Environmental Advances 4 (2021) 100041

Contents lists available at ScienceDirect

Environmental Advances

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

Patterns of mega-forest fires in east Siberia will become less predictable with climate warming

Michael Natole Jr.[a,*], Yiming Ying[a], Alexander Buyantuev[b], Michael Stessin[a], Victor Buyantuev[c], Andrei Lapenis[b]

[a] Department of Mathematics and Statistics, University at Albany, SUNY, US
[b] Department of Geography and Planning, University at Albany, SUNY, US
[c] Laboratory of Ecosystems Bioindication, Siberian Institute of Plant Physiology and Biochemistry SB RAS, Irkutsk, Russia

ARTICLE INFO

Keywords:
Burned area
Factor analysis
Fire frequency
Human influence
Machine learning
Spatial analysis
Wildfires

ABSTRACT

Very large fires covering tens to hundreds of hectares, termed mega-fires, have become a prominent feature of fire regime in taiga forests worldwide, and in Siberia in particular. Here, we applied an array of machine learning algorithms and statistical methods to estimate the relative importance of various factors in observed patterns of Eastern Siberian fires mapped with satellite data. More specifically, we tested linkages of "hot spot" ignitions with 42 variables representing landscape characteristics, climatic, and anthropogenic factors, such as human population density, locations of settlements and road networks. Analysis of data spanning seventeen years (2001–2017) showed that during low or moderately high fire seasons, models with full set of variables predict locations of fires with a very high probability (AUC = 95%). Sensitivity, or the ratio of correctly predicted fire pixels to the total number of pixels analyzed, declined to 30–40% during warm and dry years of increased fire activity, especially in models driven by anthropogenic variables only. This analysis demonstrates that if warming in Eastern Siberia continues, forest fires will become not only more frequent but also less predictable. We explain this by examining model performance as a function of either temperature or precipitation. This effect from climate makes it nearly impossible to segregate ignition points from locations, which were burnt several hours or even several days earlier. An increase in secondary burnt locations makes it difficult for machine learning algorithms to establish causality links with anthropogenic and other groups of variables.

1. Introduction

Regions throughout the world are experiencing an increase in frequency and intensity of wildfires resulting in devastating impacts on ecosystems and human-dominated landscapes (Abatzoglou and Williams, 2016; Cascio, 2018; Ertugrul et al., 2019; Shvidenko et al., 2011; Westerling et al., 2006). Environmental problems, such as forest mortality and degraded air quality, and the increasing economic cost of combatting fires, demand for better wildfire forecasting, especially in the wake of the recently emerged mega-fires that take the form of ecological disasters since they burn with high intensity and affect vast tracts of land, impact vegetation, wildlife habitat, carbon sequestration, and other ecosystem services (Bowman et al., 2017; Stephens et al., 2014; Turner, 2010). Understanding principle factors that lead to a wildfire of that magnitude should enable developing better and more cost-effective strategies to prevent and mitigate mega-fires. Both ecological and anthropogenic factors drive the conditions that allow wildfires to occur (Hantson et al., 2016). In particular, the wildland-urban interface (Spyratos et al., 2007) is prone to increased frequency of fires since human activity contributes significantly to not only the cause of an ignition, but it amplifies the impact of other contributing factors through policy and land use. Therefore, understanding the driving factors of wildfires is paramount for developing effective fire management strategies (Finney, 2005; Syphard and Keeley, 2015).

Forecasting the occurrence of a wildfire requires predicting the availability of two essential components: readily available fuel and the possibility of an ignition event (Allen et al., 2002). Human activity has had a significant impact on both of these factors. Forest clearing and other human activities, such as arson, have not only contributed to an increase in the frequency of fires, but the magnitude of them by making areas more flammable (Kukavskaya et al., 2013; Lindenmayer et al., 2020). Even more so, climate change has had substantial impacts on fire regimes throughout the world (Ertugrul et al., 2019; 2021; Goss et al., 2020). The role of climate is critical for forecasting fire events (Pereira et al.,

* Corresponding author.
E-mail address: mnatole.jr@me.com (M. Natole Jr.).

https://doi.org/10.1016/j.envadv.2021.100041
Received 30 December 2020; Received in revised form 25 February 2021; Accepted 1 March 2021

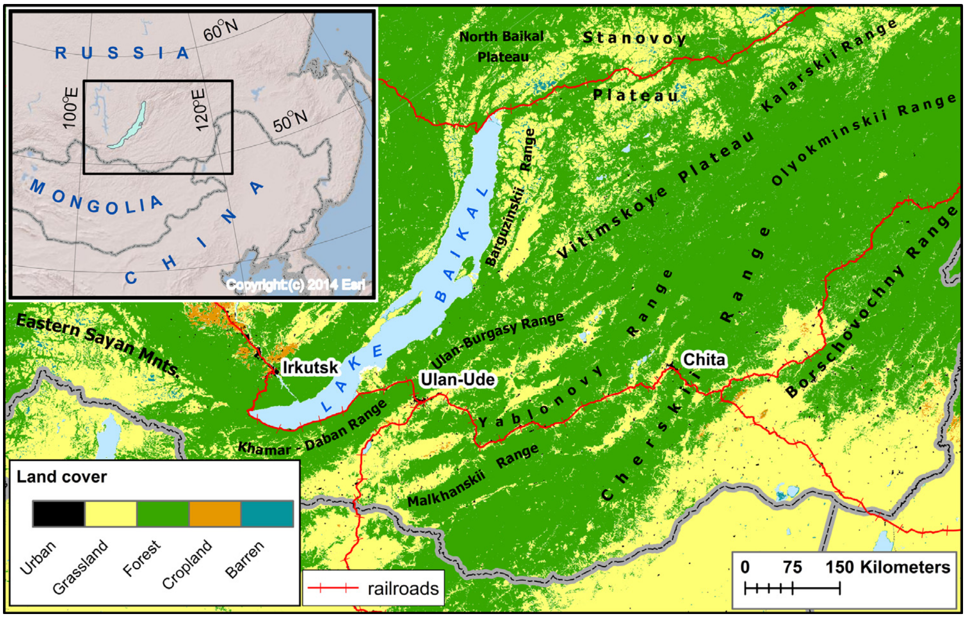

Fig. 1. Study Area. Uneven, south-west to north-east stripes of forest steppe (yellow color) reflect on complex topography.

2020). As climate conditions enable significant fire events to occur, the impact of human activity is far reaching (Cullen et al., 2021; Jones et al., 2020; Robinne et al., 2018). The disturbance in the fire regime caused by human activity has caused damages to the ecological balance of many regions around the world. To account for the importance of human activity, ignitions can either be classified as human or natural with both top-down and bottom-up drivers determining the spatial patterns (Massada et al., 2013). Factors such as weather and climate are considered top-down drivers while bottom-up drivers are local variables that affect fuel sources. Common examples are landscape features (slope, aspect, etc.) and proximity to human created structures (roads, towns), as well as population. Although human activity is a principle driver of such events, an ignition source and fuel availability are attributed to a combination of climatic, anthropogenic, and ecological factors. Wildfire models which use various methods typically consider a combination of such factors.

Using various groups of data, researchers have applied simulation modeling, traditional statistical modeling, and machine learning methods to gain a better understanding of the factors that contribute to wildfires and develop more effective models. The most notable one has been statistical methods, specifically, logistic regression (Cardille et al., 2001; Catry et al., 2010; Massada et al., 2009) has been applied extensively. Other such statistical methods include generalized linear mixed models (GLMM) (Finney et al., 2009) and geographic weighted regression (Koutsias et al., 2010). More recently, there has been a push in using machine learning methods (Massada et al., 2013; Parisien and Moritz, 2009). Those include classification and regression trees (CART), which are more capable of capturing the complex relationships between variables (Breiman, 2001; De'ath and Fabricius, 2000; Sturtevant and Cleland, 2007), and boosted regression trees (BRT) using a random subset (Friedman, 2002) applied at multiple scales in wildfire studies in the United States (Parisien and Moritz, 2009). More robust methods, such as Random Forest (RF) (Breiman, 2001; Prasad et al., 2006) and Maximum Entropy (MaxEnt) (Phillips et al., 2006), have also been adopted. Unlike the previous methods, the MaxEnt model uses presence only data. Neural networks have also been used for wildfire ignition prediction (Chuvieco et al., 2003; De Vasconcelos et al., 2001), but determining the significant factors using this approach is challenging (Cheng and Wang, 2008; Satir et al., 2016). Support vector machines (Vapnik et al., 1997) is a popular method that has also recently been applied to wildfire analysis (Jaafari and Pourghasemi, 2019).

A significant drawback of many previous studies is that they all use modeling methods based on a mixture of predictor variables. However, predictor variables describing complex ecological systems should not be combined in a simple linear fashion. Anthropogenic, climate, and other natural factors are often reciprocally related and amplified by each other. In 2003, for example, East Siberia experienced one of the most devastating fire seasons in recent history (Figs. 1 and 2). This anomaly was preceded by extremely dry winter months from December 2002 to March 2003 (Sitnov and Mokhov, 2018). In adjacent northern boreal forests of China, where climatic conditions were about the same, however, fires in 2003 did not reach the same magnitude as in the south of East Siberia (Huang et al., 2009). These contrasting responses of two geographically close regions to similar weather patterns demonstrate the importance of fire prediction based on such factors as topography, anthropogenic features, as well as forest management practices. Topographic slope, aspect, and elevation, for example, often determine the type of vegetation cover and susceptibility of a landscape to fires. Even more so, fire history of a region can play a role in the fire regime. Regions where rainfall has been decreasing, for example, are becoming more susceptible to ignitions by possessing abundant fuel that would not have been readily available previously. Shifts in human population have made some regions hotbeds for fire activity due to the impact of human actions and policy changes (Abatzoglou and Williams, 2016; Balch et al., 2017).

Therefore, to better understand principle wildfire drivers and their interactions it is necessary to have as much control over variables as possible. Here we chose the region of the boreal forest within the Russian Federation subjected to the same forest management practices typical of the south of East Siberia with vast open areas and low population densities (Fig. 1). At the same time, we also split variables into nearly uniform groups, including climatic, anthropogenic, topographic, and landscape variables. It should be expected, however, that all these factors might correlate with fire frequency and areas burned. Yet, the strength of this connection should vary and reflect on the importance of some of these variables in the formation of spatial patterns of forest fires.

The goal of our study was to investigate the significance of natural and human drivers of wildfires occurring in the south of East Siberia in the hopes of determining better forecasting methodologies. The region is known for its high fire activity, which is hypothesized to increase due to recent climate trends (Groisman et al., 2013) and rapid institutional transformations entailing changes in land use and related policies in this

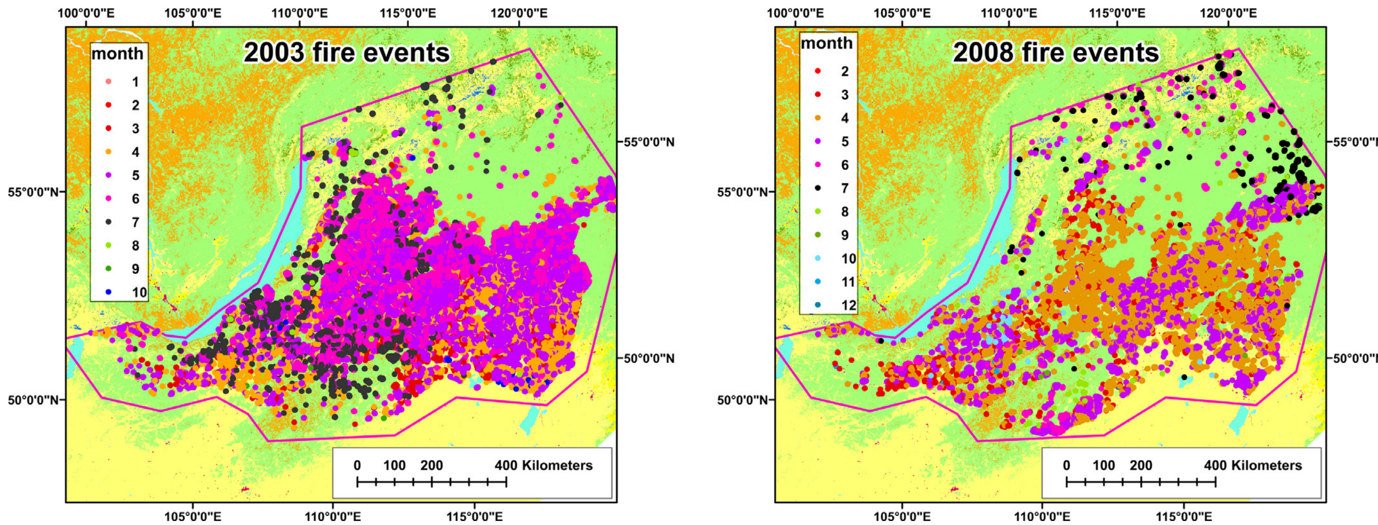

Fig. 2. Spatial patterns of observed fires colored by month. Year 2003 had the most fire events compared to all other years observed. This unprecedented number of forest mega-fires during 2003 was noticed in the entire zone of boreal forests of Northern Eurasia and Northern America (Sitnov and Mokhov, 2018). Most probable cause for this anomaly was extremely low winter precipitation (mean maximum snow water equivalent) and hot summer temperatures (Groisman et al., 2013). In all seasons fires were more frequent in the southern part of selected region. The same figures demonstrate importance of mountain ridges extending from south-west to northeast (2008).

Table 1
Predictor variables used in modeling. The dependent variable is categorical and characterizes presence or absence of fire in any given 10x10 km pixel.

Type	Name	Type	Name
Anthropogenic	Distance to Road	**Landscape**	% Forest
	Distance to Town		% Grasslands
	Population		Slope
	% Railway		Elevation
	% Highway		Aspect
	% Unpaved Road	**Climatic**	Monthly average temperature (Jan - Dec)
	% Trail		Monthly total precipitation (Jan - Dec)
	% Seasonal Road		
	% Bridges		
	% City		
	% Village		
	% Seasonal Settlement		
	% Abandoned Settlement		

area (Kukavskaya et al., 2013; Shvidenko et al., 2011). More specifically, we tested the overall response of fire events to important explanatory variables we were able to identify, as well as the explanatory power of three groups of variables individually (Table 1). The first group, which combines climatic variables, consisted of twelve-monthly average temperatures and total monthly precipitation. The anthropogenic group was composed of variables related to human activities, such as distance to road, distance to town, population, etc. The third group, which we refer to as the landscape group, comprised of topographic features such as slope, aspect, elevation, and two landscape characteristics - percent of forest and percent of grassland. We employed four statistical and machine learning algorithms (models), each driven by the full set of variables and by three individual groups of variables, to describe spatial distribution of fires in the region (Fig. 3) during the 2001-2017 fire seasons. Thus, each of the seventeen fire seasons was modeled sixteen times: four models, each forced by four groups of variables. Each simulation consisted of training with cross-validations and testing runs. The purpose of these simulations was a) to test performance of various mathematical models and b) estimate relative importance of anthropogenic, climatic and landscape factors in formation of geographic patterns of wildfires. During these experiments, however, we found that model performance greatly depended on the state of fire emergency caused by inter-annual climatic fluctuations.

2. Materials and methods

2.1. Study area

Southern Eastern Siberia (Fig. 1) is characterized by a high frequency of wildfires during dry conditions of mid-spring and throughout the summer (Krylov et al., 2014). Furthermore, fire risk has remarkably increased because of recent political and socio-economic transitions in this region. The dissolution of the Soviet Union and consequential institutional transformations in the 1990s are believed to contribute to the increase of fires. One reason for such an increase is the lack of control and fire management while another reason is the economically motivated arson by local timber dealers with the purpose of salvage logging largely spurred by the increased timber trade between Russia and China (Narins, 2015). Illegal logging in this area has been the inseparable component of this increased pressure on forest resources, which has led to an increase in fire hazard (Kukavskaya et al., 2013). This predominantly mountainous region characterized by extremely continental climate with long and cold winters and short and hot summers is occupied by southern taiga transitioning into the grassland biome. Forests are dominated by coniferous species - Scots pine (*Pinus silvestris*) and Siberian pine (*P. sibirica*) and larch trees, *Larix sibirica* with some *L. gmelinii*. The ecotone zone between boreal forest and grassland is prone

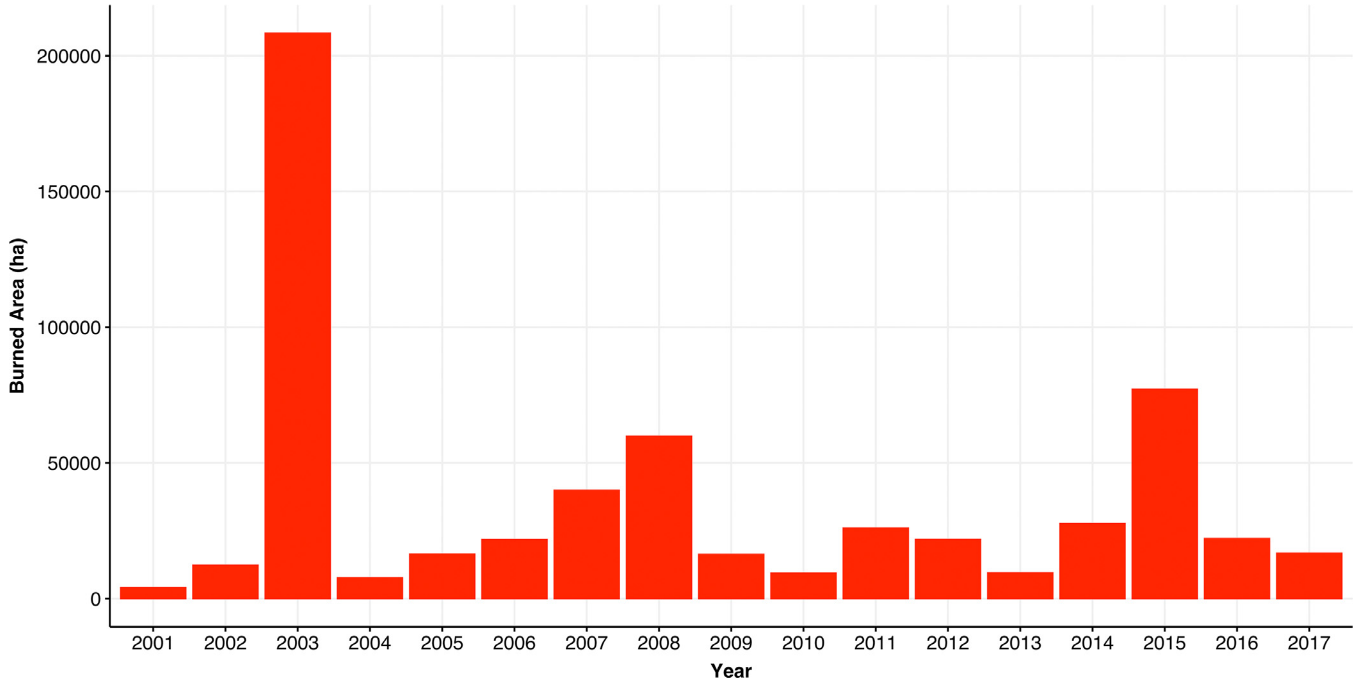

Fig. 3. Burned area (in ha) for wildfires from 2001 to 2017.

Table 2
Summary of total ignition points and fire area per each year versus climatic factors. Maximum and minimum values during 2001–2017 are highlighted in bold font.

Year	Number of ignition points	Burn area [ha]	Total annual precipitation [mm]	Mean annual temperature [oC]	Index of Arctic Oscillations
2001	**2508**	**4028**	357.2	4.1	0.16
2002	5308	12,321	361.6	3.2	0.07
2003	**111,552**	**208,303**	366.8	3.8	0.15
2004	3800	7687	333	3.6	0.19
2005	7417	16,372	379.7	4	0.38
2006	8784	21,734	356.4	4.8	0.14
2007	8951	39,906	354.1	2.6	0.27
2008	34,496	59,878	449.1	3.2	0.18
2009	6906	16,300	420.5	4.7	0.33
2010	2833	9430	377.3	5.1	1.04
2011	11,154	26,037	350.6	3.6	0.53
2012	15,154	21,796	**496.1**	5	0.18
2013	4534	9514	442.2	4.2	0
2014	9934	27,629	357.2	3.3	0.07
2015	24,224	77,160	**320**	3	**0.63**
2016	12,898	22,155	366.7	4	0.11
2017	5399	16,711	376.8	3.1	0.26
Min	**2508**	**4028**	**320**	5.1	1.04
Max	**111,552**	**208,303**	**496.1**	2.6	**0.63**

to rapid successions and boundary shifts following forest disturbances (Soja et al., 2007). Being in proximity to major transportation routes this area has been historically populated and economically developed, but mainly along those routes. Although wildfires are the dominant ecosystem disturbance, fires also create considerable threats to humans and economy in the area. Human activities in the wilderness, mainly industrial logging (Kukavskaya et al., 2013) and mineral resource extraction, had in turn been a significant impact on the region's fire regime. The summary of the 2001–2017 fire history and its spatial patterns is provided in Fig. 3 and Table 2.

2.2. Data and variables

We used the FIRMS active fire data product available as the 2001–2017 archive of ignition points estimated using thermal data from the MODIS instrument (Giglio et al., 2016). Burned area estimates were obtained from the MCD64A1 MODIS product available at 500-m resolution (Giglio et al., 2015) by using the Google Earth Engine to extract individual monthly data layers and clipping them to our study area. Land cover data was obtained from the MODIS MCD12Q1 product (Friedl and Sulla-Menashe, 2015) by aggregating the IGBP classification into broader categories (Fig. 3). Climate variables included air temperature and precipitation obtained from the NCEP/NCAR Reanalysis 1 data (Kalnay et al., 1996). For temperature, we used monthly mean values, which were also converted to annual mean temperatures. Monthly precipitation values were additionally aggregated to seasonal (January-May) precipitation. For each fire season, we considered climate variables of the same year and the previous year.

We used the ASTER Global Digital Elevation Model product (ASTGTM) available at 30-m spatial resolution (Spacesystems and

/Japan, 2019) to create topographic variables-elevation, slope and aspect. Finally, standard digital topographic data for the area included GIS vector layers of roads, trails, and settlements with population. A total of 42 predictor variables (Table 1) describing climatic, land cover, topographic and anthropogenic characteristics were included to determine which factors are most significant in predicting fire ignition and how are those factors related. All selected variables were rasterized to the same grain size of 10X10 km. Categorical variables (see the full list in Table 1) were transformed using a one-hot encoder and then averaged to determine a proportion for that particular cell size. For each year, a background dataset of about 100,000 locations was generated. The dataset was split with 80% for training and the remaining 20% for validation. For each of the methods used, we separated the variables into anthropogenic, climate, and landscape models. Using this approach, we were able to compare the relationship between the different classes of variables using the statistical and machine learning models.

2.3. Statistical analysis

We applied the the most commonly used statistical model, generalized linear model (GLM), to determine the effect of climate variables on anthropogenic and natural models (Hastie and Tibshirani, 1990). In this work, we used the logit link with binomial response since our dependent variable is binary:

$$\log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \sum_{i=1}^{p} X_i\beta_i + \epsilon, \tag{1}$$

where $\mu = E(Y)$ is the probability of ignition occurrence, i.e. a 1, and $1-\mu$ is the probability of a background (no ignition) point. β_0 is the intercept, p is the number of explanatory variables X_i and ϵ is the error term. For variable importance, we conducted a t-test to determine which variables were significant at the 5% level and ranked variables using the z -values. To test for spatial autocorrelation, we inspected the semi-variogram of residuals. GLM results were produced using the *glm* package in R program (Team, 2013).

2.4. Machine learning methods

Machine learning methods are being increasingly used for wildfire analysis (Gholamnia et al., 2020; Sulova and Jokar Arsanjani, 2021). For our models, we considered three different modeling approaches. The first method is random forest (RF) (Breiman, 2001). It is based on the idea of constructing a large number of decision trees. Each tree in the ensemble makes a prediction and the outcome that occurs most frequently is the chosen label. RF is considered a robust algorithm because it can overcome instability issues resulting from the use of a single tree result. In our work, the trees were tuned to determine the best number of trees to grow, the number of variables at each node, and the maximum number of nodes (Friedman, 2002). To determine the model error and variable importance, the generated trees were tested by calculating the mean square error (MSE) for each variable using the results from the out-of-bag samples (data not chosen for the modeling process). The difference in the MSE for each tree is averaged and normalized across the trees. A large increase in MSE indicates the significance of the variable (Cutler et al., 2007). Furthermore, a large gap in the increase in MSE can be used to group variables to different levels of significance or provide a cutoff for insignificant variables. RF models were created using the *randomForest* package in R (Liaw et al., 2002).

Maximum entropy (MaxEnt) is different from random forest in that it is a presence-only machine learning algorithm (Phillips et al., 2006). The method compares predictor variables at the presence locations against a sample of absence points and is based on the principle of maximum entropy (Elith* et al., 2006). A key assumption made by the model is that the average value of a constraint function is within an acceptable margin of error from the empirical average of the respective variable across all the presence locations. MaxEnt, therefore, selects the most uniform probability distribution as the best way to represent the data and each occurrence location is assigned a probability (Phillips et al., 2006). For species distribution modeling in ecology, MaxEnt has typically performed the best among presence-only methods (Elith* et al., 2006; Pearson et al., 2007). Variable importance is considered by observing the increase of the regularized training gain through the various iterations of model creation. The training gain is defined as the increase in the probability of a presence at training locations. In other words, variables that maximize their presence probability are considered the most important. Previous works have used a standalone software package; however, this work uses the *maxent* software in R (Phillips et al., 2017).

Support vector machines are popular techniques for machine learning and data mining tasks (Vapnik et al., 1997). In this method, the approach is to construct a line that separates two classes and also optimizes the distance between those two classes. For our work, we consider ϵ-SV regression (SVR) (Smola and Schölkopf, 2004). The main idea behind this method is to find a line of best fit that results in preselected error. A major consideration in SVR is whether data distribution is linear or non-linear. In most cases, a natural extension is to apply a kernel function for nonlinear data. Examples of such kernel functions include linear, polynomial, Gaussian, or the hyperbolic tangent. The use of support vector machines in wildfire forecasting has recently been applied (Jaafari and Pourghasemi, 2019; Syifa et al., 2020). For this work, we used the *kernlab* software package found in R (Karatzoglou et al., 2016). However, a significant drawback of this method is the difficulty of determining the significant model parameters.

As a way to measure model performance across the different methods, we calculated the area under the ROC curve (AUC) (Hanley and McNeil, 1982). The AUC metric is defined as the probability of ranking the minority class samples over the majority class samples. For this work, we want to rank ignition points over background locations. A significant advantage of using AUC is that the metric is considered threshold-independent because it evaluates models against all possible thresholds (Franklin, 2010). Values range between 0.5 and 1.0, where 0.5 is considered random guessing and 1.0 is perfect prediction. At the same time, AUC does not account for prevalence or different misclassification costs arising from false-negative and false-positive diagnoses (Halligan et al., 2015). Therefore, we calculated two more characteristics - model sensitivity, or the portion of successfully predicted fires, and specificity, or the portion of correctly predicted locations without fires. These two characteristics were estimated based on observed locations of fires and randomly selected locations without fires from the entire region. The number of ignition points and their exact locations significantly varied across 2001–2017 time period causing spatial patterns of fires to notably change (Fig. 3). Thus, model sensitivity and specificity reflect well on successful identification of specific fire patterns while the AUC, as mentioned above, characterizes overall probability of ranking individual fire points versus majority of locations which do not have fires.

3. Results

3.1. Performance of models

Performances of all machine learning and statistical models are summarized in AUC graphs (Fig. 4) while sensitivity and specificity are shown inTables 3–5. Additional graphs with AUC variability during the 2001–2017 period can be found in the supplementary materials (Figures S1-S4). As it is visible from these figures, performances of models constructed for each separate group of variables – anthropogenic, climatic, and landscape - were generally lower than the performances of the same models driven by the full set of variables. Depending on the model used the employment of the full set of variables allowed us to predict from 85% to 95% of observed ignition pixels (Fig. 4). It is important to note that AUC of 0.5 signifies a lack of meaningful prediction. Besides, models were tuned for each year separately. Therefore, such high val-

Table 3
Sensitivity and specificity, respectively, for each model year using climate variables.

Climate	GLM		RF		MaxEnt		SVR	
Year	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
2001	0.92	0.51	0.92	0.56	0.94	0.45	0.96	0.43
2002	0.8	0.57	0.88	0.54	0.8	0.59	0.94	0.59
2003	0.51	0.95	0.54	0.95	0.46	0.95	0.58	0.96
2004	0.81	0.61	0.89	0.56	0.9	0.32	0.92	0.59
2005	0.77	0.66	0.81	0.65	0.8	0.58	0.86	0.62
2006	0.82	0.5	0.76	0.65	0.76	0.58	0.81	0.74
2007	0.81	0.8	0.65	0.92	0.74	0.82	0.79	0.91
2008	0.47	0.92	0.6	0.85	0.48	0.87	0.67	0.9
2009	0.83	0.7	0.9	0.62	0.7	0.73	0.89	0.75
2010	0.91	0.32	0.95	0.35	0.8	0.52	0.9	0.57
2011	0.56	0.82	0.63	0.76	0.65	0.71	0.75	0.81
2012	0.71	0.81	0.65	0.89	0.59	0.86	0.82	0.91
2013	0.81	0.69	0.87	0.58	0.73	0.84	0.9	0.64
2014	0.69	0.74	0.67	0.75	0.65	0.77	0.77	0.82
2015	0.54	0.89	0.59	0.84	0.54	0.85	0.67	0.91
2016	0.72	0.82	0.71	0.86	0.78	0.67	0.83	0.88
2017	0.71	0.54	0.79	0.57	0.73	0.43	0.85	0.63

Table 4
Sensitivity and specificity, respectively, for each model year using anthropogenic variables.

Anthro.	GLM		RF		MaxEnt		SVR	
Year	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
2001	0.81	0.59	0.82	0.67	0.85	0.62	1	0.04
2002	0.66	0.53	0.82	0.48	0.41	0.82	0.97	0.54
2003	0.25	0.94	0.27	0.96	0.34	0.91	0.75	0.55
2004	0.77	0.41	0.71	0.61	0.83	0.37	0.99	0.51
2005	0.65	0.73	0.6	0.82	0.66	0.76	0.62	0.84
2006	0.59	0.73	0.66	0.71	0.7	0.63	0.8	0.82
2007	0.52	0.87	0.55	0.89	0.62	0.83	0.78	0.91
2008	0.45	0.88	0.43	0.9	0.41	0.91	0.71	0.9
2009	0.58	0.83	0.66	0.84	0.57	0.9	0.93	0.73
2010	0.75	0.65	0.82	0.51	0.8	0.54	0.95	0.56
2011	0.57	0.74	0.58	0.76	0.64	0.69	0.79	0.81
2012	0.56	0.78	0.57	0.86	0.57	0.86	0.9	0.8
2013	0.75	0.77	0.74	0.81	0.71	0.86	0.73	0.75
2014	0.58	0.7	0.59	0.75	0.59	0.75	0.79	0.83
2015	0.29	0.88	0.3	0.95	0.35	0.87	0.68	0.9
2016	0.59	0.58	0.72	0.69	0.51	0.85	0.84	0.81
2017	0.4	0.7	0.68	0.59	0.47	0.71	0.95	0.61

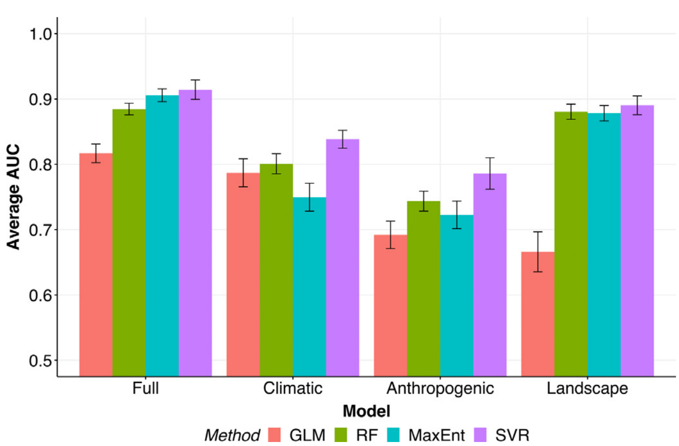

Fig. 4. Average performance and a 95% confidence interval for each model and group of variables.

ues of AUC should be taken with caution and the clear understanding of how experiments were conducted. Some insights in the limitation of our methods can be found through examination of prediction maps shown in supporting materials (Figures S5-S8). These maps demonstrate that for some years distribution of fires looks very uniform without reflecting on patterns visible in the data such as the alignment of most fires along major mountain ranges (Fig. 1). Yet, AUC values for these years can be quite high (Figure S1).

Among all models, the GLM resulted in the worst performance for any group of variables (AUC in the range from 60% to 80%). An indicator of poor performance of GLM model driven by the full set of variables is its low model sensitivity and specificity which, for some years, drop down to 50% (Tables 3–5). GLM suffered from overfitting issues in most years analyzed. On the other hand, SVR performed the best across all groups of variables and for most of the years. These findings are further corroborated by comparing prediction maps using the various methods provided in the supplementary materials (Figs. S5–S8). Apparently, spatial patterns of fire predictions generated by GLM look very homogenous and are less consistent with data in comparison with any of the machine learning algorithms. More specifically, GLM results were biased towards more uniform fire fields across the entire region oriented in the north-south or east-west directions. Therefore, despite AUC values greater than 60%, GLM provided little to no information as to where fires would actually occur. In contrast to GLM, practically all machine learning models driven by the complete set of variables accurately predicted increased fire activity in the southeast portion of the study area as well as helped to appreciate topographic effects of Yablonoy, Cherskii and Borschovochny mountain ranges on fires geography (Figs. 1 and

Table 5
Sensitivity and specificity, respectively, for each model year using landscape variables.

Land.	GLM		RF		MaxEnt		SVR	
Year	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
2001	0.83	0.6	0.94	0.43	0.93	0.48	0	1
2002	0.35	0.87	0.87	0.61	0.77	0.74	0.96	0.74
2003	0.28	0.97	0.53	0.97	0.5	0.98	0	1
2004	0.6	0.61	0.88	0.66	0.89	0.63	0.98	0.72
2005	0.62	0.76	0.83	0.83	0.83	0.84	0.85	0.83
2006	0.57	0.84	0.82	0.84	0.78	0.89	0.87	0.9
2007	0.56	0.83	0.79	0.92	0.78	0.93	0.84	0.95
2008	0.32	0.96	0.73	0.93	0.72	0.93	0.78	0.94
2009	0.56	0.75	0.84	0.81	0.79	0.88	0.92	0.81
2010	0.75	0.65	0.87	0.77	0.84	0.83	0.89	0.8
2011	0.34	0.94	0.8	0.84	0.78	0.87	0.8	0.85
2012	0.49	0.85	0.83	0.83	0.77	0.91	0.87	0.91
2013	0.73	0.76	0.89	0.81	0.89	0.82	0.91	0.77
2014	0.46	0.94	0.8	0.89	0.78	0.9	0.84	0.88
2015	0.29	0.96	0.74	0.92	0.67	0.85	0.77	0.96
2016	0.43	0.85	0.85	0.78	0.86	0.78	0.93	0.85
2017	0.41	0.8	0.85	0.76	0.85	0.74	0.86	0.78

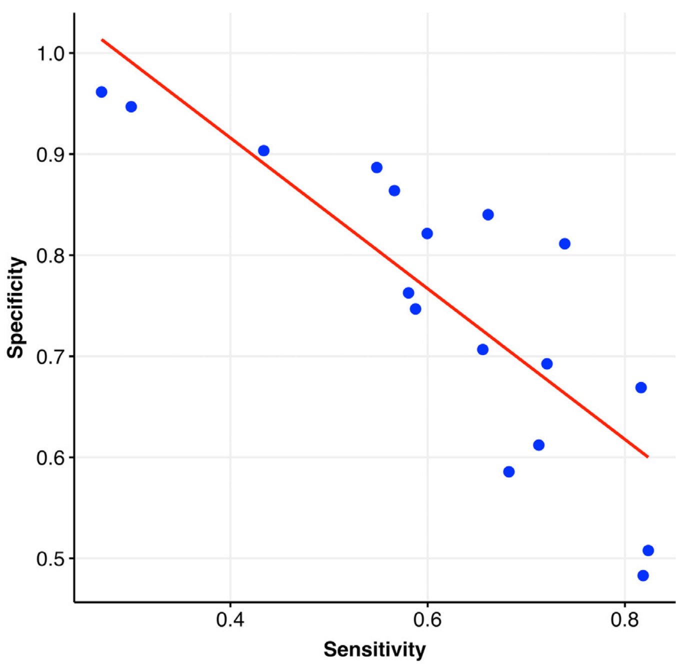

Fig. 5. The tradeoff between model's sensitivity and specificity for experiments driven by anthropogenic variables (RF algorithm). The same inverse relation was found for all models and all groups of variables. Low model sensitivity coincides with warm, dry years with large number of fires. The same years, however, have relatively high specificity. The correlation coefficient is 69%.

Table 6
Summary of sensitivity and specificity for each model year and group of variables (average ± standard deviation).

Group of Variables	Model			
Anthropogenic	GLM	RF	MaxEnt	SVR
Sensitivity	0.57±0.16	0.62±0.16	0.59±0.16	0.84±0.12
Specificity	0.73±0.14	0.75±0.15	0.76±0.15	0.70±0.22
Climate	**GLM**	**RF**	**MaxEnt**	**SVR**
Sensitivity	0.73±0.14	0.75±0.13	0.71±0.13	0.82±0.11
Specificity	0.70±0.17	0.70±0.17	0.68±0.18	0.75± 0.16
Landscape	**GLM**	**RF**	**MaxEnt**	**SVR**
Sensitivity	0.51±0.17	0.82±0.09	0.79±0.10	0.77±0.30
Specificity	0.82±0.12	0.80±0.13	0.82±0.12	0.86±0.09

S5). Such an effect is observed as stripes of fires oriented southwest to northeast (Fig. 2). It should be mentioned, that years with this particular characteristic on prediction maps are the years with model sensitivity exceeding 70–80%. At the same time, the years with more monotonic distribution of predicted fires are years with relatively high model specificity. Thus, model sensitivity and model specificity demonstrate inverse relations. This tradeoff between two performance characteristics can be found across all models and all groups of variables (Fig. 5).

Relative performance of models driven by individual groups of variables gives some additional insights into wildfire predictability (Fig. 4). The employment of only anthropogenic variables results in the worst overall performance. Typically, this group was able to predict fires, depending on the model and the year, at the AUC around 70–75% across all models (Fig. 5). Furthermore, during warm and dry years, when the number of ignitions increases, the sensitivity of models driven by anthropogenic variables dropped to 30–40% (Tables 3–6). This is consistent with prediction maps showing the lack of clear spatial patterns in fire locations during years with large number of fires (Figs. S5–S8).

Climatic factors showed some improvement in comparison with anthropogenic ones (Figs S1–S4). The nature of this phenomenon is not completely clear. The landscape group of variables showed further improvement by outperforming anthropogenic and climatic variables resulting in the AUC of machine learning algorithms from 80% to 90%. Yet, the AUC of GLM driven by landscape variables only resulted in the lowest of all groups of AUC of about 67% (Fig. 4). In other words, landscape variables performed rather well in machine learning algorithms and extremely poor in GLM. The slope or type of vegetation have highly nonlinear interactions with fires in the very rugged terrain of this region (Fig. 1). Therefore, naturally prone to bias caused by nonlinear interactions, machine learning algorithms are more sensitive to spatial variability in regional topography than the generalized linear models are. As it was stated above, models driven by a complete list of variables outperformed any model forced by any specific group, including landscape. At the same time, models driven by landscape variables perform nearly as good as models driven by the full set of variables (Figs. S1–S4). This observation demonstrates importance of landscape characteristics as the most powerful predictors of spatial patterns in fires. A relatively high performance of models driven by anthropogenic and climatic groups of variables, when landscape characteristics are not considered, confirms the hypothesized dependence of anthropogenic (roads, settlements etc.) and climatic (temperature and precipitation) variables on landscape features.

Table 7
Summary of top 5 factors for GLM, RF, and MaxEnt. Variables are presented in the order of their significance as estimated within each model by three independent criteria (Tables S1–S3). No variable importance shown for SVR because of the nature of this machine learning algorithm (see text for details).

Models	Top 5 Ranked Variables			
	Climatic		Anthropogenic	Landscape
	Temperature	Precipitation		
GLM Model	*April*	Oct	*Road Distance*	*Forest*
	Dec	Feb	*Town Distance*	*Elevation*
	Sep	*Aug*	Population	*Grassland**
	Nov	*May*	*Unpaved*	*Slope*
	Aug	*Jun*	*Trails*	*Aspect*
RF Model	May	*May*	*Road Distance*	*Grassland**
	Apr	*Jun*	*Town Distance*	*Forest*
	Aug	Jul	Population	*Elevation*
	Feb	Nov	*Unpaved*	*Slope*
	Sep	*Aug*	*Trails*	*Aspect*
MaxEnt Model	Jun	*Jun*	*Road Distance*	*Forest*
	Sep	*May*	*Town Distance*	*Slope*
	Apr	*Aug*	Seasonal Road	*Elevation*
	Dec	Apr	*Trails*	*Grassland**
	Feb	Feb	*Unpaved*	*Aspect*

3.2. Significance of variables

For GLM, RF, and MaxEnt, we used standard approaches to determine the significant factors as outlined in our methodology. To review, we examined the z-score, percent increase in MSE, and percent contribution for the respective method (See Tables S1–S3). The results are summarized in Table 7. Independent of machine learning algorithms, we found that monthly average temperatures in April and September, as well as total precipitation in May, June, and August, were within the top five significant climatic factors. Among anthropogenic variables proximity to roads and towns, as well as type of roads (paved or unpaved), were among five top significant factors across all models. Significant landscape features found in every model include percent of forest and grassland inside pixels, as well as elevation, slope and aspect. Percent of grassland cover within each pixel, however, has strong negative correlation with the percent of forest area, and, therefore, was not considered as an independent variable.

3.3. Contribution of anthropogenic, climatic and landscape variables into spatial and temporal variability of fires

Two groups of variables, anthropogenic and landscape, reveal the spatial heterogeneity of the study area. Since these variables have not changed during the study period they cannot serve as drivers of temporal variability in fires. On the other hand, inter-annual climatic fluctuations across this large geographic area are expected to explain spatial-temporal patterns of fire activity. Regional weather patterns, in turn, are significantly dependent on landscape characteristics. Temperature, for example, is higher at southerly slopes and decreases with increasing elevation. Orientation and altitude of major mountain systems also have significant effects on spatial patterns and sums of local precipitation events. To address the above dependencies, we compared the performance of machine learning algorithms driven by anthropogenic and landscape variables with that of the same models driven by climatic variables only. Depending on an individual year, GLM and machine learning models driven by these groups of variables can be trained to predict locations of pixels with fires with the AUC ranging from 60% (nearly uniform distribution of predicted fires) to 90% (similar to observed, visible spatial patters) (Figures S2-S4). We found, however, that for years with high AUC for models driven by climatic variables, the performance of those same models driven by anthropogenic or landscape variables increased as well. The same trend can be seen if we compare performance of models driven by all variables with that forced only by climatic variables. In essence, these comparisons demonstrate that during some years all groups of variables perform better than during other years. During the extremely wet 2012 the total number of detected ignitions was 15,154, and the total burnt area was 21,796 ha (Table 2). At the same time, during the overall dry year of 2015 when total precipitation was only 320 mm, the number of ignitions was 24,244 while the total burnt area (77,160 ha) was nearly three times greater than in 2012 (Table 2). The same relation exists for the warmest and the coldest years, as well as for years with the lowest and the highest Arctic Oscillation (AO) index (Table 2). This index reflects on the intensity of Arctic vortex, specifically, the barometric pressure difference between the Atlantic part of the Arctic and central Eurasia (Thompson and Wallace, 2001). During years with positive AO, the Atlantic part of the Arctic is exposed too low, while central Eurasia experiences higher barometric pressure. Therefore, during years with positive AO values dry and warm springs and summers persist over East Siberia and the number of fires here increases (Kim et al., 2020). Such a finding is also supported by our results (Table 7).

Generally, we see that year-to-year variability in the number of fires, or the area burned, are explained by variability in climatic conditions, such as mean annual temperature, total precipitation, or the AO index. At the same time, we note another relationship of fires with climatic variables. More specifically, the performance of all models, especially those driven by anthropogenic variables only, decreases with the increase in temperature or AO and increases with the increase in precipitation. In other words, forecasting fire spatial patterns improves during years when cold and wet conditions occur simultaneously (Fig. 6).

4. Discussion

One clear finding of our study is that machine learning significantly outperforms traditional statistical methods (Figs. S1 and S2). This can be explained by the weaker performance of GLM models, which suffer from overfitting problems. Examining confusion matrices for the training data showed that the GLM algorithm is insufficient in correctly labeling training data. For a given year, the method is only able to accurately predict either background (unburned) or burned locations, but not both. This phenomenon causes very large variability in sensitivity and specificity, which ranges from 0.4 to 0.9 for different years (Fig. S8). The second factor of poor GLM performance is that the assumption of independence of data points does not always hold true. Semi-variograms of residuals (Figs S9 and S10) reveal that many of the proposed variables are spatially autocorrelated. This is to be expected for some variables. Temperature, for example, depends on elevation and the particular location of a fire event, i.e. whether it occurs in the far north or in the south of the study area. Therefore, relatively high performance of GLM driven by climatic variables alone might be misleading and result from strong spatial autocorrelation between surface temperature and precipitation. Yet, although at smaller year to year variability in performance, machine learning algorithms follow this pattern between model sensitivity and specificity too.

Overall, machine learning algorithms RF, MaxEnt, and SVR, all performed better compared to GLM. For RF and MaxEnt this, most likely, stems from their ability to identify the background and presence points. As in previous works (Massada et al., 2013), RF and MaxEnt achieved similar performance in properly labeling the training data. However, SVR is able to identify the classes significantly better than any other methods considered. This can be explained by high sensitivity and specificity of the SVR model. Compared to RF and MaxEnt, this model was able to accurately classify a greater proportion of presence and background points. Proper use of SVR requires a time-consuming process of parameter selection with cross validation, which is critical for the method to achieve its highest performance. Therefore, parameter tuning for this method results in high computation time in comparison to other machine learning methods.

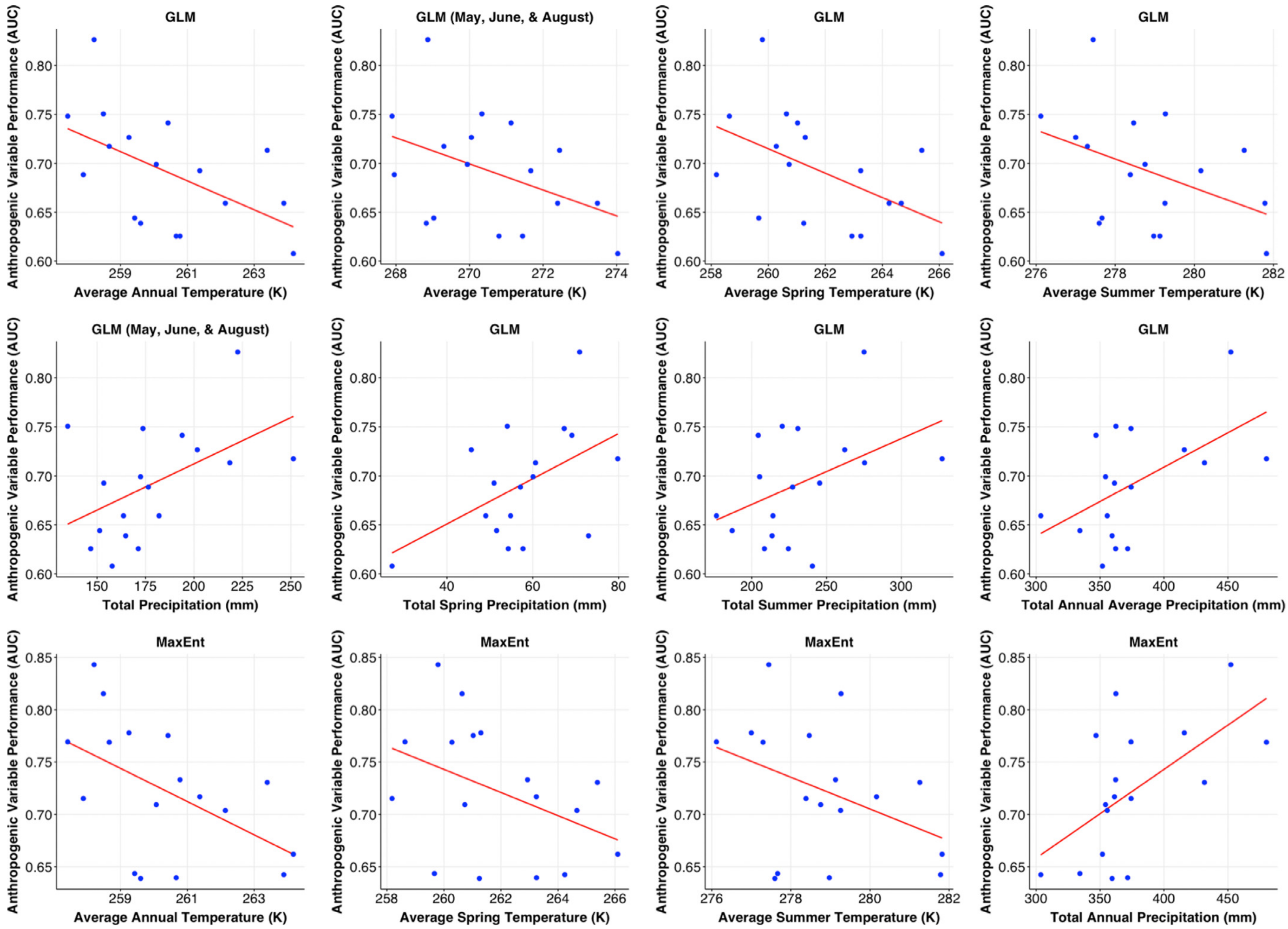

Fig. 6. Scatterplots of models driven by anthropogenic group of variables against temperature and precipitation for individual years.

Machine learning methods and generalized linear models all converged to a very similar ranking of individual variables. Furthermore, this ranking was not sensitive to specific metrics of variable significance (Tables 3–5, Figs. S1–S3). Machine learning algorithms identify proximity to towns and roads as the two most significant anthropogenic factors. Both the RF and MaxEnt, however, demonstrate that an additional significant factor can be the size of the human population. This is consistent with previous studies and confirms that human presence is among the primary triggers of fires, especially in the wildland-urban interface (Massada et al., 2009; Spyratos et al., 2007). Our study shows that most intentional or unintentional ignitions take place in proximity of unpaved roads and trails used by either tourists or loggers, the pattern is consistent across the 17 years of analysis. A recent spatial analysis of a single fire season conducted by Greenpeace in 2019 also revealed that the majority of fires in this region have started in close proximity to roads and built-up space (https://unearthed.greenpeace.org/2019/05/28/russia-wildfires-siberia-map/). It was also noted that some fires were the result of prescribed burning.

Our analyses also reveal that the total annual area of burnt forests in southern East Siberia is significantly influenced by patterns of atmospheric circulation, specifically the intensity of Arctic Oscillations (AO), which is consistent with recently published results (Kim et al., 2020). We found that winter precipitation, which depends greatly on the phase of the AO, serve as a good indicator of forest area burnt annually (Fig. 3). Overall, the mean annual temperature of our study area demonstrates a visible positive correlation with AO values ($r^2 = 0.44$). Nonetheless, this connection does not override, as it was mentioned earlier, the dependence of fire intensity and their spatial patterns, on non-climatic factors. Rather, the AO index, along with other climatic variables, provide a broad scale context for highly heterogeneous spatial patterns of fires, which are controlled by landscape and anthropogenic factors at much finer scales. We found that landscape variables, such as percent of forest cover, slope and elevation, are significant predictors of fire locations. Higher forest cover in each pixel and connectivity of forested area increase fuel availability and allow for a more rapid spread of fire and increase fire frequency. Most fires usually occur in the area at intermediate elevation ranges of about 700–1300 m with frequency decreasing in valleys and basins and at higher elevations.

Climate variables that appear to be the most significant determinants of local fire activity are average air temperature in April, August, and September and total precipitation in the months of May, June, and August. It is important to separate the significance of these climatic variables as controls of fire activity from that of the AO index, mean annual temperature or mean regional winter precipitation, being the strongest controls of the overall number and area of annual fires.

According to our results, monthly mean temperature and precipitation of late spring and early summer months are good predictors of spatial patterns in spring and summer forest fires. At the same time, the importance of September temperature and August precipitation provides evidence of the second and typically less intensive period of autumn fires.

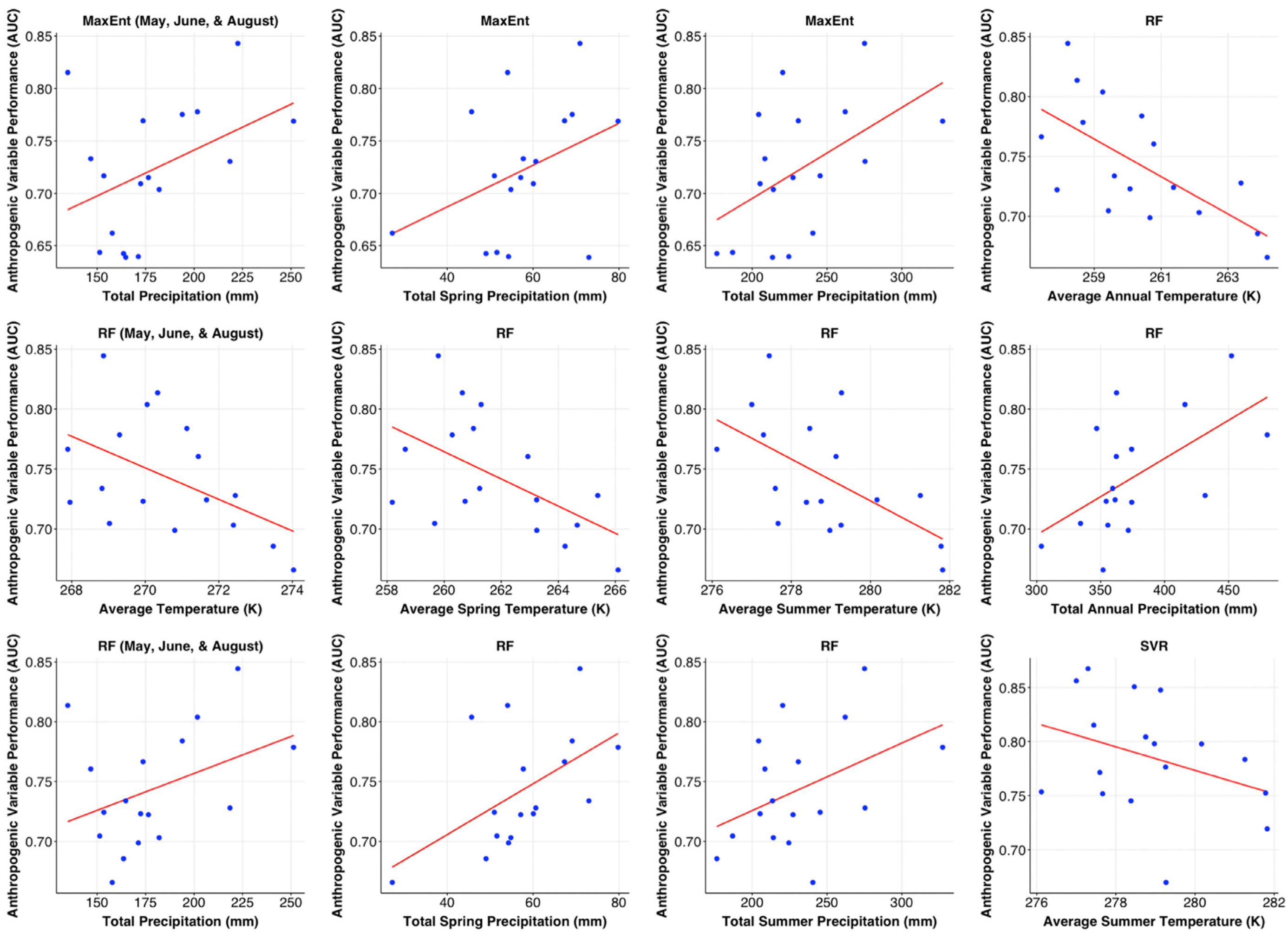

Fig. 6. Continued

The study region is known for customary grass burning, often in privately owned lots, especially in the south of the area predominantly occupied by the forest-steppe ecotone and agricultural ecosystems. These "controlled" burnings are supposed to convert the last year's dry grass into a mineral fertilizer and increase grassland productivity. Often such intentional fires in the forest-steppe region cause more harm by igniting bordering forests. Therefore, it is not surprising to observe a much higher fire frequency in the south, not only in grassland landscapes but in forested areas as well. Furthermore, April and May fires are more frequent in the south, while late summer (July, August and September) are more frequent in the north of the region (Figs. 2 and S5).

During the last 50 years the mean annual temperature in Eastern Siberia has been increasing at the rate of 0.3–°C per decade, which is about two times faster than the average rate of 0.25 per decade for the entire Northern Hemisphere. April and September temperatures showed growth at the rate which is higher than the mean annual temperature (Groisman et al., 2013). Precipitation over the same period of time showed a wave like pattern: increase from 1940s to late 1990s and the strong decline after 2000 (Groisman et al., 2013). During recent several years, however, East Siberia have been experiencing winter cooling which is most likely caused by the intensification of AO and reduction of the Arctic sea ice. Such pattern is known as the "warm Arctic - cold Siberia" (WACS) (Jin et al., 2020). However, our findings do not suggest any winter cooling in this area (Fig. 1). On the opposite, - December temperature continued to increase at a very high rate of about 1.6°C over the last two decades (Table S8). Furthermore, according to the consensus of IPCC global circulation models mean annual temperature of Eastern Siberia will most likely continue to grow at the rate higher than the average rate of the Northern Hemisphere, while precipitation will continue to decline with a high degree of probability (Change, 2014).

Another important result of our study is that during warm and/or dry years, the accuracy of model predictions, especially those driven by anthropogenic variables, decreases (Fig. 6). This effect is critically important for explaining year-to-year variability in models' sensitivity and specificity (Fig. 7). Sensitivity of all models declines dramatically after fire emergency extended over more than one percent of the region's area (Fig. 7). This decline is somewhat compensated by the distinguishable increase in model specificity, i.e. the ability to successfully predict locations without fires. Such an increase, however, is not related to a better predictability of spatial patterns of fires. On the opposite, during dry and hot years, fire activity increases, and overall model performance expressed as AUC decreases (Fig. 6). Consequently, the overall predictability of fire patterns declines with increase in the fraction of the area under fires.

We hypothesize this finding might reflect inter-annual changes of sensitivity of fires to anthropogenic and landscape factors caused by variability in regional weather. It is reasonable to suggest that during moist and cold years spatial patterns of fires do not depend on proximity to settlements or roads as much as during dry and hot years, when even a small ignition could quickly lead to a much larger fire. By the same reason, the type of vegetation has lower impact on fire patterns during wet years, and the dependence is somewhat stronger during drier years.

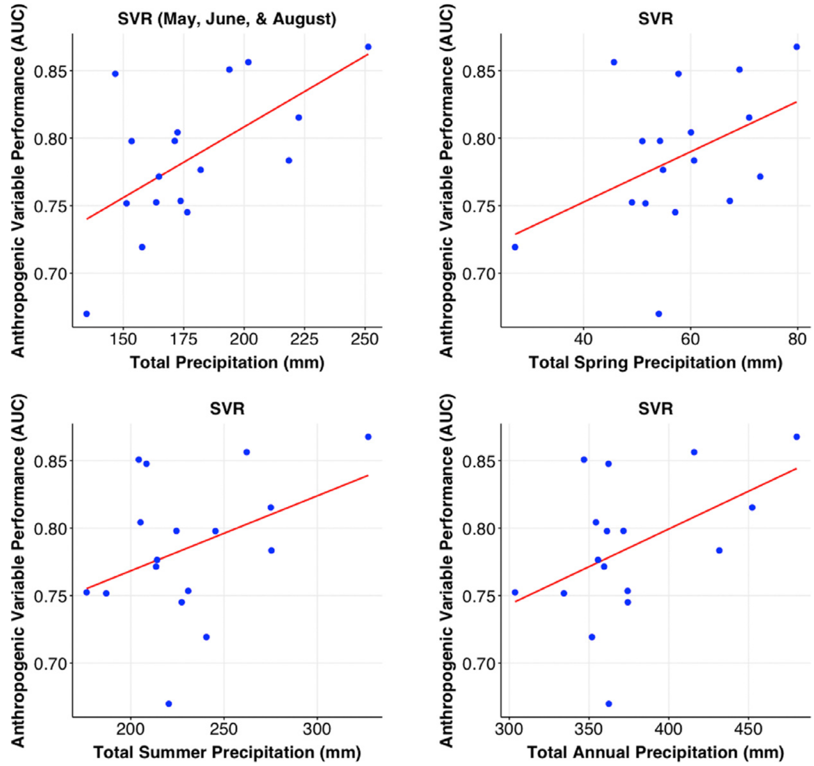

Fig. 6. Continued

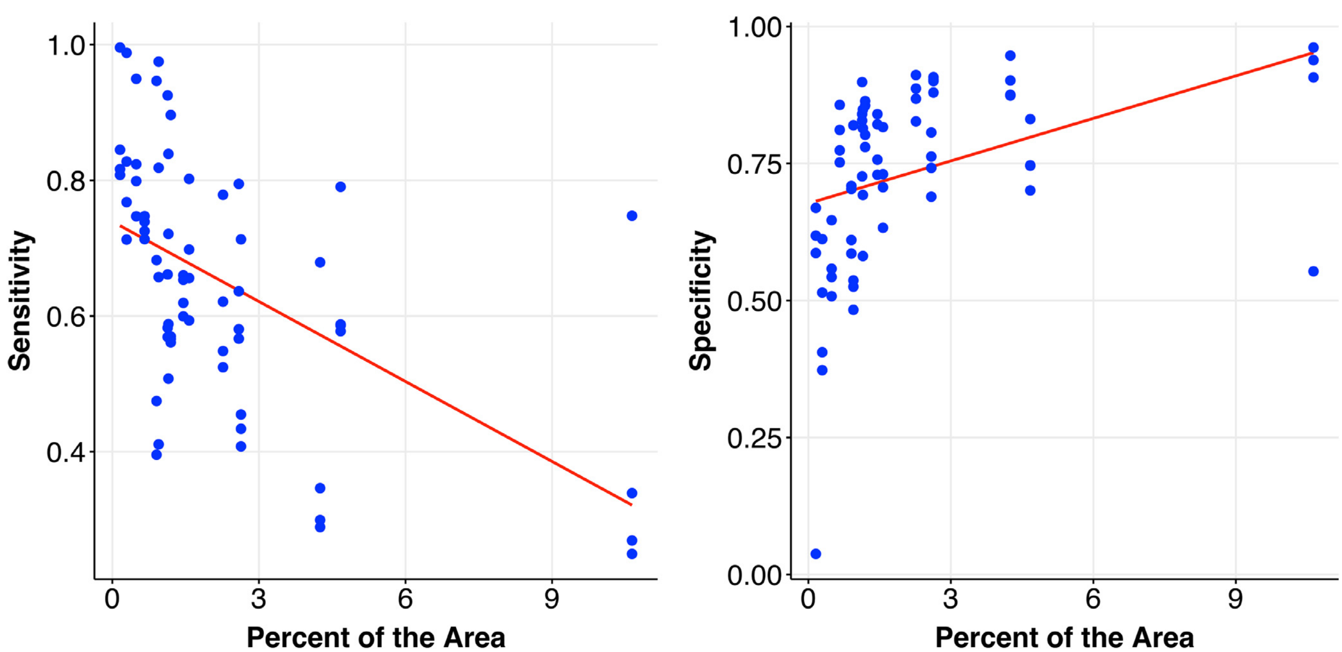

Fig. 7. Scatterplots of sensitivity ($r^2 = 0.29$) (left) and specificity ($r^2 = 0.15$) (right) for GL, RF, MaxEnt and SVM models driven by anthropogenic variables as function of the percent of the region covered with fires. Each point represents single year of modeling by one of four methods we used. Similar relations were found for all other groups of variables such as landscape, climatic and the full set of variables. If area occupied by the fires exceeds one percent, the results of the modeling for this specific year demonstrate drastic decline in sensitivity and increase in specificity of calculations.

It is likely that mega-fires absorb smaller fire events to form even larger agglomerates of fire, which makes the separation of background from ignition locations less robust. During the most intensive fire season of 2003, for example, the distribution of ignition locations looked more uniform and did not depend on elevation as it was the case in the less intensive 2008 season (Fig. 2). We conclude that during seasons of high fire activity geographic patterns are diffused by mega-fires, which makes their association with proximity to roads or towns less discernible. Such a pattern-erasing "inferno" effect of mega-fires reduces the accuracy of machine learning algorithms and makes forest fires less predictable.

5. Conclusion

In our work, we analyzed wildfires in the south of East Siberia using both statistical and machine learning models with the goal to advance future forecasting techniques. In the initial analysis, we showed that our machine learning methods performed better than our statistical method and examined the most significant variables for determining wildfire prediction. We then conducted a regression analysis of the impact of climate on model performance. We can conclude that as burnt areas occupy more than 1% of the territory, the accuracy of the model severely declines regardless of the method used. Therefore, as regions become warmer, models will be incapable of producing reliable forecasts that can be used by government agencies to better prepare for such devastating events. A drawback of our work is we did not have information on ignition points and had to use locations of burnt areas instead. Therefore, it should be reasonable to suggest that substantial improvement in the prediction of the geography of mega-fires can be achieved only with the incorporation of additional information on location and ignition points. Information from more frequently passing satellites or airplanes and drones will be critical in developing better models. In any case, improvements in the forecast of mega-fires will depend not so much on improvements in mathematical models but on development of new, more accurate methods of remote-sensing as well as ground truth verification methods such as web-based cameras and other fire-detecting sensors.

Funding

The work of Yiming Ying is in part supported by the National Science Foundation (NSF) under Grant no. 1816227. The work of Alexander Buyantuev is in part supported by the National Science Foundation (NSF) under Grant no. 1531511. The work of Andrei Lapenis was supported by the National Science Foundation grant number 1455544.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Michael Natole Jr.: Investigation, Formal analysis, Supervision. **Yiming Ying:** Data curation, Formal analysis, Writing - original draft. **Alexander Buyantuev:** Data curation, Formal analysis, Writing - original draft. **Michael Stessin:** Data curation, Formal analysis, Writing - original draft. **Victor Buyantuev:** Data curation, Formal analysis, Writing - original draft. **Andrei Lapenis:** Data curation, Formal analysis, Writing - original draft.

Acknowledgments

Data: The base layers of topographic data was provided on the condition that we cannot distribute them to others, either for pay or free of charge. The data are not our property. We were provided it only for this project. Those interested should contact the Siberian branch of the Federal Agency of Cartography and Geodesy, the Ministry of economic development of Russian Federation.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.envadv.2021.100041

References

Abatzoglou, J.T., Williams, A.P., 2016. Impact of anthropogenic climate change on wildfire across western us forests. Proc. Natl. Acad. Sci. USA 113 (42), 11770–11775.
Allen, C.D., Savage, M., Falk, D.A., Suckling, K.F., Swetnam, T.W., Schulke, T., Stacey, P.B., Morgan, P., Hoffman, M., Klingel, J.T., 2002. Ecological restoration of southwestern ponderosa pine ecosystems: a broad perspective. Ecol. Appl. 12 (5), 1418–1433.
Balch, J.K., Bradley, B.A., Abatzoglou, J.T., Nagy, R.C., Fusco, E.J., Mahood, A.L., 2017. Human-started wildfires expand the fire niche across the united states. Proc. Natl. Acad. Sci. USA 114 (11), 2946–2951.
Bowman, D.M., Williamson, G.J., Abatzoglou, J.T., Kolden, C.A., Cochrane, M.A., Smith, A.M., 2017. Human exposure and sensitivity to globally extreme wildfire events. Nat. Ecol. Evol. 1 (3), 1–6.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Cardille, J.A., Ventura, S.J., Turner, M.G., 2001. Environmental and social factors influencing wildfires in the upper midwest, United Atates. Ecol. Appl. 11 (1), 111–127.
Cascio, W.E., 2018. Wildland fire smoke and human health. Sci. Total Environ. 624, 586–595.
Catry, F.X., Rego, F.C., Bação, F.L., Moreira, F., 2010. Modeling and mapping wildfire ignition risk in portugal. Int. J. Wildland Fire 18 (8), 921–931.
Change, I.P.O.C., 2014. Climate Change 2013: the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
Cheng, T., Wang, J., 2008. Integrated spatio-temporal data mining for forest fire prediction. Trans. GIS 12 (5), 591–611.
Chuvieco, E., Allgöwer, B., Salas, J., 2003. Integration of physical and human factors in fire danger assessment. In: Wildland Fire Danger Estimation and Mapping: the Role of Remote Sensing Data. World Scientific, pp. 197–218.
Cullen, A.C., Axe, T., Podschwit, H., 2021. High-severity wildfire potential–associating meteorology, climate, resource demand and wildfire activity with preparedness levels. Int. J. Wildland Fire 30 (1), 30–41.
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.
De Vasconcelos, M.P., Silva, S., Tome, M., Alvim, M., Pereira, J.C., 2001. Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. Photogramm. Eng. Remote Sens. 67 (1), 73–81.
De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81 (11), 3178–3192.
Elith*, J., H. Graham*, C., P. Anderson, R., Dudik, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., et al., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29 (2), 129–151.
Ertugrul, M., Ozel, H.B., Varol, T., Cetin, M., Sevik, H., 2019. Investigation of the relationship between burned areas and climate factors in large forest fires in the çanakkale region. Environ. Monit. Assess. 191 (12), 1–12.
Ertugrul, M., Varol, T., Ozel, H.B., Cetin, M., Sevik, H., 2021. Influence of climatic factor of changes in forest fire danger and fire season length in Turkey. Environ. Monit. Assess. 193 (1), 1–17.
Finney, M., Grenfell, I.C., McHugh, C.W., 2009. Modeling containment of large wildfires using generalized linear mixed-model analysis. Forest Sci. 55 (3), 249–255.
Finney, M.A., 2005. The challenge of quantitative risk analysis for wildland fire. For. Ecol. Manag. 211 (1-2), 97–108.
Franklin, J., 2010. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press.
Friedl, M., Sulla-Menashe, D., 2015. Mcd12q1 modis/terra+ aqua Land Cover Type Yearly l3 Global 500m Sin Grid v006 [data set], 10. NASA EOSDIS Land Processes DAAC.
Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367–378.
Gholamnia, K., Gudiyangada Nachappa, T., Ghorbanzadeh, O., Blaschke, T., 2020. Comparisons of diverse machine learning approaches for wildfire susceptibility mapping. Symmetry 12 (4), 604.
Giglio, L., Justice, C., Boschetti, L., Roy, D., 2015. Mcd64a1 modis/terra+ aqua Burned Area Monthly l3 Global 500m sin Grid v006 [data set]. NASA EOSDIS Land Processes DAAC, Sioux Falls, SD, USA.
Giglio, L., Schroeder, W., Justice, C.O., 2016. The collection 6 modis active fire detection algorithm and fire products. Remote Sens. Environ. 178, 31–41.
Goss, M., Swain, D.L., Abatzoglou, J.T., Sarhadi, A., Kolden, C.A., Williams, A.P., Diffenbaugh, N.S., 2020. Climate change is increasing the likelihood of extreme autumn wildfire conditions across california. Environ. Res. Lett. 15 (9), 094016.
Groisman, P.Y., Blyakharchuk, T.A., Chernokulsky, A.V., Arzhanov, M.M., Marchesini, L.B., Bogdanova, E.G., Borzenkova, I.I., Bulygina, O.N., Karpenko, A.A., Karpenko, L.V., et al., 2013. Climate changes in Siberia. In: Regional Environmental Changes in Siberia and Their Global Consequences. Springer, pp. 57–109.
Halligan, S., Altman, D.G., Mallett, S., 2015. Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: a discussion and proposal for an alternative approach. Eur. Radiol. 25 (4), 932–939.
Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve.. Radiology 143 (1), 29–36.
Hantson, S., Arneth, A., Harrison, S.P., Kelley, D.I., Prentice, I.C., Rabin, S.S., Archibald, S., Mouillot, F., Arnold, S.R., Artaxo, P., et al., 2016. The status and challenge of global fire modelling. Biogeosciences 13 (11), 3359–3375.
Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models, 43. CRC press.
Huang, S., Siegert, F., Goldammer, J., Sukhinin, A., 2009. Satellite-derived 2003 wildfires in southern Siberia and their potential influence on carbon sequestration. Int. J. Remote Sens. 30 (6), 1479–1492.

Jaafari, A., Pourghasemi, H.R., 2019. Factors influencing regional-scale wildfire probability in Iran: an application of random forest and support vector machine. In: Spatial Modeling in GIS and R for Earth and Environmental Sciences. Elsevier, pp. 607–619.
Jin, C., Wang, B., Yang, Y.-M., Liu, J., 2020. “Warm arctic-cold siberia” as an internal mode instigated by north atlantic warming. Environ. Res. Lett. 47 (9). e2019GL086248
Jones, M.W., Smith, A., Betts, R., Canadell, J.G., Prentice, I.C., Le Quéré, C., 2020. Climate change increases risk of wildfires. Sci. Brief Rev. 116, 117.
Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al., 1996. The ncep/ncar 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77 (3), 437–472.
Karatzoglou, A., Smola, A., Hornik, K., 2016. Kernlab: Kernel-based machine learning lab. R package version 0.9.
Kim, J.-S., Kug, J.-S., Jeong, S.-J., Park, H., Schaepman-Strub, G., 2020. Extensive fires in southeastern Siberian permafrost linked to preceding arctic oscillation. Sci. Adv. 6 (2), eaax3308.
Koutsias, N., Martínez-Fernández, J., Allgöwer, B., 2010. Do factors causing wildfires vary in space? Evidence from geographically weighted regression. GISci. Remote Sens. 47 (2), 221–240.
Krylov, A., McCarty, J.L., Potapov, P., Loboda, T., Tyukavina, A., Turubanova, S., Hansen, M.C., 2014. Remote sensing estimates of stand-replacement fires in russia, 2002–2011. Environ. Res. Lett. 9 (10), 105007.
Kukavskaya, E.A., Buryak, L., Ivanova, G., Conard, S., Kalenskaya, O., Zhila, S.V., McRae, D.J., 2013. Influence of logging on the effects of wildfire in siberia. Environ. Res. Lett. 8 (4), 045034.
Liaw, A., Wiener, M., et al., 2002. Classification and regression by randomforest. R News 2 (3), 18–22.
Lindenmayer, D.B., Kooyman, R.M., Taylor, C., Ward, M., Watson, J.E., 2020. Recent australian wildfires made worse by logging and associated forest management. Nat. Ecol. Evol. 4 (7), 898–900.
Massada, A.B., Radeloff, V.C., Stewart, S.I., Hawbaker, T.J., 2009. Wildfire risk in the wildland–urban interface: a simulation study in northwestern Wisconsin. For. Ecol. Manag. 258 (9), 1990–1999.
Massada, A.B., Syphard, A.D., Stewart, S.I., Radeloff, V.C., 2013. Wildfire ignition-distribution modelling: a comparative study in the Huron–Manistee national forest, michigan, usa. Int. J. Wildland fire 22 (2), 174–183.
Narins, T.P., 2015. Dynamics of the Russia–China forest products trade. Growth Change 46 (4), 688–703.
Parisien, M.-A., Moritz, M.A., 2009. Environmental controls on the distribution of wildfire at multiple spatial scales. Ecol. Monogr. 79 (1), 127–154.
Pearson, R.G., Raxworthy, C.J., Nakamura, M., Townsend Peterson, A., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in madagascar. J. Biogeogr. 34 (1), 102–117.
Pereira, M.G., Parente, J., Amraoui, M., Oliveira, A., Fernandes, P.M., 2020. The role of weather and climate conditions on extreme wildfires. In: Extreme Wildfire Events and Disasters. Elsevier, pp. 55–72.
Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3–4), 231–259.
Phillips, S. J., Dudík, M., Schapire, R. E., 2017. Maxent software for modeling species niches and distributions (version 3.4. 1). Tillgänglig från url: http://biodiversityinformatics. amnh. org/open_source/maxent.
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.
Robinne, F.-N., Burns, J., Kant, P., Flannigan, M., Kleine, M., de Groot, B., Wotton, D., 2018. Global fire challenges in a warming world. IUFRO.
Satir, O., Berberoglu, S., Donmez, C., 2016. Mapping regional forest fire probability using artificial neural network model in a mediterranean forest ecosystem. Geomatics Nat. Hazards Risk 7 (5), 1645–1658.
Shvidenko, A., Shchepashchenko, D., Vaganov, E., Sukhinin, A., Maksyutov, S.S., McCallum, I., Lakyda, I., 2011. Impact of wildfire in Russia between 1998–2010 on ecosystems and the global carbon budget. In: Doklady Earth Sciences, 441. Springer, pp. 1678–1682.
Sitnov, S., Mokhov, I., 2018. A comparative analysis of the characteristics of active fires in the boreal forests of eurasia and north America based on satellite data. Izvestiya Atmos. Ocean. Phys. 54 (9), 966–978.
Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14 (3), 199–222.
Soja, A.J., Tchebakova, N.M., French, N.H., Flannigan, M.D., Shugart, H.H., Stocks, B.J., Sukhinin, A.I., Parfenova, E., Chapin III, F.S., Stackhouse Jr, P.W., 2007. Climate-induced boreal forest change: predictions versus current observations. Glob. Planet. Change 56 (3-4), 274–296.
Spacesystems, A., /Japan, U.S., 2019. Aster Global Digital Elevation Model v003 [data set]. NASA EOSDIS Land Processes DAAC.
Spyratos, V., Bourgeron, P.S., Ghil, M., 2007. Development at the wildland–urban interface and the mitigation of forest-fire risk. Proc. Natl. Acad. Sci. USA 104 (36), 14272–14276.
Stephens, S.L., Burrows, N., Buyantuyev, A., Gray, R.W., Keane, R.E., Kubian, R., Liu, S., Seijo, F., Shu, L., Tolhurst, K.G., et al., 2014. Temperate and boreal forest mega-fires: characteristics and challenges. Front. Ecol. Environ. 12 (2), 115–122.
Sturtevant, B.R., Cleland, D.T., 2007. Human and biophysical factors influencing modern fire disturbance in northern wisconsin. Int. J. Wildland Fire 16 (4), 398–413.
Sulova, A., Jokar Arsanjani, J., 2021. Exploratory analysis of driving force of wildfires in australia: an application of machine learning within google earth engine. Remote Sens. 13 (1), 10.
Syifa, M., Panahi, M., Lee, C.-W., 2020. Mapping of post-wildfire burned area using a hybrid algorithm and satellite data: the case of the camp fire wildfire in california, usa. Remote Sens. 12 (4), 623.
Syphard, A.D., Keeley, J.E., 2015. Location, timing and extent of wildfire vary by cause of ignition. Int. J. Wildland Fire 24 (1), 37–47.
Team, R. C., 2013. R: a language and environment for statistical computing.
Thompson, D.W., Wallace, J.M., 2001. Regional climate impacts of the northern hemisphere annular mode. Science 293 (5527), 85–89.
Turner, M.G., 2010. Disturbance and landscape dynamics in a changing world. Ecology 91 (10), 2833–2849.
Vapnik, V., Golowich, S.E., Smola, A.J., 1997. Support vector method for function approximation, regression estimation and signal processing. In: Advances in Neural Information Processing Systems, pp. 281–287.
Westerling, A.L., Hidalgo, H.G., Cayan, D.R., Swetnam, T.W., 2006. Warming and earlier spring increase western us forest wildfire activity. Science 313 (5789), 940–943.