Ecological Indicators 64 (2016) 72–84

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques

Zohre Sadat Pourtaghi [a], Hamid Reza Pourghasemi [b,*], Roberta Aretano [c], Teodoro Semeraro [c]

[a] Department of Environment Management Engineering, College of Natural Resources, Yazd University, Yazd, Iran
[b] Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz, Iran
[c] Laboratory of Landscape Ecology, Department of Biological and Environmental Sciences and Technologies, University of Salento, Lecce, Italy

ARTICLE INFO

Article history:
Received 8 July 2015
Received in revised form 8 December 2015
Accepted 18 December 2015
Available online 7 January 2016

Keywords:
Forest fire susceptibility mapping
Boosted regression tree
Generalized additive model
Random forest
GIS
Iran

ABSTRACT

Forests are living dynamic systems and these unique ecosystems are essential for life on earth. Forest fires are one of the major environmental concerns, economic, and social in the worldwide. The aim of current research is to identify general indicators influencing on forest fire and compare forest fire susceptibility maps based on the boosted regression tree (BRT), generalized additive model (GAM), and random forest (RF) data mining models in the Minudasht Township, Golestan Province, Iran. According to expert opinion and literature review, fifteen condition factors on forest fire have been selected in the study area. These are slope degree, slope aspect, elevation, topographic wetness index (TWI), topographic position index (TPI), plan curvature, wind effect, annual temperature and rainfall, soil texture, distance to roads, rivers, and villages, normalized difference vegetation index (NDVI), and land use. Forest fire locations were identified using MODIS images, historical records, and extensive field checking. 106 (≈70%) locations, out of 151 forest fires identified, were used for models building/training, while the remaining 45 (≈30%) cases were used for the models validation.

BRT, GAM, and RF data mining models were used to distinguish between presence and absence of forest fires and its mapping. These algorithms were used to perform feature selection in order to reveal the variables that contribute more to forest fire occurrence. Finally, for validation of models, the area under the curve (AUC) for forest fire susceptibility maps was calculated. The validation of results showed that AUC for three mentioned models varies from 0.7279 to 0.8770 (AUC_{BRT} = 80.84%, AUC_{GAM} = 87.70%, and AUC_{RF} = 72.79%,). Results indicated that the main drivers of forest fire occurrence were annual rainfall, distance to roads, and land use factors. The results can be applied to primary warning, fire suppression resource planning, and allocation work.

1. Introduction

Forests are living dynamic systems with a biologically diverse and complicated structure (Aleemahmoodi Sarab et al., 2013). Forests in the world are a source of ecosystem services essential for human well-being, playing a crucial role in the persistence of vital processes in the environment, including weather adjustment, carbon preservation (UNEP, 2007) and regulating the climate and the carbon cycle. However, forest ecosystems are increasingly threatened by fires caused by natural and anthropogenic factors (Chen et al., 2012; Naebi, 2003; Ghomi Motazeh et al., 2013). Fires can be a destructive ecological factor causing many negative effects in various aspects of life, including natural environment, economics, and health (Herawati et al., 2006; Ghomi Motazeh et al., 2013), but a good plan can provide suitable tools for ecosystem management. In this content, fire prevention must be paid special attention (Ghomi Motazeh et al., 2013). Fire is a natural power that affects vegetation communities over time and, as a natural process, it has a significant effect on conserving ecosystems health. Since twentieth century, increasing fires due to human inattention turned fires to a main threat for forests (Nasi et al., 2002; Ghomi Motazeh et al., 2013).

It is possible both to control nature and also to provide the fire risk map and thereby to minimize the fire frequency and avert damage (Jaiswal et al., 2002). A precise evaluation of forest fire problems and decision on solutions can be satisfactory when fire risk mapping is available (Jaiswal et al., 2002; Erten

* Corresponding author.
E-mail addresses: hamidreza.pourghasemi@yahoo.com, hr.pourghasemi@shirazu.ac.ir (H.R. Pourghasemi).

http://dx.doi.org/10.1016/j.ecolind.2015.12.030
1470-160X/

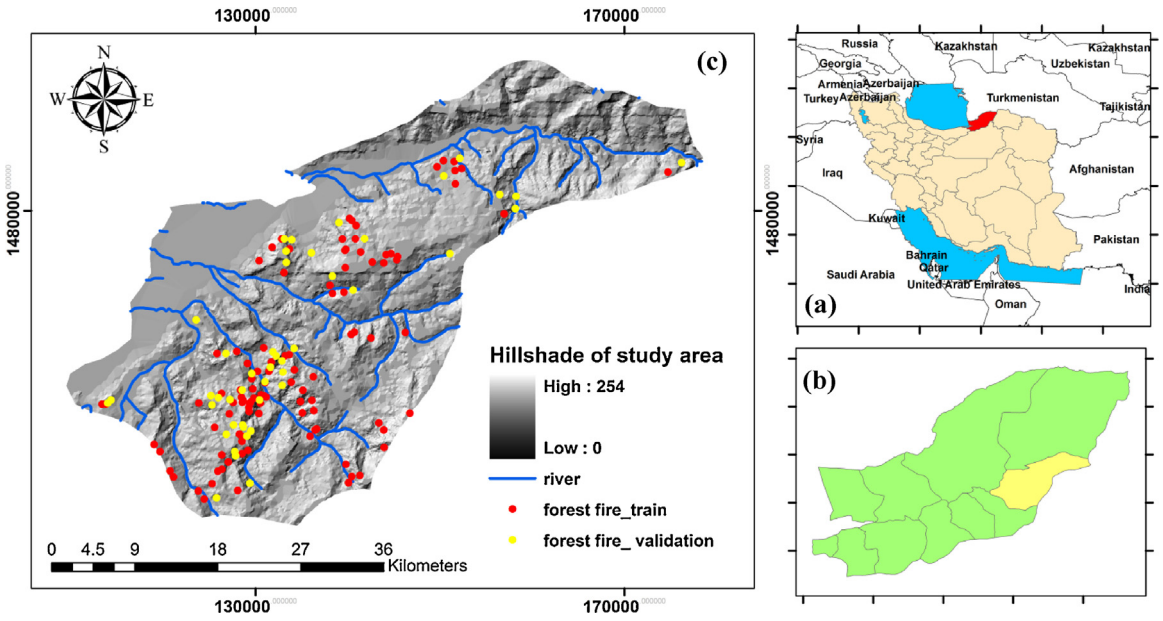

Fig. 1. Location of the study area; (a) Iran map (b) Golestan Province map; (c) Forest fire location map with hill shaded map.

et al., 2004). Various methods and algorithm for fire hazard zoning by Remote Sensing (RS) and Geospatial Information System (GIS) have been presented. Many studies have been executed to produce forest fire risk maps using GIS and RS techniques (Chuvieco and Congalton, 1989; Prosper-Laget et al., 1995; Chuvieco and Sales, 1996; Jaiswal et al., 2002; Erten et al., 2004; Wulder and Franklin, 2006; Pradhan et al., 2007; Razali, 2007; Saklani, 2008; Chuvieco et al., 2010; Adab et al., 2013; Ariapour and Shariff, 2014; Arpacia et al., 2014; Eskandari and Chuvieco, 2015; Salvatia and Ferrara, 2015; Pourghasemi, 2015; Aretano et al., 2015). Vasconcelos et al. (1995) noted that vegetation, topography, climatology, and fire history are considerable components of hazard in order to assess forest fire risk. Moreover, Pradhan et al. (2007) emphasized that normalized differential vegetation index (NDVI), soil, slope, aspect, and land use were the efficient factors to assess fire risk hazard. Janbaz Ghobadi et al. (2012) used topography, vegetation, slope, aspect, NDVI, and meteorology factors to provide forest fire risk map. Recently, new statistical techniques, namely data mining techniques, have been developed to assess fire risk (Chen et al., 1996; Fayyad et al., 1996; Zhu and Davidson, 2007; Gutiérrez et al., 2009). These techniques allow analyzing complex multivariate problems and building predictive models from large datasets. The main advantages of these models are: the estimation of the potential spatial distribution of a phenomenon, the anticipation of future changes on its allocation and the establishment of the importance of each predictor in determining the distribution of the target variable. A vast number of techniques to construct predictive models exist, such as generalized linear model (GLM), logistic multiple regression (LMR), generalized additive model (GAM), classification and regression trees (CART), and artificial neural network (ANN). The random forest (RF) is a development of classification and regression tree (CART) methods (Breiman, 2001; Shataee et al., 2011). Additionally, the boosted regression tree (BRT) is a combination of statistical and machine learning techniques and an extension of CART, as promising technique utilized in ecological modeling (Aertsen et al., 2010; Shataee et al., 2011). Stojanova et al. (2006) used data mining techniques such as logistic regression, and decision trees to predict forest fires in Slovenia. Leuenberger et al. (2013) used random forest analysis to map forest fire occurrences in Swiss Alps. Woolford et al. (2009) used generalized additive model to assess a spatio-temporal model for forest fires caused by people in a portion of boreal forest in northeastern Ontario, Canada.

High risk of wildfire events in forested areas exists in Iran. About seven percent of Iran area is covered by forests/and wooded lands. According to FAO (food and agriculture organization) database on forest/other wooded land fires in Iran (Movaghati et al., 2008), the number of fires per year equals 130 and the average burnt area per year is 5400 ha (Allard, 2003; Movaghati et al., 2008). However, fires are not largely monitored and enough detection facilities are not available. The development of various models based on simulation and mathematics in developing countries such as Iran should be considered, because traditional methods are now slowly given up and other ancient philosophies are less fit to the current world situation (Zeki and Keles, 2005; Ghomi Motazeh et al., 2013). Forest fire in Golestan Province in Iran, is still one of the most natural hazard problems. According to the last decade studies, 9068 ha of the forests have been burnt by fire (Janbaz Ghobadi et al., 2012). Thus, the main purpose of this study is the comparison of BRT, GAM, and RF statistical and decisions tree based regression models for forest fire modeling. In particular, specific aims are: (i) to identify and evaluate the importance of the general indicators involved; (ii) to develop a model capable of predicting forest fire location using BRT, GAM, and RF data mining models in the Minudasht Township of Golestan Province, Iran; (iii) to implement the model into a geographical information system to offer a data mining methodology for forest fire susceptibility mapping in Iran. These maps could be used for early warning, fire suppression resource planning, and allocation works for designing strategies to prevent forest fire or to employ specific measures for controlling occurred forest fire. Furthermore, these maps could provide guidance on how changes in land use or climate will influence the distribution and density of forest fires, especially in Iran County.

2. Study area

The study area, as shown in Fig. 1, lies in the eastern part of Golestan Province, in northern Iran. The geographical location of the study area is between latitudes 37°00′27″ to 37°27′53″ N, and longitudes 55°14′00″ to 56°00′39″ E. It covers an area about 1531 km^2. The elevation of the study area ranges of 100–2500 m above sea level. The climate of Minudasht is moderate and