling and Software*, *101*, 194–203.

Li, Y., Yang, X., Cai, H., Xiao, L., Xu, X., & Liu, L. (2015). Topographical characteristics of agricultural potential productivity during cropland transformation in China. *Sustainability (Switzerland)*, *7*(1), 96–110. https://doi.org/10.3390/su7010096

Liao, D., & Valliant, R. (2012). Variance inflation factors in the analysis of complex survey data. *Survey Methodology*, *38*(1), 53–62.

Linn, R. R., Winterkamp, J. L., Weise, D. R., & Edminster, C. (2010). A numerical study of slope and fuel structure effects on coupled wildfire behaviour, (1994), 179–201.

Lozano, F. J., Suárez-Seoane, S., & de Luis, E. (2007). Assessment of several spectral indices derived from multi-temporal Landsat data for fire occurrence probability modelling. *Remote Sensing of Environment*, *107*(4), 533–544. https://doi.org/https://doi.org/10.1016/j.rse.2006.10.001

Lozano, F. J., Suárez-Seoane, S., Kelly, M., & Luis, E. (2008). A multi-scale approach for modeling fire occurrence probability using satellite data and classification trees: A case study in a mountainous Mediterranean region. *Remote Sensing of Environment*, *112*(3), 708–719. https://doi.org/https://doi.org/10.1016/j.rse.2007.06.006

MacGregor, D. G., & González-Cabán, A. (2011). Decision modeling for analyzing fire action outcomes. *Research Paper PSW-RP-258*.

Martell, D. L., Otukol, S., & Stocks, B. J. (1987). A logistic model for predicting daily people-caused forest fire occurrence in Ontario. *Canadian Journal of Forest Research*, *17*(5), 394–401. https://doi.org/10.1139/x87-068

Martín, Y., Zúñiga-Antón, M., & Rodrigues Mimbrero, M. (2019). Modelling temporal variation of fire-occurrence towards the dynamic prediction of human wildfire ignition danger in northeast Spain. *Geomatics, Natural Hazards and Risk*, *10*(1), 385–411. https://doi.org/10.1080/19475705.2018.1526219

Martínez, J., Vega-Garcia, C., & Chuvieco, E. (2009). Human-caused wildfire risk rating for prevention planning in Spain. *Journal of Environmental Management*, *90*(2), 1241–1252.

https://doi.org/https://doi.org/10.1016/j.jenvman.2008.07.005

Masala, F., Bacciu, V., Sirca, C., & Spano, D. (2008). Fire – weather relationship in the Italian peninsula, (1993), 55–61.

McKenzie, D., Peterson, D. L., & Agee, J. K. (2000). Fire frequency in the interior Columbia river basin: Building regional models from fire history data. *Ecological Applications*, *10*(5), 1497–1516. https://doi.org/10.1890/1051-0761(2000)010[1497:FFITIC]2.0.CO;2

Mell, W., Jenkins, M. A., Gould, J., & Cheney, P. (2007). A physics-based approach to modelling grassland fires. *International Journal of Wildland Fire*, *16*(1), 1–22. Retrieved from https://doi.org/10.1071/WF06002

Mendoza, G. A., & Martins, H. (2006). Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms. *Forest Ecology and Management*, *230*(1), 1–22. https://doi.org/https://doi.org/10.1016/j.foreco.2006.03.023

Meyn, A., White, P. S., Buhk, C., & Jentsch, A. (2007). Environmental drivers of large, infrequent wildfires: the emerging conceptual model. *Progress in Physical Geography: Earth and Environment*, *31*(3), 287–312. https://doi.org/10.1177/0309133307079365

Mhawej, M., Faour, G., & Adjizian-Gerard, J. (2015). Wildfire Likelihood's Elements: A Literature Review. *Challenges*, *6*(2), 282–293. https://doi.org/10.3390/challe6020282

Montiel, C., & Kraus, D. (2010). *Best Practices of Fire Use - Prescribed Burning and Suppression Fire Programmes in Selected Case-Study Regions of Europe*. *Best Practices of Fire Use - Prescribed Burning and Suppression Fire Programmes in Selected Case-Study Regions in Europe*.

Moreira, F., Rego, F. C., & Ferreira, P. G. (2001). Temporal (1958 – 1995) pattern of change in a cultural landscape of northwestern Portugal : implications for fire occurrence, *2*(Ccrn 1995), 557–558.

Mouillot, F., & Field, C. B. (2005). Fire history and the global carbon budget: a 1°× 1° fire history reconstruction for the 20th century. *Global Change Biology*, *11*(3), 398–420. https://doi.org/10.1111/j.1365-2486.2005.00920.x

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, *37*(2), 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Nasiri, M. (2013). Determining the priority of effective factors on forest fire from analytical hierarchy process. *Journal of Applied Biological Sciences*, *7*(1), 61–64.

Nhongo, E. J. S., Fontana, D. C., Guasselli, L. A., & Bremm, C. (2019). Probabilistic modelling of wildfire occurrence based on logistic regression, Niassa Reserve, Mozambique. *Geomatics, Natural Hazards and Risk*, *10*(1), 1772–1792. https://doi.org/10.1080/19475705.2019.1615559

O'Laughlin, J. (2005). Conceptual model for comparative ecological risk assessment of wildfire effects on fish, with and without hazardous fuel treatment. *Forest Ecology and Management*, *211*(1), 59–72. https://doi.org/https://doi.org/10.1016/j.foreco.2005.01.028

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., & Pereira, J. M. C. (2012a). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management*, *275*, 117–129. https://doi.org/10.1016/J.FORECO.2012.03.003

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., & Pereira, J. M. C. (2012b). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management*, *275*, 117–129. 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. Retrieved from https://doi.org/10.1071/WF09139

Parente, J., Pereira, M. G., Amraoui, M., & Fischer, E. M. (2018). Heat waves in Portugal: Current regime, changes in future climate and impacts on extreme wildfires. *Science of the Total Environment*, *631–632*, 534–549. https://doi.org/10.1016/j.scitotenv.2018.03.044

Parisien, M.-A., Parks, S. A., Krawchuk, M. A., Flannigan, M. D., Bowman, L. M., & Moritz, M. A. (2011). Scale-dependent controls on the area burned in the boreal forest of Canada, 1980–2005. *Ecological Applications*, *21*(3), 789–805. https://doi.org/10.1890/10-0326.1

Pastor, E., Zárate, L., Planas, E., & Arnaldos, J. (2003). Mathematical models and calculation systems for the study of wildland fire behaviour. *Progress in Energy and Combustion Science*, *29*(2), 139–153. https://doi.org/10.1016/S0360-1285(03)00017-0

Paton, D. (2014). *Wildfire Hazards, Risks, and Disasters*. *Elsevier* (Vol. 7). https://doi.org/10.1016/0196-9781(86)90210-X

Pausas, J. G. (2004). Peninsula (Mediterranean Basin). *Climatic Change*, *63*, 337–350.

Pham, B. T., Tien Bui, D., Pourghasemi, H. R., Indra, P., & Dholakia, M. B. (2017). Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theoretical and Applied Climatology*, *128*(1), 255–273. https://doi.org/10.1007/s00704-015-1702-9

Podur, J. J., Martell, D. L., & Stanford, D. (2010). A compound Poisson model for the annual area burned by forest fires in the province of Ontario. *Environmetrics*, *21*(5), 457–469. https://doi.org/10.1002/env.996

Portier, J., Gauthier, S., Robitaille, A., & Bergeron, Y. (2018). Accounting for spatial autocorrelation improves the estimation of climate, physical environment and vegetation's effects on boreal forest's burn rates. *Landscape Ecology*, *33*(1), 19–34. https://doi.org/10.1007/s10980-017-0578-8

Pourghasemi, H. reza, Beheshtirad, M., & Pradhan, B. (2016). A comparative assessment of prediction capabilities of modified analytical hierarchy process (M-AHP) and Mamdani fuzzy logic models using Netcad-GIS for forest fire susceptibility mapping. *Geomatics, Natural Hazards and Risk*, *7*(2), 861–885. https://doi.org/10.1080/19475705.2014.984247

Pourtaghi, Z. S., Pourghasemi, H. R., Aretano, R., & Semeraro, T. (2016). Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecological Indicators*, *64*, 72–84. https://doi.org/10.1016/J.ECOLIND.2015.12.030

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*, *9*(2), 181–199. https://doi.org/10.1007/s10021-005-0054-1

Preisler, H. K., Westerling, A. L., Gebert, K. M., Munoz-Arriola, F., & Holmes, T. P. (2011). Spatially explicit forecasts of large wildland fire probability and suppression costs for California. *International Journal of Wildland Fire*, *20*(4), 508–517. Retrieved from https://doi.org/10.1071/WF09087

R Core Team. (2019). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria. Retrieved from https://www.r-project.org/

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice. *Journal of Economic Literature*, *44*(3), 631–661. Retrieved from http://www.aeaweb.org/articles?id=10.1257/jel.44.3.631

Rihan, W., Zhao, J., Zhang, H., Guo, X., Ying, H., Deng, G., & Li, H. (2019). Wildfires on the Mongolian Plateau: Identifying drivers and spatial distributions to predict wildfire probability. *Remote Sensing*, *11*(20). https://doi.org/10.3390/rs11202361

Roloff, G. J., Mealey, S. P., Clay, C., Barry, J., Yanish, C., & Neuenschwander, L. (2005). A process for modeling short- and long-term risk in the southern Oregon Cascades. *Forest Ecology and Management*, *211*(1), 166–190. https://doi.org/https://doi.org/10.1016/j.foreco.2005.02.006

Rose, A. N., McKee, J. J., Urban, M. L., & Bright, E. A. (2018). LandScan 2017. Oak Ridge, TN: Oak Ridge National Laboratory SE - July 1, 2018. Retrieved from https://landscan.ornl.gov/

Salis, M., Ager, A. A., Alcasena, F. J., Arca, B., Finney, M. A., Pellizzaro, G., & Spano, D. (2014). Analyzing seasonal patterns of wildfire exposure factors in Sardinia, Italy. *Environmental Monitoring and Assessment*, *187*(1), 4175. https://doi.org/10.1007/s10661-014-4175-x

Salis, M., Ager, A. A., Finney, M. A., Arca, B., & Spano, D. (2014). Analyzing spatiotemporal changes in wildfire regime and exposure across a Mediterranean fire-prone area. *Natural Hazards*, *71*(3), 1389–1418. https://doi.org/10.1007/s11069-013-0951-0

Satir, O., Berberoglu, S., & Donmez, C. (2016). Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem. *Geomatics, Natural Hazards and Risk*, *7*(5), 1645–1658. https://doi.org/10.1080/19475705.2015.1084541

Schimanke, S., Isaksson, L., Edvinsson, L., Undén, P., Ridal, M., Moigne, P. L., … Glinton, M. (2019). UERRA data user guide, 49. Retrieved from https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-uerra-europe-single-levels?tab=overview

Schmoldt, D. L. (2001). Application of Artificial Intelligence to Risk Analysis for Forested Ecosystems BT - Risk Analysis in Forest Management. In K. von Gadow (Ed.) (pp. 49–74). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-2905-5_3

Schultz, M. G., Heil, A., Hoelzemann, J. J., Spessa, A., Thonicke, K., Goldammer, J. G., … van Het Bolscher, M. (2008). Global wildland fire emissions from 1960 to 2000. *Global Biogeochemical Cycles*, *22*(2), 1–17. https://doi.org/10.1029/2007GB003031

Scoccimarro, E. (2011). Extreme Events in High Resolution CMCC Regional and Global Climate Models. *SSRN Electronic Journal*.

Shokouhi, B. V., & Eslami, M. (2019). Fuzzy logic based burned severity classification and mapping with Landsat-8 data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(2/W16), 259–265. https://doi.org/10.5194/isprs-archives-XLII-2-W16-259-2019

Sikder, I. U., Mal-Sarkar, S., & Mal, T. K. (2006). Knowledge-Based Risk Assessment Under Uncertainty for Species Invasion. *Risk Analysis*, *26*(1), 239–252. https://doi.org/10.1111/j.1539-6924.2006.00714.x

Sirca, C., Casula, F., Bouillon, C., García, B. F., Fernández Ramiro, M. M., Molina, B. V., & Spano, D. (2017). A wildfire risk oriented GIS tool for mapping Rural-Urban Interfaces. *Environmental Modelling and Software*, *94*, 36–47. https://doi.org/10.1016/j.envsoft.2017.03.024

Song, H.-S., & Lee, S.-H. (2017). Effects of wind and tree density on forest fire patterns in a mixed-tree species forest. *Forest Science and Technology*, *13*(1), 9–16. https://doi.org/10.1080/21580103.2016.1262793

Soukup, T., Sousa, A., & Langanke, T. (2017). CLC2018 Technical Guidelines, (3436), 0–60. Retrieved from https://land.copernicus.eu/pan-european/corine-land-cover

Suryabhagavan, K. V, Alemu, M., & Balakrishnan, M. (2016). GIS-based multi-criteria decision analysis for forest fire susceptibility mapping : a case study in Harenna forest , southwestern Ethiopia, *57*(1), 33–43.

Swetnam, T. W., & Betancourt, J. L. (1998). Mesoscale Disturbance and Ecological Response to Decadal Climatic Variability in the American Southwest. *Journal of Climate*, *11*(12), 3128–3147. https://doi.org/10.1175/1520-0442(1998)011<3128:MDAERT>2.0.CO;2

Syphard, A. D., Radeloff, V. C., Keeley, J. E., Hawbaker, T. J., Clayton, M. K., Stewart, S. I., & Hammer, R. B. (2007). Human influence on California fire regimes. *Ecological applications*, *17*(5), 1388-1402. https://doi.org/10.1890/06-1128.1

Syphard, A. D., Radeloff, V. C., Keuler, N. S., Taylor, R. S., Hawbaker, T. J., Stewart, S. I., & Clayton, M. K. (2008). Predicting spatial patterns of fire on a southern California landscape. *International Journal of Wildland Fire*, *17*(5), 602–613. https://doi.org/10.1071/WF07087

Thompson, M. P., & Calkin, D. E. (2011). Uncertainty and risk in wildland fire management: A review. *Journal of Environmental Management*, *92*(8), 1895–1909. https://doi.org/10.1016/j.jenvman.2011.03.015

Thompson, M. P., Calkin, D. E., Gilbertson-Day, J. W., & Ager, A. A. (2011). Advancing effects analysis for integrated, large-scale wildfire risk assessment. *Environmental Monitoring and Assessment*, *179*(1), 217–239. https://doi.org/10.1007/s10661-010-1731-x

Tien Bui, D., Tuan, T. A., Klempe, H., Pradhan, B., & Revhaug, I. (2016). Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, *13*(2), 361–378. https://doi.org/10.1007/s10346-015-0557-6

Torres, F. T. P., Romeiro, J. M. N., Santos, A. C. de A., de Oliveira Neto, R. R., Lima, G. S., & Zanuncio, J. C. (2018). Fire danger index efficiency as a function of fuel moisture and fire behavior. *Science of the Total Environment*, *631–632*, 1304–1310. https://doi.org/10.1016/j.scitotenv.2018.03.121

Vadrevu, K. P., Eaturu, A., & Badarinath, K. V. S. (2010). Fire risk evaluation using multicriteria analysis—a case study. *Environmental Monitoring and Assessment*, *166*(1), 223–239. https://doi.org/10.1007/s10661-009-0997-3

Valentina Bacciu, Grazia Pellizzaro, M. S. (2011). Estimating vegetation fire emissions from Sardinian wildland fires (2005-2009). In *Conference: International Conference on Fire Behaviour and Risk*.

Vilar del Hoyo, L., Martín Isabel, M. P., & Martínez Vega, F. J. (2011). Logistic regression models for human-caused wildfire risk estimation: analysing the effect of the spatial accuracy in fire occurrence data. *European Journal of Forest Research*, *130*(6), 983–996. https://doi.org/10.1007/s10342-011-0488-2

Westerling, A. L., Hidalgo, H. G., Cayan, D. R., & Swetnam, T. W. (2006). Warming and earlier spring increase Western U.S. forest wildfire activity. *Science*, *313*(5789), 940–943. https://doi.org/10.1126/science.1128834

Wilson. J.P., G. J. C. (2000). Primary topographic attributes. In *Terrain Analysis: Principles and Applications* (pp. 51–85).

Wotton, B. M., & Flannigan, M. D. (1993). Length of the fire season in a changing climate. *The Forestry Chronicle*, *69*(2), 187–192. https://doi.org/10.5558/tfc69187-2

Zawadzki Zygmunt, Marcin Kosinski, Krzysztof Slomczynski, D. S., & Rcpp, L. (2019). Package ‘ FSelectorRcpp .’ Retrieved from https://cran.r-project.org/web/packages/FSelectorRcpp/index.html)

Non-exclusive licence to reproduce thesis and make thesis public

I, Behzad Valipour Shokouhi,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

"Mapping Wildfire Susceptibility of Sardinia Island, Italy", supervised by Dr. Evelyn Uuemaa

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Behzad Valipour Shokouhi

Tartu, **16.01.2020**