TSECfire v1.0: Quantifying Wildfire Drivers and Predictability in Boreal Peatlands Using a Two-Step Error-Correcting Machine Learning Framework

Rongyun Tang[1], Mingzhou Jin[1], Jiafu Mao[2], Daniel Ricciuto[2], Anping Chen[3], Yulong Zhang[1]

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

Correspondence to: 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, a quantitative understanding of natural and anthropogenic influences on the changes in BP fires remains elusive. Here, we quantified the predictability of BP fires and their primary controlling factors from 1997 to 2016 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) non-parametric models outperformed parametric models in predicting fire occurrences, and the machine learning model of Random Forest performed the best with the area under the Receiver Operating Characteristic curve ranging from 0.83 to 0.93 across multiple fire data sets, 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 BP fire prediction and disentangle the primary factors determining BP fires, which are critical for predicting future fire risks under climate change.

1 Introduction

The carbon-rich boreal peatlands (BPs) cover only ~2% of the Earth's surface (Gorham, 1991) but have accumulated ~20%–40% (450 ± 150 PgC) of the global soil carbon, historically playing 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 induce carbon emissions into the atmosphere through increasing decomposition of peat soil organic matter and fire combustions (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 smouldering—that can transition from one to the other at different phases. It is noteworthy that compared to flaming combustion, smouldering combustion is easier to ignite, harder to suppress, more persistent in low temperature and high moisture peat (Huang and Rein, 2019). Besides releasing CO2, smouldering produces more CO, CH4, smokes, 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 smouldering combustion remains poorly understood, despite recent efforts on using experimental, statistical, and computational tools to investigate smouldering 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, smouldering 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 smouldering fires, therefore, our current understanding of BP fires and their predictability are still very limited, hampering the peat fire hazard mitigation and firefighting.

Most studies ascribe the ignition and propagation of flaming fires to the joint impact of heat source, fire-favor climate, fuel, and anthropogenic factors. Flameless smouldering peatland fires are not an exception although upland flaming fires and underground smouldering in BPs are fundamentally different in their chemical and physical aspects (Certini, 2014; Costafreda-Aumedes et al., 2017; Rabin et al., 2017). 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 smouldering fires. Importantly, Yuan et al. (2021) suggested that the smouldering 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 microorganism revealed that peat self-heating reactions (soil respiration and microorganism growth) could happen at temperatures as low as 25–55 ºC (Ranneklev and Bååth, 2003), while temperature could reach 500–700 ºC during smouldering (Hurley et al., 2015). The dramatic changes in micro process of smouldering 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, the use of 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, and 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. Only 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 knowledges and understanding of peat fire combustions, the solution of the primal and inverse problems, and the unavailable large scale peat soil and peat burning characteristic data are still obstacles in building the 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 stirred the application of a broad range of ML algorithms in wildfire science research, such as fire detection, fire weather exploration, fire behaviour prediction, fire impacts 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 be still full of challenges, and further studies are needed to deal with such commission and omission problems and 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 on severely skewed fire data, and a series of sensitivity tests were performed on each of multiple fire data sets to address possible drivers of BP fires. Specific research goals include to (1) examine the performances of multiple ML algorithms on 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 correct predicting errors to improve the ML predicting accuracy that is suppressed by the 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. All data sets were re-gridded to 1° × 1° with a monthly time resolution.

2.1 Response Variables

To evaluate ML framework robustness for difference response variables, five fire data sets 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 products MCD45A1, and 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° × 0.25° (Giglio et al., 2013; Randerson et al., 2012). The FireCCI5.1 BA data set 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 data sets were derived from the MODIS Terra and Aqua satellites 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 2016, this study used the monthly 0.5° × 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) 2m windspeed (WIN) were included as feature inputs. 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° × 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 were 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° × 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) were interpolated and re-gridded into a monthly scale at a spatial resolution of 0.5° × 0.5° (Klein Goldewijk et al., 2017).

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. All data sets were re-gridded to 1° × 1° with a monthly time resolution.

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

3.2 The Two-step Hierarchy Machine Learning Framework

Given our assumption of predominantly smouldering fires within our defined research areas, we introduced an ML framework specifically designed to predict 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 non-linear 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, 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. We also conducted a range of factor-control simulations using a method akin to backward selection to investigate the key contributors to boreal peatland (BP) fire occurrences and to understand the BP-smoldering fire mechanisms. The two-step ML framework is detailed in Figure 1.

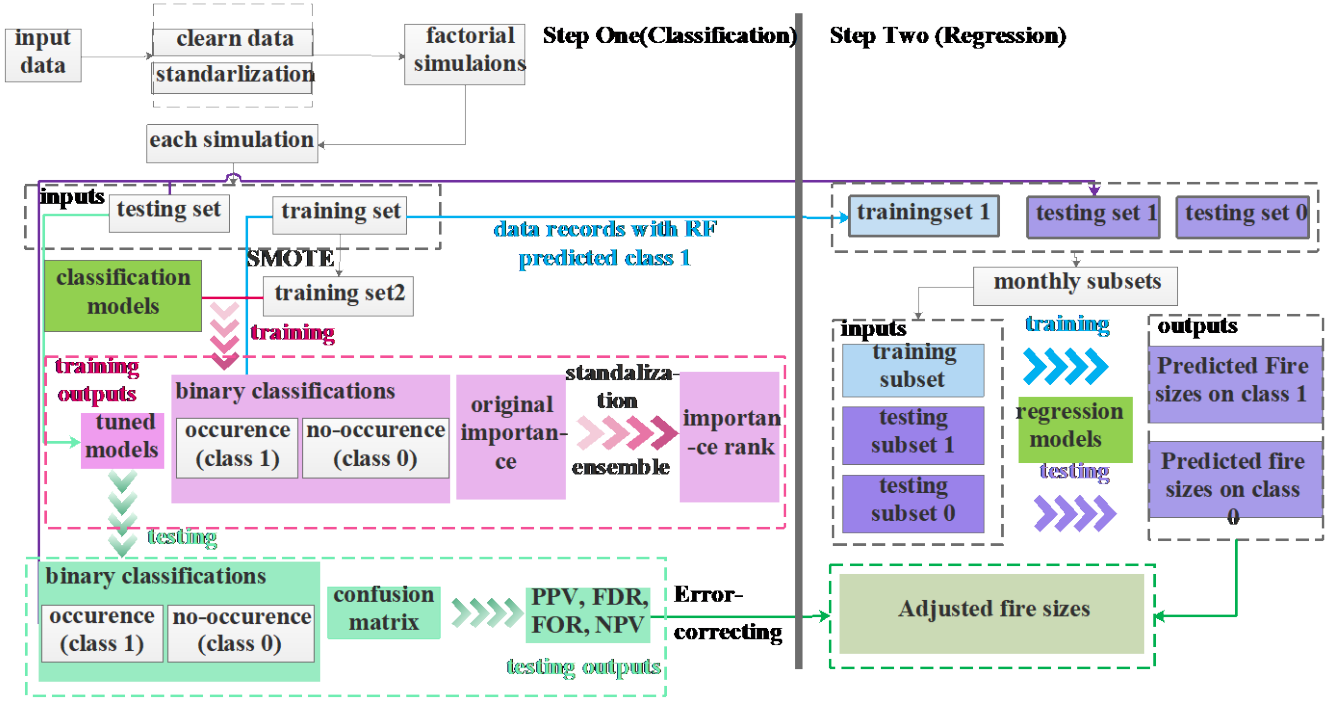


Figure 1. The two-step ML framework, where PPV, FDR, FOR, and NPV stand for positive predictive value, false discovery rate, false omission rate, and negative predictive value, respectively. SMOTE stands for the oversampling algorithm– Synthetic Minority Oversampling Techniques. The error-correcting process is detailed in the Methods part.

We began by pre-processing the data, which encompassed data integration, treatment of missing values, and standardization. Subsequently, the data was divided into 70% for training and 30% for testing. Given the nature of BP fires as predominantly extreme events, there's 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, in our case, fire occurrences. 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's 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's important to note 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 don't 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 the 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, isn't 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, in the initial step, we classified the training dataset based on C emissions/Burned area. Values equal to 0 indicate the no-fire months, represented as class 0. Conversely, values 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/Burned area > 0, or class 1), we employed the regression model R_{mf} to predict the fire impact:

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

Where X_{mf} is the explanatory data in month m with fire; and Y_{mf} is the predictive variable (C emission/burned area) at month m.

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

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

$where\ X_{mn}$ is the explanatory data in month m without fires occurrences.

Every month (m) 's 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 Linear Regressor to Stacking Regressor.

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

$$R^r_{mf}(X_{mf}) = y^r_{mf}, \quad (3)$$

$$R^r_{mn}(X_{mn}) = y^r_{mn} = 0, \quad (4)$$

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}:

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

Possible uncertainties related to fire size predictions based on months-with-fire is that none fires actually occurred, which could be expressed by EP_{mn}:

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

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

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

Possible uncertainties related to fire size predictions on months-without-fire is that fire events did happen in reality, which could be expressed by:

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

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. Four evaluation metrics from classification are:

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

$$False\ Discovery\ rate\ (FDR) = \frac{FP}{TP+FP} \quad , \tag{10}$$

$$False\ omission\ rate\ (FOR) = \frac{False\ Negative\ (FN)}{FN+True\ nagative(TN)} \quad , \tag{11}$$

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

Applying classification evaluation metrics to the actual predictions (P) and potential wrong classification caused predictions (EPs), we could obtain the error-corrected prediction AP'^{r}_{mp} for the record of $(p, m) \in X'$ (in the testing set).

$$AP'^{r}_{mp} = \begin{cases} PPV \times P^{r}_{mf} + FDR \times EP^{r}_{mn}, \ If\ Z'_{mp} = 1 \\ NPV \times P^{r}_{mn} + FOR \times EP^{r}_{mf}, \ If\ Z'_{mp} = 0 \end{cases} \tag{13}$$

Where $(f, n) \in p, and\ Z'_{mp}$ 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), windspeed (WIN), cloud cover percentage (CLD), climate extremes (e.g., FRS and DTR), and anthropogenic activities (e.g., POPD). All the simulation setup is listed in Table 1. During the first round of simulation, we incorporated all features, labeled as 'ALL'. As discussed in Section 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 benefits 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.

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

Our two-step hierarchy ML framework has been designed with multi-datasets and multi-algorithms to validate the framework's predictive performance. Yet, we hadn't 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-square (R2), were documented and are detailed in Table 2.

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 the AUC value of 1 at the training stage and the 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% 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 West Siberia (WS) (Figure S2). 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 the boreal area and possibly implies that WS is more fire-prone than the HBL (Figures S3–S6[a-1], [a-2]). FireCCI, GFED, and MCD64A1 showed good consistencies 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 (Figures S7[a-1], [a-2]). 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 (Figures S11).

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) (Figure S1 and Figure 2). For RF and BAG, the reproducing accuracy rate (i.e., true positive rate and true negative rate) was higher than 90% for the FireCCI data (Figure S1). The inaccurate predictions of KNN, LogReg, SVM, and GNB were significantly influenced by the overestimated fire occurrence (namely false positive) during fire season (April–October), as shown in Figure 2. Without a prescribed underlying function, the nonparametric RF and BAG models exhibited advantages over other ML algorithms on reproducing peatland fire distributions spatially (Figures S3–S6[b-1], [b-2], [c-1], [c-2], [d-1], [d-2]) and temporally (Figure 2). Therefore, the predictions of fire occurrence from the best-performed RF were employed as the basis of fire impact predictions.

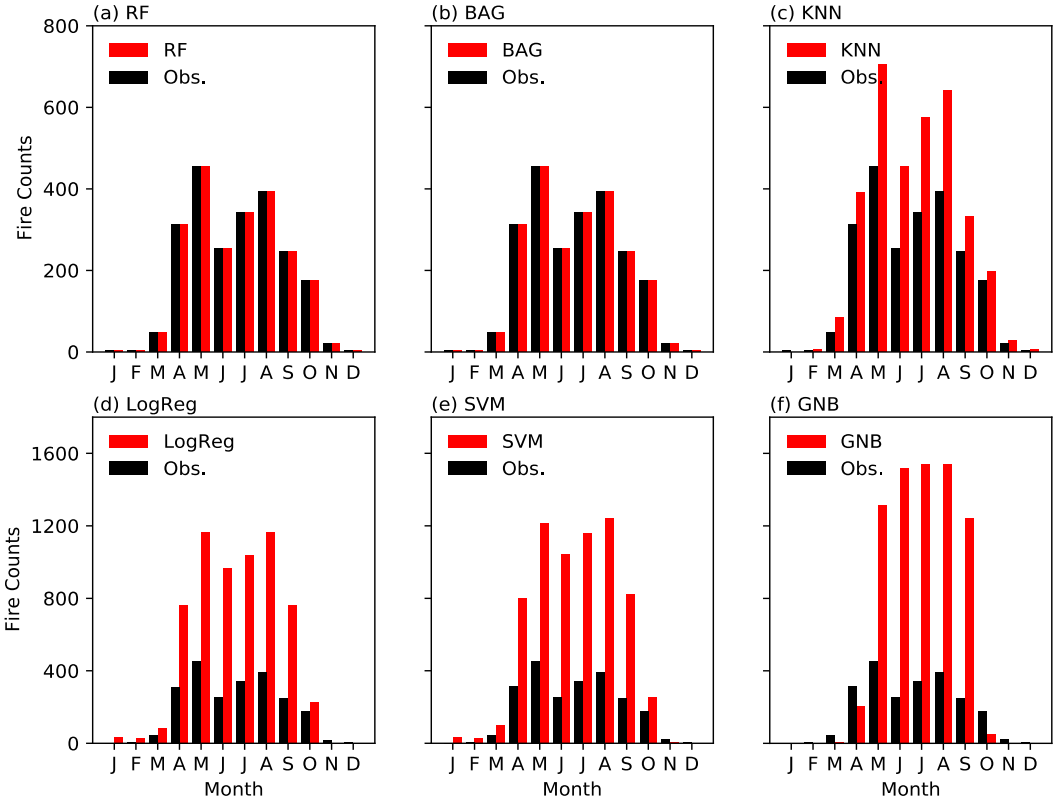


Figure 2. Seasonality of observational and predicted fire counts from the six ML algorithms with the FireCCI BA dataset.

4.2 Fire Impacts Predictability

ML regression models exhibit moderate predictabilities of fire sizes (Figure S22). ML classification at Step One and regression models at Step Two overestimated fire size during fire season (Figures S19–S21) for the monthly aggregated fire impacts. This study developed an error-correcting technique to tackle the error propagation and overestimation during fire reason and achieved satisfying performance (Figure 3, Figures S19–S21).

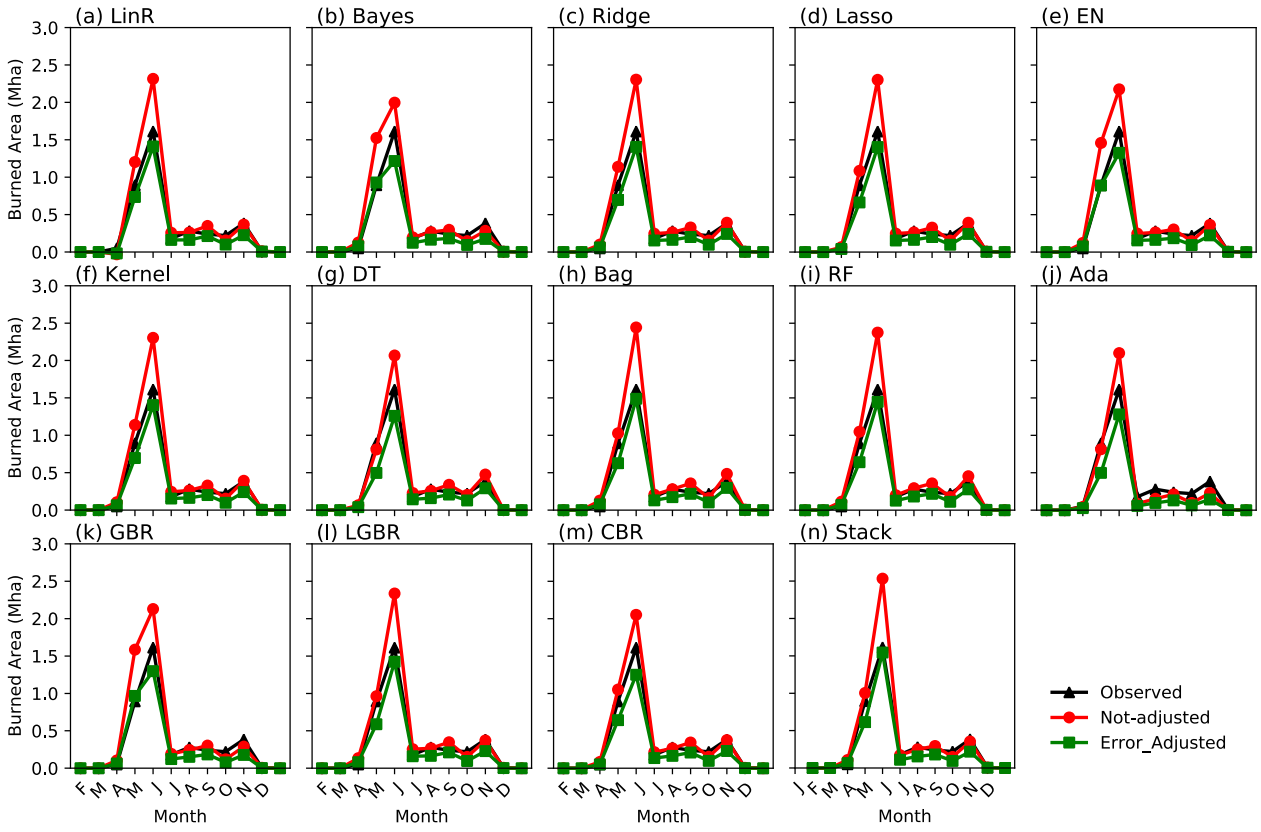


Figure 3. Seasonality of the observed, not-adjusted, and error-adjusted FireCCI BA based on the testing phase from multiple ML regression models: (a) linear, (b) Bayesian linear, (c) ridge, (d) LASSO, (e) elastic net, (f) kernel ridge, (g) decision tree, (h) Bagging, (i) RF, (j) AdaBoost, (k) gradient boosting, (l) light gradient boosting, (m) CatBoost, and (n) stacking.

WS has more fire counts and thus higher carbon emissions than the HBL (Figures S16–18). 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 (Figure S18). The error-correcting technique could slightly lower the overestimation in the HBL (Figure S18) but could greatly lower the overestimation temporally, especially in July (Figure S21). Meanwhile, the underestimation of fire impacts in June remained a common problem for all 14 regression models (Figure 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 (Figures S17[d], [e], [f]) were more accurate than those from FireCCI BA (Figures S16[a], [b], [c]), particularly in the HBL, where the BA is less than 50 km^2 (Figures S16–S18[a-1], [b-1]). Figures S16–S18 [a-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 (Figures S16–S18[a-1], [b-1]). Different BA datasets can also have temporal inconsistencies. FireCCI BA exhibits its fire season from March to May, whereas GFED BA exhibits 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 (Figure S30). Therefore, it is important to determine whether ML is applicable for various datasets.

Overall, the 14 tested regression models were able to well reproduce fire impact magnitudes and seasonality for the FireCCI BA, GFED BA, and GFED carbon (Figures 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 (Figures 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, 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 significantly 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 (Figure 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 (Figure s3[a], [c], [d], [e]). The third sect 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 simulation 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 (Figure 3[i]). Such ranks were also indicated by other simulations in Figures 3[a], [c], [d], [f], and [h]. Thus, this study found that BP fires were significantly affected by temperature, air dryness, frost, and GPP (Figure 3[a]), which collectively account for more than 80% of the predictive interpretability (Figure 3[a]). Moreover, BP fires were not sensitive to PRE, soil moisture, windspeed, and human activities.

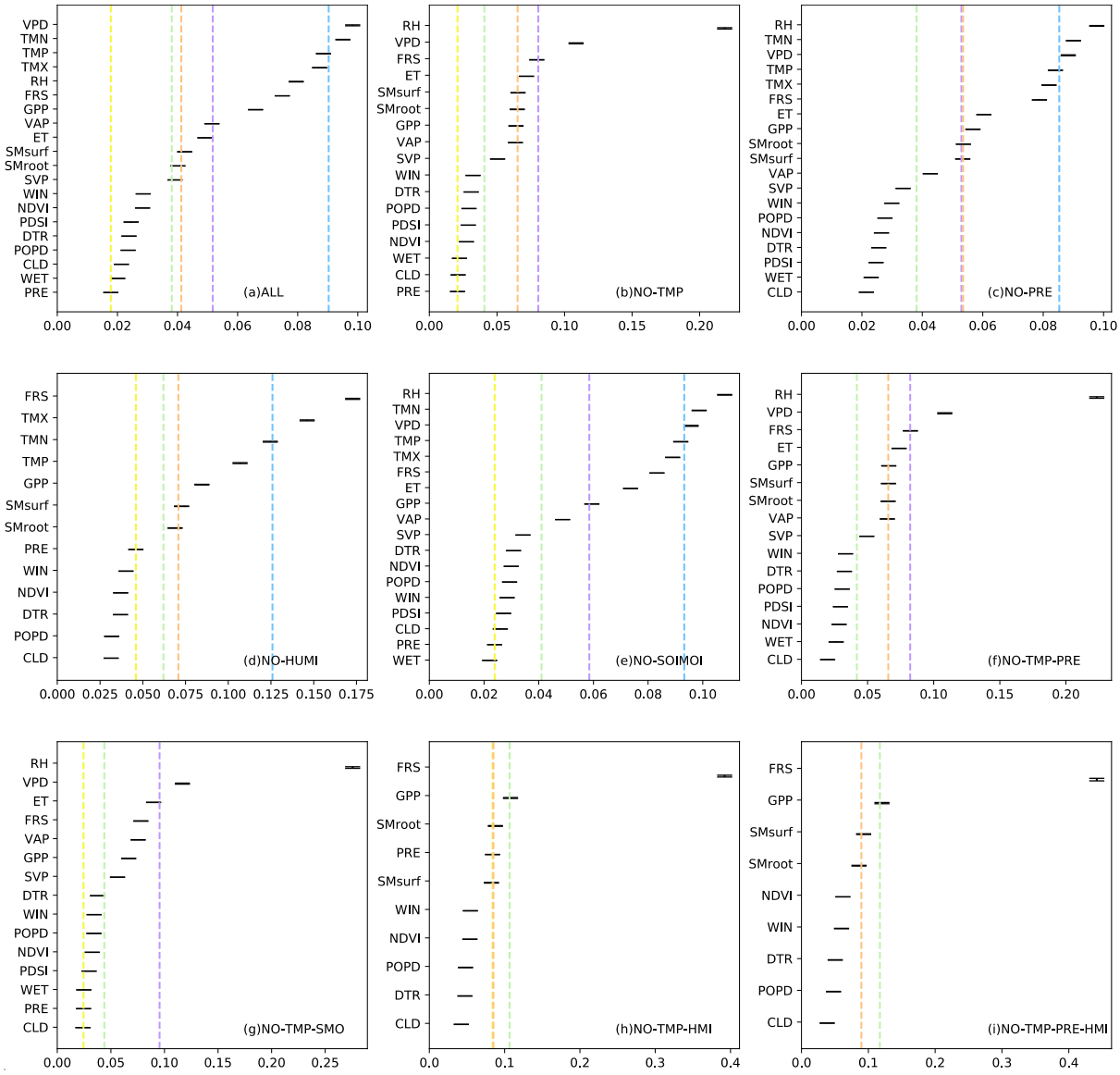


Figure 4. The bar plot stands for the factor importance rank of multiple simulation scenarios using FireCCI BA as the target variable in which the importance was determined by standardized mean and uncertainty range (minimum and maximum) from multiple ML algorithms; the dashed vertical line indicates the group mean importance of temperature (blue), PRE (yellow), air dryness (purple), soil moisture (orange), and other factors (green).

The feature importance ranks were not only validated by FireCCI BA but also by GFED BA, GFED carbon, MCD45A1, and MCD64A1. The rankings from GFED BA and GFED carbon were highly consistent with those from FireCCI (Figures 4, S12, and S13), in which temperature, air dryness, frost, and GPP were more important than PRE, soil moisture, windspeed, 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 moisture was more significant than GPP (Figures S14 and S15).

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 the peat fires than other factors, such as PRE, wind speed, and other natural and human factors. The importance of soil moisture and GPP were both ranked in the middle, 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.

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. Although ML algorithms have been extensively used in wildfire research (e.g., fire spots detection, predictive models development) (Coffield et al., 2019; Jain et al., 2020; Sayad et al., 2019; Wang and Wang, 2020; Yu et al., 2020), few studies explicitly describe what the criteria they would use to choose the ML models. Predicting accuracy may depend on the modeling algorithms and the input data. 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 six times as many records as fire classes) (Figures 1 and 2). 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), and well-identified fire months, suggesting relatively low Type I and Type II errors.

To address the extreme data imbalance, this study used both preprocessing (oversampling) and postprocessing (error correcting) in the two-step ML framework to improve predictability. In Step One, the SMOTE algorithm significantly improved the recall rate by ~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 learning algorithms, our two-step hierarchical framework demonstrated superiority in mitigating the overfitting issue. However, predicting 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 (Figures S19–S21).

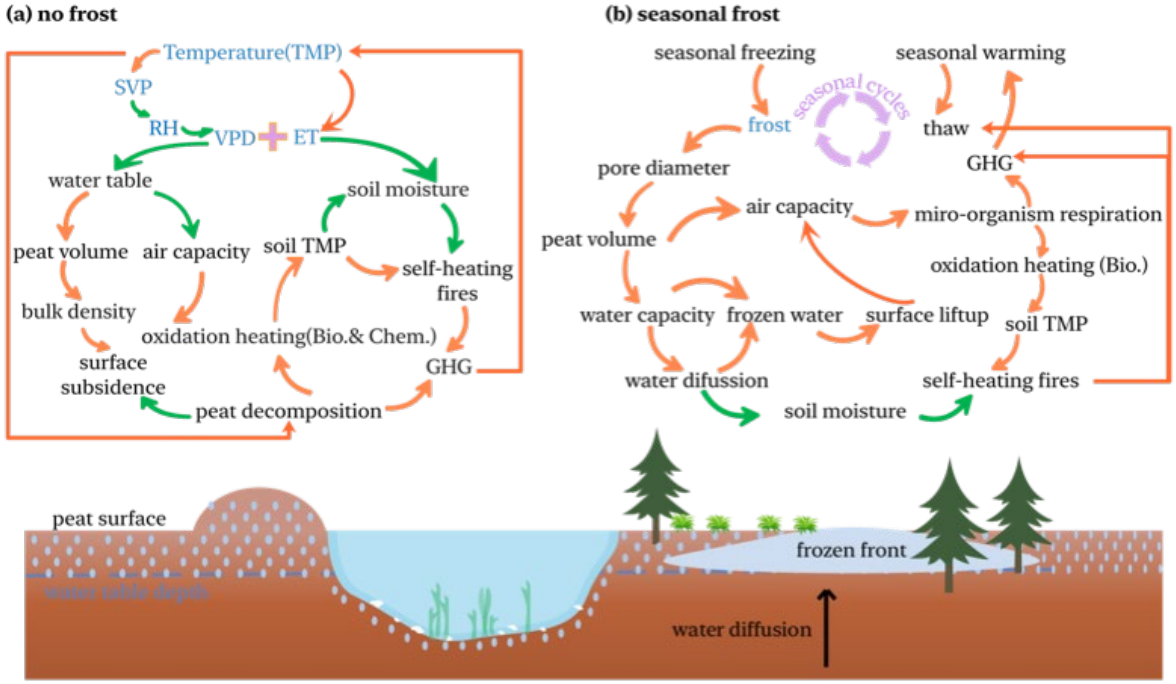
Although our hierarchical machine-learning framework showed somewhat supremacy and robustness, predicting fire sizes is fraught with challenges, stemming from many aspects. In addition to the severely imbalanced data, comprehensive ground-based data in BP regions is sparse, making satellite validation difficult, especially given that the smouldering nature of these fires often eludes satellite detection. Thus, it indicates that the satellite-based data sets are not perfect when applying them in studying broad-scale smouldering fires, especially in terms of carbon emissions due to its 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've incorporated a great number of factors believed to influence smouldering 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 smouldering fires. By identifying the most influential among these, we sought to align them with existing scientific theories to understand their potential roles in driving smouldering fires but not to quantify all the causalities due to the absence of measurement on 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, smouldering 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 the 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 (Figure 4 and Figure S12-

13). 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.



435 **Figure 5 Processes in which environmental factors participate for self-heating peatland fires. ML-identified primary factors are marked in blue; green arrows indicate negative correlation between the connected two factors, and orange arrows indicate positive correlation between the connected two factors.**

The 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 (Figure 5[a]), rising
440 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 year 2000, leading to a sharp rise in VPD on a global scale (Yuan et al., 2019). Such a warming-induced increase of 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
445 the soil moisture (Helbig et al., 2020; Novick et al., 2016), the water table turns out to be the water supplier in response to the rising VPD, which consequently results in the decrease of 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), and inducing surface subsidence with a concomitant decrease of bulk density and an increase of peat oxidations and decomposition (Leifeld et al.,
450 2011; Whittington and Price, 2006). These changes ultimately lead to more carbon being released into the atmosphere and

the formation of dryer and more flammable peat soil (Figure 5[a]). 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) (Figure 5[b]).

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 frosts (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 spreads, though FRS has been understudied in previous studies. Dixon et al. (2018) revealed that the seasonal frost layer alters Spring water balance, induces drier Spring, and enhances risks of deep smouldering. 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 dryer, 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 micro-organisms 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 (Figure 5[b]).

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 (Figures 5[a] and 5[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 smouldering peat fire as a single product. Fires detected by satellites could be a mixture of peatland surface flaming fires and smouldering fires because the detected radiant signature of smouldering is much weaker than that of flaming fires (Rein and Huang, 2021). In addition, for peat fire C emissions, it has been estimated largely by multiplying detected burned area by a range of parameters, such as average burning depth, combustion completeness, emission factors of major

carbon species. Those estimated parameters may induce large uncertainties due to the limited ability of optical satellites to detect underground smouldering 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 on peat drying and smouldering (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%. Temperature and air dryness were identified 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.

Code and data availability: The model code and data that support the findings of this study are archived on Zenodo (https://zenodo.org/record/7754018#.ZBi62uyZPK0) under the GNU General Public License V2.0 or later. Data used for this model are all publicly available. Detailed data sources are listed in the Supplementary Information.

Autor Contributions: MJ, JM, and RT conceived the research ideas. RT wrote the initial draft of the manuscript and performed the modelling and analysis. All authors contributed to preparation of the paper, edit the paper, draft revision, and provide scientific suggestions.

Competing interests: The authors declare that they have no conflict of interest.

Acknowledgements. This work was supported by the Terrestrial Ecosystem Science Scientific Focus Area (TES SFA) project and the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computing Scientific Focus Area (RUBISCO SFA) project funded through the Earth and Environmental Systems Sciences Division of the Biological and Environmental Research Office in the Office of Science of the US Department of Energy (DOE). Oak Ridge National Laboratory is supported by the Office of Science of the DOE under Contract DE-AC05-00OR22725. This work was also supported by Institute for a Secure & Sustainable Environment (ISSE) from University of Tennessee at Knoxville. We

acknowledge the support from the high performance and scientific computing platform of ISAAC hosted by the University of Tennessee.

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, 2021.

Bedia, J., Herrera, S., and Gutiérrez, J. M.: Assessing the predictability of fire occurrence and area burned across phytoclimatic regions in Spain, Natural Hazards and Earth System Sciences, 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, Journal of Geophysical Research: Atmospheres, 121, 4468–4486, https://doi.org/10.1002/2015JD024546, 2016.

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 Ecology, 11, 106–118, https://doi.org/10.4996/fireecology.1101106, 2015.

Certini, R. by G.: Fire on Earth: an Introduction, International Journal of Wildland Fire, 23, 897, https://doi.org/10.1071/wfv23n6_br, 2014.

Che Azmi, N. A., Mohd Apandi, N., and A. Rashid, A. S.: 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, International Journal of 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.

Dijk, H. van and Boekel, P.: Effect of drying and freezing on certain physical properties of peat., Netherlands Journal of Agricultural Science, 13, 248–260, https://doi.org/10.18174/njas.v13i3.17481, 1965.

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, Hydrology and Earth System Sciences Discussions, 1–30, https://doi.org/10.5194/hess-2017-678, 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, Ecological Applications, 15, 1317–1330, https://doi.org/10.1890/04-0739, 2005.

Farquad, M. A. H. and Bose, I.: Preprocessing unbalanced data using support vector machine, Decision Support Systems, 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., 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., P Harrison, S., 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, 7, 1997.

French, N. H. F., Goovaerts, P., and Kasischke, E. S.: Uncertainty in estimating carbon emissions from boreal forest fires, Journal of Geophysical Research D: Atmospheres, 109, 1–12, https://doi.org/10.1029/2003JD003635, 2004.

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), Journal of Geophysical Research: Biogeosciences, 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 Sensing of Environment, 217, 72–85, https://doi.org/10.1016/j.rse.2018.08.005, 2018.

Gorham, E.: Northern Peatlands: Role in the Carbon Cycle and Probable Responses to Climatic Warming, Ecological Applications, 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., Colin Prentice, I., 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.

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, Atmospheric Environment, 154, 247–259, 2017.

585 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,
590 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. Chang., 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,
595 https://doi.org/10.1029/2021EA001873, 2021.

Huang, X. and Rein, G.: Upward-and-downward spread of smoldering peat fire, Proceedings of the Combustion Institute, 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
600 permafrost thaw, PNAS, 117, 20438–20446, https://doi.org/10.1073/pnas.1916387117, 2020.

Hurley, M. J., Gottuk, D. T., Jr, J. R. H., Harada, K., Kuligowski, E. D., Puchovsky, M., Torero, J.́ L., Jr, J. M. W., and WIECZOREK, C. J.: SFPE Handbook of Fire Protection Engineering, Springer, 3510 pp., 2015.

Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S. W., and Flannigan, M. D.: A review of machine learning applications in wildfire science and management, Environmental Reviews, https://doi.org/10.1139/er-2020-0019, 2020.

605 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, 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, PNAS, 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
610 linked to preceding Arctic Oscillation, Science Advances, 6, eaax3308, https://doi.org/10.1126/sciadv.aax3308, 2020.

Klein Goldewijk, K., Beusen, A., Doelman, J., and Stehfest, E.: Anthropogenic land use estimates for the Holocene – HYDE 3.2, Earth System Science 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, Environmental Research Letters, 13, https://doi.org/10.1088/1748-
615 9326/aa9ea8, 2018.

Leifeld, J., Müller, M., and Fuhrer, J.: Peatland subsidence and carbon loss from drained temperate fens, Soil Use and Management, 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.

Lin, S., Cheung, Y. K., Xiao, Y., and Huang, X.: Can rain suppress smoldering peat fire?, Science of The Total Environment, 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, Environmental Research, 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 Sensing of Environment, 236, 111493, https://doi.org/10.1016/j.rse.2019.111493, 2020.

Loisel, J., Gallego-Sala, A. V., Amesbury, M. J., Magnan, G., Anshari, G., Beilman, D. W., Benavides, J. C., Blewett, J., Camill, P., Charman, D. J., Chawchai, S., Hedgpeth, A., Kleinen, T., Korhola, A., Large, D., Mansilla, C. A., Müller, J., van Bellen, S., West, J. B., Yu, Z., Bubier, J. L., Garneau, M., Moore, T., Sannel, A. B. K., Page, S., Väliranta, M., Bechtold, M., Brovkin, V., Cole, L. E. S., Chanton, J. P., Christensen, T. R., Davies, M. A., De Vleeschouwer, F., Finkelstein, S. A., Frolking, S., Gałka, M., Gandois, L., Girkin, N., Harris, L. I., Heinemeyer, A., Hoyt, A. M., Jones, M. C., Joos, F., Juutinen, S., Kaiser, K., Lacourse, T., Lamentowicz, M., Larmola, T., Leifeld, J., Lohila, A., Milner, A. M., Minkkinen, K., Moss, P., Naafs, B. D. A., Nichols, J., O'Donnell, J., Payne, R., Philben, M., Piilo, S., Quillet, A., Ratnayake, A. S., Roland, T. P., Sjögersten, S., Sonnentag, O., Swindles, G. T., Swinnen, W., Talbot, J., Treat, C., Valach, A. C., and Wu, J.: Expert assessment of future vulnerability of the global peatland carbon sink, Nat. Clim. Chang., 11, 70–77, https://doi.org/10.1038/s41558-020-00944-0, 2021.

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, 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, Geoscientific Model Development, 10, 1903–1925, https://doi.org/10.5194/gmd-10-1903-2017, 2017.

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, Hydrology and Earth System Sciences, 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, Hydrology and Earth System Sciences, 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, Nature Clim Change, 6, 1023–1027, https://doi.org/10.1038/nclimate3114, 2016.

Ohlemiller, T. J.: Modeling of smoldering combustion propagation, Progress in Energy and Combustion Science, 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, Philosophical Transactions of the Royal Society B: Biological Sciences, 371, https://doi.org/10.1098/rstb.2015.0176, 2016.

Pinzon, J. E. and Tucker, C. J.: A non-stationary 1981-2012 AVHRR NDVI3g time series, Remote Sensing, 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, Hydrological Processes, 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., Colin Prentice, I., Sitch, S., Harrison, S., and Arneth, A.: The Fire Modeling Intercomparison Project (FireMIP), phase 1: Experimental and analytical protocols with detailed model descriptions, Geoscientific Model Development, 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, Journal of Geophysical Research G: Biogeosciences, 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, Applied and Environmental Microbiology, 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, Environmental Health Perspectives, 124, 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 Journal, 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), 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 48–51, https://doi.org/10.1109/ISRITI51436.2020.9315359, 2020.

Rothermel, R. C.: A mathematical model for predicting fire spread in wildland fuels, Res. Pap. INT-115. Ogden, UT: US Department of Agriculture, Intermountain Forest and Range Experiment Station. 40 p., 115, 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 Sensing of Environment, 83, 263–286, https://doi.org/10.1016/S0034-4257(02)00077-9, 2002.

Sayad, Y. O., Mousannif, H., and Al Moatassime, H.: Predictive modeling of wildfires: A new dataset and machine learning approach, Fire Safety Journal, 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 Management, 5, 81–91, 2014.

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, Nature Geoscience, 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.

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, Atmospheric Chemistry and Physics, 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, Nature Communications, 11, https://doi.org/10.1038/s41467-020-16692-w, 2020.

Yu, Z., Loisel, J., Brosseau, D. P., Beilman, D. W., and Hunt, S. J.: Global peatland dynamics since the Last Glacial Maximum, Geophysical Research Letters, 37, 1–5, https://doi.org/10.1029/2010GL043584, 2010.

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), 2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE), 1–4, https://doi.org/10.1109/ICHVE49031.2020.9279835, 2020.