Geoscientific
Model Development

Development and technical paper

Quantifying wildfire drivers and predictability in boreal peatlands using a two-step error-correcting machine learning framework in TeFire v1.0

Rongyun Tang[1,2], **Mingzhou Jin**[1], **Jiafu Mao**[3], **Daniel M. Ricciuto**[3], **Anping Chen**[4], **and Yulong Zhang**[1]

[1]Institute for a Secure and Sustainable Environment and Department of Industrial and Systems Engineering,
University of Tennessee, Knoxville, TN 37996, USA
[2]Department of Biological & Agricultural Engineering, University of Arkansas, Fayetteville, AR 72701, USA
[3]Environmental Sciences Division and Climate Change Science Institute,
Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA
[4]Department of Biology and Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO 80523, USA

Correspondence: Mingzhou Jin (jin@utk.edu) and Jiafu Mao (maoj@ornl.gov)

Abstract. Wildfires are becoming an increasing challenge to the sustainability of boreal peatland (BP) ecosystems and can alter the stability of boreal carbon storage. However, predicting the occurrence of rare and extreme BP fires proves to be challenging, and gaining a quantitative understanding of the factors, both natural and anthropogenic, inducing BP fires remains elusive. Here, we quantified the predictability of BP fires and their primary controlling factors from 1997 to 2015 using a two-step correcting machine learning (ML) framework that combines multiple ML classifiers, regression models, and an error-correcting technique. We found that (1) the adopted oversampling algorithm effectively addressed the unbalanced data and improved the recall rate by 26.88 %–48.62 % when using multiple datasets, and the error-correcting technique tackled the overestimation of fire sizes during fire seasons; (2) nonparametric models outperformed parametric models in predicting fire occurrences, and the random forest machine learning model performed the best, with the area under the receiver operating characteristic curve ranging from 0.83 to 0.93 across multiple fire datasets; and (3) four sets of factor-control simulations consistently indicated the dominant role of temperature, air dryness, and climate extreme (i.e., frost) for boreal peatland fires, overriding the effects of precipitation, wind speed, and human activities. Our findings demonstrate the efficiency and accuracy of ML techniques in predicting rare and extreme fire events and dis-

entangle the primary factors determining BP fires, which are critical for predicting future fire risks under climate change.

1 Introduction

Carbon-rich boreal peatlands (BPs) cover only $\sim 2\%$ of the Earth's surface (Gorham, 1991) but have accumulated $\sim 20\%$–40% (450 ± 150 Pg C) of the global soil carbon, historically having a net cooling effect on the global radiation balance (Hugelius et al., 2020; Page and Hooijer, 2016; Scharlemann et al., 2014). This major land carbon pool, however, is highly vulnerable to current global warming, which tends to introduce carbon emissions into the atmosphere through increasing decomposition of peat soil organic matter and fire combustion (Turetsky et al., 2014). In particular, BP fire regimes have been undergoing pronounced changes over recent decades in terms of fire extent, frequency, and duration (Field and Raupach, 2004; Kelly et al., 2013). In BPs, there are two types of wildfires – surface flaming and underground smoldering – that can transition from one to the other at different phases. It is noteworthy that compared to flaming combustion, smoldering combustion is easier to ignite, harder to suppress, and more persistent in low-temperature and high-moisture peat (Huang and Rein, 2019). Besides releasing CO_2, smoldering produces more CO, CH_4, smoke, and even

gaseous mercury (Haynes et al., 2017; Urbanski et al., 2008), altering global carbon balance and threatening public health (Liu et al., 2015; Reid et al., 2016). Yet smoldering combustion remains poorly understood, despite recent efforts to use experimental, statistical, and computational tools to investigate smoldering ignition, spread, extinction, fuel types, burning depth, and emission estimation (Che Azmi et al., 2021; French et al., 2004; Rein and Huang, 2021). As a consequence, smoldering is not fully characterized in prevalent wildfire physical models (Rabin et al., 2017), although peatland fires are thought to be modulated by heat transfer and water content (Frandsen, 1997; Ohlemiller, 1985). Without an improved understanding of smoldering fires, our current understanding of BP fires and their predictability is still very limited, hampering peat fire hazard mitigation and firefighting.

Most studies ascribe the ignition and propagation of flaming fires to the joint impact of a heat source, fire-favoring climate, fuel, and anthropogenic factors. Flameless smoldering peatland fires are not an exception, although upland flaming fires and underground smoldering in BPs are fundamentally different in chemical and physical aspects (Costafreda-Aumedes et al., 2017; Rabin et al., 2017; Scott et al., 2013). However, compared to our understanding of flaming fires and their drivers and burning processes (Rothermel, 1972), we still know very little about key factors controlling smoldering fires. Importantly, Yuan et al. (2021) suggested that the smoldering process is a series of exothermic and often nonlinear events that include three key steps: biological reaction, chemical oxidative reaction, and drying. However, quantifying the exothermic process is not easy. For example, experiments using phospholipid fatty acid (PLFA)-based microorganisms revealed that peat self-heating reactions (soil respiration and microorganism growth) could happen at temperatures as low as 25–55 °C (Ranneklev and Bååth, 2003), while temperature could reach 500–700 °C during smoldering (Hurley et al., 2015). The dramatic changes in micro-processes of smoldering reactions consequently bring difficulties and uncertainties in measuring parameters for physical models. Furthermore, without a clear understanding of nonlinear interactions of climate, heat transformation, and fire, traditional bottom-up statistic models can be clueless.

Rather than traditional linear models, more complicated process-based physical models and data-driven statistical models – including machine learning (ML) techniques – have been extensively used to explore the environmental determinants and predictability of peat wildfires (Bedia et al., 2014; Burgan and Rothermel, 1984; Castelli et al., 2015). Process-based fire models are primarily based on well-established mathematical or physical laws that can describe fire processes, but these models may struggle with uncertain initiation and boundary conditions, as well as model parameters (Hantson et al., 2016). According to the Fire Modelling Intercomparison Project (Rabin et al., 2017), most fire schemes in current land surface models focus on forest fire occurrence,

spread, distinction, and associated impact assessment. Few models (e.g., the Community Land Model, Li et al., 2013; Rabin et al., 2017) explicitly characterize peatland fire impacts with constrains from climate (e.g., BP wetness and tropical dryness), peat fraction, water table depth, and grid cell area (Li et al., 2013). Substantial gaps in the knowledge and understanding of peat fire combustion, the solution to primal and inverse problems, and the unavailable large-scale peat soil and peat burning characteristic data are still obstacles in building peat fire combustion theory and parameterizing peat fire in process-based models (Grishin et al., 2009). Unlike general statistic models which require assumptions and unlike physical models which are supported by physical mechanisms, ML models require very few assumptions and can achieve high performance in solving nonlinear fitting and predictions (Jain et al., 2020). These benefits have led to the application of a broad range of ML algorithms in wildfire science research, such as fire detection, fire weather exploration, fire behavior prediction, fire impact evaluation, and fire management (Jain et al., 2020). ML algorithms are not only used to attribute the primary causes of fires (Yu et al., 2020) but also applied to model evaluation and diagnosis (Forkel et al., 2019). However, the majority of ML research focuses on forest fires, and just a small number of recent studies have used ML in the study of BP fires. For example, Rudiyanto et al. (2018) applied artificial intelligence in peatland monitoring and mapping with the support of remote sensing data, while some others investigated peat fire risk prediction and attribution with different ML methods (Bali et al., 2021; Horton et al., 2021; Rosadi et al., 2020). However, it is noteworthy that the recall or precision rate of peat fires was typically low in these ML studies, despite generally high ($> 70\%$) prediction accuracies (Bali et al., 2021; Horton et al., 2021; Rosadi et al., 2020). These low recall or precision rates (i.e., high type I and type II errors) are likely caused by unbalanced fire data, which also indicated that predicting severely unbalanced fire by single models could still be full of challenges, and further studies are needed to deal with such commission and omission problems as well as to improve the predictability of peat fires.

For that reason, by collating and harmonizing monthly climate-, vegetation-, soil-, and human-related variables from 1997 to 2015, we created a two-step ML framework with various ML classification and regression techniques to evaluate the model reproducibility and predictability for severely skewed fire data, and a series of sensitivity tests was performed on each of multiple fire datasets to address possible drivers of BP fires. Specific research goals include to (1) examine the performances of multiple ML algorithms in reproducing and predicting fire occurrence, fire counts, and fire impacts (i.e., burned area and carbon emissions); (2) diagnose dominant environmental controls on peatland fire activities; and (3) quantify uncertainties in the two-step ML framework and to correct predicting errors and improve the ML

predicting accuracy that is suppressed by severely skewed input data.

2 Data

Multiple sources of environmental data – including climate-, vegetation-, soil-, and human-related data – and multiple fire products were used in this study, as listed in Table S1 in the Supplement. All datasets were re-gridded to $1° \times 1°$ with a monthly time resolution, covering the period from 1997 to 2015.

2.1 Response variables

To evaluate ML framework robustness for difference response variables, five fire datasets were used in this study: the Global Fire Emission Database (GFED) version 4.1s (GFED4.1s) carbon emissions, the GFED4.1s burned area (BA), the Fire Climate Change Initiative (FireCCI) version 5.1 (FireCCI5.1) BA, the Moderate Resolution Imaging Spectroradiometer (MODIS) active fire product MCD45A1, and the product MCD64A1 burning date. The monthly BA fraction and carbon emissions from GFED4.1s span from 1997 to 2016 with a spatial resolution of $0.25° \times 0.25°$ (Giglio et al., 2013; Randerson et al., 2012; van der Werf et al., 2017). The FireCCI5.1 BA dataset ranges from 2001 to present and has a spatial resolution of 250 m at monthly or biweekly temporal resolutions (Chuvieco et al., 2018; Lizundia-Loiola et al., 2020). Monthly MCD45A1 and MCD64A1 burn date datasets were derived from the MODIS Terra and Aqua satellite products at a spatial resolution of 500 m. MCD45A1 was derived from surface reflectance dynamics by a bidirectional reflectance-distribution function-based change detection approach (Roy et al., 2002), whereas MCD64A1 was produced by a burn-sensitive vegetation index algorithm based on a combination of reflectance data and active fire observations (Giglio et al., 2018). Because only burn dates were provided, both MCD45A1 and MCD64A1 were only applied for evaluating fire occurrences rather than fire impacts.

2.2 Explanatory variables

2.2.1 Meteorology data

To reflect the climate from 1997 to 2015, this study used the monthly $0.5° \times 0.5°$ gridded Climatic Research Unit (CRU) time series data version 4.04 (Harris et al., 2020). CRU data provide meteorological variables, including mean temperature (TMP), temperature minimum (TMN), temperature maximum (TMX), cloud cover (CLD), diurnal temperature range (DTR), ground frost frequency (FRS), wet day frequency (WET), evapotranspiration (ET), precipitation (PRE), and vapor pressure (VP). Additionally, the CRU Palmer Drought Severity Index (PDSI) and the Modern-Era

Retrospective analysis for Research and Applications Version 2 (MERRA-2) 2 m wind speed (WIN) were included as feature inputs (Gelaro et al., 2017). Using the CRU saturated VP (SVP) and relative humidity (RH), we also calculated the VP deficit (VPD) based on the transforming formulations shown in Table S1.

2.2.2 Vegetation data

Monthly third-generation Global Inventory Monitoring and Modeling System (GIMMS-3g) NDVI from 1982 to 2015 with a spatial resolution of $0.83° \times 0.83°$ was used to characterize the vegetation growth condition (Pinzon and Tucker, 2014). The 8 km gridded monthly GIMMS-3g gross primary productivity (GPP) from 1982 to 2016 was also included in this study to characterize the fuel availability (Madani and Parazoo, 2020).

2.2.3 Soil moisture data

To estimate the effects of soil moisture on BP fire initiation and expansion, the Global Land Evaporation Amsterdam Model (version 3.3) surface soil moisture (SMsurf) and root-zone soil moisture (SMroot) were used (Martens et al., 2017; Miralles et al., 2011). These two datasets, which range from 1980 to 2018, were gridded at a spatial resolution of $0.5° \times 0.5°$ for each month.

2.2.4 Human activity data

The population density data were used as a proxy for human activities. The History Database of the Global Environment (version 3.2) was interpolated and re-gridded into a monthly scale at a spatial resolution of $0.5° \times 0.5°$ (Klein Goldewijk et al., 2017).

3 Methods

3.1 Study area

Our study focuses on boreal peatland areas with a minimum of 30 % histosol soil content. This criterion was set in place to ensure the dominance of conditions favoring smoldering over other types of fires. Histosols, which are organic-rich soils commonly found in boreal regions, are the result of the accumulation of partially decomposed plant material. They typically form in environments such as bogs and fens, where a high-level water table and an abundance of sphagnum moss and other vegetation contribute to peat formation. The survival and growth of trees can be challenging in those environmental conditions due to factors such as high acidity, a lack of essential nutrients, and waterlogged environments, though certain adaptive species such as black spruce can thrive. We limited the peatland area to regions with more than 30 % histosol content, aiming to ensure the presence of adequate soil

fuel for smoldering while limiting the aboveground fuel such as forests or grasses. By doing so, we also addressed the limitations of satellite-based fire products which differentiate between subsurface smoldering fires and surface fires.

3.2 The two-step hierarchy machine learning framework

Given our assumption of predominantly smoldering fires within our defined research areas, we introduced an ML framework specifically designed to predict rare and extreme fire events. The datasets fed into the framework were selected to ensure relevancy to our study objective. In this research, we proposed a two-step error-correcting ML framework that integrates imbalanced data processing, classification, regression, and error-correcting techniques. Given that over 70 % of months record no fire events, our framework aims to address this data imbalance. Furthermore, it seeks to adapt to the intricate nonlinear nature of extreme BP fires and enhance prediction accuracy. The likelihood of wrong predictions was expressed by evaluation metrics from step one, denoting a broad-based uncertainty for the framework system, that are used at step two to correct the error being propagated to fire size prediction. The evaluation accuracy results are listed in Tables S2–S5 in the Supplement. We also conducted a range of factor-control simulations using a method akin to backward selection to investigate the key contributors to BP fire occurrences and to understand the BP smoldering fire mechanisms. The two-step ML framework is detailed in Fig. 1.

We began by pre-processing the data, which encompassed data integration, treatment of missing values, and standardization. Subsequently, the data were divided into 70 % for training and 30 % for testing. Given the nature of BP fires as predominantly rare and extreme events, there is a notable imbalance in months with and without fire occurrences; only 20 % of the data records in GFED BA indicate fire occurrences. To combat overfitting arising from this dataset imbalance, we employed the synthetic minority oversampling technique (SMOTE). This algorithm enhances the training set by producing synthetic samples from the minority class – namely, the fire occurrence presence. It operates by selecting a minority class instance and determining its k-nearest neighbors (typically five for our model). From these neighbors, one is randomly selected, and synthetic samples are generated between the chosen instance and its neighbors. This method persists until there is an approximate balance between records of fire occurrence presence and absence.

In step one, we applied six prevalent classification algorithms to classify monthly fire occurrences for each grid. These algorithms include logistic regression (LogR), linear support vector machines (SVMs), random forest (RF), bagging (BAG), k-nearest neighbors (KNN), and Gaussian naïve Bayes (GNB). Each algorithm determines the likelihood of a fire "occurrence" using unique computational methods.

While the in-depth mechanics of these algorithms extend beyond the scope of this study, it is important to note that by leveraging this probability, we ultimately derived binary classifications, which serve as our fire occurrence predictions, indicating the presence or absence of fires each month at every geographical location. Subsequently, the algorithms rank the key factors influencing peat fire occurrences to identify the most contributive feature subset. For RF and BAG, feature importance is calculated based on the mean decrease in node impurity, specifically using the Gini index, adjusted by the probability of samples reaching each node. In the cases of LogR and SVM, feature importance is assessed through the coefficients present in LogR's decision functions and linear SVM's weights. Unlike the others, the KNN and GNB classifiers do not provide straightforward methods for feature importance evaluation. Instead, this study leverages a permutation approach that determines importance based on the loss function and the rise in prediction error upon feature shuffling. Due to the varied range of feature importance values obtained from these methods, normalization was applied for uniform comparison. By processing these values using their normalized absolute value, a consistent comparison was achieved. The mean and standard variation of these normalized values from different ML models help define the relative significance of driving factors and the variances between models.

In step two, we employed regression models to estimate fire sizes (or impacts) based on the fire occurrence determinations from step one. Leveraging the monthly fire occurrence predictions from the most efficient ML classifier, we extracted the relevant fire data to predict fire sizes, which encompassed burned area and C emissions. For months with no fire occurrences, fire impacts were initially assessed as zero, with subsequent error correction. We conducted the experiments with 14 regression techniques, including simple linear (LinR), ridge (ridge), least absolute shrinkage and selection operator (LASSO), adaptive boosting (Ada), gradient boosting (GBR), bagging (Bag), random forest (RF), Bayesian regression (Bayes), elastic net (EN), kernel ridge (kernel), decision tree (DT), CatBoost (CBR), and light gradient boosting (LGBR) regressions, to predict the extent of fire impacts. A core presumption of our method was the absolute accuracy of all classifications when setting up regression model inputs. This assumption, however, is not foolproof given the potential for misclassifications. To quantify the likelihood of such errors, we analyzed the confusion matrix from the classification phase. Subsequently, multiple uncertainty assessment matrices were integrated during the regression phase to rectify any propagated errors. The specifics of this error correction process are elaborated upon below.

3.3 Error corrections

More specifically, at the first step, we defined the fire occurrence classes for the training dataset according to the value of
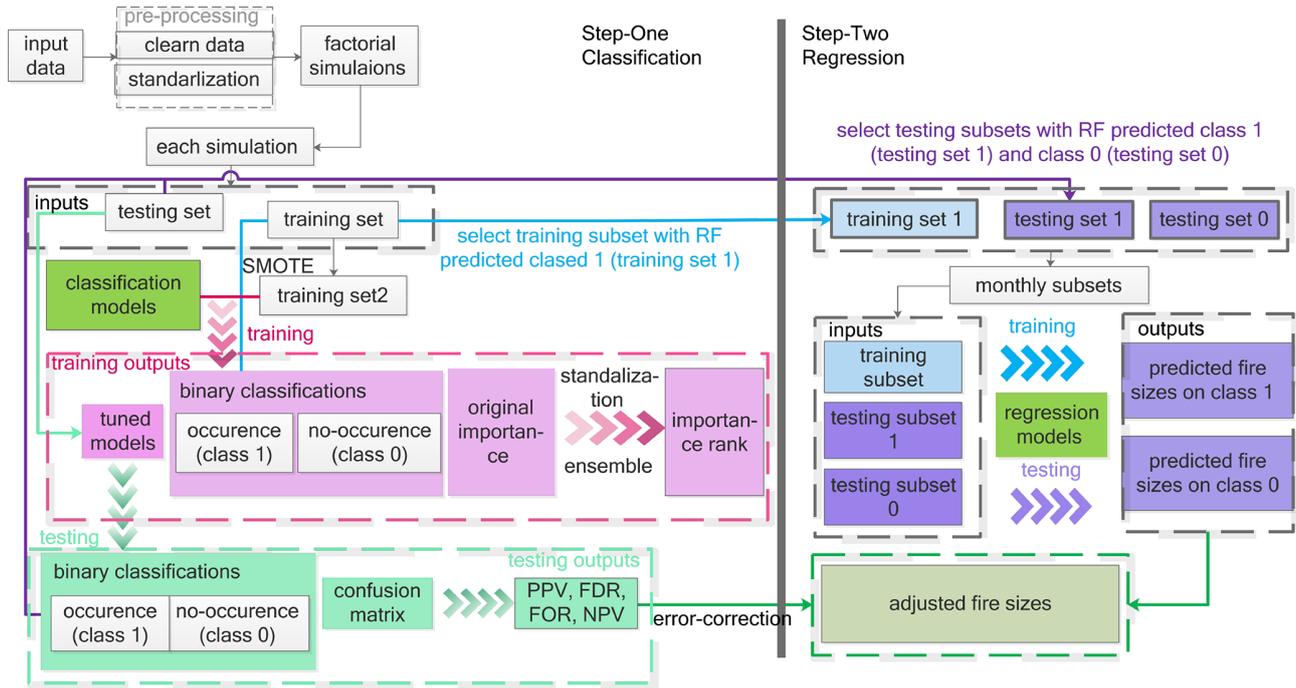
Figure 1. The two-step ML framework, where PPV, FDR, FOR, and NPV stand for positive predictive value, false discovery rate, false omission rate, and negative predictive value, respectively. SMOTE represents the oversampling algorithm – synthetic minority oversampling technique. The error-correcting process is detailed in the Methods part.

C emissions and burned area. Data records with a value equal to 0 indicate the no-fire months, represented as class 0. Conversely, data records with a value exceeding 0 mark the fire months, designated as class 1. This allowed us to distinguish between months with fires (X_{mf}) and without fires (X_{mn}).

For months showing fire activity (C emissions or burned area > 0, namely class 1), we employed the regression model R_{mf} to predict the fire impact:

$$R_{mf}(X_{mf}) = Y_{mf}, \tag{1}$$

where X_{mf} represents the explanatory data in month m with fire, and Y_{mf} is the predictive variable (C emission to burned area) in month m.

For fire-free months, we utilized the regression mode R_{mn}, predicting no impact:

$$R_{mn}(X_{mn}) = 0, \tag{2}$$

where X_{mn} represents the explanatory data in month m without fires occurrences.

Every month's (m) training dataset was bifurcated into fire months X_{mf} and non-fire months X_{mn} while the testing dataset was segmented into X'_{fm} (with fires) and X'_{nm} (without fires). Using the same input data, 14 different regression techniques were employed, ranging from a linear regressor to stacking regressor.

For each month (m) in {1, 2, 3…12} and every regression model (R^r) in {$R^1 R^2, R^3, \ldots R^{15}$}, we created regression models R^r_{mf} for months with fires and R^r_{mn} for fire-free

months:

$$R^r_{mf}(X_{mf}) = y^r_{mf}, \tag{3}$$
$$R^r_{mn}(X_{mn}) = y^r_{mn} = 0. \tag{4}$$

Then, with testing data, we predict fire size by employing model R^r_{mf}.

We then predicted fire impacts based on the classification and addressed potential uncertainties related to both fire and non-fire months. For fire month m (class 1) in testing data, the predicted fire size P^r_{mf} is

$$P^r_{mf} = R^r_{mf}(X'_{mf}). \tag{5}$$

A possible uncertainty related to fire size predictions based on months with fire is that no fires actually occurred, which could be expressed by EP_{mn},

$$EP^r_{mn} = R^r_{mn}(X'_{mf}) = 0, \tag{6}$$

while for months without fires (class 0), the original predicted fire size P^r_{mn} is

$$P^r_{mn} = R^r_{mn}(X'_{mn}) = 0. \tag{7}$$

A possible uncertainty related to fire size predictions in months without fire is that fire events did happen in reality, which could be expressed by

$$EP^r_{mf} = R^r_{mf}(X'_{mn}). \tag{8}$$

To correct for inaccuracies, we integrated four evaluation metrics from classification – positive predictive value (PPV), false discovery rate (FDR), false omission rate (FOR), and negative predictive value (NPV). By applying these metrics to actual and potentially misclassified predictions, we acquired an error-adjusted prediction. The four evaluation metrics from classification are as follows.

$$\text{Positive predictive value (PPV)} = \frac{\text{True positive (TP)}}{\text{TP} + \text{false positive (FP)}} \quad (9)$$

$$\text{False discovery rate (FDR)} = \frac{\text{FP}}{\text{TP} + \text{FP}} \quad (10)$$

$$\text{False omission rate (FOR)} = \frac{\text{False negative (FN)}}{\text{FN} + \text{true negative (TN)}} \quad (11)$$

$$\text{Negative predictive value (NPV)} = \frac{\text{TN}}{\text{FN} + \text{TN}} \quad (12)$$

Applying classification evaluation metrics to the actual fire size predictions (Ps) and erroneous predictions (EPs) that may arise from potentially incorrect classification, we could obtain the error-corrected prediction AP'^r_{mp} for the record of $(p, m) \in X'$ (in the testing set).

$$\text{AP}'^r_{mp} = \begin{cases} \text{PPV} \times P^r_{m\text{f}} + \text{FDR} \times \text{EP}^r_{mn}, & \text{if } Z'_{mp} = 1 \\ \text{NPV} \times P^r_{mn} + \text{FOR} \times \text{EP}^r_{m\text{f}}, & \text{if } Z'_{mp} = 0 \end{cases} \quad (13)$$

Here, $(f, n) \in p$, and Z'_{mp} stands for the original classification prediction in testing data.

3.4 Analysis and validation

Feature importance was further validated through factorial simulations, categorizing features with similar physical implications. We designed experiments to include all attributes and selectively exclude certain features to gauge the relative importance of grouped factors. The temperature-related group contains TMP, TMN, and TMX; PRE is the only PRE-related feature. The air-dryness-related group includes SVP, VAP, VPD, RH, WET, ET, and PDSI. The soil-moisture-related features are SMsurf and SMroot, and the "Others" group includes features representing vegetation biomass (e.g., GPP and NDVI), wind speed (WIN), cloud cover percentage (CLD), climate extremes (e.g., FRS and DTR), and anthropogenic activities (e.g., POPD). The complete simulation setup is listed in Table 1. During the first round of simulation, we incorporated all features, labeled as ALL. As discussed in Sect. 4.3, features within the "Others" group typically rank lower. Despite this, we retained this group in subsequent factor-controlling simulations. The rationale behind this decision is that the "Others" group comprises diverse features (like wind speed and vegetation types) not captured by the primary four feature categories: temperature, precipitation, air dryness, and soil moisture. This diverse feature set is a benefit for reverse verification, ensuring a more comprehensive analysis. In the second round of simulation, each

run excluded one feature group to discern the most influential among the four. For instance, by omitting the TMP (temperature) group in the NO-TMP simulation, we gauged the significance of remaining groups. This process was repeated, resulting in simulations like NO-PRE, NO-HUMI, and NO-SOM. Notably, the temperature group consistently ranked the highest in several evaluations. The third simulation set aimed to rank the relative importance of the PRE, air dryness, soil moisture, and "Others" groups, considering the temperature group had already emerged as the most influential. As such, the temperature group was excluded from all third-set simulations. We further designed simulations like NO-TMP-PRE, NO-TMP-SOM, and NO-TMP-HUMI to probe the comparative significance of groups. The air dryness group topped the ranks in this set, with PRE consistently at the bottom. Subsequently, in the fourth set, we introduced the NO-TMP-PRE-HUMI simulation to examine the comparative weight between soil moisture and other factors pertaining to vegetation and human activities.

Our two-step hierarchy ML framework has been designed with multi-datasets and multi-algorithms to validate the framework's predictive performance. Yet, we had not gauged its robustness against the direct application of machine learning algorithms. To rigorously evaluate the relative efficacy of our framework, we set up a comparative experiment using the same 14 regression models we mentioned earlier and, additionally, the extreme gradient boosting (XGBR). These models were trialed without the hierarchical structure our framework introduces. Performance metrics, including the mean squared error (MSE), mean absolute error (MAE), and the R-squared value (R^2), were documented and are detailed in Table S7.

4 Results

4.1 Fire occurrence predictability

The averaged area under the receiver operating characteristic curve (AUC), which indicates the diagnostic ability of classification ranged from 0.70 ± 0.03 (MCD64A1, the No-TMP-PRE-HUMI simulation) to 0.88 ± 0.05 (MCD45A1, the ALL simulation) for multiple MLTs (Table S3). The ALL simulation had an AUC value of 1 at the training stage and an AUC value of 0.72–0.93 at the testing stage. The RF algorithm showed the best predictive performance for fire occurrences (i.e., fire counts) (Table S4) and provided a basis for fire impact prediction. Among all datasets, MCD45A1 had the highest recall rate (0.94) and highest precision (0.96), indicating that a few months were incorrectly classified (Table S4). MCD64A1 had the lowest recall rate and precision rate, indicating discrepancies among different data sources. Using the SMOTE oversampling algorithm, the testing recall rate was effectively improved at an average rate of 26.88 %

Table 1. Simulation experiments for assessing environmental factor cluster impacts on ML predictability.

Simulations		Explanatory variable groups				
		Temperature related	Precipitation related	Air dryness related (i.e., humidity)	Soil moisture related	Others
First	ALL	Yes	Yes	Yes	Yes	Yes
Second	NO-TMP	No	Yes	Yes	Yes	Yes
	NO-PRE	Yes	No	Yes	Yes	Yes
	NO-HUMI	Yes	Yes	No	Yes	Yes
	NO-SOM	Yes	Yes	Yes	No	Yes
Third	NO-TMP-PRE	No	No	Yes	Yes	Yes
	NO-TMP-SOM	No	Yes	Yes	No	Yes
	NO-TMP-HUMI	No	Yes	No	Yes	Yes
Fourth	NO-TMP-PRE-HUMI	No	No	No	Yes	Yes

and with the highest growth of 48.62 % for the FireCCI BA dataset (Tables S5 and S6).

Besides evaluation metrics, the spatial disparities of predicted fires from MLTs and multiple datasets were also examined against corresponding observations. The BP with a histosol fraction greater than 30 % is mainly located in the Hudson Bay Lowland (HBL) and western Siberia (WS) (Fig. S2 in the Supplement). Observations from FireCCI BA, GFED BA, GFED carbon emissions, and MCD64A1 fire detection consistently show that there were fewer than 60 fire events in the HBL region from 1997 to 2015, but the fire count in WS during the same time period ranged from 30 to more than 150. This demonstrates the spatial disparity of peatland fire occurrences in boreal areas and possibly implies that WS is more fire-prone than the HBL (Figs. S3–S6a-1 and a-2 in the Supplement). FireCCI, GFED, and MCD64A1 showed good consistency among these three products with respect to the data distribution. Unlike these three datasets, MCD45A1 had higher estimation and lower spatial heterogeneity of fire counts in BP (Fig. S7a-1 and a-2 in the Supplement). The more evenly distributed data in MCD45A1 may be the primary reason why MCD45A1 had the highest predicting accuracy and best performance in reproducing the distribution of fire counts spatially and temporally in the testing stage (Fig. S11 in the Supplement).

Predictability discrepancies were also compared among multiple ML algorithms. The validation results demonstrate that the bootstrap-based ML algorithms (i.e., RF, BAG, and KNN) – in which there is no requirement for data distribution assumption, and resampling supports the inference of the population distribution – had better predictability than other algorithms (i.e., LogR, linear SVM, and GNB) (Figs. S1 in the Supplement and 2). For RF and BAG, the reproducing accuracy rate (i.e., true positive rate and true negative rate) was over 90 % with the FireCCI data (Fig. S1). The inaccurate predictions of KNN, LogReg, SVM, and GNB

were significantly influenced by the overestimated fire occurrence (namely false positive) during the fire season (April–October), as shown in Fig. 2. Without a prescribed underlying function, the nonparametric RF and BAG models exhibited advantages over other ML algorithms in reproducing peatland fire distributions spatially (Figs. S3–S6b-1, b-2, c-1, c-2, d-1, and d-2) and temporally (Fig. 2). Therefore, the predictions of fire occurrence from the best-performing RF were employed as the basis of fire impact predictions.

4.2 Fire impact predictability

ML regression models exhibit moderate predictabilities of fire sizes (Fig. S22 in the Supplement). Both ML classification at step one and regression models at step two overestimated fire size during the fire season (Figs. S19–S21 in the Supplement). This study developed an error-correcting technique to tackle the error propagation and overestimation during the fire season and achieved satisfying performance (Fig. 3 and S19–S21).

In the WS area, there are more occurrences of fire events and thus higher carbon emissions compared to those observed in the HBL area (Figs. S16–18 in the Supplement). The predicted carbon emissions from the stacked ML algorithms were overall consistent with the observations in WS and western Canada but had overestimations in the HBL (Fig. S18). The error-correcting technique could slightly lower the overestimation in the HBL (Fig. S18) but could greatly lower the overestimation temporally, especially in July (Fig. S21). Meanwhile, the underestimation of fire impacts in June remained a common problem for all 14 regression models (Fig. S21).

GFED BA and FireCCI BA were used to determine the reliability of fire impact predictions within the two-step ML framework. In terms of spatial reproducibility, the predictions from GFED BA (Fig. S17d–f) were more accurate than
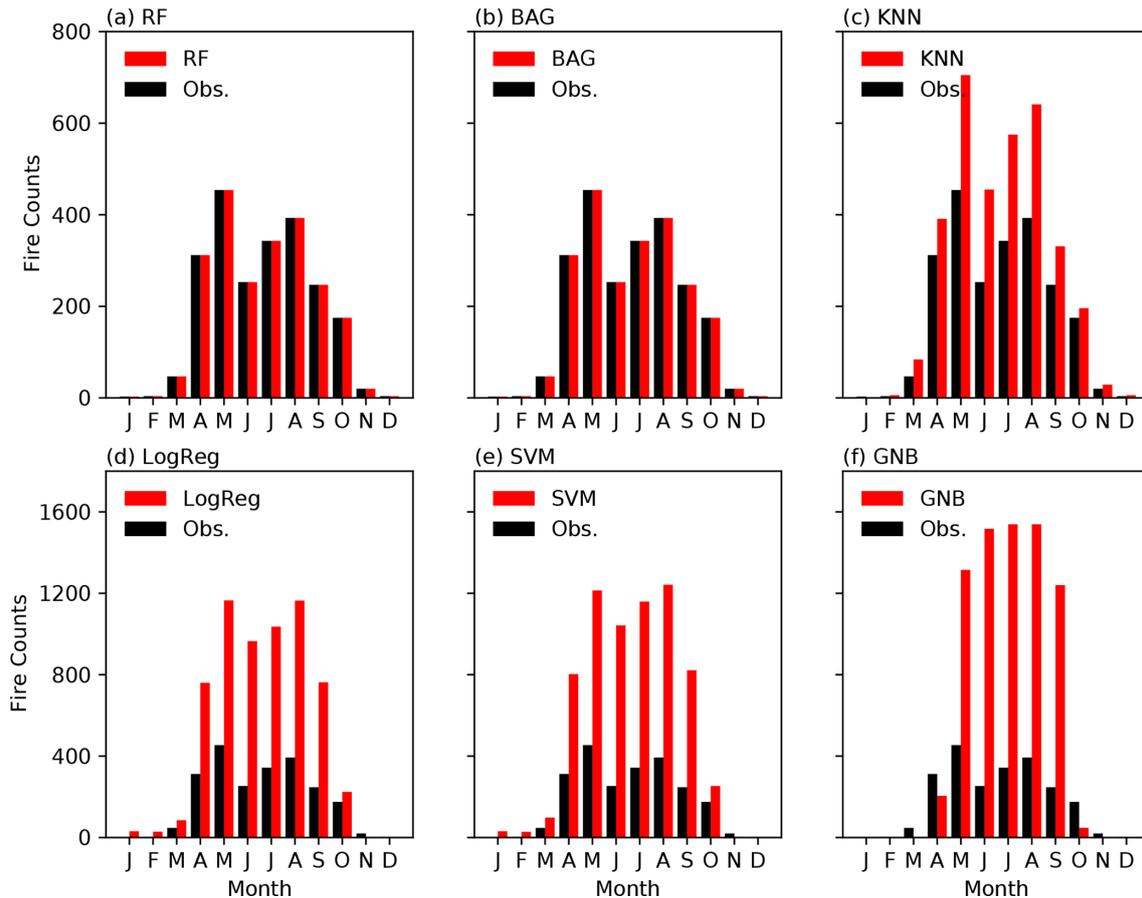
Figure 2. Seasonality of observational and predicted fire counts from the six ML algorithms with the FireCCI BA dataset.

those from FireCCI BA (Fig. S16a–c), particularly in the HBL, where the BA is less than $50\,km^2$ (Figs. S16–S18a-1 and b-1). Figures S16–S18a-2 and b-2 show that the framework underestimated burned area in northern WS and overestimated burned area in the northern HBL for FireCCI BA (Figs. S16–S18a-1 and b-1). Different BA datasets can also have temporal inconsistencies. FireCCI BA has its fire season from March to May, whereas GFED BA has its fire season from March to October. Despite the fact that April and May were the fire peak months according to both FireCCI BA and GFED BA, the burned areas predicted by the framework based on FireCCI and GFED show differences in BP. According to FireCCI, the predicted entire burned area in May has about $55\,792\,km^2$, whereas the prediction based on GFED is only about $12\,183\,km^2$. A further investigation shows that GFED BA has a bimodal distribution, while FireCCI BA is unimodally distributed (Figs. S19 and S20 in the Supplement). Therefore, it is important to determine whether ML is applicable for various datasets.

Overall, the 14 tested regression models were able to reproduce fire impact magnitudes and seasonality well for FireCCI BA, GFED BA, and GFED carbon emissions (Figs. S19–21). Those ML regression models appear to overestimate the fire effects, including carbon emissions and burned area, during fire season. However, the error-correcting approach could successfully reduce this bias (Figs. S19–21). Discrepancies among model predictabilities were small. For example, for the FireCCI data, the decision tree had the best performance with estimations that were 4.05 % higher than the observations, whereas bagging had the worst performance with estimations that were 10.84 % higher than the observations. Such small biases and discrepancies verified the reproducibility and predictability of the two-step ML framework.

4.3 Validation

The predicting results with direct application of machine learning algorithms are presented in Table S7. We could see that the majority of the models display subpar performance, exhibiting a considerable bias and a low value of explained variances during both training and testing phases. By employing our hierarchical ML framework, the explained variance typically exceeds 50 % in the testing phase. In contrast, in this instance, the variance hovers around 1 %, which is sig-
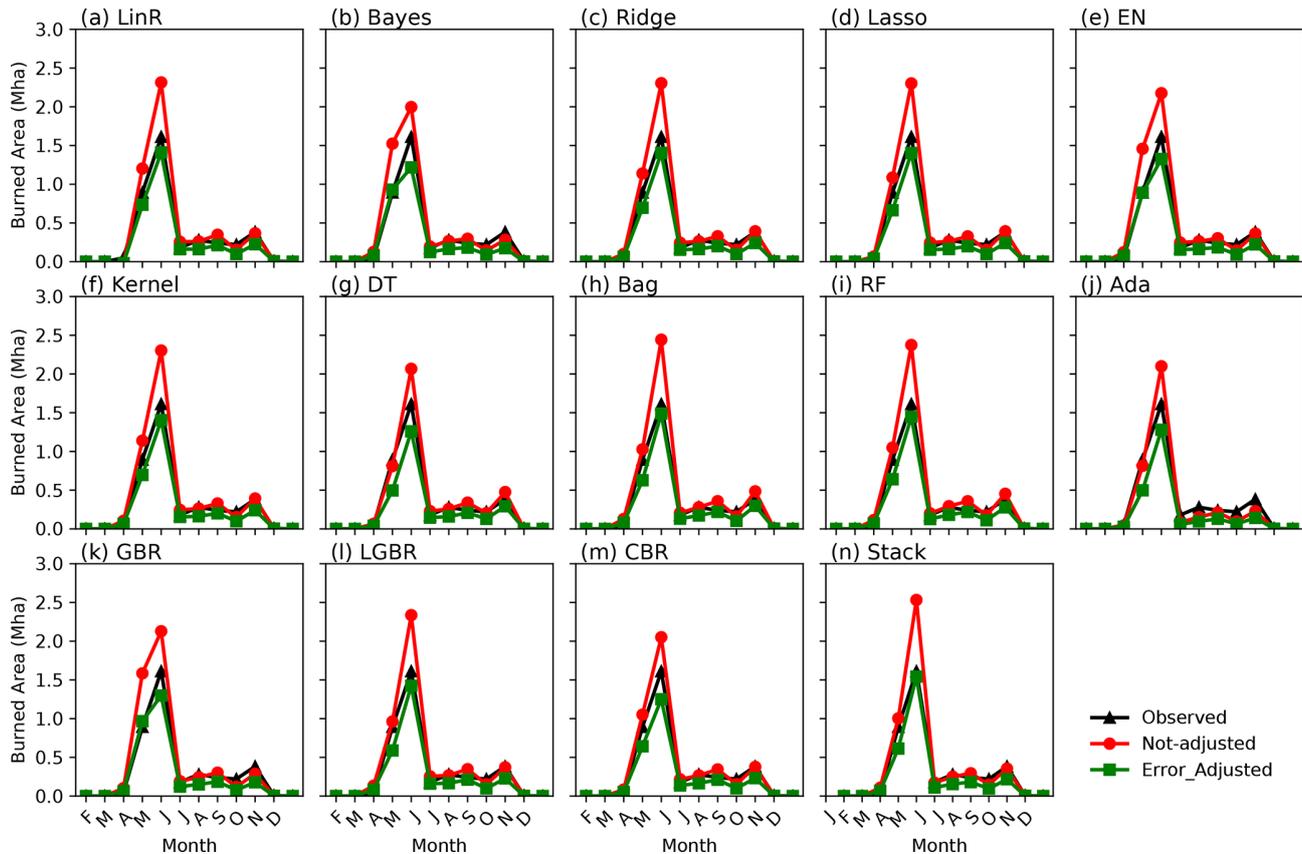
Figure 3. Seasonality of the observed, non-adjusted, and error-adjusted FireCCI BA based on the testing phase from multiple ML regression models: **(a)** linear, **(b)** Bayesian linear, **(c)** ridge, **(d)** LASSO, **(e)** elastic net, **(f)** kernel ridge, **(g)** decision tree, **(h)** bagging, **(i)** RF, **(j)** AdaBoost, **(k)** gradient boosting, **(l)** light gradient boosting, **(m)** CatBoost, and **(n)** stacking.

nificantly lower compared to the results obtained using our framework.

Notably, the RF, BGA, CBR, LGBR, and stack models demonstrate commendable performance during the training phase but falter significantly during testing. This decline suggests typical overfitting issues, largely attributed to severe data imbalances, underscoring the poor predictability of these models.

4.4 Primary causes of BP fires

To exclude feature collinearity, four sets of simulations in Table 1 were designed by opting out grouped features to confirm the importance ranking of features (Fig. 4). The first two sets of simulations (i.e., ALL, NO-TEMP, NO-PRE, NO-HUMI, and NO-SOIMOI) showed that the temperature-related feature group had the highest importance (Fig. S3a and c–e in the Supplement). The third set of simulations, which removed two feature groups, showed that the air dryness had the highest importance among the remaining four feature groups, namely the PRE, air dryness, soil dryness, and other groups. PRE was found to be the third-ranked feature according to the first three sets of simulations. The last

set of simulations was conducted to compare the relative importance of soil moisture and the other human and natural features and found that frost (FRS) and vegetation biomass (GPP) in the other human and natural features group were more important than soil moisture (Fig. 4i). Such ranks were also indicated by other simulations in Fig. 4a, c, d, f, and h. Thus, this study found that BP fires were significantly affected by temperature, air dryness, frost, and GPP (Fig. 4a), which collectively account for more than 80 % of the predictive interpretability (Fig. 4a). Moreover, BP fires were not sensitive to PRE, soil moisture, wind speed, and human activities.

The feature importance ranks were validated not only by FireCCI BA but also by GFED BA, GFED carbon, MCD45A1, and MCD64A1. The rankings from GFED BA and GFED carbon emissions were highly consistent with those from FireCCI (Figs. 4, S12, and S13 in the Supplement), in which temperature, air dryness, frost, and GPP were more important than PRE, soil moisture, wind speed, and other natural and anthropogenic factors. Feature rank discrepancies were found when the ML algorithms were applied to MCD64A1 and MCD45A1, for which the top three features were still air dryness, temperature, and FRS, but soil
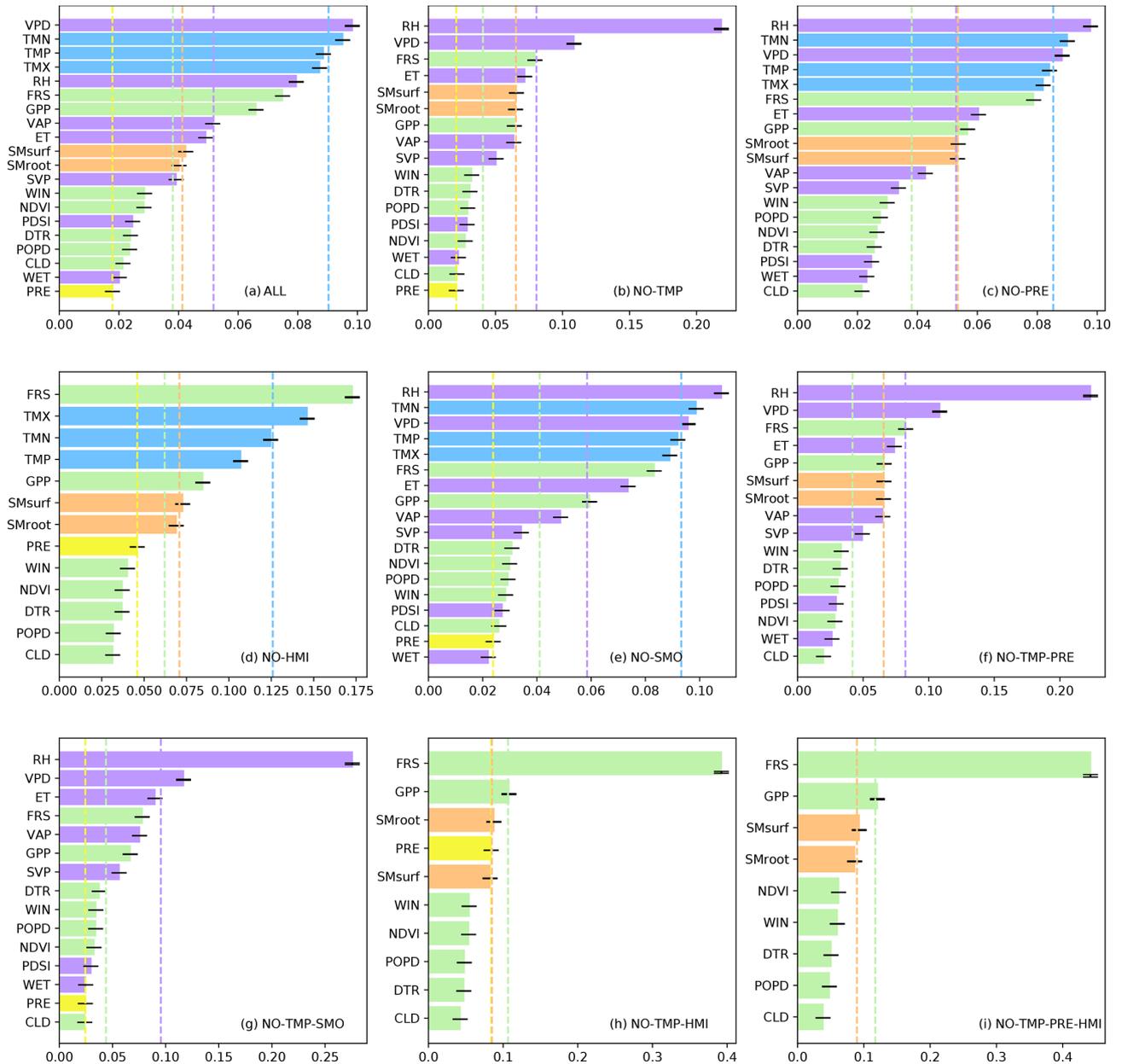
Figure 4. The bar plot stands for the factor importance rank of multiple simulation scenarios using FireCCI BA as the target variable in which the importance was determined by the standardized mean and uncertainty range (minimum and maximum) from multiple ML algorithms; the dashed vertical line indicates the group mean importance of temperature (blue), PRE (yellow), air dryness (purple), soil moisture (orange), and other factors (green).

moisture was more significant than GPP (Figs. S14 and S15 in the Supplement).

Collectively, the multisource datasets and multi-feature simulation experiments consistently suggested that air-dryness-related variables (RH, VPD, and VAP), temperature-related variables (TMN, TMP, TMX), and FRS play more important roles in peat fires than other factors, such as PRE, wind speed, and other natural and human factors. In terms of importance, soil moisture and GPP were both ranked at

a middle level, but their relative rankings could not be determined because soil moisture was considered more significant than GPP according to MCD64A1 and MCD45A1 but GPP was viewed as more important based on FireCCI, GFED BA, and GFED carbon emission datasets.

5 Discussion and limitations

5.1 ML predictability

In this study, a two-step error-correcting framework was built to investigate the BP fire predictability and the individual impacts from meteorological, vegetational, soil, and anthropogenic factors. Machine learning algorithms are increasingly utilized in wildfire research for various applications (Coffield et al., 2019; Jain et al., 2020; Sayad et al., 2019; Wang and Wang, 2020; Yu et al., 2020). However, the literature is insufficient on detailed criteria for selecting appropriate ML models for these tasks, and the interpretability of these machine learning algorithms continues to be a significant challenge for scientific research where understanding the decision-making process of the models for causality purposes is crucial (Li et al., 2023; Buch et al., 2023). To achieve higher predictive accuracy, uncertainties from ML algorithms and the input data are discussed. In this study, results from six classification models and 14 regression models indicate that nonparametric ML algorithms, including RF, bagging, and KNN, outperformed the other employed parametric models, such as LogR, linear SVM, and GNB, by overcoming the severe imbalance of fire data (the non-fire classes have 6 times as many records as fire classes) (Figs. 2 and 3). Unlike parametric models that are highly restricted to specified functional forms and a fixed number of parameters, nonparametric models can fit various functional forms, and the number of parameters grows with the size of the training set, promoting the performance of model predictability.

In BPs, it is challenging to predict fire occurrence because of the extremely unbalanced fire data. Several previous studies have employed ML to investigate peatland fire predictability. For example, Rosadi et al. (2020) employed a variety of ML algorithms to predict fire occurrence in peatland and used accuracy as the only evaluation metric. Such an evaluation method could fail to measure fire predictability once the fire data are imbalanced. According to another study that predicted peatland fire occurrence in Canada (Bali et al., 2021), the recall rates were very high (0.82–0.99) but the precision metrics were very low (0.002–0.05), which indicates a high type I error. In our study, RF regressions yielded high precision metrics (0.56–0.96) and recall rate (0.6–0.94), as well as clearly identified fire months, suggesting relatively low type I and type II errors.

To address the extreme data imbalance, this study used both pre-processing (oversampling) and post-processing (error correcting) in the two-step ML framework to improve predictability. In step one, the SMOTE algorithm significantly improved the recall rate by $\sim 26.88\,\%$–$48.66\,\%$ across all fire datasets. Processing approaches (e.g., oversampling and undersampling) were also found beneficial in earlier studies for certain ML algorithms (Farquad and Bose, 2012; Malik et al., 2021; Zhou et al., 2020). Through a comparative experiment against the direct application of machine learn-

ing algorithms, our two-step hierarchical framework demonstrated superiority in mitigating the overfitting issue. However, predicting rare and extreme fire sizes remains a challenge. To quantify and reduce uncertainty sources and error propagations in ML frameworks, procedures are typically highly tailored for specific research challenges and ML algorithms (Jiang and Nachum, 2020; Pan et al., 2019; Wang et al., 2020). In our two-step ML framework, applying evaluation metrics from the classification step (step one) in error correcting effectively lowered the overestimated BA and carbon emissions during fire season (Figs. S19–S21).

Although our hierarchical machine learning framework showed some supremacy and robustness, predicting fire sizes is fraught with challenges, stemming from many aspects. In addition to severely imbalanced data, comprehensive ground-based data in BP regions are sparse, making satellite validation difficult, especially given that the smoldering nature of these fires often eludes satellite detection. Thus, it indicates that satellite-based datasets are not perfect when applying them in studying broad-scale smoldering fires, especially in terms of carbon emissions due to high correlation with the burning depth. However, applying them at a regional scale could be acceptable, as the uncertainties are generally comparable due to spatial homogeneity. Additionally, the role of intricate local factors, like peat depth and moisture content, is pivotal in influencing fire behavior, and lacking these datasets can affect predictions. Moreover, the evolving dynamics of climate change and unpredictable human activities, such as land use changes, introduce further variations, making effective BP fire prediction a multidisciplinary challenge.

5.2 Primary driving factors of peatland fires

ML-derived statistical correlations do not necessarily indicate causality, and biophysical or biochemical principles are thus needed to further examine whether such relationships are reasonable (Schölkopf et al., 2021). We have incorporated a great number of factors believed to influence smoldering fires into our model, as one of our intentions behind developing this machine learning framework was to complement existing tools in identifying potential drivers of smoldering fires. By identifying the most influential factors, we sought to align them with existing scientific theories to understand their potential roles in driving smoldering fires but not to quantify all the causalities due to the absence of measurement for particular factors. Thus, we will connect our discoveries to existing findings to discuss theoretical support rather than implementing specific analysis at the current stage.

In this study, four sets of ML simulations were designed to determine the primary driving factors of peatland fires by removing feature groups sequentially. The results revealed that the feature importance rank exhibited general consistency in multiple fire datasets. PRE in boreal or sub-arctic

regions is primarily in the form of snow rather than rainfall due to cold weather (Behrangi et al., 2016), which has little impact on BP fires. Moreover, smoldering fires can persist for a long time (months to years) even in rainfall weather (Lin et al., 2020). This low importance was verified by our ML simulations. Similarly, in sparsely populated boreal peatland, human activities showed a marginal effect. Factorial simulations consistently demonstrated that temperature (i.e., minimum, maximum, and average values), air-dryness-related variables (e.g., RH, VPD, VAP, ET), and FRS were the primary factors driving the BP wildfire activities (Figs. 4 and S12–S13). Although these factors eventually lead to dry and combustible conditions for peatland fire occurrence and propagation, the processes in which they play roles are quite different.

BP fires are intimately tied to weather, and warming appears to increase ignitions, fire frequency, and fire severity (Duffy et al., 2005; Flannigan et al., 2005; Kohlenberg et al., 2018). In peatlands without frost (Fig. 5a), rising temperatures increase saturation vapor pressure (SVP) and continually induce an increase in vapor pressure deficit (VPD) if actual atmospheric VP does not increase as much as SVP. A recent investigation indicates that RH (i.e., ratio of actual water VP to SVP) has plunged rapidly since the year 2000, leading to a sharp rise in VPD on a global scale (Yuan et al., 2019). Such a warming-induced increase in VPD increases evapotranspiration (ET) more in peatlands than in forests with a simulated percentage of up to 30 % (Helbig et al., 2020). Because atmospheric demand (i.e., VPD) dominates the limitation of ET over the soil moisture (Helbig et al., 2020; Novick et al., 2016), the water table turns out to be the water supplier in response to rising VPD, which consequently results in a decrease in water table depth. The water table depth decrease tends to change the physical characteristics of peat in many aspects, such as by lowering the capacity of water storage, causing the peat volume to shrink and volumetric soil moisture to decrease (Price and Schlotzhauer, 1999), as well as inducing surface subsidence with a concomitant decrease in bulk density and an increase in peat oxidation and decomposition (Leifeld et al., 2011; Whittington and Price, 2006). These changes ultimately lead to more carbon being released into the atmosphere and the formation of drier and more flammable peat soil (Fig. 5a). In peatlands with frost, frost heaving deepens the active layers (Jones et al., 2015; Wang et al., 2020), changes the hydrological and thermal properties of peatland, promotes microbial and chemical exothermic reactions, strengthens peatland dryness, and consequently facilitates more frequent peatland fires (Kim et al., 2020) (Fig. 5b).

Our ML-based sensitivity simulations demonstrated the power of using big data to determine the primary causes of peat fires: temperature, atmospheric dryness (e.g., RH, VAP, VPD, ET), and frost (i.e., FRS). These simulations also helped identify the less important factors and processes. For example, wind speed and population density were ranked at

the bottom, suggesting that human activities may not be the main causes of peatland fire occurrence and that wind speed, unlike forest fires, does not significantly affect peatland fire spread. Another intriguing discovery is that the simulations in this study consistently revealed the important role that FRS has played in causing peatland fires and their spread, though FRS has been understudied in previous studies. Dixon et al. (2018) revealed that the seasonal frost layer alters spring water balance, induces a drier spring, and enhances risks of deep smoldering. More specifically, ground-freezing frost can greatly change the structure and properties of peatlands. During the water icing process, the pore diameter is enlarged, which consequently results in peat volume expansion, water tension decreases, water storage capacity increases, and air capacity surges (Dijk and Boekel, 1965). As the air capacity increases, the oxidation of the soil organic carbon is likely to increase. This oxidation produces heat and makes the soil temperature increase, which can start peatland fires by self-ignition (Arief et al., 2019; Restuccia et al., 2017). During the seasonal freezing process, soil water diffuses vertically from the bottom unfrozen layer to the upper frozen layer (frost front) (Nagare et al., 2012). After cycles of freezing and thawing (i.e., frost heaving), surface peat soil becomes drier, and the freezing surface becomes thicker in the form of surface lift above the water table. At low temperatures, heat generated from respiration and the growth of microorganisms dominates heat generated from chemical oxidation in the peat decomposition (Yuan et al., 2021). If frost heaving causes the peatland to dry out year by year, exothermic processes from biological reactions may intensify chemical oxidation with high temperatures and thus induce spontaneous peatland fires (Fig. 5b).

Collectively, the important factors uncovered by the ML framework indicated two peatland fire mechanisms that suit two types of peat soil: unfrozen and seasonal frozen peatlands (Fig. 5a and b). Temperature, air dryness, and facilitated warming and drying in an underground environment may start fires in unfrozen peatlands. For seasonal freezing and thawing in seasonally frozen peatland, frost heaving induces a deep drying and oxygen-rich underground environment and may speed up exothermic progress in biological reactions, thereby promoting peatland fire occurrences.

There are several limitations of this study. Because of a lack of gridded burned depth data and bulk density, this ML-based work could not predict and evaluate peat fire severity. The satellite-based fire datasets used in this study do not provide underground smoldering peat fire as a single product. Fires detected by satellites could be a mixture of peatland surface flaming fires and smoldering fires because the detected radiant signature of smoldering is much weaker than that of flaming fires (Rein and Huang, 2021). In addition, peat fire C emissions have been estimated largely by multiplying detected burned area by a range of parameters, such as average burning depth, combustion completeness, and emission factors of major carbon species. Those estimated param-
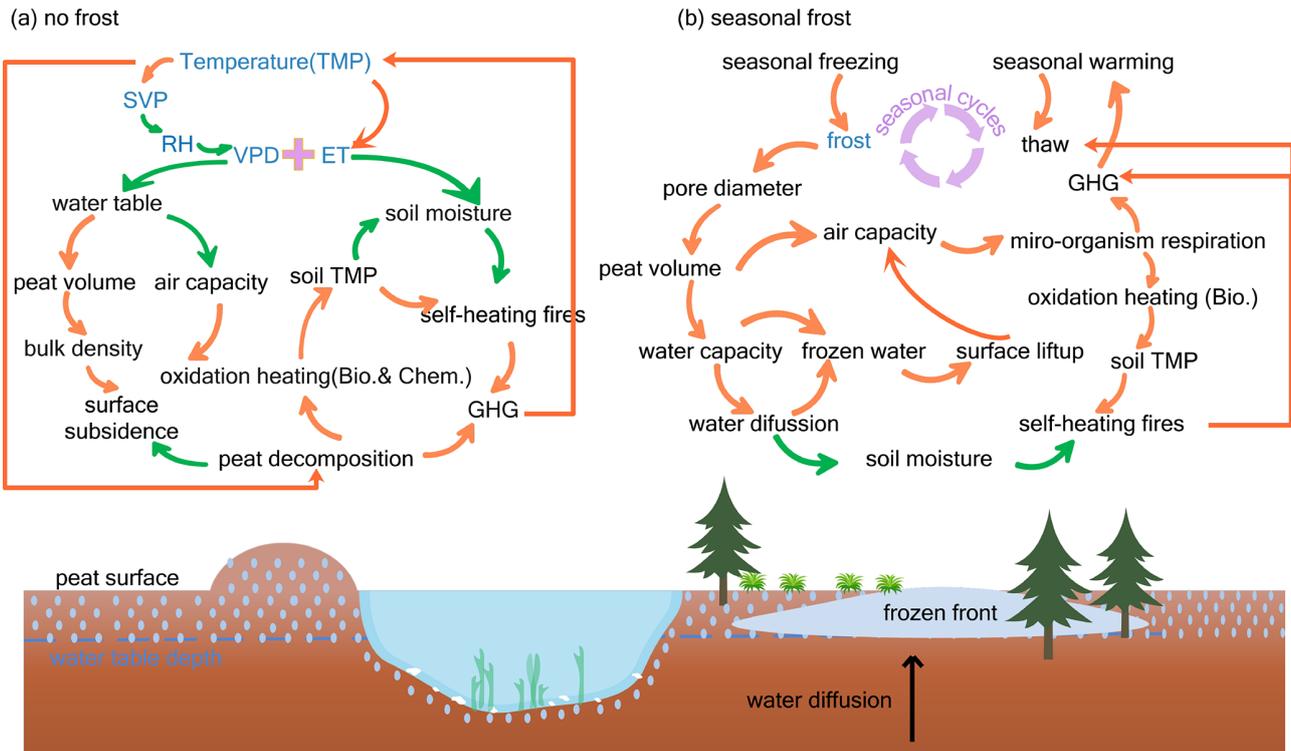
Figure 5. Processes in which environmental factors participate in self-heating peatland fires. ML-identified primary factors are marked in blue; green arrows indicate negative correlation between the two connected factors, and orange arrows indicate positive correlation between the two connected factors.

eters may induce large uncertainties due to the limited ability of optical satellites to detect underground smoldering and burning depth (Graham et al., 2022). The limited data availability, such as vegetation types (moss and vascular plants), burning depth, bulk density, water table depth, and soil temperature, makes ML algorithms limited in fully accounting for all contributing factors. Moreover, since the relationships identified by the ML framework do not automatically imply causality, the underlying physical mechanisms still need to be further validated by future experimental work or theoretical analyses, such as the overriding control of temperature-related variables on inducing boreal peatland fires and the mechanism by which frost impacts peat drying and smoldering (Dixon et al., 2018).

6 Conclusion

This study constructed a two-step error-correcting ML framework to explore the predictability of peatland fire occurrences and impacts (including burned area and C emissions). Major climate, vegetation, soil, and human factors that possibly induce BP fires were included in a range of factorial simulations. The framework successfully predicted the fire counts (occurrences) and fire impacts with accuracy, in general, greater than 80 %, demonstrating the framework's

utility in predicting rare and extreme fire events. Temperature and air dryness were suggested to dominate the fires in unfrozen BPs, while FRS was determined to dominate fire in frozen BPs through the impacts of frost heaving (seasonal freezing–thawing) on changing thermal–hydrological characteristics of peat soil. Our research provides preliminary insights into the overriding impacts of temperature (including temperature-related air dryness and frost heaving) on BP fires via big data and ML. To overcome the ML's limitations in inferencing causality from data association and to further validate the underlying physical mechanisms in BPs fire, more field data (such as peat soil properties and peat burning properties), as well as additional site experimental, statistical, or computational works, are needed in the future.

be obtained from https://e4ftl01.cr.usgs.gov/MOTA/ (Roy et al., 2002; Giglio et al., 2018). The CRU TS 4.04 data are archived at https://crudata.uea.ac.uk/cru/data/hrg/ (Harris et al., 2020). MERRA-2 2 m wind speed data can be found at https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2/M2T1NXSLV.5.12.4/ (Gelaro et al., 2017). GPP data are available at https://doi.org/10.3334/ORNLDAAC/1789 (Madani and Parazoo, 2020). GIMMS3g NDVI data are obtained from https://data.tpdc.ac.cn/en/data/9775f2b4-7370-4e5e-a537-3482c9a83d88/ (The National Center for Atmospheric Research, 2018). GLEAM soil moisture data can be accessed at https://www.gleam.eu/ (Martens et al., 2017; Miralles et al., 2011). HYDE population density data are available at https://doi.org/10.17026/dans-25g-gez3 (Klein Goldewijk, 2017). The boreal peatland map is obtained from https://doi.org/10.17043/hugelius-2020-peatland-1 (Hugelius et al., 2020).

For a more detailed list of data information, please refer to Supplement Table S1.

Author contributions. MJ, JM, and RT conceived the research ideas. RT wrote the initial draft of the paper and performed the modeling and analysis. All authors contributed to preparation of the paper, editing the paper, draft revision, and providing scientific suggestions.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Special issue statement. This article is part of the special issue "The role of fire in the Earth system: understanding interactions with the land, atmosphere, and society (ESD/ACP/BG/GMD/N-HESS inter-journal SI)". It is a result of the EGU General Assembly 2020, 4–8 May 2020.

Review statement. This paper was edited by Yilong Wang and reviewed by three anonymous referees.

References

Arief, A. T., Nukman, and Elwita, E.: Self-Ignition Temperature of Peat, J. Phys. Conf. Ser., 1198, 042021, https://doi.org/10.1088/1742-6596/1198/4/042021, 2019.

Bali, S., Zheng, S., Gupta, A., Wu, Y., Chen, B., Chowdhury, A., and Khim, J.: Prediction of Boreal Peatland Fires in Canada using Spatio-Temporal Methods, Climate Change AI. ICML 2021 Workshop on Tackling Climate Change with Machine Learning. Climate Change AI, https://www.climatechange.ai/papers/icml2021/12 (last access: 19 January 2023), 2021.

Bedia, J., Herrera, S., and Gutiérrez, J. M.: Assessing the predictability of fire occurrence and area burned across phytoclimatic regions in Spain, Nat. Hazards Earth Syst. Sci., 14, 53–66, https://doi.org/10.5194/nhess-14-53-2014, 2014.

Behrangi, A., Christensen, M., Richardson, M., Lebsock, M., Stephens, G., Huffman, G. J., Bolvin, D., Adler, R. F., Gardner, A., Lambrigtsen, B., and Fetzer, E.: Status of high-latitude precipitation estimates from observations and reanalyses, J. Geophys. Res.-Atmos., 121, 4468–4486, https://doi.org/10.1002/2015JD024546, 2016.

Buch, J., Williams, A. P., Juang, C. S., Hansen, W. D., and Gentine, P.: SMLFire1.0: a stochastic machine learning (SML) model for wildfire activity in the western United States, Geosci. Model Dev., 16, 3407–3433, https://doi.org/10.5194/gmd-16-3407-2023, 2023.

Burgan, R. E. and Rothermel, R. C.: BEHAVE: fire behavior prediction and fuel modeling system–FUEL subsystem, U. S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT, https://doi.org/10.2737/INT-GTR-167, 1984.

Castelli, M., Vanneschi, L., and Popovič, A.: Predicting burned areas of forest fires: An artificial intelligence approach, Fire Ecol., 11, 106–118, https://doi.org/10.4996/fireecology.1101106, 2015.

Che Azmi, N. A., Mohd Apandi, N., and Rashid, A. S. A.: Carbon emissions from the peat fire problem – a review, Environ. Sci. Pollut. Res., 28, 16948–16961, https://doi.org/10.1007/s11356-021-12886-x, 2021.

Chuvieco, E., Pettinari, M. L., Lizundia-Loiola, J., Storm, T., and Padilla Parellada, M.: ESA Fire Climate Change Initiative (Fire_cci): MODIS Fire_cci Burned Area Pixel product, version 5.1 (3.1), https://doi.org/10.5285/58F00D8814064B79A0C49662AD3AF537, 2018.

Coffield, S. R., Graff, C. A., Chen, Y., Smyth, P., Foufoula-Georgiou, E., and Randerson, J. T.: Machine learning to predict final fire size at the time of ignition, Int. J. Wildland Fire, 28, 861–873, https://doi.org/10.1071/WF19023, 2019.

Costafreda-Aumedes, S., Comas, C., and Vega-Garcia, C.: Human-caused fire occurrence modelling in perspective: a review, Int. J. Wildland Fire, 26, 983–998, https://doi.org/10.1071/WF17026, 2017.

Dixon, S. J., Lukenbach, M. C., Kettridge, N., Devito, K. J., Petrone, R. M., Mendoza, C. A., and Waddington, J. M.: Seasonally frozen soil modifies patterns of boreal peatland wildfire vulnerability, Hydrol. Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/hess-2017-678, in review, 2018.

Duffy, P. A., Walsh, J. E., Graham, J. M., Mann, D. H., and Rupp, T. S.: Impacts of large-scale atmospheric-ocean variability on Alaskan fire season severity, Ecol. Appl., 15, 1317–1330, https://doi.org/10.1890/04-0739, 2005.

ESA: Fire_cci Burned Area dataset, ESA [data set], https://geogra.uah.es/fire_cci/firecci51.php (last access: 12 October 2021), 2018.

Farquad, M. A. H. and Bose, I.: Preprocessing unbalanced data using support vector machine, Decis. Support Syst., 53, 226–233, https://doi.org/10.1016/j.dss.2012.01.016, 2012.

Field, C. B. and Raupach, M. R. (Eds.): The global carbon cycle: integrating humans, climate, and the natural world, Island Press, Washington, 526 pp., ISBN 1-55963-526-6 (cloth: alk. paper), ISBN-10 1559635274, ISBN-13 978-1559635271 (pbk.: alk. paper), 2004.

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

Forkel, M., Andela, N., Harrison, S. P., Lasslop, G., van Marle, M., Chuvieco, E., Dorigo, W., Forrest, M., Hantson, S., Heil, A., Li, F., Melton, J., Sitch, S., Yue, C., and Arneth, A.: Emergent relationships with respect to burned area in global satellite observations and fire-enabled vegetation models, Biogeosciences, 16, 57–76, https://doi.org/10.5194/bg-16-57-2019, 2019.

Frandsen, W. H.: Ignition probability of organic soils, Can. J. Forest Res., 27, 1471–1477, 1997.

French, N. H. F., Goovaerts, P., and Kasischke, E. S.: Uncertainty in estimating carbon emissions from boreal forest fires, J. Geophys. Res.-Atmos., 109, 1–12, https://doi.org/10.1029/2003JD003635, 2004.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017 (data available at: https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2/M2T1NXSLV.5.12.4/, last access: June 2021).

Giglio, L., Randerson, J. T., and Van Der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), J. Geophys. Res.-Biogeo., 118, 317–328, https://doi.org/10.1002/jgrg.20042, 2013.

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., and Justice, C. O.: The Collection 6 MODIS burned area mapping algorithm and product, Remote Sens. Environ., 217, 72–85, https://doi.org/10.1016/j.rse.2018.08.005, 2018 (data available at: https://e4ftl01.cr.usgs.gov/MOTA/, last access: 20 January 2020).

Gorham, E.: Northern Peatlands: Role in the Carbon Cycle and Probable Responses to Climatic Warming, Ecol. Appl., 1, 182–195, https://doi.org/10.2307/1941811, 1991.

Graham, L. L. B., Applegate, G. B., Thomas, A., Ryan, K. C., Saharjo, B. H., and Cochrane, M. A.: A Field Study of Tropical Peat Fire Behaviour and Associated Carbon Emissions, Fire, 5, 62, https://doi.org/10.3390/fire5030062, 2022.

Grishin, A. M., Yakimov, A. S., Rein, G., and Simeoni, A.: On physical and mathematical modeling of the initiation and propagation of peat fires, J. Eng. Phys. Thermophy., 82, 1235–1243, https://doi.org/10.1007/s10891-010-0293-7, 2009.

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., Bachelet, D., Ciais, P., Forrest, M., Friedlingstein, P., Hickler, T., Kaplan, J. O., Kloster, S., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Melton, J. R., Meyn, A., Sitch, S., Spessa, A., van der Werf, G. R., Voulgarakis, A., and Yue, C.: The status and challenge of global fire modelling, Biogeosciences, 13, 3359–3375, https://doi.org/10.5194/bg-13-3359-2016, 2016.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Sci. Data, 7, 109, https://doi.org/10.1038/s41597-020-0453-3, 2020 (data available at: https://crudata.uea.ac.uk/cru/data/hrg/, last access: June 2021).

Haynes, K. M., Kane, E. S., Potvin, L., Lilleskov, E. A., Kolka, R. K., and Mitchell, C. P.: Gaseous mercury fluxes in peatlands and the potential influence of climate change, Atmos. Environ., 154, 247–259, 2017.

Helbig, M., Waddington, J. M., Alekseychik, P., Amiro, B. D., Aurela, M., Barr, A. G., Black, T. A., Blanken, P. D., Carey, S. K., Chen, J., Chi, J., Desai, A. R., Dunn, A., Euskirchen, E. S., Flanagan, L. B., Forbrich, I., Friborg, T., Grelle, A., Harder, S., Heliasz, M., Humphreys, E. R., Ikawa, H., Isabelle, P.-E., Iwata, H., Jassal, R., Korkiakoski, M., Kurbatova, J., Kutzbach, L., Lindroth, A., Löfvenius, M. O., Lohila, A., Mammarella, I., Marsh, P., Maximov, T., Melton, J. R., Moore, P. A., Nadeau, D. F., Nicholls, E. M., Nilsson, M. B., Ohta, T., Peichl, M., Petrone, R. M., Petrov, R., Prokushkin, A., Quinton, W. L., Reed, D. E., Roulet, N. T., Runkle, B. R. K., Sonnentag, O., Strachan, I. B., Taillardat, P., Tuittila, E.-S., Tuovinen, J.-P., Turner, J., Ueyama, M., Varlagin, A., Wilmking, M., Wofsy, S. C., and Zyrianov, V.: Increasing contribution of peatlands to boreal evapotranspiration in a warming climate, Nat. Clim. Change, 10, 555–560, https://doi.org/10.1038/s41558-020-0763-7, 2020.

Horton, A. J., Virkki, V., Lounela, A., Miettinen, J., Alibakhshi, S., and Kummu, M.: Identifying Key Drivers of Peatland Fires Across Kalimantan's Ex-Mega Rice Project Using Machine Learning, Earth and Space Science, 8, e2021EA001873, https://doi.org/10.1029/2021EA001873, 2021.

Huang, X. and Rein, G.: Upward-and-downward spread of smoldering peat fire, P. Combust. Inst., 37, 4025–4033, https://doi.org/10.1016/j.proci.2018.05.125, 2019.

Hugelius, G., Loisel, J., Chadburn, S., Jackson, R. B., Jones, M., MacDonald, G., Marushchak, M., Olefeldt, D., Packalen, M., Siewert, M. B., Treat, C., Turetsky, M., Voigt, C., and Yu, Z.: Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw, P. Natl. Acad. Sci. USA, 117, 20438–20446, https://doi.org/10.1073/pnas.1916387117, 2020.

Hugelius, G., Loisel, J., Chadburn, S., Jackson, R. B., Jones, M.,MacDonald, G., Marushchak, M., Olefeldt, D., Packalen, M., Siewert, M. B., Treat, C., Turetsky, M., Voigt, C., and Yu, Z.: Maps of northern peatland extent, depth, carbon storage and nitrogen storage. Dataset version 1, Bolin Centre Database [data set], https://doi.org/10.17043/hugelius-2020-peatland-1, 2020.

Hurley, M. J., Gottuk, D. T., Hall, J. R., Harada, K., Kuligowski, E. D., Puchovsky, M., Torero, J., Watts, J. M., and Wieczorek, C. J.: SFPE Handbook of Fire Protection Engineering, Springer, 3510 pp., https://doi.org/10.1007/978-1-4939-2565-0, 2015.

Jain, P., Coogan, S. C., Subramanian, S. G., Crowley, M., Taylor, S., and Flannigan, M. D.: A review of machine learning applications in wildfire science and management, Environ. Rev., 28, 478–505, https://doi.org/10.1139/er-2020-0019, 2020.

Jiang, H. and Nachum, O.: Identifying and correcting label bias in machine learning, in: International Conference on Artificial Intelligence and Statistics, 702–712, 2020.

Jones, B. M., Grosse, G., Arp, C. D., Miller, E., Liu, L., Hayes, D. J., and Larsen, C. F.: Recent Arctic tundra fire initiates widespread thermokarst development, Sci. Rep.-UK, 5, 15865, https://doi.org/10.1038/srep15865, 2015.

Kelly, R., Chipman, M. L., Higuera, P. E., Stefanova, I., Brubaker, L. B., and Hu, F. S.: Recent burning of boreal forests exceeds fire regime limits of the past 10 000 years, P. Natl. Acad. Sci. USA, 110, 13055–13060, https://doi.org/10.1073/pnas.1305069110, 2013.

Kim, J.-S., Kug, J.-S., Jeong, S.-J., Park, H., and Schaepman-Strub, G.: Extensive fires in southeastern Siberian permafrost linked to preceding Arctic Oscillation, Science Advances, 6, eaax3308, https://doi.org/10.1126/sciadv.aax3308, 2020.

Klein Goldewijk, C. G. M.: Anthropogenic land-use estimates for the Holocene; HYDE 3.2, V1, DANS Data Station Archaeology [data set], https://doi.org/10.17026/dans-25g-gez3, 2017.

Klein Goldewijk, K., Beusen, A., Doelman, J., and Stehfest, E.: Anthropogenic land use estimates for the Holocene – HYDE 3.2, Earth Syst. Sci. Data, 9, 927–953, https://doi.org/10.5194/essd-9-927-2017, 2017.

Kohlenberg, A. J., Turetsky, M. R., Thompson, D. K., Branfireun, B. A., and Mitchell, C. P. J.: Controls on boreal peat combustion and resulting emissions of carbon and mercury, Environ. Res. Lett., 13, 035005, https://doi.org/10.1088/1748-9326/aa9ea8, 2018.

Leifeld, J., Müller, M., and Fuhrer, J.: Peatland subsidence and carbon loss from drained temperate fens, Soil Use Manage., 27, 170–176, https://doi.org/10.1111/j.1475-2743.2011.00327.x, 2011.

Li, F., Levis, S., and Ward, D. S.: Quantifying the role of fire in the Earth system – Part 1: Improved global fire modeling in the Community Earth System Model (CESM1), Biogeosciences, 10, 2293–2314, https://doi.org/10.5194/bg-10-2293-2013, 2013.

Li, F., Zhu, Q., Riley, W. J., Zhao, L., Xu, L., Yuan, K., Chen, M., Wu, H., Gui, Z., Gong, J., and Randerson, J. T.: Attention-Fire_v1.0: interpretable machine learning fire model for burned-area predictions over tropics, Geosci. Model Dev., 16, 869–884, https://doi.org/10.5194/gmd-16-869-2023, 2023.

Lin, S., Cheung, Y. K., Xiao, Y., and Huang, X.: Can rain suppress smoldering peat fire?, Sci. Total Environ., 727, 138468, https://doi.org/10.1016/j.scitotenv.2020.138468, 2020.

Liu, J. C., Pereira, G., Uhl, S. A., Bravo, M. A., and Bell, M. L.: A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke, Environ. Res., 136, 120–132, https://doi.org/10.1016/j.envres.2014.10.015, 2015.

Lizundia-Loiola, J., Otón, G., Ramo, R., and Chuvieco, E.: A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data, Remote Sens. Environ., 236, 111493, https://doi.org/10.1016/j.rse.2019.111493, 2020.

Madani, N. and Parazoo, N. C.: Global Monthly GPP from an Improved Light Use Efficiency Model, 1982–2016, ORNL DAAC, https://doi.org/10.3334/ORNLDAAC/1789 [data set], 2020.

Malik, A., Rao, M. R., Puppala, N., Koouri, P., Thota, V. A. K., Liu, Q., Chiao, S., and Gao, J.: Data-Driven Wildfire Risk Prediction in Northern California, Atmosphere, 12, 109, https://doi.org/10.3390/atmos12010109, 2021.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev., 10, 1903–1925, https://doi.org/10.5194/gmd-10-1903-2017, 2017 (data available at: https://www.gleam.eu/, last access: June 2021).

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15, 453–469, https://doi.org/10.5194/hess-15-453-2011, 2011.

Nagare, R. M., Schincariol, R. A., Quinton, W. L., and Hayashi, M.: Effects of freezing on soil temperature, freezing front propagation and moisture redistribution in peat: laboratory investigations, Hydrol. Earth Syst. Sci., 16, 501–515, https://doi.org/10.5194/hess-16-501-2012, 2012.

Novick, K. A., Ficklin, D. L., Stoy, P. C., Williams, C. A., Bohrer, G., Oishi, A. C., Papuga, S. A., Blanken, P. D., Noormets, A., Sulman, B. N., Scott, R. L., Wang, L., and Phillips, R. P.: The increasing importance of atmospheric demand for ecosystem water and carbon fluxes, Nat. Clim. Change, 6, 1023–1027, https://doi.org/10.1038/nclimate3114, 2016.

Ohlemiller, T. J.: Modeling of smoldering combustion propagation, Prog. Energ. Combust., 11, 277–310, https://doi.org/10.1016/0360-1285(85)90004-8, 1985.

Page, S. E. and Hooijer, A.: In the line of fire: The peatlands of Southeast Asia, Philos. T. Roy. Soc. B, 371, 20150176, https://doi.org/10.1098/rstb.2015.0176, 2016.

Pan, Z., Du, H., Ngiam, K. Y., Wang, F., Shum, P., and Feng, M.: A Self-Correcting Deep Learning Approach to Predict Acute Conditions in Critical Care, arXiv:1901.04364 [cs, stat], 2019.

Pinzon, J. E. and Tucker, C. J.: A non-stationary 1981-2012 AVHRR NDVI3g time series, Remote Sens.-Basel, 6, 6929–6960, https://doi.org/10.3390/rs6086929, 2014.

Price, J. S. and Schlotzhauer, S. M.: Importance of shrinkage and compression in determining water storage changes in peat: the case of a mined peatland, Hydrol. Process., 13, 2591–2601, https://doi.org/10.1002/(SICI)1099-1085(199911)13:16<2591::AID-HYP933>3.0.CO;2-E, 1999.

Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., Kaplan, J. O., Li, F., Mangeon, S., Ward, D. S., Yue, C., Arora, V. K., Hickler, T., Kloster, S., Knorr, W., Nieradzik, L., Spessa, A., Folberth, G. A., Sheehan, T., Voulgarakis, A., Kelley, D. I., Prentice, I. C., Sitch, S., Harrison, S., and Arneth, A.: The Fire Modeling Intercomparison Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions, Geosci. Model Dev., 10, 1175–1197, https://doi.org/10.5194/gmd-10-1175-2017, 2017.

Randerson, J. T., Chen, Y., Van Der Werf, G. R., Rogers, B. M., and Morton, D. C.: Global burned area and biomass burning emissions from small fires, J. Geophys. Res.-Biogeo., 117, 1–23, https://doi.org/10.1029/2012JG002128, 2012.

Ranneklev, S. B. and Bååth, E.: Use of Phospholipid Fatty Acids To Detect Previous Self-Heating Events in Stored Peat, Appl. Environ. Microb., 69, 3532–3539, https://doi.org/10.1128/AEM.69.6.3532-3539.2003, 2003.

Reid, C. E., Brauer, M., Johnston, F. H., Jerrett, M., Balmes, J. R., and Elliott, C. T.: Critical review of health impacts of wildfire smoke exposure, Environ. Health Persp., 124, 1334–1343, https://doi.org/10.1289/ehp.1409277, 2016.

Rein, G. and Huang, X.: Smouldering wildfires in peatlands, forests and the arctic: Challenges and perspectives, Current Opinion in Environmental Science & Health, 24, 100296, https://doi.org/10.1016/j.coesh.2021.100296, 2021.

Restuccia, F., Huang, X., and Rein, G.: Self-ignition of natural fuels: Can wildfires of carbon-rich soil start by self-heating?, Fire Safety J., 91, 828–834, https://doi.org/10.1016/j.firesaf.2017.03.052, 2017.

Rosadi, D., Andriyani, W., Arisanty, D., and Agustina, D.: Prediction of Forest Fire Occurrence in Peatlands using Machine Learning Approaches, in: 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 10 December 2020, Yogyakarta, Indonesia, 48–51, https://doi.org/10.1109/ISRITI51436.2020.9315359, 2020.

Rothermel, R. C.: A mathematical model for predicting fire spread in wildland fuels, Research Paper INT-115, US Department of Agriculture, Intermountain Forest and Range Experiment Station, Ogden, UT 84401, 40 pp., 1972.

Roy, D. P., Lewis, P. E., and Justice, C. O.: Burned area mapping using multi-temporal moderate spatial resolution data – a bi-directional reflectance model-based expectation approach, Remote Sens. Environ., 83, 263–286, https://doi.org/10.1016/S0034-4257(02)00077-9, 2002 (data available at: https://e4ftl01.cr.usgs.gov/MOTA/, last access: 20 September 2019).

Rudiyanto, Minasny, B., Setiawan, B. I., Saptomo, S. K., and McBratney, A. B.: Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands, Geoderma, 313, 25–40, https://doi.org/10.1016/j.geoderma.2017.10.018, 2018.

Sayad, Y. O., Mousannif, H., and Al Moatassime, H.: Predictive modeling of wildfires: A new dataset and machine learning approach, Fire Safety J., 104, 130–146, https://doi.org/10.1016/j.firesaf.2019.01.006, 2019.

Scharlemann, J. P., Tanner, E. V., Hiederer, R., and Kapos, V.: Global soil carbon: understanding and managing the largest terrestrial carbon pool, Carbon Manag., 5, 81–91, 2014.

Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., and Bengio, Y.: Toward Causal Representation Learning, Proc. IEEE, 109, 612–634, https://doi.org/10.1109/JPROC.2021.3058954, 2021.

Scott, A. C., Bowman, D. M., Bond, W. J., Pyne, S. J., and Alexander, M. E.: Fire on earth: an introduction, John Wiley & Sons, ISBN-13 978-1119953562, 2013.

Tang, R.: Tangetal2023, Zenodo [code], https://doi.org/10.5281/zenodo.10072144, 2023.

The National Center for Atmospheric Research: Global GIMMS NDVI3g v1 dataset (1981–2015). National Tibetan Plateau/Third Pole Environment Data Center, https://data.tpdc.ac.cn/en/data/9775f2b4-7370-4e5e-a537-3482c9a83d88/ (last access: June 2021), 2018.

Turetsky, M. R., Benscoter, B., Page, S., Rein, G., Van Der Werf, G. R., and Watts, A.: Global vulnerability of peatlands to fire and carbon loss, Nat. Geosci., 8, 11–14, https://doi.org/10.1038/ngeo2325, 2014.

Urbanski, S. P., Hao, W. M., and Baker, S.: Chapter 4 Chemical Composition of Wildland Fire Emissions, in: Developments in Environmental Science, vol. 8, edited by: Bytnerowicz, A., Arbaugh, M. J., Riebau, A. R., and Andersen, C., Elsevier, 79–107, https://doi.org/10.1016/S1474-8177(08)00004-1, 2008.

van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997–2016, Earth Syst. Sci. Data, 9, 697–720, https://doi.org/10.5194/essd-9-697-2017, 2017.

van Dijk, H. and Boekel, P.: Effect of drying and freezing on certain physical properties of peat, Neth. J. Agr. Sci., 13, 248–260, https://doi.org/10.18174/njas.v13i3.17481, 1965.

Vrije Universiteit Amsterdam: Global Fire Emissions Database, Version 4.1 (GFED4.1s), Vrije Universiteit Amsterdam [data set], https://www.geo.vu.nl/~gwerf/GFED/GFED4/ (last access: 13 June 2019), 3 July 2015.

Wang, S. S.-C. and Wang, Y.: Quantifying the effects of environmental factors on wildfire burned area in the south central US using integrated machine learning techniques, Atmos. Chem. Phys., 20, 11065–11087, https://doi.org/10.5194/acp-20-11065-2020, 2020.

Wang, T., Yang, D., Yang, Y., Piao, S., Li, X., Cheng, G., and Fu, B.: Permafrost thawing puts the frozen carbon at risk over the Tibetan Plateau, Science Advances, 6, eaaz3513, https://doi.org/10.1126/sciadv.aaz3513, 2020.

Whittington, P. N. and Price, J. S.: The effects of water table draw-down (as a surrogate for climate change) on the hydrology of a fen peatland, Canada, Hydrol. Process., 20, 3589–3600, https://doi.org/10.1002/hyp.6376, 2006.

Yu, Y., Mao, J., Thornton, P. E., Notaro, M., Wullschleger, S. D., Shi, X., Hoffman, F. M., and Wang, Y.: Quantifying the drivers and predictability of seasonal changes in African fire, Nat. Commun., 11, 2893, https://doi.org/10.1038/s41467-020-16692-w, 2020.

Yuan, H., Restuccia, F., Rein, G., Yuan, H., Restuccia, F., and Rein, G.: Spontaneous ignition of soils: a multi-step reaction scheme to simulate self-heating ignition of smouldering peat fires, Int. J. Wildland Fire, 30, 440–453, https://doi.org/10.1071/WF19128, 2021.

Yuan, W., Zheng, Y., Piao, S., Ciais, P., Lombardozzi, D., Wang, Y., Ryu, Y., Chen, G., Dong, W., Hu, Z., Jain, A. K., Jiang, C., Kato, E., Li, S., Lienert, S., Liu, S., Nabel, J. E. M. S., Qin, Z., Quine, T., Sitch, S., Smith, W. K., Wang, F., Wu, C., Xiao, Z., and Yang, S.: Increased atmospheric vapor pressure deficit reduces global vegetation growth, Science Advances, 5, 1–13, https://doi.org/10.1126/sciadv.aax1396, 2019.

Zhou, W., Chen, W., Zhou, E., Huang, Y., Wei, R., and Zhou, Y.: Prediction of Wildfire-induced Trips of Overhead Transmission Line based on data mining, in: 2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE), 6–10 September 2020, Beijing, China, 1–4, https://doi.org/10.1109/ICHVE49031.2020.9279835, 2020.