Science of Remote Sensing 7 (2023) 100088

Contents lists available at ScienceDirect

Science of Remote Sensing

journal homepage: www.sciencedirect.com/journal/science-of-remote-sensing

Global fire modelling and control attributions based on the ensemble machine learning and satellite observations

Yulong Zhang [a,b,c,*], Jiafu Mao [b,**], Daniel M. Ricciuto [b], Mingzhou Jin [a], Yan Yu [d], Xiaoying Shi [b], Stan Wullschleger [b], Rongyun Tang [a], Jicheng Liu [e,f]

[a] Institute for a Secure & Sustainable Environment, University of Tennessee, Knoxville, TN, USA
[b] Environmental Sciences Division and Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[c] Eastern Forest Environmental Threat Assessment Center, Southern Research Station, USDA Forest Service, Research Triangle Park, NC, USA
[d] Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, China
[e] ESSIC/CISESS, University of Maryland, College Park, MD, USA
[f] NOAA NESDIS Center for Satellite Applications and Research, College Park, MD, USA

ARTICLE INFO

Keywords:
Contemporary fire dynamics
Global fire modeling
Fire attributions
Ensemble machine learning
Remote sensing
Climate-vegetation-human controls

ABSTRACT

Contemporary fire dynamics is one of the most complex and least understood land surface phenomena. Global fire controls related to climate, vegetation, and anthropogenic activity are usually intertwined, and difficult to disentangle in a quantitative way. Here, we leveraged an ensemble of five machine learning (ML) models and multiple satellite-based observations to conduct global fire modeling for three fire metrics (burned area, fire number, and fire size), and quantified driving mechanisms underlying annual fire changes in a spatially resolved manner for the period 2003–2019. Ensemble learning is a meta-approach that combines multiple ML predictions to improve accuracy, robustness, and generalization performance. We found that the optimized ensemble ML well reproduced annual dynamics of global burned area ($R^2 = 0.90$, $P < 0.001$), total fire numbers ($R^2 = 0.86$, $P < 0.001$), and averaged fire size ($R^2 = 0.70$, $P < 0.001$). Additionally, the ensemble ML captured key spatial patterns of multi-year mean magnitudes, annual variabilities, anomalies, and trends for different fire metrics. Our ML-based fire attributions further highlighted the dominant role of enhanced anthropogenic activity in reducing global burned area (−1.9 Mha/yr, $P < 0.01$), followed by climate control (−1.3 Mha/yr, $P < 0.01$) and insignificant positive vegetation control (0.4 Mha/yr, $P = 0.60$). Spatially, climate dominated a much larger burned area (53.7%) than human (23.4%) or vegetation control (22.9%); however, the counteracting effects from regional wetting and drying trends weakened the net climate impacts on global burned area. The fire number and fire size exhibited similar spatial control patterns with burned area; globally, however, fire number tended to be more affected by climate while fire size more influenced by human activities. Overall, our study confirmed the feasibility and efficiency of ensemble ML in global fire modeling and subsequent control attributions, providing a better understanding of contemporary fire regimes and contributing to robust fire projections in a changing environment.

1. Introduction

Global wildfire is an ubiquitous component of the Earth system (Archibald et al., 2013; Bowman et al., 2009; Jones et al., 2022). Each year, wildfires burn an average of 450 million hectares of land vegetation (Lizundia-Loiola et al., 2020), a size equivalent to 1.5 times of total area of India. Wildfire, as a major disturbance, has the potential to rapidly change ecosystem structures and functions (Lasslop et al., 2020; McLauchlan et al., 2020), or even permanently reshape ecosystem successional trajectories and biodiversity levels (He et al., 2019; Kelly et al., 2020). As an immediate response, wildfires release significant levels of air chemical pollutants (Urbanski et al., 2008), aerosols (May et al., 2014), and greenhouse gases (Ross et al., 2013), thus affecting regional air quality, threatening public health, and even changing short-term and long-term climate (Andela et al., 2017). It was estimated that global fire-induced carbon emissions were around 2.2 Pg C yr^{-1} (Van Der Werf

* Corresponding author. Institute for a Secure & Sustainable Environment, University of Tennessee, Knoxville, TN, USA.
** Corresponding author.
E-mail addresses: ylzhang.planet@gmail.com (Y. Zhang), maoj@ornl.gov (J. Mao).

https://doi.org/10.1016/j.srs.2023.100088
Received 30 November 2022; Received in revised form 25 April 2023; Accepted 25 April 2023
Available online 26 April 2023

et al., 2017), contributing to about 23% of total fossil fuel emissions (Friedlingstein et al., 2022), thus potentially exacerbating climate warming (Lasslop et al., 2019).

Global fire regimes that describe distributions and characteristics of fire size, frequency, intensity, and seasonality are heterogeneous due to the interplay of fuel resources, fire weather conditions, and effective natural or anthropogenic ignitions (Archibald et al., 2013; Krawchuk and Moritz, 2011). Broadly, high frequency fires tend to occur in areas of intermediate vegetation productivity (e.g., tropical savannah) under conditions of abundant fuel loading, dry weather and frequent ignitions (Krawchuk and Moritz, 2011); whereas low or moderate frequency fires in high-biomass regions are generally limited either by moisture conditions, e.g., for rain forests (Cochrane, 2003) or by ignitions for boreal and temperate forests (Veraverbeke et al., 2017). Despite conducive dry condition, wildfires in arid and semi-arid regions are often limited by fuel availability (Balch et al., 2013). Anthropogenic activities (e.g., deforestation, cropland expansion, fire management) also play a key role in intensifying or suppressing fires (Andela et al., 2017; Bowman et al., 2011). Understanding contemporary fire dynamics and their key controls linked to climate, vegetation and human activity is critical for detecting and projecting fire-regime trajectories (Kelley et al., 2019), thus better serving future climate projection and mitigation (Bowman et al., 2020).

Satellite measurements provide an unrivaled capability for monitoring global fire dynamics (Giglio et al., 2013, 2018). Recent research based on multiple satellite datasets revealed a 25% decline in global burned area between 1998 and 2015, mostly in African tropical savannahs and Asia and Australia semi-arid grasslands (Andela et al., 2017; Forkel et al., 2019). However, other regions such as Western United States (Balch et al., 2018; Williams et al., 2019) and boreal Canada (Hanes et al., 2019) showed considerable increases in total burned area. Such global heterogeneous fire changes are expected to be regulated by a series of controls from climate, vegetation, and human (Bowman et al., 2011; Krawchuk and Moritz, 2011). However, satellite observations cannot directly quantify the effects of these controls, which can mainly be distinguished through wildfire models (Hantson et al., 2016; Kelley et al., 2019).

Currently, two types of wildfire models are widely used for fire control attribution: process-based fire-enabled dynamic global vegetation models (DGVMs) and data-driven statistical models. DGVMs account for the detailed physical processes of fire formation, including ignition, spread, duration, suppression, and extinction (Hantson et al., 2016; Li et al., 2012), and thus can be used to simulate and assess fire dynamics at different spatial and temporal scales. However, benchmarking evaluation indicated that the state-of-the-art DGVMs from the Fire Modeling Intercomparison Project (FireMIP) are largely incapable of capturing recent inter-annual variation and declining trend in global burned area, despite being able to generally reproduce the spatial pattern and seasonality in burned area (Andela et al., 2017; Hantson et al., 2020). This modeling deficiency in wildfire simulations may suggest incomplete fire functional representation, inaccurate model inputs or biased model parameters for these DGVMs (Hantson et al., 2020; Riley and Thompson, 2016; Zhu et al., 2022).

Data-driven statistical models build empirical fire relationships with key drivers based on fire observations, and are arguably better at reproducing observed fire dynamics albeit with highly simplified mechanistic representations of fire processes (Andela et al., 2017; Forkel et al., 2017; Kelley et al., 2019). For example, Forkel et al. (2017) developed a burned area model of SOFIA (Satellite Observations to predict FIre Activity), which accounted for linear effects from climatic, environmental, and socioeconomic controls on fire, and showed better performance than the process-based DGVM of JSBACH-SPITFIRE. Similarly, Kelley et al. (2019) refined the linear combination framework by including fire controls from fuel continuity, fuel moisture, potential ignition, and anthropogenic suppression. This statistical model reproduced the spatial extent of annual burning and associated trends well and was further used to explore bioclimatic and human controls on global fire regimes.

Despite the feasibility of data-driven statistical models, a modeling structure based on assumptions of linear or fixed non-linear forms (e.g., logistic model) may violate the dynamic, non-linear characterize of fire controls from climate, vegetation, and human activity (Bowman et al., 2011; Hantson et al., 2016). Machine learning (ML)-based fire modeling with the benefit of coping with non-linear fitting might overcome this shortage if well guided by expert knowledge (Aldersley et al., 2011; Jain et al., 2020; Yu et al., 2022; Zhu et al., 2022). Based on a deep neural network, Zhu et al. (2022) developed a surrogate model for a global Earth system model and significantly improved the accuracy of burned area prediction when compared to the original process-based fire model. Yu et al. (2022) corrected future fire carbon emissions simulated by 13 Earth system models based on historical observations and three ML models and highlighted the elevated global socioeconomic risks from wildfire. Although these studies demonstrated the promise of ML-based fire modeling, only limited fire metrics (mostly burned area) and ML approaches were involved. The ensemble learning is a general meta-approach to combine multiple ML predictions (Dietterich, 2000) and has advantages of improved prediction accuracy and robustness, reduced overfitting, and improved model stability (Dong et al., 2020; Sagi and Rokach, 2018). However, the ensemble learning also has some limitations, such as increased computational complexity, reduced interpretability, and increased model complexity (Sagi and Rokach, 2018). Previous fire modeling studies based on the ensemble learning were mainly conducted at local scales (Jain et al., 2020; Tuyen et al., 2021). The feasibility of the ensemble learning on global fire modeling has not been well tested and the relevant global constraints on contemporary fire dynamics, particularly at spatial scales, remained understudied.

ML models are commonly regarded as a “black box”, capable of representing the behavior of complicated systems but not straightforwardly interpretable by humans (Rudin, 2019). The lack of interpretability makes ML-based fire attribution a challenging task (Jain et al., 2020; Roscher et al., 2020). One commonly used approach is the permutation feature importance, which measures the change degree of a model score (e.g., R^2, RMSE) when a single feature is randomly shuffled or removed (Breiman, 2001). For example, Aldersley et al. (2011) indicated the major roles of high temperature, intermediate annual rainfall, and prolonged dry periods in determining global burned area frequency based on the feature ranking importance in a random forest. Using a spatially resolved permutation approach, (Forkel et al., 2019) further highlighted strong fire controls from vegetation properties and socio-economic variables (e.g., human population density) that are poorly represented in DGVMs. However, fire attribution based on permutation importance only reflects the generic operating rules of a specific ML model, in most cases, without the local level information (e.g., Aldersley et al., 2011), and is unable to quantify the intrinsic predictive value (both sign and magnitude) given the change of a feature (e.g., Forkel et al., 2019). Recently, some localized explainability techniques within ML models have been developed to overcome this shortage, such as Shapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017) and local interpretable model-agnostic explanations (LIME) (Ribeiro et al., 2016). However, these new methods are usually kernel-based and not effectivity suitable for explaining the ensemble of multiple ML models.

To address the above-identified knowledge gaps, in this study, we performed global fire modeling on three fire metrics, including burned area, fire number and fire size, using an ensemble of five advanced MLs and multi-source global observations, and further explored global fire annual controls related to climate, vegetation, and human changes from 2003 to 2019 through the ML-based factorial simulations. Our study is expected to enhance our understanding of contemporary fire dynamics at multiple dimensions, and further to contribute to robust fire projections in a warming climate.

2. Materials and methods

2.1. Global datasets

Global-scale fire predictors used in this study include eight climate variables (reflecting energy, heat, atmospheric and soil dryness), five vegetation variables (relating to fuel amount and vegetation composition), four human variables (relating to social, economic, and agricultural activities), and two terrain variables (including elevation and slope) from 2003 to 2019. These data are either compiled from ground observations, or derived from remote sensing measurements, which are listed in Table S1. The fire predictors here were identified based on previous global fire modeling studies (Forkel et al., 2017; Kelley et al., 2019; Yu et al., 2022) and assumed to provide multi-dimensional information on fire controls, which are related to fire weather, fuel dryness, fuel loading, natural or anthropogenic ignitions and human suppression (Bowman et al., 2011; Kelly et al., 2020).

To avoid the carry-over effect from the instantaneous post-fire biomass change (e.g., leaf area index [LAI] decreases after fire), we chose the past-year averaged LAI to model current year fire metrics rather than using the current year LAI. To further account for the effect of fuel accumulation in previous years especially in arid/semi-arid areas due to rainfall (Abatzoglou et al., 2018; Tang et al., 2021), we also included the past two-year LAI as vegetation input. Given the lightning data used in our study is seasonal climatology data with no interannual dynamics, it was made static during each simulation. Since the global population density is five-year interval data, their temporal gaps were linearly interpolated. Global GDP data is only available until 2015 (Kummu et al., 2018), and we treated GDP after 2015 as constant 2015 values.

Global fire data (or target data) used in this study included annual burned area, fire number and fire size, which are from satellite-based ESA-CCI51; this fire product combines MODIS near-infrared band and active fire information from thermal channels through a hybrid approach (Lizundia-Loiola et al., 2020). Compared to other global fire products (e.g., GFEDv4s and FireCCILT11), ESA-CCI51 has better data consistency since 2000 and is more sensitive to detecting smaller burned patches due to the highest original spatial resolution (i.e., 250 m) (Lizundia-Loiola et al., 2020). All the global data were aggregated to a standard grid scale of 0.25x0.25° for the subsequent ML fire modeling. For the original continuous variables (e.g., LAI), we conducted the aggregation based on the spatial averaged method. For those concrete variables (i.e., 250 m land cover types from ESA-CCI51), we calculated their area fraction at the 0.25-degree pixel using the area-weighted method (i.e., actual area/pixel area). All those pre-processed global products are used to feed the machine learning-based global fire modeling and fire control attributions mentioned below.

2.2. Machine-learning modeling framework

We developed global fire models for three fire metrics (burned area, fire size and fire number) based on an ensemble learning approach, which includes five state-of-the-art machine learning (ML) models. Ensemble learning is a widely used meta-approach in machine learning by combining predictions from multiple models (Caruana et al., 2004; Sagi and Rokach, 2018), and could, theoretically and practically, obtain better predictive performance than any of its individual members if the base models are accurate and diverse (Dietterich, 2000). Here, we incorporated five popular MLs, including Random Forest, CatBoost, XgBoost, LightGBM and Neural Network. Among these MLs, Random Forest adopts a bootstrap method called bagging to combine randomly generated trees to make the target prediction (Breiman, 2001); CatBoost, XgBoost, and LightGBM use gradient boosting method to iteratively fit a sequence of decision trees, but substantially differ in algorithm details (T. Chen and Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018). For example, CatBoost adopts the symmetric trees, or balanced trees while both XGBoost and LightGBM are asymmetric trees. In the process of constructing trees, LightGBM grows leaf-wise (horizontally) while XGBoost grows level-wise (vertically). Neural Network (also known as Multi-Layer Perceptron) is a deep ML inspired by biological neural networks (Abiodun et al., 2018). Overall, selected MLs with different model structures and parameters are expected to have broad capacity to potentially capture complex interrelations between fire metrics and their relevant drivers.

To build fire models for three fire metrics based on the ensemble ML, we included 19 factors related to climate, vegetation, human and terrain as model inputs (Table S1). For each fire metric, we built a global model based on all global samples and 14 regional models based on regional samples from 14 GFED fire zones (Fig. 1), and then evaluated their model performances. To construct the ensemble ML for each fire zone, all five single MLs mentioned above would be optimized firstly based on the grid search method and their model predictions are then combined to yield a final "ensemble" value according to their derived weights:

$$\hat{y} = \sum_{j=1}^{n} \left(w_j * p_{j,x} \right) \tag{1}$$

Where $\hat{y}$ is the estimated ensemble value; n is the number of ML models involved (n <= 5 for this study); $p_{j,x}$ is the predicted output value for the jth ML model based on the input of x; w_j is the weight of the jth model prediction. The weights w_j are non-negative and sum to 1, reflecting the relative importance of each base model in the ensemble. The optimal weights can be learned by minimizing a loss function that measures the discrepancy between the ensemble's predictions and the true values of the target variable. Here we adopted the root mean squared error (RMSE) as the loss function:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{2}$$

Where n is the sample size; $\hat{y}_i$ and y_i are respectively ith estimated ensemble value (seen in Eq. (1)) and observed value.

There are several approaches to calculate the weight of the base model for the ensemble, such as averaging, bagging, boosting, and stacking (Caruana et al., 2004; Sagi and Rokach, 2018). Here we chose the intuitive and computational cost-effective method of averaging. Rather than the simple averaging method (assuming the equal weight among base models), we adopted the forward stepwise ensemble selection (Caruana et al., 2004), which is a greedy search-based weighted averaging method. This method has been approved to be efficient and robust, and particularly useful to avoid overfitting issue by implementing selection with replacement, sorted ensemble initialization and bagged ensemble selection (Caruana et al., 2004). More details on the ensemble model construction and evaluation can be found in the following section.

2.3. Model training and validation

To develop fire models, we first constructed global fire training and testing datasets at the annual scale from 2003 to 2019. To reduce the influence of non-burnable surface (e.g., water bodies, desert, and barren land), land grids with multi-year mean vegetation fraction less than 5% were masked out (25.8% of total land surface; shown by light brown color in Fig. 1) according to the CCI51 land cover data. Those vegetated areas that did not register any fire signals during the study period (2003–2019; 26.5% of the total land surface; shown by gray color in Fig. 1) were also excluded in our analysis to prevent potential overfitting of non-fires in our fire models. Finally, to balance between data representativeness and computing efficiency, we randomly sampled 50% of the total burned grids over the vegetated area (i.e., ~60000 grid cells accounting for 23.5% of total land surface or 32% of total vegetated

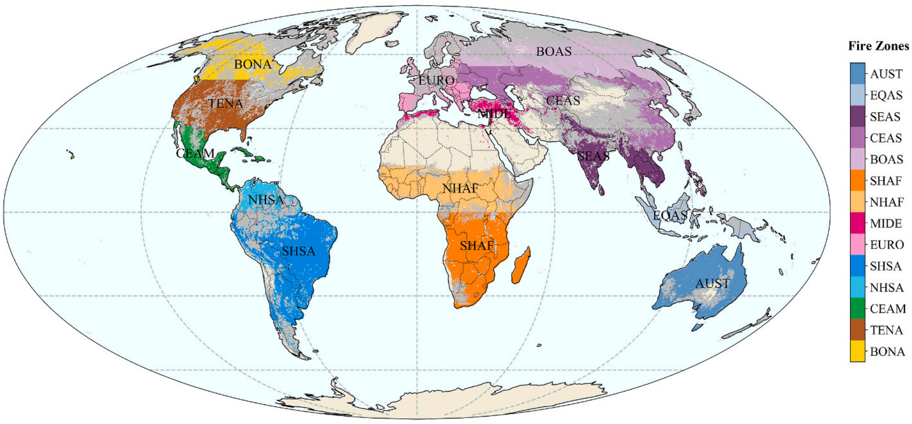

Fig. 1. Fire zones used in this study. Masked land area is unburnable non-vegetated area. The 14 fire zones include: BONA (Boreal North America), TENA (Temperate North America), CEAM (Central America), NHSA (Northern Hemisphere South America), SHSA (Southern Hemisphere South America), EURO (Europe), MIDE (Middle East), NHAF (Northern Hemisphere Africa), SHAF (Southern Hemisphere Africa), BOAS (Boreal Asia), CEAS (Central Asia), SEAS (Southeast Asia), EQAS (Equatorial Asia), and AUST (Australia & New Zealand). Non-vegetated area is labeled as light brown color, while non-fire regions over vegetated area are labeled as gray color. Original source: https://globalfiredata.org/pages/data/#ancillary.(For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

area) and obtained ~ 1 million yearly records for the three fire metrics and 19 relevant influencing factors. All the data were normalized based on their mean and standard deviation. We then randomly split 80% of the total data sites in each fire zone as training dataset and 20% as testing dataset.

Based on the training dataset, we then developed one global model and 14 regional models for each of three fire metrics (burned area, fire number and fire size). During the training process, five-fold cross validation was adopted to train the selected 5 MLs and their ensemble. Here, we optimized each ML and their ensemble by minimizing the cost function of RMSE (see Eq. (2)) based on an open source hyperparameter optimization framework (i.e., OPTUNA) (Akiba et al., 2019). The final ensemble of fire prediction from five trained MLs was created based on the weighted averaging approach (see Eq. (1) (Caruana et al., 2004);). Please note that if MLs could not increase or even dampen the performance of the ensemble, their weights would be set as 0 and excluded in the final ensemble. When the optimized hyperparameter ranges were identified through cross validation, they would be re-trained based on the whole training dataset, and then evaluated based on the testing dataset in terms of R^2 and RMSE. All the training and testing processes were conducted based on the automated machine learning Python package mljar (Version 0.11.5; https://github.com/mljar/mljar-supervised) in a Linux environment. 15 clusters with a total of 240 CPUs were reserved in the high performance and scientific computing platform of ISAAC hosted by the University of Tennessee to conduct the model optimization and evaluation processes.

2.4. Global fire simulation and evaluation

Using the developed global and regional ML models, we ran spatial simulations of annual burned area, fire number and fire size from 2003 to 2019. We then evaluated global ML-based simulations for three fire metrics based on the fire product from CCI51 in four aspects: (1) Multiple-year mean; (2) Annual variability in terms of standard deviation; (3) Annual anomaly relative to the multiple-year mean (here setting the year of 2019 as an example), and (4) Annual trend (quantified by Theil-Sen's slope). Global burned area simulated by seven fire models from FireMIP (Rabin et al., 2017) during 2003–2013 were also collected to compare our ML-based simulations. These models included CLM, JULES-INFERNO, CTEM, JSBACH-SPITFIRE, LPJ-GUESS--GlobFIRM, LPJ-GUESS-SIMFIRE-BLAZE and LPJ-GUESS-SPITFIRE.

2.5. Fire controls through factorial simulations

Based on the optimized ensemble ML and model inputs, we quantified individual fire controls related to climate, vegetation, and human on three fire metrics through factorial simulation design (Table 1). Specifically, we first conducted a simulation by allowing all three groups of

Table 1
Factorial design to disaggregate fire controls related to climate, vegetation, and human. The symbol '+' indicates that the input variable changes along time, while the symbol '- indicates that the input variable is fixed as the level in the initial year of 2003. The direct fire controls from climate, vegetation and human are 'All – $CLM_{control}$', 'All – $VEG_{control}$', and 'All – HUM0', respectively. The joint fire control between climate and vegetation, vegetation and human, and climate and vegetation are 'All – $CLM\_VEG_{control}$', 'All – $VEG\_HUM_{control}$', and 'All – $CLM\_VEG_{control}$', respectively.

Simulations	Climate	Vegetation	Human
All	+	+	+
$CLM_{control}$	−	+	+
$VEG_{control}$	+	−	+
$HUM_{control}$	+	+	−
$CLM\_VEG_{control}$	−	−	+
$VEG\_HUM_{control}$	+	−	−
$CLM\_HUM_{control}$	−	+	−

factors to change along the time (i.e., 'All'). Then, we performed a series of simulations by holding one or two groups of factors at the level of the initial year of 2003 but allowing the remaining group(s) of factors to vary with time. Finally, the direct control of one group or the combined control of two groups of factors could be derived from the difference between 'All' and sub-simulations. For example, 'All – $CLM_{control}$' is the direct control from climate, while 'All – $VEG\_HUM_{control}$' is the joint control from vegetation and human (Table 1). Similar simulations for 17 model inputs that have the annual dynamic information (bold variables in Table S1) were further conducted to quantify their direct control effects.

To derive the spatial pattern of major fire control among three groups of drivers (i.e., climate [CLM], vegetation [VEG] and human [HUM]), we compared their effects on annual trend in the specific fire metric. Here the annual trend was calculated through the non-parametric Theil-Sen estimator, which is particularly robust to outliers, while the statistical significance level was derived from Kendall's tau-b coefficient (Fernandes and Leblanc, 2005). Six forms of major fire controls based on their signs (i.e., positive and negative) were identified: positive (CLM+) and negative (CLM-) climate controls, positive (VEG+) and negative (VEG-) vegetation controls, positive (HUM+) and negative (HUM-) human controls. For each group of the fire control, we further identified the specific dominant factor (among all factors for that group in Table S17) by performing the similar trend comparison shown above.

3. Results

3.1. Grid-level model performance evaluation

We developed a global model (GlobeModel) and 14 regional models

(RegionModel) for three fire metrics, namely burned area, fire number, and average fire size. This was based on a random sample of grid-level data obtained from global burned area (Fig. 1) using a total of 21 feature inputs (Table S1). During the training process, we performed five-fold cross-validation and evaluated the performance of five different machine learning models (Fig. 2). Our analysis revealed that LightGBM performed the best, whereas Neural Network exhibited the lowest performance with XGBoost, Random Forest and CatBoost falling in between (Fig. 2). However, the ensemble ML model that incorporated predications from multiple relevant ML models outperformed each individual model in terms of R^2 and RMSE for both GlobeModel and RegionModel (Fig. 2). Therefore, this ensemble model strategy was chosen for use in our study.

In GlobeModel, LightGBM contributed the highest weight (0.67) in the ensemble ML, followed by XGBoost (0.33), while the remaining models were not included (Table 2). However, in RegionModels, the weights of individual models exhibited a varied pattern (Table 2). For example, over the boreal North America (BONA), RegionModel was dominated by XGBoost (0.67) and CatBoost (0.33), while other models were excluded. Overall, LightGBM, XGBoost and CatBoost carried the most significant weights in the ensemble MLs for RegionModels. The ensemble ML demonstrated comparable performance on both the training (80%) and testing (20%) datasets, as evidenced by the R^2 and RMSE values presented in Table 2. This indicates the robustness of our fire ML-based modeling framework. However, some fire zones with small or moderate fire fractions (such as BONA, TENA and CEAS) showed relatively poor performances likely due to the lower frequency of fire signals when compared to the higher fire fraction zones (such as NHAF and SHAF) (Fig. 2 & Table 2). The grid-level model performance for fire number simulation was similar to that for burned area (Fig. S1).

3.2. Assessment of global level model simulations

Globally, the ensemble ML for RegionModels effectively reproduced the annual dynamics of global burned area (R^2 = 0.90, $P < 0.001$; Fig. 3a), total fire numbers (R^2 = 0.86, $P < 0.001$; Fig. 3b), and average fire size (R^2 = 0.70, $P < 0.001$; Fig. 3c). RegionModel demonstrated slightly better performance than GlobeModel in simulating these fire metrics at both global (Fig. 3) and regional scales (Fig. S2). Both GlobeModel and RegionModel accurately captured the annual variations and long-term declining trends observed in global burned area (Fig. 3a) and total fire number (Fig. 3b), and insignificant trend in global averaged fire size (Fig. 3c), although the magnitudes of these variations and trends were somewhat underestimated. In contrast, FireMIP models displayed large inter-model variability and significant negative discrepancies when compared to satellite-based observations, particularly in simulating global burned area (Fig. 3d and S3).

Spatially, the ensemble ML (hereafter referred to as RegionModel) simulated the magnitude and extent of annual burned area reasonably well, especially for the major fire zones in Africa and northern Australia (Fig. 4a & b). The negative anomalies of burned areas in 2019 (Fig. 4e & f) and long-term declining trends in annual burned areas over those major fire zones (Fig. 4g & h) were also captured by the ensemble ML, although the magnitudes were slightly underestimated. In terms of the annual variability of burned area, the ensemble ML replicated the observed overall pattern of spatial gradient but tended to underestimate the magnitudes (Fig. 4c & d). Global fire number (Fig. S4) and averaged fire size (Fig. S5) simulated by the ensemble ML exhibited similar pattern of model performance with burned area.

3.3. Global fire controls derived from the ensemble ML

Through factorial simulations based on the optimized regional ensemble ML, we fully examined fire controls related to climate, vegetation, and human factors on annual fire trends. From 2003 to 2019, human control showed significant negative influence on global burned area (−1.9 Mha/yr, $P < 0.01$), followed by climate control (−1.3 Mha/yr, $P < 0.01$), and negligible positive vegetation control (0.4 Mha/yr, $P = 0.60$) (Fig. 5a and b). For global fire number, climate control (−8792/yr, $P < 0.01$) contributed slightly more negatively than human control (−5666/yr, $P < 0.01$), whereas vegetation control was insignificant (Fig. 5c and d). For the global averaged fire size, only human control showed significant negative influence (−0.37 ha/fire/yr), while two other controls showed insignificant positive influences (Fig. 5e and f). Overall, all three control groups together produced significant declining trends in global burned area, and total fire number as well as insignificant trend in global averaged fire size (Fig. 5), which were generally consistent with the observed trends from CCI51 (Fig. 3).

Global fire controls showed heterogeneous spatial patterns for their types, signs, and magnitudes (Fig. 6). For global burned area, the major control from climate accounted for 53.7% of total burned area, followed by vegetation control (23.4%) and human control (22.9%). However, a high portion (71.7%) of human control area showed negative effect on annual burned area trend, whereas climate and vegetation controls only had 53.0% and 42.8%, respectively (Fig. 7a). Across fire zones, human controls (mostly negative) were primarily concentrated in northern and

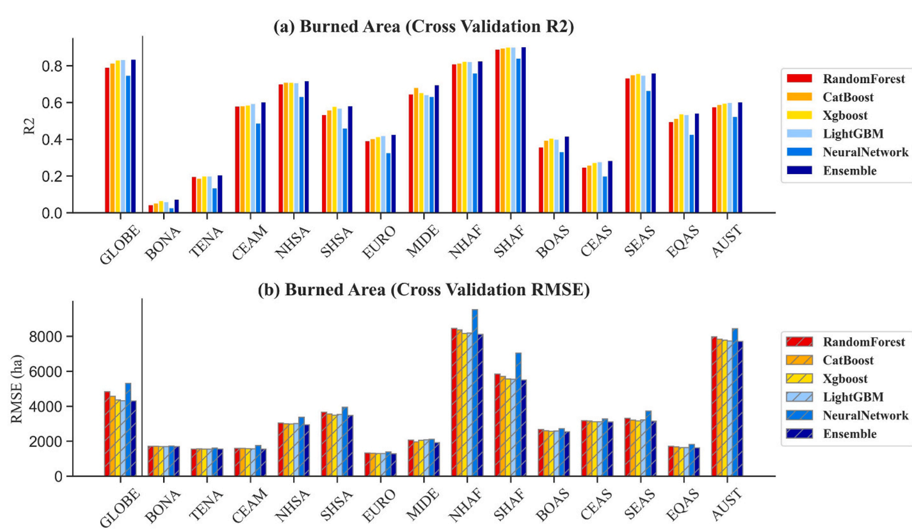

Fig. 2. Grid-level model evaluation for burned area among one global (GLOBE; left side of the vertical line) and 14 regional (right side of the vertical line) ML models based on five-fold cross validation. The abbreviations of fire zones are seen in Fig. 1.

Table 2
Model weights and performance of the ensemble MLs for burned area simulation. The abbreviations of fire zones are seen in Fig. 1. The weights of the ML model are represented using a color scheme, with darker shades of green indicating higher weight values, and the highest weight values being denoted in bold. The model performance was based on the training (80%) and testing (20%) datasets.

Fire Zones	Weight of Ensemble Models					Training		Testing	
	LightGBM	XGBoost	CatBoost	RF	NN	R^2	RMSE (ha)	R^2	RMSE (ha)
GLOBE	**0.67**	0.33	0	0	0	0.938	2123.0	0.936	2179.2
BONA	0	**0.67**	0.33	0	0	0.420	795.2	0.387	842.9
TENA	**0.5**	0.25	0.25	0	0	0.424	1108.6	0.420	1109.0
CEAM	**0.5**	0.25	0	0.25	0	0.721	1203.8	0.724	984.1
NHSA	0.2	**0.4**	**0.4**	0	0	0.921	1295.5	0.917	1331.1
SHSA	0.2	**0.6**	0.2	0	0	0.833	1881.2	0.836	1776.1
EURO	**0.5**	0.25	0.25	0	0	0.692	584.9	0.718	552.2
MIDE	0.25	0	**0.5**	0	0.25	0.825	1066.3	0.778	1078.2
NHAF	**0.4**	**0.4**	0.2	0	0	0.945	4282.8	0.947	4130.2
SHAF	**0.4**	**0.4**	0.2	0	0	0.966	3215.8	0.965	3283.0
BOAS	0.2	**0.4**	0.2	0	0.2	0.749	1252.2	0.756	1271.3
CEAS	**0.6**	0.2	0	0	0.2	0.572	1962.4	0.565	2069.1
SEAS	0.25	**0.5**	0.25	0	0	0.889	1790.1	0.907	1769.9
EQAS	**0.4**	**0.4**	0.2	0	0	0.763	954.7	0.836	792.9
AUST	**0.4**	0.2	0.2	0	0.2	0.778	5155.6	0.725	5778.2

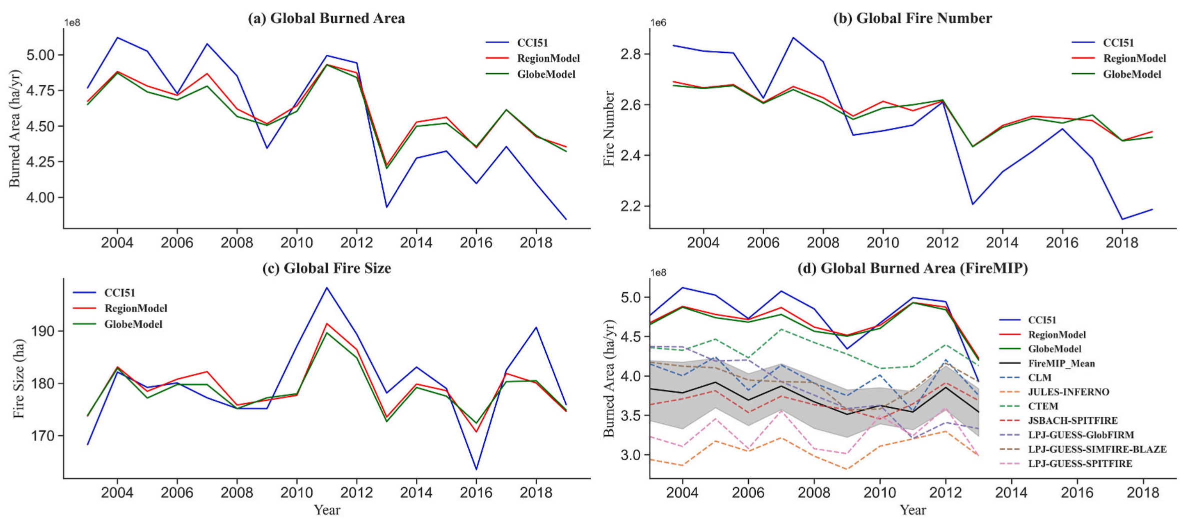

Fig. 3. Evaluations of global fire simulations from the ensemble MLs (GlobeModel & RegionModel) and FireMIP based on the satellite-based fire product from CCI51. 'FireMIP_Mean' in (d) are the mean of seven fire models from FireMIP.

southern Africa, Central Asia and North China; the majority of positive vegetation controls were distributed in Australia, Middle East, and Southeast Asia, while negative vegetation controls were scattered throughout northwestern Africa and central Europe; positive climate controls were extensively located in Boreal Asia, Boreal North America and most of United States; by contrast, negative climate controls were mostly found in the long-belt of northern Central Asia, southern South America, southern China and northern Australia (Fig. 6a).

Global fire number showed similar spatial pattern of its controls with burned area, except the slightly smaller area (2.9% of total burned area) of human control particularly in southern Africa and North China (Fig. 6c & 7b). Compared to the burned area, fire size showed less extents of vegetation control (2.8% of total burned area) and human control (2.9% of total burned area), but slightly larger area of climate control (3.5% of total burned area). Especially, negative climate controls on fire number were more extensive (9.3% of total burned area) than that on burned area (Fig. 7c), which were mainly evidenced in eastern Europe and southeastern Brazil (Fig. 6e). Regarding to the fire control strength, those area with major human controls had substantially larger magnitudes than other two controls for all three fire metrics (Fig. 6b, d & f).

The dominant fire controls are represented mechanistically by a series of proxies from climate, vegetation, and socioeconomic factors. Based on the ensemble ML, we identified that the dominant human controls on global burned area are primarily related to GDP (13.1% of total burned area), followed by population density (5.5%), urban fraction (2.1%) and cropland fraction (1.1%); for global vegetation controls, three leading factors are past one-year LAI (13.7%), past two-year LAI (5.9%), and forest fraction (1.2%); global climate controls are mainly related to annual VPD (12.5%), solar radiation (12.4%) and

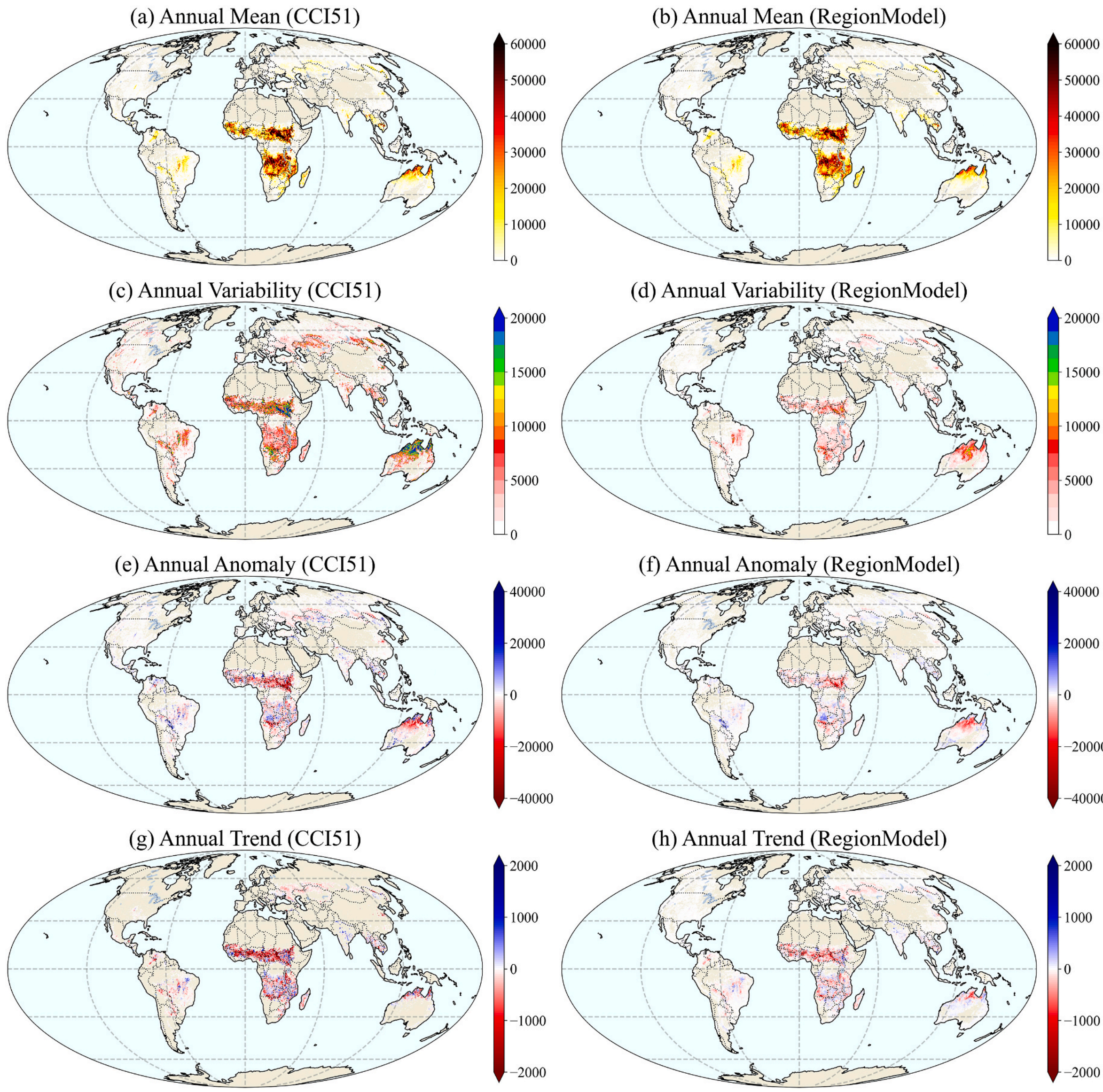

Fig. 4. Evaluation of the ensemble ML (RegionModel) in simulating global patterns of burned area in terms of multi-year amplitude, annual variability, annual anomaly in 2019 and long-term annual trend. The unit here is ha/yr.

precipitation (11.6%), respectively (Fig. 8a). Reginal socioeconomic developments reflected by increasing GDP, population density and urban expansion are mainly responsible for the negative human control on burned area, especially in northern and southern Africa, North China, and Southeast Asia (Fig. S6). Accumulated fuel loadings from the enhanced vegetation growth are the main driver for positive vegetation control, notably in Australia and India (Fig. S7). Although wide-spread warming and drying trends are associated with positive climate control, regional wetting trend is responsible for negative climate control (e.g., in central China and southeastern United States) (Fig. S8). Global controls for fire number and fire size share similar dominant factors with that for burned area in most areas (Fig. 8b and c). However, fire number showed smaller fraction that linking with GDP (10.4%), whereas fire size showed slightly larger fraction that linking with VPD (14.2%). Over the major fire zones in Africa, fire number intends to be controlled by climate and vegetation, which differs from fire size being mainly controlled by human factors (Fig. 8b and c). In other words, human management of burned areas in Africa has primarily relied on containing the size of fires rather than the number of fires.

4. Discussion

4.1. Global fire dynamics modeling based on ML

Global fire dynamics are one of the most complex land surface phenomena (Bowman et al., 2009; Hantson et al., 2016). Our ensemble ML optimized based on satellite and climate observations well reproduced annual dynamics of global burned area ($R^2 = 0.90$, $P < 0.001$; Fig. 3a), which serves the basis of our fire control attribution. In contrast, the process-based fire-enabled DGVMs from FireMIP poorly simulated inter-annual variations and trends in global burned area (Fig. 3d and S3), reflecting the need in fire process optimization or parameterization and input improvements, e.g., by integrating more effective observation-constrained fire controls and their interactions in the process models (Andela et al., 2017; Hantson et al., 2020). Some empirical fire models, e.g., based on linear regression (Andela et al., 2017) or logistic curve (Forkel et al., 2017; Kelley et al., 2019) could reasonably simulate the decline trend in global burned area by accounting for the key fire drivers particularly from the human suppression. However, most of these models adopted globally invariant

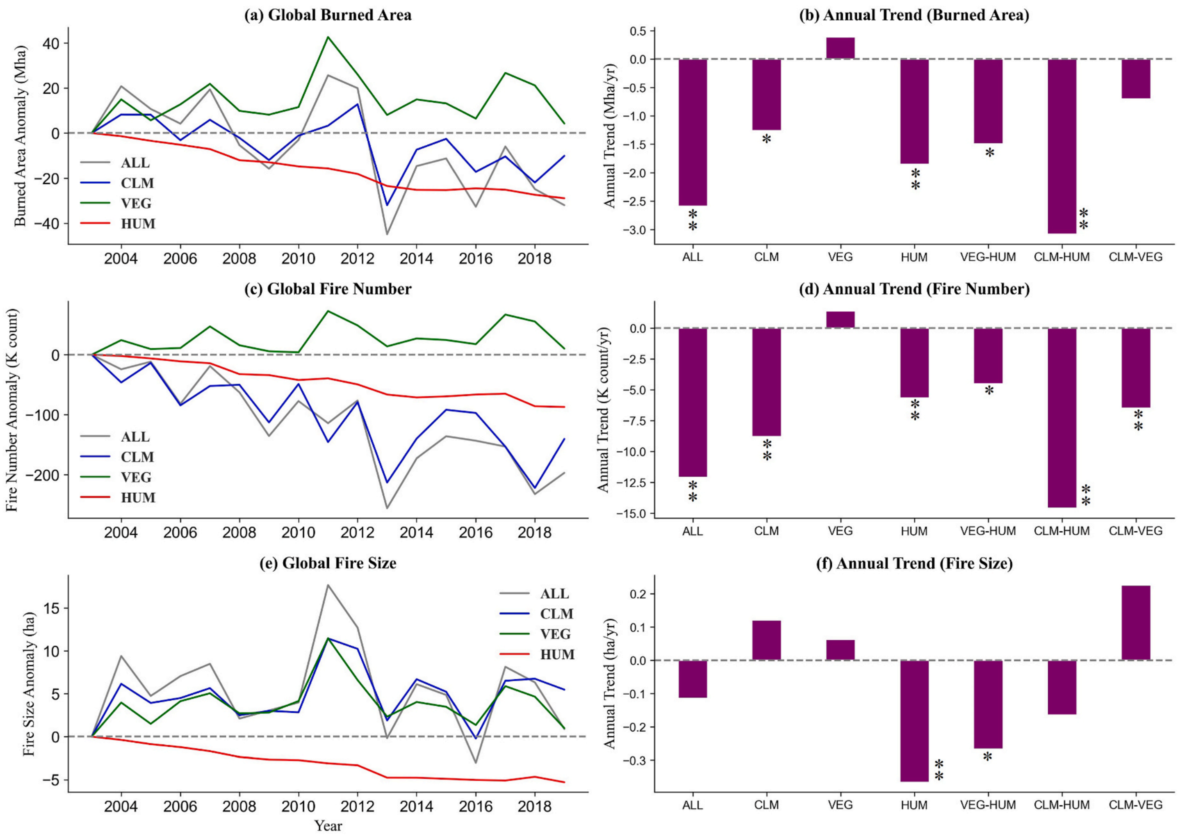

Fig. 5. Effects of climate, vegetation and human controls on annual global burned area, total fire number and averaged fire size. Scenario simulations are seen in Table 1. CLM, VEG, HUM are direct fire controls from climate, vegetation, and human, respectively. ALL is the simulated total controls. VEG-HUM, CLM-HUM and CLM-VEG in (b), (d) and (f) are joint controls from vegetation and human, climate and human, and climate and vegetation, respectively. *p < 0.05; **p < 0.01.

parameters during the model calibration process. Our evaluation for the two ML-based global and regional modeling strategies showed that regional models could better represent fire dynamics in the regional scale than the global model (Fig. S2). In addition, the empirical models rely on the assumption that the burned area is a linear combination of its key drivers, which may simplify or ignore the likely non-linear and interactive effects among fire drivers that ML-based models are capable of (Abiodun et al., 2018; Jain et al., 2020).

Based on a deep neural network and regional modeling strategy, Zhu et al. (2022) built a global surrogate fire model, which substantially improved the original process-based fire model within the land component of Department of Energy Earth system model. Depending on the selected inputs, Forkel et al. (2017) showed the random forest produced a slightly more realistic prediction of global burned area when compared to the data-driven model of SOFIA. In this study, our cross validation (Fig. 2) identified gradient boosting MLs as better models for predicting burned area than previously adopted random forest or deep neural network. Among the gradient boosting MLs, LightGBM developed by Microsoft outperformed all other models (Fig. 2) likely due to the novel algorithm integration of Gradient-based One-Side Sampling and Exclusive Feature Bundling (Ke et al., 2017). By combing all MLs together, the ensemble ML showed the best model performance in terms of R^2 and RMSE over all fire zones (Fig. 2), suggesting an alternative but highly efficient approach for global and regional fire modeling (Jain et al., 2020; Van Breugel et al., 2016).

Besides burned area, the ensemble ML also reproduced annual dynamics in global fire number and fire size, two fire metrics that were rarely involved in previous fire modeling studies (Hantson et al., 2016, 2020). During the study period, global fire number showed a significant decline trend while global averaged fire size demonstrated insignificant changes (Fig. 3), suggesting the fire number being as the primary driver of global decline in burned area (Andela et al., 2017). The ensemble ML based on the unified modeling framework reasonably captured such temporal co-influence patterns among three fire matrices. Spatially, the ensemble ML further reproduced the major patterns of annual magnitude, variability, specific year anomaly and long-term trend for all three fire metrics (Fig. 4, S2, S4 & S5), highlighting the advantage of the ensemble ML as an effective spatial-temporal diagnostic simulation tool (Abiodun et al., 2018; Jain et al., 2020).

4.2. Global fire controls revealed by ML

Global fire controls relating to climate, vegetation, and anthropogenic activity are typically entangled, and hard to separate (Forkel et al., 2017; Rabin et al., 2017). Based on the optimized ensemble ML and factorial sensitivity simulations, we fully quantified fire controls from climate, vegetation, and human on three fire metrics in a spatially resolved wall-to-wall manner. The proposed fire attribution framework based on the optimized ML and classical factor simulations could well avoid the drawbacks of commonly used explainability approaches (such as the permutation feature importance; Breiman, 2001) by providing local level information on the magnitude and sign of the prediction (or target change) given the change of a particular feature input. In addition, this framework is particularly useful for the ensemble of multiple MLs, among which the model-based feature importance scores derived from the same or varying explainability approaches (Ribeiro et al., 2016; Lundberg and Lee, 2017) are theoretically incomparable.

Overall, our ML-based fire attribution highlighted the dominant role of enhanced anthropogenic activity in reducing global burned area (−1.9 Mha/yr, $P < 0.01$; Fig. 5), which is generally consistent with earlier research (Andela et al., 2017; Kelley et al., 2019; Wu et al., 2021). However, our study indicated that human control over burned area was

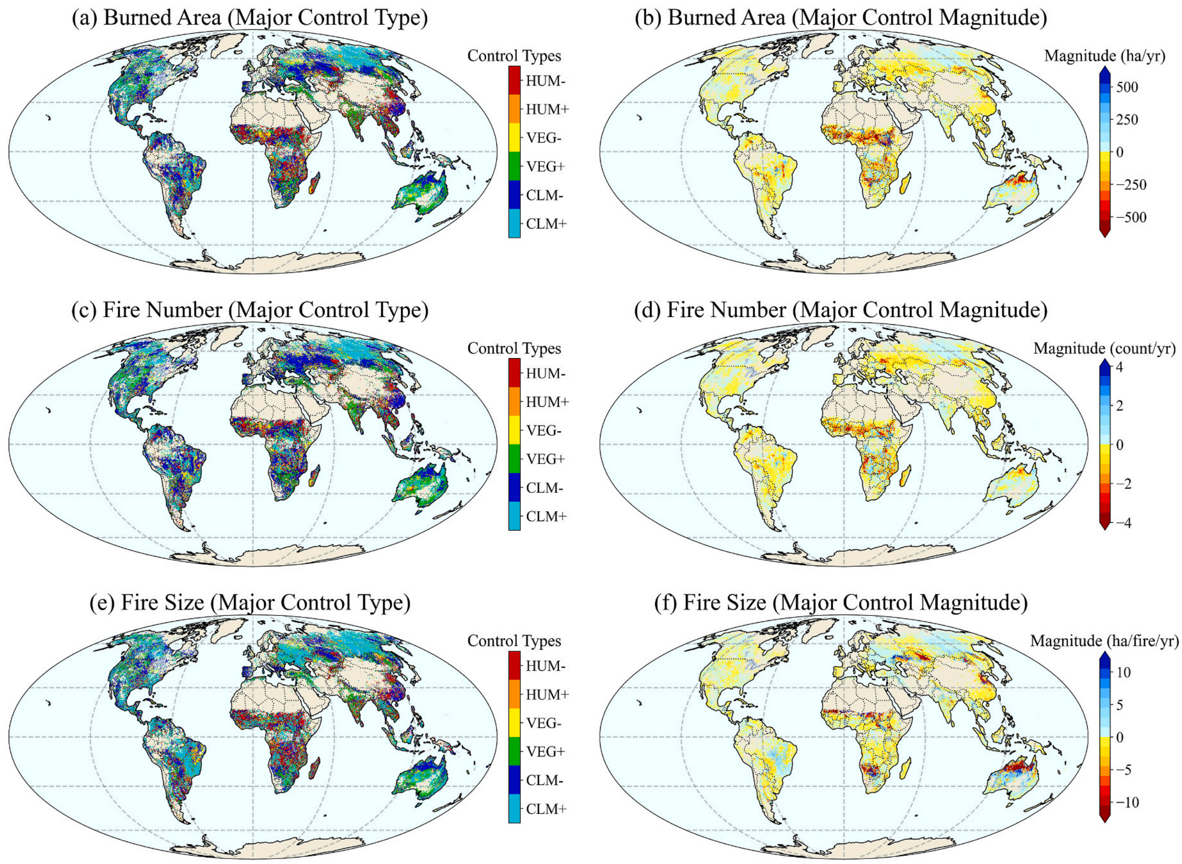

Fig. 6. Spatial patterns of major fire control types and magnitudes for burned area, fire number and fire size. Six fire control types in (a), (c) and (e) are: CLM+ (positive climate control), CLM- (negative climate control), VEG+ (positive vegetation control), VEG- (negative vegetation control), HUM+ (positive human control) and HUM- (negative human control).

mainly established through suppressing fire size rather than fire number, particularly in southern Africa (Fig. 6). Additionally, our study showed that socioeconomic factors of GDP and population density served as better proxies for human controls on burned area when compared with cropland (Fig. 8a), suggesting a more integrated mechanism for human fire control, e.g., via increasing landscape fragmentation, enhancing artificial fire suppression or land management (Forkel et al., 2019; Kelley et al., 2019). Besides northern and southern Africa, China showed relatively higher fraction of negative human control (Fig. 6), likely due to the stringent national policy on reducing air pollution and preventing wildfires (Wang, 2021).

Climate control dominated a substantially larger burned area (53.7%) than that for human control (23.4%) (Fig. 6a). However, climate control showed a relatively weaker influence on global burned area (−1.3 Mha/yr, $P < 0.01$; Fig. 5) largely due to the counteraction of negative and positive influences (Figs. 6 and 7) related to reginal wetting and drying trends (Forkel et al., 2019). Although a lesser negative influence, climate control largely regulated inter-annual variation of global burned area (Fig. 5), suggesting the strong perturbation of climate variability in enabling annual fire activity (Abatzoglou et al., 2018; Barbero et al., 2020; Tang et al., 2021). Previous studies indicated lightning as a major driver of recent large fire years in boreal forests (Felsberg et al., 2018; Veraverbeke et al., 2017). However, due to the lack of inter-annual information in the lightning climatology, climate controls in boreal regions may be underestimated in our study. That being said, our ML-based simulation indicated a wide-spread positive control from a drier and hotter climate in boreal North America and Eurasia (Figs. 6 and 7), implying the enhanced fire risk in these climate-warming sensitive areas (Coogan et al., 2020; Descals et al., 2022; Jones et al., 2022; Veraverbeke et al., 2017). Conversely, the regional wetting trend resulted in decreased burned area, especially in southeastern China and U.S., and southern Brazil (Fig. 6), suggesting the negative feedback of climate on fire (Kelley et al., 2019).

Global wide-spread vegetation greening occurred during the study period (Piao et al., 2020; Zhang et al., 2017). However, only 23.4% of total burned area was dominated by vegetation growth and mainly located in the fuel-limited arid and semi-arid regions, such as southern Australia, and middle east (Fig. 6). Most parts of India with a tropical climate also exhibited major vegetation controls (Fig. 6). A recent study indicated the greening of India was mainly caused by intensive cultivation of crops (Chen et al., 2019), which can be well captured by satellite-based LAI. Therefore, the positive vegetation control in India may be further related to the fuel loading accumulation as well as intentional ignitions associated with agricultural practices. Drier and hotter climate resulted in less burned area in certain sparsely vegetated areas, such as northern Australia and Kazakhstan (Fig. 6), which may be potentially linked to the reduced fuel loadings caused by the climate-vegetation interaction (Abatzoglou et al., 2018; Kelley et al., 2019).

4.3. Limitations and implications

Although the promising aspects of the ensemble ML for global fire modeling and analysis, several uncertainties and limitations need to be noted here. First, the ensemble ML tended to underestimate global burned area in the initial period (i.e., before 2008), but overestimate in the late period (i.e., after 2012) (Fig. 3a); such a tendency further resulted in an underestimation in annual trend and variability in global burned area (Fig. 4). This global tendency appears not to be dominated by specific major fire zones but resulted from an accumulative effect

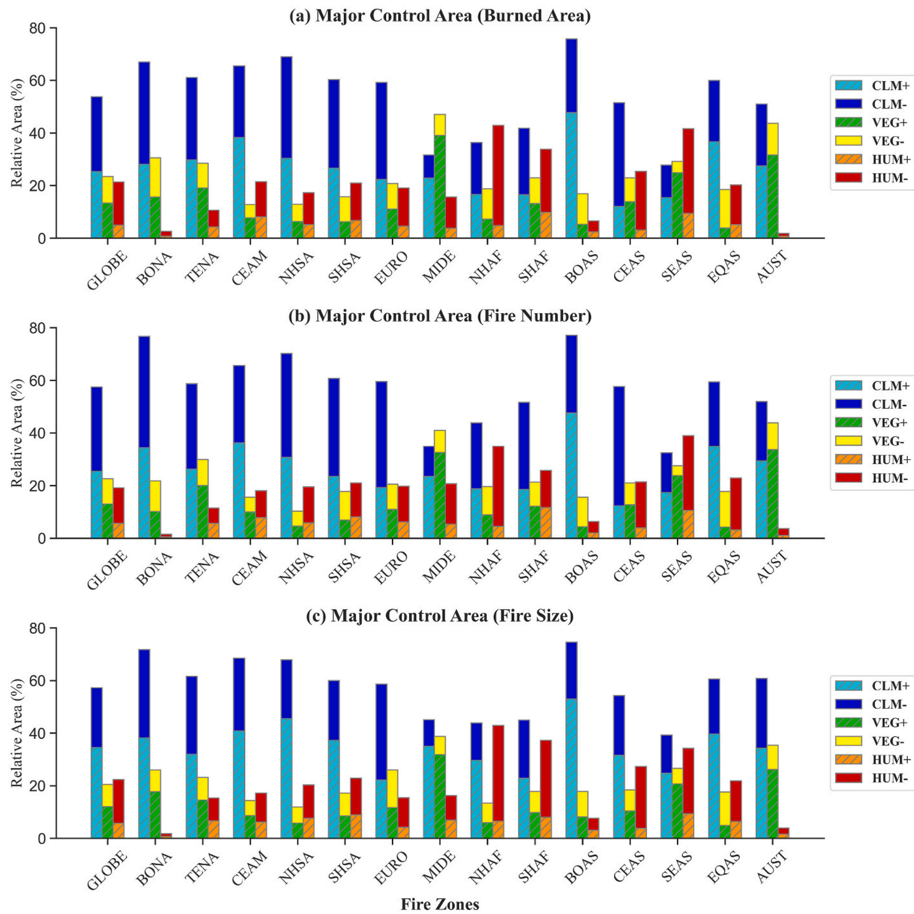

Fig. 7. Comparisons of major fire control areas for burned area, fire number and fire size. Relative areas are the percentage area relative to global burned area (for GLOBE) or total burned area in each fire zone. Six fire control types are seen in Fig. 6.

from them (Figs. S9–S12). Secondly, compared to the fire size, fire number showed larger underestimation in its annual variability, thus annual trend (Fig. 3, S4 & S5), and this underestimation further led to the underestimation of annual burned area, particularly in the major fire zones in Africa and Australia (Fig. 4). The weakened fire variability in the ensemble ML is probably due to the lack of effective and heterogenous within-grid information for anthropogenic and natural ignitions (Hantson et al., 2020). Such precise information is still difficult to obtain at high spatial resolution, although socioeconomic factors can roughly represent them (Bowman et al., 2020; Hantson et al., 2016). Thirdly, data uncertainties in forcing and satellite-based fire observations were both involved in the ML training and spatial upscaling, which may be further propagated to the uncertainties in control attributions. For example, the missing interannual variability information for lightning may be responsible for the relatively poor performance of fire simulations in boreal regions (Fig. 2), while incomplete records of GDP after 2015 may underestimate human controls in major fire zones (e.g., Africa), thus overestimate global burned area. In addition, the original satellite-based fire product (i.e., CCI51) may still miss some small fires that cannot be effectively detected by the 250-m sensor (Ramo et al., 2021). The fire modeling capacity of the ensemble ML and relevant quantification of fire controls may need to be further evaluated based on future released global high-quality data (e.g., Sentinel-based 20 m fire product (Roteta et al., 2019);).

Despite the limitations mentioned above, our ML-based global fire modeling and analysis framework does have important implications. Firstly, our study confirmed that the optimized ensemble ML is capable of reconstructing annual fire dynamics that current process-based fire-enabled DGVMs generally failed to reproduce, suggesting an effective alternative to develop ML-based surrogate fire model for Earth system model (Yu et al., 2022; Zhu et al., 2022). Although our study was focused on the annual scale, similar framework could be readily expanded to seasonal or even smaller temporal scales when input and training data are available. Secondly, through factorial simulations, our study sufficiently quantified and disentangled global fire controls related to climate, vegetation, and human for three fire metrics, which are the key components of global integral fire regime. Our efforts in understanding the modern fire dynamics could help better depict the pattern and mechanism of the contemporary global fire regime, which is still challenging and debatable (Archibald et al., 2013; Bowman et al., 2020). Thirdly, our study highlighted the human-induced negative control on global declining fire activity, but meanwhile acknowledged the more extensive climate and vegetation controls (Jones et al., 2022; Kelley et al., 2019), which depicts a complex picture of future fire regime change (Descals et al., 2022; Wu et al., 2021), particularly in the context of projected enhanced human activities and climate change (Andela et al., 2017; Bowman et al., 2020).

5. Conclusion

In this study, we developed a global model and 14 regional models for three fire metrics using an ensemble of five cutting-edge MLs and multiple satellite-based observations from 2003 to 2019. Overall, the ensemble ML from regional models well reproduced annual dynamics of global burned area ($R^2 = 0.90$, $P < 0.001$), total fire numbers ($R^2 = 0.86$, $P < 0.001$) and averaged fire size ($R^2 = 0.70$, $P < 0.001$), and further captured key patterns of their multi-year mean magnitudes, yearly variabilities, yearly anomalies, and long-term trends. Factorial

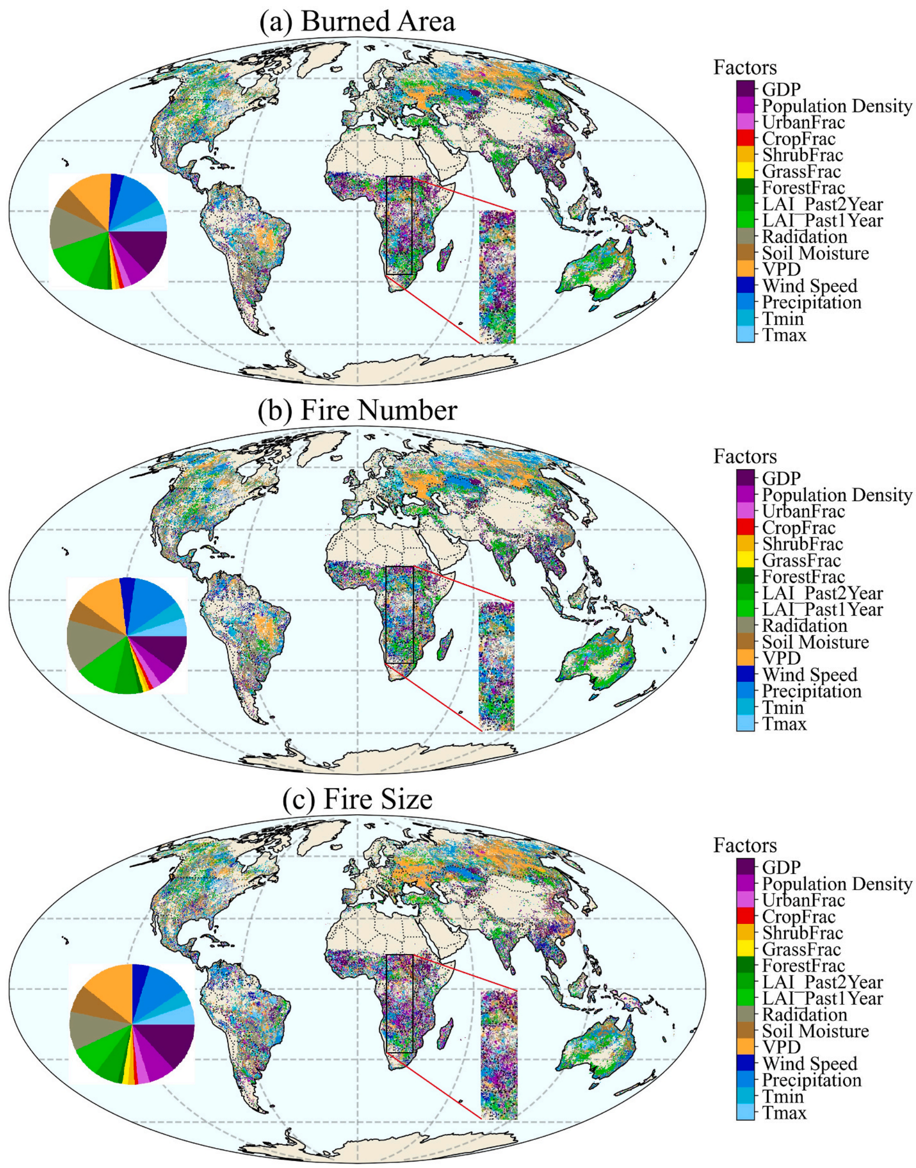

Fig. 8. Dominant factors for major climate, vegetation, and human controls on global burned area (a), fire number (b) and fire size (c). Specific factors are seen in Table S1. The inserted pie chart on the lower left panel is the area proportion for all control factors. The gradient of fire controls in the central Africa from northern to southern hemispheres (marked by black rectangle) is highlighted in each subfigure. The color schemes are used to quantify fire controls from four human factors (purple to red), five vegetation factors (yellow to green) and seven climate factors (gray to yellow and blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

simulations based on the optimized ensemble ML provided the basis for spatial and temporal fire attributions related to climate, vegetation, and human factors. During the study period, human control substantially negatively influenced the global burned area (−1.9 Mha/yr, $P < 0.01$), followed by climate control (−1.3 Mha/yr, $P < 0.01$), and insignificant positive vegetation control (0.4 Mha/yr, $P = 0.60$). However, spatially, climate control dominated a much larger burned area (53.7%) than that for human control (23.4%) and vegetation control (22.9%). The counteracting effect of regional wetting and drying trends was primarily responsible for the weaker climate control on global burned area. Although similar with burned area, global fire number tended to be more influenced by climate, but fire size was more influenced by human factors. Overall, this study provides an efficient approach for global fire modeling and subsequent control attribution based on the ensemble ML and factor simulations; and the revealed heterogeneous fire controls would enhance our understanding of modern fire regimes at multiple aspects, further leading to robust fire projections in a warming climate.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

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, the development team of the open-accessed automated machine learning Python package (mljar), and public data releases from ESA, NASA and other facilities and authors listed in Table S2. Finally, we acknowledge the constructive comments from two anonymous reviewers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.srs.2023.100088.

References

Abatzoglou, J.T., Williams, A.P., Boschetti, L., Zubkova, M., Kolden, C.A., 2018. Global patterns of interannual climate-fire relationships. Global Change Biol. 24 (11), 5164–5175. https://doi.org/10.1111/gcb.14405.

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., Arshad, H., 2018. State-of-the-art in artificial neural network applications: a survey. Heliyon 4 (11), e00938.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A Next-Generation Hyperparameter Optimization Framework, pp. 2623–2631.

Aldersley, A., Murray, S.J., Cornell, S.E., 2011. Global and regional analysis of climate and human drivers of wildfire. Sci. Total Environ. 409 (18), 3472–3481. https://doi.org/10.1016/j.scitotenv.2011.05.032.

Andela, N., Morton, D.C., Giglio, L., Chen, Y., van der Werf, G.R., Kasibhatla, P.S., DeFries, R.S., Collatz, G.J., Hantson, S., Kloster, S., Bachelet, D., Forrest, M., Lasslop, G., Li, F., Mangeon, S., Melton, J.R., Yue, C., Randerson, J.T., 2017. A human-driven decline in global burned area. Science 356 (6345), 1356–1362. https://doi.org/10.1126/science.aal4108.

Archibald, S., Lehmann, C.E.R., Gomez-Dans, J.L., Bradstock, R.A., 2013. Defining pyromes and global syndromes of fire regimes. Proc. Natl. Acad. Sci. USA 110 (16), 6442–6447. https://doi.org/10.1073/pnas.1211466110.

Balch, J.K., Bradley, B.A., D'Antonio, C.M., Gómez-Dans, J., 2013. Introduced annual grass increases regional fire activity across the arid western USA (1980–2009). Global Change Biol. 19 (1), 173–183.

Balch, J.K., Schoennagel, T., Williams, A.P., Abatzoglou, J.T., Cattau, M.E., Mietkiewicz, N.P., St Denis, L.A., 2018. Switching on the big burn of 2017. Fire 1 (1), 17.

Barbero, R., Abatzoglou, J.T., Pimont, F., Ruffault, J., Curt, T., 2020. Attributing increases in fire weather to anthropogenic climate change over France. Front. Earth Sci. 8, 104. https://doi.org/10.3389/feart.2020.00104.

Bowman, D.M.J.S., Balch, J., Artaxo, P., Bond, W.J., Cochrane, M.A., D'Antonio, C.M., DeFries, R., Johnston, F.H., Keeley, J.E., Krawchuk, M.A., Kull, C.A., Mack, M., Moritz, M.A., Pyne, S., Roos, C.I., Scott, A.C., Sodhi, N.S., Swetnam, T.W., 2011. The human dimension of fire regimes on Earth: the human dimension of fire regimes on Earth. J. Biogeogr. 38 (12), 2223–2236. https://doi.org/10.1111/j.1365-2699.2011.02595.x.

Bowman, D.M.J.S., Balch, J.K., Artaxo, P., Bond, W.J., Carlson, J.M., Cochrane, M.A., D'Antonio, C.M., DeFries, R.S., Doyle, J.C., Harrison, S.P., Johnston, F.H., Keeley, J. E., Krawchuk, M.A., Kull, C.A., Marston, J.B., Moritz, M.A., Prentice, I.C., Roos, C.I., Scott, A.C., et al., 2009. Fire in the Earth system. Science 324 (5926), 481–484. https://doi.org/10.1126/science.1163886.

Bowman, D.M.J.S., Kolden, C.A., Abatzoglou, J.T., Johnston, F.H., van der Werf, G.R., Flannigan, M., 2020. Vegetation fires in the anthropocene. Nat. Rev. Earth Environ. 1 (10), 500–515. https://doi.org/10.1038/s43017-020-0085-3.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Caruana, R., Niculescu-Mizil, A., Crew, G., Ksikes, A., 2004. Ensemble Selection from Libraries of Models, vol. 18. Twenty-First International Conference on Machine Learning - ICML '04. https://doi.org/10.1145/1015330.1015432.

Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R.K., Fuchs, R., Brovkin, V., Ciais, P., Fensholt, R., Tømmervik, H., Bala, G., Zhu, Z., Nemani, R.R., Myneni, R.B., 2019. China and India lead in greening of the world through land-use management. Nat. Sustain. 2 (2), 122–129. https://doi.org/10.1038/s41893-019-0220-7.

Chen, T., Guestrin, C., 2016. Xgboost: A Scalable Tree Boosting System, pp. 785–794.

Cochrane, M.A., 2003. Fire science for rainforests. Nature 421 (6926), 913–919.

Coogan, S.C.P., Cai, X., Jain, P., Flannigan, M.D., 2020. Seasonality and trends in human- and lightning-caused wildfires ≥ 2 ha in Canada, 1959–2018. Int. J. Wildland Fire 29 (6), 473. https://doi.org/10.1071/WF19129.

Descals, A., Gaveau, D.L., Verger, A., Sheil, D., Naito, D., Peñuelas, J., 2022. Unprecedented fire activity above the Arctic Circle linked to rising temperatures. Science 378 (6619), 532–537.

Dietterich, T.G., 2000. Ensemble Methods in Machine Learning, vols. 1–15.

Dong, X., Yu, Z., Cao, W., Shi, Y., Ma, Q., 2020. A survey on ensemble learning. Front. Comput. Sci. 14, 241–258.

Felsberg, A., Kloster, S., Wilkenskjeld, S., Krause, A., Lasslop, G., 2018. Lightning forcing in global fire models: the importance of temporal resolution. J. Geophys. Res.: Biogeosciences 123 (1), 168–177. https://doi.org/10.1002/2017JG004080.

Fernandes, R., Leblanc, S.G., 2005. Parametric (modified least squares) and non-parametric (Theil–Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors. Rem. Sens. Environ. 95 (3), 303–316.

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., Arneth, A., 2019. Emergent relationships with respect to burned area in global satellite observations and fire-enabled vegetation models. Biogeosciences 16 (1), 57–76. https://doi.org/10.5194/bg-16-57-2019.

Forkel, M., Dorigo, W., Lasslop, G., Teubner, I., Chuvieco, E., Thonicke, K., 2017. A data-driven approach to identify controls on global fire activity from satellite and climate observations (SOFIA V1). Geosci. Model Dev. (GMD) 10 (12), 4443–4476. https://doi.org/10.5194/gmd-10-4443-2017.

Friedlingstein, P., Jones, M.W., O`Sullivan, M., Andrew, R.M., Bakker, D.C., Hauck, J., Le Quéré, C., Peters, G.P., Peters, W., Pongratz, J., 2022. Global carbon budget 2021. Earth Syst. Sci. Data 14 (4), 1917–2005.

Giglio, L., Boschetti, L., Roy, D.P., Humber, M.L., Justice, C.O., 2018. The Collection 6 MODIS burned area mapping algorithm and product. Rem. Sens. Environ. 217, 72–85. https://doi.org/10.1016/j.rse.2018.08.005.

Giglio, L., Randerson, J.T., Van Der Werf, G.R., 2013. Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). J. Geophys. Res.: Biogeosciences 118 (1), 317–328.

Hanes, C.C., Wang, X., Jain, P., Parisien, M.-A., Little, J.M., Flannigan, M.D., 2019. Fire-regime changes in Canada over the last half century. Can. J. For. Res. 49 (3), 256–269. https://doi.org/10.1139/cjfr-2018-0293.

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., et al., 2016. The status and challenge of global fire modelling. Biogeosciences 13 (11), 3359–3375. https://doi.org/10.5194/bg-13-3359-2016.

Hantson, S., Kelley, D.I., Arneth, A., Harrison, S.P., Archibald, S., Bachelet, D., Forrest, M., Hickler, T., Lasslop, G., Li, F., Mangeon, S., Melton, J.R., Nieradzik, L., Rabin, S.S., Prentice, I.C., Sheehan, T., Sitch, S., Teckentrup, L., Voulgarakis, A., Yue, C., 2020. Quantitative assessment of fire and vegetation properties in simulations with fire-enabled vegetation models from the Fire Model Intercomparison Project. Geosci. Model Dev. (GMD) 13 (7), 3299–3318. https://doi.org/10.5194/gmd-13-3299-2020.

He, T., Lamont, B.B., Pausas, J.G., 2019. Fire as a key driver of Earth's biodiversity. Biol. Rev. 94 (6), 1983–2010. https://doi.org/10.1111/brv.12544.

Jain, P., Coogan, S.C.P., Subramanian, S.G., Crowley, M., Taylor, S., Flannigan, M.D., 2020. A review of machine learning applications in wildfire science and management. Environ. Rev. 28 (4), 478–505. https://doi.org/10.1139/er-2020-0019.

Jones, M.W., Abatzoglou, J.T., Veraverbeke, S., Andela, N., Lasslop, G., Forkel, M., Smith, A.J., Burton, C., Betts, R.A., van der Werf, G.R., 2022. Global and regional trends and drivers of fire under climate change. Rev. Geophys., e2020RG000726

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30.

Kelley, D.I., Bistinas, I., Whitley, R., Burton, C., Marthews, T.R., Dong, N., 2019. How contemporary bioclimatic and human controls change global fire regimes. Nat. Clim. Change 9 (9), 690–696. https://doi.org/10.1038/s41558-019-0540-7.

Kelly, L.T., Giljohann, K.M., Duane, A., Aquilué, N., Archibald, S., Batllori, E., Bennett, A. F., Buckland, S.T., Canelles, Q., Clarke, M.F., 2020. Fire and biodiversity in the anthropocene. Science 370 (6519) eabb0355.

Krawchuk, M.A., Moritz, M.A., 2011. Constraints on global fire activity vary across a resource gradient. Ecology 92 (1), 121–132.

Kummu, M., Taka, M., Guillaume, J.H., 2018. Gridded global datasets for gross domestic product and Human Development Index over 1990–2015. Sci. Data 5 (1), 1–15.

Lasslop, G., Coppola, A.I., Voulgarakis, A., Yue, C., Veraverbeke, S., 2019. Influence of fire on the carbon cycle and climate. Curr. Clim. Change Rep. 5 (2), 112–123. https://doi.org/10.1007/s40641-019-00128-9.

Lasslop, G., Hantson, S., Harrison, S.P., Bachelet, D., Burton, C., Forkel, M., Forrest, M., Li, F., Melton, J.R., Yue, C., Archibald, S., Scheiter, S., Arneth, A., Hickler, T., Sitch, S., 2020. Global ecosystems and fire: multi-model assessment of fire-induced tree-cover and carbon storage reduction. Global Change Biol. 26 (9), 5027–5041. https://doi.org/10.1111/gcb.15160.

Li, F., Zeng, X., Levis, S., 2012. A process-based fire parameterization of intermediate complexity in a Dynamic Global Vegetation Model. Biogeosciences 9 (7), 2761–2780.

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

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.

May, A., McMeeking, G., Lee, T., Taylor, J., Craven, J., Burling, I., Sullivan, A.P., Akagi, S., Collett Jr., J., Flynn, M., 2014. Aerosol emissions from prescribed fires in the United States: a synthesis of laboratory and aircraft measurements. J. Geophys. Res. Atmos. 119 (20), 11–826.

McLauchlan, K.K., Higuera, P.E., Miesel, J., Rogers, B.M., Schweitzer, J., Shuman, J.K., Tepley, A.J., Varner, J.M., Veblen, T.T., Adalsteinsson, S.A., Balch, J.K., Baker, P., Batllori, E., Bigio, E., Brando, P., Cattau, M., Chipman, M.L., Coen, J., Crandall, R., et al., 2020. Fire as a fundamental ecological process: research advances and frontiers. J. Ecol. 108 (5), 2047–2069. https://doi.org/10.1111/1365-2745.13403.

Piao, S., Wang, X., Park, T., Chen, C., Lian, X., He, Y., Bjerke, J.W., Chen, A., Ciais, P., Tømmervik, H., Nemani, R.R., Myneni, R.B., 2020. Characteristics, drivers and feedbacks of global greening. Nat. Rev. Earth Environ. 1 (1), 14–27. https://doi.org/10.1038/s43017-019-0001-x.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. Adv. Neural Inf. Process. Syst. 31.

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., et al., 2017. The Fire

Modeling Intercomparison Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions. Geosci. Model Dev. (GMD) 10 (3), 1175–1197. https://doi.org/10.5194/gmd-10-1175-2017.
Ramo, R., Roteta, E., Bistinas, I., Van Wees, D., Bastarrika, A., Chuvieco, E., Van der Werf, G.R., 2021. African burned area and fire carbon emissions are strongly impacted by small fires undetected by coarse resolution satellite data. Proc. Natl. Acad. Sci. USA 118 (9), e2011160118.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Why should i trust you?". Explaining the predictions of any classifier 1135–1144.
Riley, K., Thompson, M., 2016. An uncertainty analysis of wildfire modeling. Natural hazard uncertainty assessment: modeling and decision support. Geophys. Monogr. 223, 193–213.
Roscher, R., Bohn, B., Duarte, M.F., Garcke, J., 2020. Explainable machine learning for scientific insights and discoveries. IEEE Access 8, 42200–42216.
Ross, A.N., Wooster, M.J., Boesch, H., Parker, R., 2013. First satellite measurements of carbon dioxide and methane emission ratios in wildfire plumes. Geophys. Res. Lett. 40 (15), 4098–4102.
Roteta, E., Bastarrika, A., Padilla, M., Storm, T., Chuvieco, E., 2019. Development of a Sentinel-2 burned area algorithm: generation of a small fire database for sub-Saharan Africa. Rem. Sens. Environ. 222, 1–17. https://doi.org/10.1016/j.rse.2018.12.011.
Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1 (5), 206–215.
Sagi, O., Rokach, L., 2018. Ensemble learning: a survey. Wiley Interdiscipl. Rev.: Data Min. Knowl. Discov. 8 (4), e1249.
Tang, R., Mao, J., Jin, M., Chen, A., Yu, Y., Shi, X., Zhang, Y., Hoffman, F.M., Xu, M., Wang, Y., 2021. Interannual variability and climatic sensitivity of global wildfire activity. Adv. Clim. Change Res. 12 (5), 686–695.
Tuyen, T.T., Jaafari, A., Yen, H.P.H., Nguyen-Thoi, T., Van Phong, T., Nguyen, H.D., Van Le, H., Phuong, T.T.M., Nguyen, S.H., Prakash, I., 2021. Mapping forest fire susceptibility using spatially explicit ensemble models based on the locally weighted learning algorithm. Ecol. Inf. 63, 101292.
Urbanski, S.P., Hao, W.M., Baker, S., 2008. Chemical composition of wildland fire emissions. Dev. Environ. Sci. 8, 79–107.
Van Breugel, P., Friis, I., Demissew, S., Lillesø, J.-P.B., Kindt, R., 2016. Current and future fire regimes and their influence on natural vegetation in Ethiopia. Ecosystems 19 (2), 369–386.
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., Morton, D.C., Collatz, G.J., 2017. Global fire emissions estimates during 1997–2016. Earth Syst. Sci. Data 9 (2), 697–720.
Veraverbeke, S., Rogers, B.M., Goulden, M.L., Jandt, R.R., Miller, C.E., Wiggins, E.B., Randerson, J.T., 2017. Lightning as a major driver of recent large fire years in North American boreal forests. Nat. Clim. Change 7, 9.
Wang, P., 2021. China's air pollution policies: progress and challenges. Current Opin. Environ. Sci. & Health 19, 100227.
Williams, A.P., Abatzoglou, J.T., Gershunov, A., Guzman-Morales, J., Bishop, D.A., Balch, J.K., Lettenmaier, D.P., 2019. Observed impacts of anthropogenic climate change on wildfire in California. Earth's Future 7 (8), 892–910. https://doi.org/10.1029/2019EF001210.
Wu, C., Venevsky, S., Sitch, S., Mercado, L.M., Huntingford, C., Staver, A.C., 2021. Historical and future global burned area with changing climate and human demography. One Earth 4 (4), 517–530.
Yu, Y., Mao, J., Wullschleger, S.D., Chen, A., Shi, X., Wang, Y., Hoffman, F.M., Zhang, Y., Pierce, E., 2022. Machine learning–based observation-constrained projections reveal elevated global socioeconomic risks from wildfire. Nat. Commun. 13 (1), 1250. https://doi.org/10.1038/s41467-022-28853-0.
Zhang, Y., Song, C., Band, L.E., Sun, G., Li, J., 2017. Reanalysis of global terrestrial vegetation trends from MODIS products: browning or greening? Rem. Sens. Environ. 191, 145–155. https://doi.org/10.1016/j.rse.2016.12.018.
Zhu, Q., Li, F., Riley, W.J., Xu, L., Zhao, L., Yuan, K., Wu, H., Gong, J., Randerson, J., 2022. Building a machine learning surrogate model for wildfire activities within a global Earth system model. Geosci. Model Dev. (GMD) 15 (5), 1899–1911.