

# 1 **Interpreting machine learning prediction of fire emissions and comparison** 2 **with FireMIP process-based models**

3 Sally S.-C. Wang<sup>1</sup>, Yun Qian<sup>1</sup>, L. Ruby Leung<sup>1</sup>, Yang Zhang<sup>2</sup>

4 <sup>1</sup>Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, Washington, 99354,  
5 USA

6 <sup>2</sup> Department of Civil and Environmental Engineering, Northeastern University, Boston, Massachusetts, 02115, USA

7 *Correspondence to:* Sally S.-C. Wang ([sing-chun.wang@pnnl.gov](mailto:sing-chun.wang@pnnl.gov)) and L. Ruby Leung ([Ruby.Leung@pnnl.gov](mailto:Ruby.Leung@pnnl.gov))

8 **Abstract.** Annual burned areas in the United States have increased twofold during the past decades. With more  
9 large fires resulting in more emissions of fine particulate matter, an accurate prediction of fire emissions is critical  
10 for quantifying the impacts of fires on air quality, human health, and climate. This study aims to construct a machine  
11 learning (ML) model with game-theory interpretation to predict monthly fire emissions over the contiguous US  
12 and to understand the controlling factors of fire emissions. The **optimized ML model is used to diagnose the**  
13 **process-based models in the Fire Modeling Intercomparison Project (FireMIP) to inform future development.**  
14 Results show promising performance for the ML model, Community Land Model (CLM), and Joint UK Land  
15 Environment Simulator-Interactive Fire And Emission Algorithm For Natural Environments (JULES-INFERN0)  
16 in reproducing the spatial distributions, seasonality, and interannual variability of fire emissions over CONUS.  
17 Regional analysis shows that only the ML model and CLM simulate the realistic interannual variability of fire  
18 emissions for most of the subregions ( $r > 0.95$  for ML and  $r = 0.14 \sim 0.70$  for CLM), except for Mediterranean  
19 California, where all the models perform poorly ( $r = 0.74$  for ML and  $r < 0.30$  for the FireMIP models). Regarding  
20 seasonality, most models capture the peak emission in July over western US. However, all models except for the  
21 ML model fail to reproduce the bimodal peaks in July and October over Mediterranean California, which may be  
22 explained by the **smaller wind speeds of the atmospheric forcing data during Santa Ana wind events and** limitations  
23 in model parameterizations for capturing the effects of Santa Ana winds on fire activity. Furthermore, most models  
24 struggle to capture the spring peak in emissions in southeastern US, probably due to underrepresentation of human  
25 effects and the influences of winter dryness on fires in the models. As for extreme events, both the ML model and  
26 CLM successfully reproduce the frequency map of extreme emission occurrence but overestimate the number of  
27 months with extremely large fire emissions. Comparing the fire  $PM_{2.5}$  emissions from the ML model with process-  
28 based fire models highlights their strengths and uncertainties for regional analysis and prediction and provides  
29 useful insights on future directions for model improvements.

## 30 **1. Introduction**

31 Large fires have increased across the United States over the past two decades, especially in the western  
32 US. While the total area burned in 2020 increased by 51% compared to the 10-year average for 2010-2019, the  
33 total number of fires in 2020 is smaller than the 10-year average. This indicates the contribution of larger and more  
34 powerful fires to the growing burned areas (NIFC, 2020). Large fires can directly lead to property damages and  
35 pose a threat to human lives (Thomas et al., 2017). Meanwhile, fine particulate matter ( $PM_{2.5}$ , particles with an  
36 aerodynamic diameter smaller than and equal to  $2.5 \mu m$ ) emitted from fires not only have negative impacts on  
37 human health but also affect climate and ecosystems (Johnston et al., 2012; Ward et al., 2012; Rap et al., 2013;

38 Kaulfus et al., 2017; Liu et al., 2018; Wang et al., 2018; Stowell et al., 2019). Driven by stronger fire heating and  
39 with higher injection height, aerosols emitted from large fires can be transported to broader area and stay in  
40 atmosphere longer. Given the increasing trend of fire emissions, fire smokes may become the predominant source  
41 of PM<sub>2.5</sub> in the US in the future (Yue et al., 2013; Liu et al., 2016; Ford et al., 2018). Thus, an accurate prediction  
42 of fire emissions is imperative for investigating the impacts of historical and future fires on air quality, human  
43 health, and climate.

44 One of the widely used methods for predicting fire emission is process-based fire parameterization. These  
45 process-based models generally employ universal functions depicting non-linear relationships between fires and  
46 the input variables and apply the same functions to all grid cells in a model (Pechony and Shindell, 2009; Thonicke  
47 et al., 2010). In addition, the parameters of the process-based model are usually determined by empirical or  
48 statistical functions, assuming that the same parameters apply to all the regions or regions with limited fire  
49 observations (Crevoisier et al., 2007; Parisien et al., 2016). Recently, Zou et al. (2019) developed the Region-  
50 Specific ecosystem feedback Fire (RESFire) model that includes region- and PFT-specific fire parameterizations  
51 in subregions over the globe. Their model shows improved spatial distributions and temporal variations of fire  
52 activities compared to the CLM fire model. Process-based models are usually included in the dynamic global  
53 vegetation models (DGVMs) to simulate fire dynamics, vegetation dynamics, and biogeochemistry driven by  
54 atmospheric forcing and socio-economic data (Li et al., 2013; Knorr et al., 2016). Fire emissions, including trace  
55 gases and aerosols, are calculated from the simulated fire carbon emissions and the emission factors, with the  
56 former computed as the product of the burned area, fuel load, and combustion completeness. The process-based  
57 models in DGVM coupled with other components of Earth system models can be used to assess the impacts of  
58 environmental factors on fires and the feedback between fire emissions, land processes, and climate (Kloster et al.,  
59 2010). In 2014, the Fire Model Intercomparison Project (FireMIP) was initiated to compare nine DGVMs that  
60 include fire modules to better understand the performances of the global fire models (Rabin et al., 2017). The  
61 FireMIP enables comprehensive evaluation and comparison across various process-based models and provides a  
62 dataset of long-term fire simulations for regional and global analysis (Li et al., 2019; Hantson et al., 2020).

63 Besides process-based fire models, data-driven statistical models are also commonly used to estimate fire  
64 activities using relationships between fires and predictor variables. Multiple linear regression (MLR) is a popular  
65 simple statistical method used for fire modeling (Spracklen et al., 2009; Morton et al., 2013; Urbieto et al., 2015;  
66 Williams et al., 2019). MLR can achieve a good performance, but it fails to capture the non-linear relationships  
67 between fires and predictors, and it is sensitive to the collinearity and combinations of predictors (Littell et al.,  
68 2009). Unlike MLR, machine learning (ML) is a novel tool for advancing fire modeling, given its strengths in  
69 resolving the complex relationships between the target and predictor variables. Different ML approaches have been  
70 used to estimate fire occurrence, burned areas, or emissions at various time scales and spatial scales (Cortez and  
71 Morais, 2007; Aldersley et al., 2011; Dillon et al., 2011; Birch et al., 2015; Kane et al., 2015; Coffield et al., 2019;  
72 Wang and Wang, 2020). Even though ML models generally achieve higher accuracy than simple statistical models,  
73 their decision processes are often inscrutable, and hence lack interpretability. The development of explainable ML  
74 represents major advances for scientific applications beyond predictions (Gunning, 2017; Arrieta et al., 2020). For  
75 example, Wang et al. (2021) used the Extreme Gradient Boosting (XGBoost) algorithm and Shapley Additive  
76 explanation (SHAP) to predict wildfire burned area and revealed the relationships between burned areas and  
77 predictor variables. As process-based and data-driven models have their own advantages and weaknesses, as listed  
78 in Table 1, comparing these models and assessing their uncertainties in historical simulations and future projections  
79 are important. Yue et al. (2013) applied an MLR and a parameterization method to estimate burned areas in  
80 ecoregions of the western US and found that both models explained ~50% of the variance in the observed burned  
81 areas. Although they compared the burned areas estimated by the two methods and quantified their uncertainties in  
82 fire projections, both methods are only driven by meteorology while the effects of fuels and human activities are  
83 not considered.

84 The FireMIP dataset provides long-term simulations of multiple DGVMs with fire modules, allowing  
85 comparisons between process-based and data-driven models, with all models considering all the potential factors  
86 influencing fires, including climate, weather, vegetation, and human activities. This study aims to develop an ML  
87 model with game theory interpretation for fire emission prediction and to understand controls of fire emissions.  
88 The ML model and SHAP are then used to reveal the important factors controlling fire emissions and diagnosis the  
89 process-based FireMIP models. The ML model predicts the monthly  $PM_{2.5}$  emissions from fires during 2000-2020  
90 at a spatial resolution of  $0.25^\circ \times 0.25^\circ$  over the contiguous US (CONUS). It uses the XGBoost algorithm and  
91 incorporates various predictors, including local and large-scale meteorology, land surface characteristics, and  
92 socioeconomic variables, which are common input variables also used by the FireMIP models while some are  
93 specifically related to fire activities in CONUS. We acknowledge that different input variables between the ML  
94 and FireMIP models might cause additional uncertainty for comparison. This study aims to construct an ML model  
95 that predicts fire emissions over CONUS and utilize the ML model and SHAP to reveal the important factors  
96 contributing to fire emissions that might not be fully represented in the process-based models. In this context, the  
97 ML model and FireMIP models are optimized using different data or predictors at various scales, which enables us  
98 to use the ML to diagnose the performance of FireMIP models over CONUS through the comparisons of their  
99 performances and variable importance from the ML model. We evaluate and compare the predicted fire emissions  
100 from the ML and FireMIP models against the GFED fire emission product, focusing on spatial distributions,  
101 seasonality, and interannual variability over selected regions in CONUS. Additionally, the ML model and the  
102 SHAP importance are used to identify the important drivers of fire emissions in different regions and compare them  
103 with the corresponding parameterizations in the process-based models. Lastly, we compare the process-based and  
104 ML model performances in simulating extremely large fire emissions, including the spatial distributions of  
105 frequency and two case studies.

## 106 2. Data

### 107 2.1 Fire-induced $PM_{2.5}$ emission data

108 Monthly fire  $PM_{2.5}$  emission data is obtained from the Global Fire Emissions Database (GFED). GFED  
109 version 4 provides monthly burned area at  $0.25^\circ$  spatial resolution from 1997 to present, based on a combination of  
110 the MODIS burned area product with active fire data from the Tropical Rainfall Measuring Mission (TRMM)  
111 Visible and Infrared Scanner (VIRS) and Along-Track Scanning Radiometer (ATSR) family of sensors (Giglio et  
112 al., 2013). The GFED fire  $PM_{2.5}$  emissions are estimated by combining the burned area boosted by small fire burned  
113 area (Randerson et al., 2012) and the emission factors based on Akagi et al. (2011) with a revised version of the  
114 Carnegie-Ames-Stanford Approach (CASA) biogeochemical model that estimates fuel loads and combustion  
115 completeness for each monthly time step (van der Werf et al., 2017). The emission factors are dependent on the  
116 fire types, including savanna, boreal forest, temperate forest, tropical forest, and agriculture (van der Werf et al.,  
117 2017). We use the GFED fire  $PM_{2.5}$  emission as the target variable in the machine learning model development and  
118 for model evaluation.

119 To reduce spatial heterogeneity and help model learning, we apply the inverse distance weighting (IDW)  
120 (Bartier and Keller, 1996; Shepard, 1968) to interpolate the monthly gridded fire  $PM_{2.5}$  emission at  $0.25^\circ \times 0.25^\circ$ .  
121 The IDW method determines the value at a grid cell as the weighted average of the surrounding values within a  
122 search distance, with the weights proportional to the inverse of the distance raised to the power value  $p$ . Here we  
123 choose a value of 1 for  $p$  and a search distance of 35 km for IDW processing. Note that the total fire emitted  $PM_{2.5}$   
124 within a search distance after IDW processing is constrained to be the same as the original data. In this study, we  
125 only include grids with more than eight months of fire emissions larger than zero (in a total of 250 months),

126 encompassing 90% of the total fire emissions and ensuring sufficient data for the XGBoost model training. The  
127 interpolated fire emission is normalized based on its 21-year mean and standard deviation for each grid to reduce  
128 the skewness and improve data symmetry.

## 129 **2.2 Predictor variables**

130 We develop an empirical model at  $0.25^\circ \times 0.25^\circ$  grid resolution driven by various predictor variables at a  
131 monthly scale from January 2000 to October 2020. Given the datasets have different spatial resolutions, all the  
132 predictor variables are resampled to the spatial resolution of  $0.25^\circ \times 0.25^\circ$  by linear interpolation. The predictor  
133 variables used in the model along with their original spatial and temporal resolutions are included in Table 2. Most  
134 variables were also used in Wang et al. (2021) for developing an ML model of fire burned area over the contiguous  
135 U.S.

136 **Local meteorology:** Same as the local meteorological predictors used in Wang et al. (2021), we include monthly  
137 data of mean surface temperature, relative humidity (RH) at 2 m, daily precipitation, zonal (U) and meridional (V)  
138 components of wind at 10 m from the North American Regional Reanalysis (NARR) (Mesinger et al., 2006) and  
139 1000-hour dead fuel moisture (FM1000), Energy Release Component (ERC), and vapor pressure deficit (VPD)  
140 from the gridMET dataset (Abatzoglou and Kolden, 2013; Coffield et al., 2019). Drought is a natural phenomenon  
141 that influences fires through ignition efficiency, fuel availability, and fuel moisture. Thus, we include the monthly  
142 Standardized Precipitation Evapotranspiration Index (SPEI), a multiscale drought index based on climatic data  
143 (Vicente-Serrano et al., 2010). Given that lightning is one of the major ignition sources of fires and makes up  
144 approximately 75% of burned areas in western US (Pyne, 1984; Stephens, 2005), in this study, we add the cloud-  
145 to-ground (CG) lightning flash density from Severe Weather Data Inventory (SWDI) based on the National  
146 Lightning Detection Network (NLDN) (Cummins and Murphy, 2009; NOAA, 2006). The daily number of CG  
147 lightning flashes is summarized in  $0.1^\circ$  tiles and we aggregate the daily data to monthly scale.

148 **Large-scale meteorological patterns:** Large-scale meteorological patterns at a synoptic scale have been found to  
149 link to large fire events (Crimmins, 2006; Trouet et al., 2009; Zhong et al., 2020; Dong et al., 2021). Furthermore,  
150 it has been shown that including predictors of large-scale meteorological patterns conducive to wildfires  
151 significantly improves the prediction of burned areas over CONUS (Wang et al., 2021). Thus, we follow the  
152 methods developed by Wang et al. (2021) using the singular value decomposition (SVD) method to construct  
153 predictors representing the synoptic patterns driving fire emission variability. Note that the only difference between  
154 Wang et al. (2021) and this study is that they used wildfire burned area data and we use fire emissions to construct  
155 the SVDs. Three regions where large fires periodically occur are selected for constructing SVDs: Northern  
156 California, southern Rocky Mountains, and southeastern US, as defined in Wang et al. (2021). For each region, we  
157 calculate the daily mean fire  $PM_{2.5}$  emissions over the region and compute the day-to-day correlations between the  
158 regional mean fire  $PM_{2.5}$  emissions and the five gridded daily meteorological variables (surface temperature, 2-  
159 meter RH, U-wind and V-wind at 850 hPa, and geopotential height at 500 hPa) for all  $1^\circ \times 1^\circ$  grid cells within the  
160 large-scale domain, giving a correlation map for each meteorological variable. The correlation maps are then used  
161 to derive the SVD modes representing the large-scale meteorological patterns related to fires. Finally, we compute  
162 the monthly standard deviation of the daily SVD time series for the first two SVD modes, representing the month-  
163 to-month variations of synoptic fluctuations and atmospheric instability. The detailed methods and discussions

164 about the SVDs are provided in Wang et al. (2021). Overall, the identified SVDs for the three regions are similar  
165 to the SVDs in Wang et al. (2021) calculated using wildfire burned areas (Figs. S1-3).

166 **Land-surface properties:** We use the same set of variables in the burned area model that represent the effects of  
167 fuel and land surface states on fire emissions, including evapotranspiration (ET), surface soil moisture, land types,  
168 and topography (Wang et al., 2021). Monthly mean ET, vegetation fraction, and surface soil moisture are obtained  
169 from the North American Land Data Assimilation System (NLDAS-2) (Xia et al., 2012). Land cover data of the  
170 LAI classification scheme is obtained from the Terra and Aqua combined MODIS Land Cover Climate Modeling  
171 Grid (CMG) Version 6 data (Friedl, 2015). Since the land cover data is at yearly intervals from 2001 to 2020, we  
172 use the land cover data of 2001 for 2000. Topography data of slope and elevation is obtained from Amatulli et al.  
173 (2018).

174 Besides the above-mentioned variables that were also used in Wang et al. (2021), in this study, we consider  
175 the effect of fuel load on fire emissions, since fuel load is critical to fire emissions through its controls on fuel  
176 consumption and burned areas (Parks et al., 2012; Liu and Wimberly, 2015). As there are limited observations of  
177 fuel load, we use LAI to approximate the canopy bulk density, [which is important crown characteristics to predict](#)  
178 [crown fire spread](#), and vegetation fraction to represent the existing amount of vegetation (Keane et al., 2005). LAI  
179 is taken from MODerate resolution Imaging Spectroradiometer (MODIS) instruments (Myneni et al., 2015) and  
180 vegetation fraction is obtained from the NLDAS-2. [As LAI may not fully represent the available biomass, we also](#)  
181 [include fuel load simulated by Community Land Model \(CLM\). Monthly fuel load data from 2000 to 2015 is](#)  
182 [obtained from a simulation by CLM version 5 with biogeochemistry and prognostic crop, driven by atmospheric](#)  
183 [forcing from GSWP3v1 \(Lawrence et al., 2019\). The fuel load after 2015 is taken from a simulation under the SSP3](#)  
184 [\(shared socioeconomic pathways\) scenario. CLM fuel load is validated by comparing with the fuel-measured fuel](#)  
185 [load from the global fuel consumption database \(van der Werf et al., 2017; Van Leeuwen et al., 2014\), as shown in](#)  
186 [Fig. S4. The CLM-simulated fuel load is generally consistent with the measured fuel load for different vegetation](#)  
187 [types across CONUS based on the limited measurements.](#) Additionally, we include normalized fuel load as a  
188 predictor to capture the effects of temporal variation of fuel load, as the influence of fuel load on fire emissions is  
189 mainly attributed to its spatial variation rather than the temporal variation (Lasslop and Kloster, 2015).  
190

191 **Socioeconomic variables:** We use population density and gross domestic product (GDP) per capita to represent  
192 human effects on wildfires. The population density data is obtained from the Gridded Population of the World data  
193 collection (GPW V4) for the years 2000, 2010, 2015, and 2020, with a spatial resolution of 30 arc-second (CIESIN-  
194 Columbia University, 2017). The populations in other years are linearly interpolated between the abovementioned  
195 four years. The GDP per capita is taken from a gridded global dataset for 2000-2015 with a spatial resolution of 5  
196 arc minutes (Kummu et al., 2018). For the GDP after 2015, we use the data of 2015.

### 197 **3. Description of fire emission models**

#### 198 **3.1 Process-based fire emission models**

199 The Fire Model Intercomparison Project (FireMIP) includes a set of common fire modeling experiments  
200 from nine DGVMs driven by the same forcing data, allowing a better understanding of global fire models (Rabin  
201 et al., 2017). The FireMIP dataset provides global gridded burned area fraction and fire emissions, including carbon  
202 and 33 species of trace gases and aerosols over 1700-2012. Nine DGVMs with different fire modules are included

203 in FireMIP, including Community Land Model version 4.5 (CLM4.5) with the CLM5 fire module, Canadian  
 204 Terrestrial Ecosystem Model (CTEM), Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg with Spread  
 205 and InTensity fire model (JSBACH-SPITFIRE; hereafter referred to as JSBACH), Joint UK Land Environment  
 206 Simulator with Interactive Fire And Emission Algorithm For Natural Environments (JULES-INFERNO; hereafter  
 207 referred to as JULES), Lund-Potsdam-Jena General Ecosystem Simulator with Global FIRE Model (LPJ-GUESS-  
 208 GlobFIRM; hereafter referred to as LPJ-Glob), LPJ-GUESS with SIMple FIRE model and Blaze-Induced Land-  
 209 Atmosphere Flux Estimator (LPJ-GUESS-SIMFIRE-BLAZE; hereafter referred to as LPJ-SIM), LPJ-GUESS with  
 210 SPITFIRE model (LPJ-GUESS-SPITFIRE; hereafter referred to as LPJ-SPI), MC2, and Organizing Carbon  
 211 Hydrology In Dynamic Ecosystems with SPITFIRE model (ORCHIDEE-SPITFIRE; hereafter referred to as  
 212 ORCHIDEE) (Rabin et al., 2017).

213 The nine DGVMs in FireMIP are driven by the CRU-NCEP v5.3.2 atmospheric forcing data with a spatial  
 214 resolution of  $0.5^\circ$  and a 6-hourly temporal resolution (Wei et al., 2014; Rabin et al., 2017). Other forcing data,  
 215 including annual global atmospheric  $\text{CO}_2$  concentration, land use and land cover, and population density from 1700  
 216 to 2012 is taken from various data sources (Klein Goldewijk et al., 2010; Hurtt et al., 2011; Le Quéré et al., 2014).  
 217 Monthly cloud-to-ground lightning frequency with a resolution of  $0.5^\circ \times 0.5^\circ$  over 1901-2012 is calculated based  
 218 on the observed relationship between present-day lightning and convective available potential energy (CAPE)  
 219 anomalies (Pfeiffer et al., 2013). Fire emissions in FireMIP are calculated considering the fire carbon emissions  
 220 and vegetation characteristics based on the plant functional type (PFT) from the FireMIP historical transient control  
 221 run (SF1). SF1 breaks the simulation period into three phases: the spin-up phase in 1700, the transient phase in  
 222 1701-1900, and the transient phase in 1901-2012 (see the detailed descriptions and model settings in Rabin et al.,  
 223 2017, Li et al., 2019, and Hantsan et al., 2020). In the 1901-2012 transient phase, the models are driven by time-  
 224 varying atmospheric forcing,  $\text{CO}_2$  concentration, LULCC, population density, and lightning data. Note that the  
 225 MC2 and CTEM runs start from 1901 and 1861, while the rest of the models start from 1700. As the spatial  
 226 resolutions of the FireMIP models are different, the regridded model outputs with  $1^\circ \times 1^\circ$  resolution obtained from  
 227 Li et al. (2019) are used to compare with the GFED data and the ML model.

### 228 3.2 ML-based approach: An eXtreme Gradient Boosting (XGBoost) model

229 The eXtreme Gradient Boosting (XGBoost) is a decision-tree-based ensemble machine learning method  
 230 using the gradient boosting approach (Chen and Guestrin, 2016). The XGBoost model builds multiple decision  
 231 trees that are added subsequently and learn the errors of the previous tree to reduce the loss and obtain the best  
 232 prediction. Unlike the gradient boosting machine (GBM) that also uses the gradient boosting approach, XGBoost  
 233 utilizes a more regularized model formalization to prevent over-fitting and improve the computational efficiency.  
 234 The formula for the prediction at step  $t$  and grid location  $i$  can be defined as follows:

$$235 \quad \hat{y}_i^t = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)$$

236 where  $f_t(x_i)$  is the tree model at step  $t$ ,  $\hat{y}_i^t$  and  $\hat{y}_i^{(t-1)}$  are the predictions at steps  $t$  and  $t-1$ , and  $x_i$  are the predictor  
 237 variables. The parameters of the model  $f_t(x_i)$  are selected by optimizing the objective function that measures how  
 238 well the model fit the training data:

$$239 \quad Obj^t = \sum_{i=1}^n L^t + \Omega^t$$

240 which is composed of the loss function  $L^t$  and the regularizing term  $\Omega^t$  in each step.  $L_t$  is defined as  $l(y_i, \hat{y}_i^{t-1} +$   
 241  $f_t(x_i))$  and  $\Omega^t$  is defined as  $\gamma T + \frac{1}{2}\lambda\|\omega\|^2$ , where  $\gamma$  is the regularization term which penalizes the number of  
 242 leaves in the tree  $T$  and  $\lambda$  is the regularization term which penalizes  $\omega$ , the weights of different leaves.

243 We use grid search to choose the set of suitable hyperparameters and achieve the best ML model  
 244 performance. Grid search is a tuning technique for computing the optimal values of hyperparameters considering  
 245 a range of numbers with a given increment. The parameter set that yields the best 5-fold cross-validation score is  
 246 selected as the final set of hyper-parameters. The considered hyper-parameters, their search domains, and the final  
 247 values are denoted in Table S1.

248 The 10-fold cross-validation (CV) technique is applied to evaluate the model and avoid overfitting. First,  
 249 we randomly divide the fire emission dataset (2000-2020 over CONUS) into ten equal-sized splits. Then, we train  
 250 the model with nine splits of the data and use the trained model to predict fire emissions for the remaining one split.  
 251 This process is repeated ten times for each split. Finally, the predictions are evaluated by grids and regions using  
 252 root mean square error (RMSE), correlation coefficient (R), and the index of agreement (IoA). The IoA represents  
 253 the ratio of the mean square error and the potential error, and the value closer to 1 indicates better agreement.

### 254 3.3 Shapley additive explanations (SHAP)

255 We utilize the SHAP to identify the relative importance of the predictor variables. SHAP is a novel  
 256 approach to resolve and explain variable importance based on game theory (Lundberg and Lee, 2017). Within the  
 257 scope of game theory, the goal is a prediction for a single observation. Each predictor variable is referred to as a  
 258 “player” in this game and contributes to the goal (“payout”). For each predictor, the SHAP variable importance  
 259 measures the marginal contribution considering all possible combinations of the predictor variables. The marginal  
 260 contribution is calculated by comparing the differences between the model fit  $f_x(S \cup \{i\})$  including the predictor  $i$   
 261 and another model fit  $f_x(S)$  without predictor  $i$ . When there is more than one predictor  $i$ , the marginal contribution  
 262 also depends on the interactions with other predictors. Thus, the calculation repeats considering the whole set of  
 263 the predictors. The final contribution  $\phi_i$  of predictor  $i$  is the weighted average of all marginal contributions:

$$264 \quad \phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(F - |S| - 1)!}{F!} [f_x(S \cup \{i\}) - f_x(S)]$$

265 where  $F$  is the total number of features,  $S$  is the subset of predictors from all predictors except for predictor  $i$ ,  
 266  $\frac{|S|!(F - |S| - 1)!}{F!}$  is the weighting factor counting the number of permutations of the subset  $S$ .  $f_x(S)$  is the expected  
 267 output given the predictors subset  $S$ .  $[f_x(S \cup \{i\}) - f_x(S)]$  is the difference made by predictor  $i$ .

268 Compared to the commonly used feature importance, such as gain, or split count, SHAP is more consistent  
 269 and faithful to the model (Lundberg et al., 2019). More importantly, SHAP provides local importance that measures  
 270 the variable importance for each sample, while most of the feature importance metrics only have global importance  
 271 that measures variable contributions limited to the entire dataset. The global importance by SHAP is the average  
 272 of the absolute SHAP values for each predictor, providing an overall picture of the predominant variables  
 273 controlling fire emissions in CONUS. The local importance will be used to identify the important predictors for  
 274 large fire events in the ML model and diagnose the deficiency of the process-based models.

## 275 4. Results

### 276 4.1 XGBoost model performance and variable importance

277 Table 3 shows the whole CONUS and regional model performance, including RMSE, IoA, and correlation.  
278 The model performs well at grid level over CONUS, with an RMSE of 0.16 g/m<sup>2</sup> and an IoA of 0.84. Figure 1a  
279 shows the map of correlation between the observed and predicted monthly fire emission time series for each grid  
280 over CONUS. Overall, the results indicate the ML model can reproduce the interannual variability of fire emissions  
281 at 0.25° resolution over CONUS, with a mean correlation of 0.58 and more than 70% of the grids having  
282 correlations larger than 0.4. To better assess model performance in different regions, Table 3 summarizes the model  
283 performance for several selected regions: (1) western forest area, (2) Mediterranean California, (3) southwestern  
284 US, and (4) southeastern US (color boxes in Fig. 1a). The regions where fires frequently occur are selected by the  
285 similarity of ecoregions, vegetation types, and fire regimes. Figs. 1b-e show the time series of observed and  
286 predicted fire PM<sub>2.5</sub> emissions averaged over several regions. Generally, the ML model reproduces the interannual  
287 variability of fire emissions for the selected regions ( $r=0.84-0.98$ ). Among these regions, Mediterranean California  
288 has the smallest correlation coefficient and largest RMSE compared to other regions, which can be explained by  
289 the fact that fires in this region interact with multiple factors, including human activity, complex terrain, and Santa  
290 Ana winds (Syphard et al., 2008; Yue et al., 2014). The interactions between fires and these factors pose  
291 uncertainties and challenges in fire prediction over this region. It is also worth noting that the ML model captures  
292 the large fire events in September 2020 in Oregon and California but underestimates the peak values by ~30%  
293 (Figs. 1b and 1c). In addition, we also test the ML model's ability to provide accurate predictions on unseen data  
294 (i.e., generalization) by using data from 2000 to 2019 as a training set and data from 2020 as a testing set. As shown  
295 in Fig. S5, the ML model can reproduce the spatial patterns of fire emissions well but underestimates the emissions  
296 of the peak in September 2020. The results are within our expectations because the ML model generally fails to  
297 make accurate predictions for the data outside of the training domain or has large uncertainties in extrapolation  
298 (Tsubaki and Mizoguchi, 2020; Hooker, 2004). Since 2020 features the largest fire emissions in the study period,  
299 we conducted another test using 2000-2017 and 2019-2020 to train the ML model and test on the data of 2018. We  
300 selected 2018 because 2018 had the largest fires on record before 2020. The ML successfully reproduces the  
301 temporal variability of fire emissions ( $r=0.92$ ) and captures the peak in Aug 2018, as well as the spatial distributions  
302 of fire emissions ( $r=0.52$ ).

303 To improve understanding of the ML prediction, we utilize the SHAP method to quantify the contributions  
304 of each predictor variable to the prediction and identify the key contributing factors of fire PM<sub>2.5</sub> emission. SHAP  
305 importance is chosen because it provides not only global importance but also local importance that helps understand  
306 which variables have larger contributions to specific events or regions. Here, we first demonstrate the global  
307 importance that considers all the samples. Fig. 2 shows the 20 most important variables for the model ranked by  
308 the absolute mean SHAP values. The SHAP value for a feature indicates its contribution to the prediction, so larger  
309 absolute mean SHAP values indicate larger contributions to the fire emissions. Among the top 10 variables, seven  
310 of them are local meteorological variables, indicating local meteorology is the predominant control of fire  
311 emissions, as these variables control fire activity directly (Liu and Wimberly, 2015; Abatzoglou et al., 2016; Wang  
312 et al., 2021). Besides local meteorology, the predictors of large-scale meteorology (SVD1\_SElag2 and  
313 SVD2\_SElag2) are identified as the eighth and tenth important variables, showing that meteorology is not only  
314 important at local scale but also at synoptic scale (Trouet et al., 2009; Pollina et al., 2013; Dong et al., 2021).  
315 Finally, in addition to meteorology, fuel load is identified as the fifth important variable in the model, as fuel load  
316 affects emission through controlling burned area and fuel consumption (Seiler and Crutzen, 1980). Considering the  
317 important variables in different regions, the selected regions in western US (western forest area, Mediterranean  
318 California, and southwestern US) generally share the common top 10 variables (Fig. S6). Over western US,

319 predictors controlling fuel dryness and fuel amount, including RH, fuel moisture (FM1000), ERC, vegetation  
320 fraction, and fuel load, contribute more to fire emissions. On the other hand, large-scale meteorological patterns  
321 (SVDs\_SElag2) are more important for fire emissions in southeastern US.

322 As the dominant drivers differ for different temporal scales, we aggregate the monthly SHAP values to  
323 obtain annual and seasonal time series of SHAP values for each variable. The annual and seasonal time series are  
324 the averaged SHAP values over the study period for each year and month, respectively. Fig. S7 shows the mean  
325 |SHAP| values at seasonal and interannual time scale for the whole CONUS. Considering both the mean |SHAP|  
326 and larger correlations ( $r > 0.5$ ) between the annual/seasonal time series of SHAP and mean fire emissions,  
327 temperature, VPD, RH, and ERC are the dominant variables controlling the seasonal variation of fire emissions.  
328 These factors have relatively stronger seasonality than other variables (e.g., VPD is usually higher in the summer).  
329 On the other hand, large-scale circulation patterns, including SVD1\_SElag2, SVD2\_SElag2, and SVD1\_RM, are  
330 important variables controlling both the seasonal and interannual variability of fire emissions, while SVD2\_RM  
331 and SVD2\_NCA mainly control interannual variability. Some identified large-scale meteorology has significant  
332 seasonality (e.g., SVDs\_SElag2 are predominant in spring and SVD1\_RM is strongest in summer), and most of  
333 them have interannual variability, as shown in Fig. 8. Overall, the SHAP analysis shows different dominant  
334 predictors for fire emissions at various time scales.

## 335 4.2 General comparison between GFED, ML, and FireMIP models

336 This section compares the performance of the ML and FireMIP models benchmarked against observations  
337 from GFED, and the evaluations are based on spatial distributions, seasonality, and interannual variability of fire  
338  $PM_{2.5}$  emissions. Since the spatial resolutions of the GFED data, ML models, and FireMIP models are different,  
339 they are all regridded to  $1^\circ \times 1^\circ$  using bilinear interpolation. Note that the simulation period of FireMIP models  
340 ends in 2012, so we use the overlapping period of 2000–2012 for comparison and exclude the MC2 model because  
341 its simulation ends in 2008.

### 342 4.2.1 Spatial distributions of fire $PM_{2.5}$ emissions and sensitivities to RH and temperature

343 Fig. 3 compares the observed and simulated spatial distributions of long-term mean monthly fire  $PM_{2.5}$   
344 emissions averaged over 2000–2012. Among the models, the ML model, CLM, and JULES have better performance  
345 in reproducing the spatial distributions of fire emissions over CONUS, with a correlation coefficient of 0.83, 0.52,  
346 0.40, respectively. The ML model shows the best agreement with GFED, though it overestimates fire emissions  
347 over Northern California. Both CLM and JULES simulate more  $PM_{2.5}$  emissions over southeastern US, and JULES  
348 overestimates fire emissions in Northern California. Some other models, such as CTEM, JSBACH, and LPJ-SIP,  
349 tend to overestimate fire emissions over central US (e.g., Great Plains and Texas). LPJ-SIP captures the hotspots  
350 of fire emissions over western US and southeastern US, but it simulates much more  $PM_{2.5}$  emissions over the Rocky  
351 Mountain and northeastern US. In terms of the total amount of  $PM_{2.5}$  emissions, all models except ORCHIDEE-  
352 SPITFIRE overestimate  $PM_{2.5}$  emissions (8.33–79.49 Tg), compared to the GFED estimate of 4.98 Tg during 2000-  
353 2012 over CONUS (Table 4).

354 The overestimations in some models may be explained by the sensitivities of fire emissions to individual  
355 meteorological variables. Fig. 4 shows the slopes for the dependence of annual mean fire  $PM_{2.5}$  emissions on annual  
356 mean RH from the CRUNCEP atmospheric forcing data for GFED and the ten models based on linear regression.  
357 Since the ML model uses NARR meteorology as predictors, we also include sensitivities of the fire emissions  
358 predicted by the ML model to the NARR meteorology (Fig. 4b). Almost all models capture the negative dependence  
359 of  $PM_{2.5}$  emissions on RH over western US ( $r = -0.06 \sim -0.84$ ), but the sensitivities in the models are much stronger  
360 (steeper negative slope) than in GFED. For temperature, positive sensitivity is shown over western US in GFED

361 (Fig. 5), with the most significant slope in northern California. The sensitivities to temperature in models agree  
362 with the observed sensitivities ( $r=-0.06\sim 0.64$ ), but some models show much stronger sensitivities over western,  
363 central, and southeastern US. Generally, the spatial distributions of the long-term mean fire emissions shown in  
364 Fig. 3 match well with the spatial distributions of sensitivities to RH or temperature, suggesting an important role  
365 of the sensitivities in the model biases of predicting fire emissions. However, the correspondence of large fire  
366 emissions to the sensitivities to RH or temperature shows regional differences. For instance, in western US, the  
367 stronger sensitivities to both RH and temperature correspond to the overestimations in this region for most models,  
368 including the ML model, CLM, CTEM, JSBACH, JULES, LPJ-SIM, and LPJ-SPI (Figs. 4 and 5). On the other  
369 hand, over central US, larger  $PM_{2.5}$  emissions simulated by CTEM and JSBACH only correspond to stronger  
370 sensitivity to temperature (Fig. 5). Similar to central US, in southeastern US, the overestimations in CLM and  
371 JULES only correlate with stronger sensitivity to temperature (Fig. 5). Regional differences in the correspondences  
372 between the predicted fire emissions and their sensitivity to meteorology can be explained by several factors. For  
373 western US, the overestimations of fire emissions correspond to both stronger sensitivities to RH and temperature,  
374 given that fire activities are sensitive to fuel aridity that is controlled by temperature and fuel moisture (Abatzoglou  
375 and Williams, 2016; Holden et al., 2018). As for southeastern US, fuels in this region typically burn at higher RH  
376 and the interannual RH variation (standard deviation) is smaller (Balch et al., 2017; Brey et al., 2018). With higher  
377 RH values and less variation in RH, the fire emissions in southeastern US show weaker sensitivity to RH than to  
378 temperature in observation (Table S2). The above analysis shows that the overestimation of fire emissions in the  
379 models may be attributed to the stronger sensitivities to meteorology. However, fire activities are controlled by  
380 meteorology and other factors such as vegetation and human, so the analysis of fire emission sensitivity to  
381 meteorology only provides a potential explanation to the overestimation of fire emissions in the models (Forkel et  
382 al., 2019).

#### 383 4.2.2 Seasonality and interannual variability over CONUS

384 In addition to evaluating spatial distributions, it is also important to compare the models' ability to  
385 reproduce the temporal variability of fire emissions. As the models may systematically over- or underestimate fire  
386 emissions, we normalize the emissions by the mean and standard deviation and focus only on its temporal  
387 variability. Fig. 6a shows the seasonality of normalized fire  $PM_{2.5}$  emission over CONUS. Most models capture the  
388 seasonality of fire emission successfully ( $r>0.85$ ), except LPJ-SIM which simulates peak emission in August-  
389 October ( $r=0.65$ ). Among the models, the ML model has the highest correlation coefficient between prediction and  
390 observation from GFED ( $r=0.98$ ) and successfully reproduces the peak in August. The seasonal peaks simulated  
391 by the FireMIP models are broader and flatter than the peak in GFED, with an early peak in June-July continuing  
392 to September (Fig. 6a).

393 In terms of interannual variability (Fig. 6b), the ML model, CLM, and JULES perform better than other  
394 models, with larger correlation coefficient between simulated and observed fire  $PM_{2.5}$  emissions ( $r=0.87, 0.71,$  and  
395  $0.55$  for ML, CLM, and JULES, respectively; Table 4). Other models have relatively poor performance in capturing  
396 the interannual variability. The interannual variability of fire emissions shows several peaks in 2002, 2007, and  
397 2012 (black line in Fig. 6b), when western US contributes 76% of the total emissions to the peaks in these years.  
398 Almost all models except ORCHIDEE capture the peak in 2012. However, most models miss the peaks in 2002  
399 and 2007. Among all models, LPJ-Glob model simulates the peaks in the two years, while ML, JULES, and CLM  
400 only capture the largest emission in 2007 (Fig. 6b).

### 401 4.2.3 Seasonality and interannual variability by regions

402 As the temporal variability of fire activities varies by region, we compare the performance between GFED  
403 and the ML and FireMIP models by the regions defined in Sec. 4.1. Fig. 7 shows the seasonality and interannual  
404 variability of normalized fire  $PM_{2.5}$  emission over western forest area, Mediterranean California, southwestern US,  
405 and the southeastern US. All models generally capture the seasonality of the western forest area peaking in summer,  
406 with correlation coefficients larger than 0.8 (Table 4). Even though the FireMIP models generally reproduce the  
407 peaks in summer, the predicted peaks are broad and flat, indicating a relatively longer fire season starting in June  
408 and ending in September (Fig. 7a). When looking at the interannual variability, we find that the ML model has the  
409 best performance with a correlation coefficient of 0.93, and it successfully captures the largest fire emission in  
410 2007. CLM, JULES, and LGJ-Glob perform better than the rest of the models ( $r=0.70, 0.60,$  and  $0.51$  for CLM,  
411 JULES, and LPJ-Glob, respectively; Table 4), but all of them still miss the peaks in 2007 and overestimate fire  
412 emissions in 2001 and 2003 (Fig. 7b). The emission peak in 2007 is mainly attributed to the large fires in Idaho,  
413 which were associated with synoptic weather patterns characterized by positive geopotential height and temperature  
414 anomalies over the Pacific Coast and western US (Zhong et al., 2020). Consistent with prior findings, SHAP  
415 importance shows that in the ML model SVD predictors (SVD\_NCA and SVD\_RM in July and August 2007 Fig.  
416 8a) are the dominant factors of fire emissions in 2007 (contribute 27% and 28% for July and August 2007,  
417 respectively), which are characterized by high pressure, low RH, and northeasterly winds over western US (Figs.  
418 S1 and S2). Thus, the underestimation of peak emission in 2007 may be explained by the fact that the influences  
419 of large-scale meteorology on fire activity are not fully considered in the FireMIP models, which are point models  
420 driven only by local atmospheric forcing.

421 In Mediterranean California, the seasonality of fire emissions shows a bimodal pattern, peaking in August  
422 and October. The peak in October is mainly due to the extremely large fires associated with Santa Ana winds in  
423 2003 and 2007 (Keeley et al., 2009; Yue et al., 2014). The ML model simulates a flatter peak from July to October,  
424 while all the FireMIP models except ORCHIDEE capture the first emission peak in summer but fail to simulate the  
425 large fire emission in October (Fig. 7c). The underestimation associated with the Santa Ana winds is also shown in  
426 the interannual time series in Fig. 7d. Several models, including LPJ-Glob, CTEM, LPJ-SPI, and JULES, capture  
427 the peak in 2007 but only the ML model predicts both peaks in 2003 and 2007 even though the peak in 2003 is  
428 underestimated. According to the SHAP importance from the ML model, the peak emissions in October 2003 and  
429 October 2007 are mainly contributed by the SVD predictors and ERC (SVD2\_NCA and SVD1\_RM together  
430 contribute 20% to the fire emissions for October 2003 and SVDs\_SElag2 and SVD2\_RM together contribute 31%  
431 to the fire emissions for October 2007) and ERC (15% and 18% for October 2003 and 2007, respectively) (Fig.  
432 8b). The results indicate that the ML model captures the effect of synoptic weather patterns on fire activity by  
433 including the SVD predictors. Even though the wind speed is included in simulating fire spread in the FireMIP  
434 models, the spatial resolutions of the models and/or the atmospheric forcing data may not be fine enough to resolve  
435 the strengthened offshore winds through the complex terrain, and subsequently, they may not well capture the  
436 effects of Santa Ana winds on fires. [As shown in Fig. S8, the wind speeds from NARR are significantly larger than  
437 from CRUNCEP for the strong wind days \(daily wind speed > 4.5 m/s\) over southwestern California \(116-119 °W,  
438 32.6-34.8 °N\) during 2000-2012 October as well as the during Oct 2003 and 2007 \(Table S3\). The results indicate  
439 the lower wind speeds in the CRUNCEP atmospheric forcing used in FireMIP may partially explain the model  
440 biases for the events associated with Santa Ana winds.](#) Besides the above-mentioned shortfall, all the models have  
441 problems reproducing the interannual variability of the fire emissions over Mediterranean California, with very  
442 low correlations ( $r<0.25$ ) for the FireMIP models and a relatively low correlation ( $r=0.72$ ) for the ML model (Table  
443 4; Fig. 7d). The poor performance for this region may be due to the complex relations between fires and multiple  
444 factors, including meteorology, complex terrain, fuel, and human, which may not be fully represented in the models  
445 (Mann et al., 2016; Radeloff et al., 2018).

446 Both the ML model and LPJ-SIM successfully reproduce the seasonality of fire emission in southwestern  
447 US peaking in June ( $r=0.99$  and  $0.94$  for ML and LPJ-SIM, respectively), while other models simulate relatively  
448 smooth seasonality (Fig. 7e and Table 4). The ML model, LPJ-SIM, and ORCHIDEE have better performance for  
449 the interannual variability, with correlation coefficients of  $0.95$ ,  $0.40$ , and  $0.45$ , respectively (Table 4). However,  
450 most FireMIP models show larger variability in fire emissions than the GFED, and they all fail to capture the  
451 extremely large fire emission in 2011 (Fig. 7f). The peak fire emission in 2011 over southwestern US was caused  
452 by extremely low atmospheric moisture along with moderately high temperature, leading to record-breaking VPD  
453 and wildfire activities (Williams et al., 2015). To explain why the FireMIP models fail to capture the peak of 2011,  
454 we compare the VPD calculated from CRUNCEP data and the VPD data from gridMET used in the ML model. As  
455 Fig. S9 shows, CRUNCEP shows smaller positive anomalies of VPD over southwestern US in 2011 summer, while  
456 gridMET data demonstrates a significantly larger VPD anomaly. The biases in CRUNCEP data may partially  
457 explain the underestimations in all FireMIP models.

458 For southeastern US, the seasonal cycle of fire  $PM_{2.5}$  emissions displays a bimodal pattern, peaking in  
459 spring (March-April) and fall (September and October) (Fig. 7g). Most models fail to reproduce the bimodal fire  
460 emissions, but the ML model, LPJ-SIM, and LPJ-SPI can capture the bimodal pattern. Although LPJ-SIM and LPJ-  
461 SPI predict the bimodal peaks, the first peak simulated by LPJ-SIM shows a one-month delay, and the second peak  
462 simulated by LPJ-SIM and LPJ-SPI is one month early and one month late, respectively (Fig. 7g). In addition, the  
463 ML model, CLM, and JSBACH reproduce the interannual variability of fire  $PM_{2.5}$  emissions relatively well ( $r=0.96$ ,  
464  $0.57$ , and  $0.72$  for the ML model, CLM, and JSBACH, respectively) (Table 4 and Fig. 7h). Interestingly, CLM and  
465 JSBACH can capture several peaks in 2007, 2010, and 2011 but they do not simulate seasonality correctly, which  
466 may be explained by the underestimation in spring compensated by the overestimations in summer related to  
467 abnormal dryness or drought.

### 468 **4.3 Performance in modeling extreme events**

469 Fire activity in the US is becoming more hazardous, particularly over western US, due to more frequent  
470 hotter and drier conditions as climate continues to warm (Williams et al., 2019). Thus, it is necessary to assess  
471 whether the ML model and process-based models can capture the extreme events in terms of their magnitude,  
472 frequency, timing, and location, which is essential to future projection and adaptation. As CLM performs relatively  
473 well among the FireMIP models, we select CLM for comparison with the ML model at  $1^\circ \times 1^\circ$  resolution, focusing  
474 on the spatial patterns of extreme event frequency and two case studies with extremely large fire emissions.

#### 475 **4.3.1 Frequency of extreme event occurrence**

476 Fig. 9 shows the frequency maps of months with large fire emissions during 2000-2012 for GFED, the ML  
477 model, and CLM. Large fire emission is defined as monthly fire  $PM_{2.5}$  emissions greater than the 95<sup>th</sup> percentile of  
478 fire  $PM_{2.5}$  emission considering all the grids over CONUS in 2000-2012. GFED shows hot spots with a higher  
479 frequency over northern California, the Pacific Northwest, and southeastern US, with total counts ranging from 15  
480 to 105 (Fig. 9a). The ML model captures the spatial patterns ( $r=0.74$ ), but it overestimates the number of months  
481 by a factor of two to three compared to GFED, especially over western US (Fig. 9b). The spatial patterns of large  
482 fire emission occurrence simulated by CLM are generally consistent with the observed distribution by GFED  
483 ( $r=0.35$ ). However, it overestimates the frequency, particularly over Idaho and northeastern US, and simulates more  
484 significant numbers of months with extreme events over large spatial extents, may be due to its coarse spatial  
485 resolution (Fig. 9c).

487 To evaluate how well the models simulate the large fire emissions, we compare model performance for the  
488 two recent cases reported to be the largest fire events during 2000-2012, including the fires in southern US in 2011  
489 and western US in 2012. During 2011, a severe drought leading to large wildfires was observed over southern US,  
490 including Arizona, New Mexico, and Texas (NOAA, 2012; Wang et al., 2015). Fig. 10 shows the maps of annual  
491 mean fire  $PM_{2.5}$  emissions over southern US from GFED, the ML model, and CLM. GFED shows the largest fire  
492 emissions close to the border of Arizona and New Mexico in conjunction with other small hotspots over New  
493 Mexico, Texas, and Louisiana (Fig. 10a). The ML model overall reproduces the spatial distributions of the fire  
494 emissions ( $r=0.96$ ) and captures the largest fire emission in Arizona and New Mexico in 2011 (Fig. 10b). However,  
495 CLM does not capture the hotspots observed in GFED over Arizona and New Mexico but simulates larger fire  
496 emissions in Louisiana instead (Fig. 10c). In terms of the time series, the ML model reproduces the temporal  
497 variability of fire emissions and successfully captures the peak of total fire  $PM_{2.5}$  emissions in June 2011 ( $r=0.98$ ;  
498 Figs. 10d and 10e). Although CLM simulates the peak in June, it overestimates fire emissions in the following  
499 months by a factor of 4 ( $r=0.52$ ; Figs. 10d and 10e).

500 In 2012, western US experienced several major wildfires (NOAA, 2013). The warm and dry conditions led  
501 to large wildfires in California, Oregon, New Mexico, and Colorado (Fig. 11). Both the ML model and CLM  
502 capture the hotspots with large fire emissions (Fig. 11b and 11c) and have correlation coefficients of 0.56 and 0.37,  
503 respectively. However, the ML model tends to overestimate fire emissions, especially in areas surrounding the  
504 grids with extremely large fire emissions (Fig. 11b). CLM misses some large fire emissions in Colorado and New  
505 Mexico and underestimates the larger fire emissions in several hotspots (Fig. 11a), which may be explained by its  
506 coarse resolution. The time series of normalized fire  $PM_{2.5}$  emissions in 2012 show one peak in August. The ML  
507 model captures the peak and presents a high correlation between the simulated and observed normalized and total  
508  $PM_{2.5}$  fire emissions ( $r=0.98$ ). CLM captures the peak in August but overestimates emissions in September and  
509 October ( $r=0.84$ ; Figs. 11d and 11e). [To test model generalization, we train the model using data of 2000-2009 and  
510 2013-2020 and test on 2010-2012 and compare the ML performance with CLM \(Figs. S10-11\). The ML model  
511 performance is as good as the 10-fold cross-validation, demonstrating that the ML model performs well on  
512 predicting unseen data.](#)

## 513 5. Discussion and conclusions

514 This study provides the first assessment to evaluate the performance of data-driven and process-based  
515 models in predicting fire  $PM_{2.5}$  emissions over CONUS. We first demonstrate that the developed ML model  
516 performs well in predicting monthly fire  $PM_{2.5}$  emissions nationwide at grid cells of  $0.25^\circ \times 0.25^\circ$  resolution from  
517 2000 to 2020, with an RMSE of  $0.16 \text{ g/m}^2$  and IoA of 0.84. The ML model outperforms prior statistical models  
518 predicting fire activities at similar spatial and temporal scales (Carvalho et al., 2008; Bedia et al., 2014).  
519 Considering the performance at a regional scale, the ML model reproduces the interannual variability of fire  
520 emissions for the selected regions, with correlation coefficients ranging from 0.84 to 0.98. Therefore, the ML model  
521 has a promising performance in predicting fire emission over CONUS at a relatively fine spatial resolution.  
522 Compared to the wildfire burned area model in Wang et al. (2021), the fire emission model in this study shows  
523 slight degradations in capturing the interannual variability of fire emission at grid level (e.g., percentage of grids  
524 with correlations larger than 0.4). This may be explained by the fact that the fire emission model may not effectively  
525 resolve the relationships between fires and predictors when more grids with less fire occurrence are included (i.e.,  
526 more zeros or unburned grids) without reliable information about ignition. As a side note, both burned area and  
527 emission ML models have relatively poor performance over Mediterranean California, indicating the challenges in

528 modeling fire activities in this region where the terrain and land use are complex. The SHAP variable importance  
529 shows that meteorology at both local and synoptic scale as well as fuel loads are important variables controlling  
530 fire emissions over CONUS. Regional analysis of predictors indicates that fuel dryness such as fuel moisture and  
531 energy release component (ERC) and fuel load are important for predicting fire emissions in western US, while  
532 large-scale meteorological patterns (SVDs\_SElag2) contribute more to fire emissions in southeastern US.

533 We then compare the simulated fire PM<sub>2.5</sub> emissions from the ML model and FireMIP models against  
534 GFED from 2000 to 2012 at the spatial resolution of 1° × 1°. The ML model, CLM, and JULES reproduce the  
535 spatial distribution more reasonably than the rest of the FireMIP models (r=0.83, 0.52, and 0.40 for the ML, CLM,  
536 and JULES, respectively). Both CLM and JULES simulate more fire PM<sub>2.5</sub> emissions over southeastern US, which  
537 can be explained by several reasons. First, it has been shown that the satellite-observed burned areas in southeastern  
538 US are much smaller than the burned areas estimated from the ground-based fire records, which might have resulted  
539 from the small prescribed and agricultural fires (Hu et al., 2016; Nowell et al., 2018). In addition, large differences  
540 exist among different satellite estimated fire PM<sub>2.5</sub> emissions in southeastern US (Li et al., 2019). As a consequence,  
541 these studies highlighted uncertainties about the GFED estimated burned area and emission over southeastern US.  
542 Second, cropland fires are one of the predominant fire types in this region. Among the FireMIP models, CLM is  
543 the only model that simulates cropland fires (Li et al., 2013). For JULES, even though it does not simulate cropland  
544 fires, it treats croplands as natural grasslands. The emission factors of grasslands and croplands used in the FireMIP  
545 models are larger than in GFED4s, thereby causing larger fire PM<sub>2.5</sub> emissions in southeastern US in CLM and  
546 JULES (van der Werf et al., 2017; Li et al., 2019). Furthermore, Li et al. (2019) noted that CLM4.5 simulates  
547 higher fuel loads in croplands than the CASA model used by GFED4s, leading to higher fire carbon emissions  
548 estimated by CLM than by GFED. It is worth noting that the ML model incorporates fuel load simulated by CLM4.5  
549 but it predicts fire emissions closer to GFED4s, indicating a smaller sensitivity of fire emission to fuel load in the  
550 ML model. The overestimation of fire PM<sub>2.5</sub> emissions can also be explained by the sensitivity to meteorology. The  
551 spatial distributions of the long-term mean fire emissions shown in Fig. 3 correlate with the spatial distributions of  
552 sensitivities to RH and/or temperature, with regional differences. For western US, large fire emissions are  
553 associated with stronger sensitivities to both RH and temperature in the ML and most FireMIP models. For central  
554 and southeastern US, overestimation of fire PM<sub>2.5</sub> emissions only corresponds to stronger sensitivity to temperature  
555 in some FireMIP models.

556 Besides comparing model performance aggregated over CONUS, we analyze the model performance for  
557 several regions, including the western forest area, Mediterranean California, southwestern US, and southeastern  
558 US. For the western forest area, the ML model performs well in capturing both seasonality and interannual  
559 variability of fire PM<sub>2.5</sub> emissions, with correlation coefficients of 0.98 and 0.96, respectively. In contrast, the  
560 FireMIP models generally reproduce the seasonality well but do not simulate the interannual variability well,  
561 especially underestimating the peak in 2007, which related to large-scale meteorological patterns favorable for fires  
562 in Pacific Northwest (Zhong et al., 2020). For Mediterranean California, all FireMIP models only capture the first  
563 peak in August but fail to simulate the second peak in October, which is caused by large fires related to Santa Ana  
564 winds in 2003 and 2007. Such lack of peak emission is also shown in the interannual variability, as all FireMIP  
565 models show limited ability to simulate the peaks in these two years. By contrast, the ML model successfully  
566 predicts the bimodal seasonality and the large fire emissions related to the Santa Ana winds in 2003 and 2007. The  
567 underestimation of the peak in the FireMIP models may be attributed to the underrepresentation of the effects of  
568 large-scale meteorology in the two regions, as the ML model and SHAP importance show that SVD predictors  
569 have larger contributions to the fire emissions in both events. The results of the two regions in the western US  
570 suggest that fire parameterization in the FireMIP models could be improved by including the effects of regional  
571 and large-scale meteorology (e.g., Santa Ana winds) on fire activity (Yue et al., 2014). Modeling the effect of Santa  
572 Ana winds on wildfires may be particularly challenging as the offshore Santa Ana winds exhibit variability related

573 to both synoptic scale pressure anomaly over the Great Basin and local thermodynamic forcing associated with  
574 strong desert-ocean temperature gradient (Hughes and Hall, 2010).

575 As for southwestern US, the ML model and LPJ-SIM estimate the peak in June ( $r=0.99$  and  $0.94$  for ML  
576 and LPJ-SIM, respectively), which highly agrees with the GFED observation. Interestingly, most FireMIP models  
577 fail to capture the extremely large fire emission in the 2011 summer mainly due to the low biases of VPD anomalies  
578 in CRUNCEP (Tang et al., 2017). Unlike southwestern US, the seasonality of southeastern US has peaks in March-  
579 April and September-October. The two peaks of fire emissions correspond to wildfires (Mar-Apr and Sep),  
580 cropland fires (Feb-Mar and Aug-Oct), and prescribed fires (Feb-Apr and Oct) that include burnings for pest  
581 controls and land cleaning (Knapp et al., 2009; Lin et al., 2014). Most models fail to reproduce the bimodal fire  
582 emissions, but the ML model, LPJ-SIM, and LPJ-SPI can capture the bimodal pattern. Even though the seasonality  
583 of fires over this region is not simulated accurately, the CLM and JSBACH well reproduce the interannual  
584 variability of fire  $PM_{2.5}$  emissions and predict the peaks. The FireMIP models' shortfall in reproducing the bimodal  
585 seasonality can be explained by two reasons. First, the relationships between human and fire spread implemented  
586 in the process-based models may not be realistic compared to the observed relationships. Parisien et al. (2016)  
587 demonstrated the large spatial variability of human impacts on burned areas in North America, which is not well  
588 represented in the FireMIP models (Li et al., 2019). Second, drier conditions in winter would promote fires in  
589 springtime (Wear and Greis, 2013; Wang et al., 2021), which may not be directly considered in the FireMIP models  
590 but are incorporated as SVD predictors in the ML model. Overall, the representations of the effects of human and  
591 large-scale meteorology on fires may explain why the models simulate the seasonality incorrectly in southeastern  
592 US. In addition to the comparison of general model performance, we also compare the ability of the data-driven  
593 and processed-based models in predicting extremely large fire emissions. Both ML and CLM models reproduce  
594 the spatial pattern of extreme fire events and reasonably simulate the historical events of large fires in southwestern  
595 and western US.

596 It is known that different fire emission inventories have their uncertainties and prior studies have compared  
597 fire emission inventories over the globe or CONUS (Urbanski et al., 2018; Liu et al., 2020). The GFED fire  
598 emissions used in this study are known to underestimate the fire emission peak in springtime over the southeastern  
599 US, which may be explained by the fact that other products such as FINN or QFED capture more small fire activity  
600 compared to the GFED approach (Koplitz et al., 2018; Carter et al., 2020). Although FINN can capture more small  
601 fires, it underestimates the intensity of large fires for some cases, which has been attributed to the cloud coverage  
602 on daily scale detection (Paton-Walsh et al., 2012). QFED and GFAS, which estimate emissions using fire radiative  
603 power (FRP) from satellites, are also more sensitive to small fires than GFED. However, QFED tends to estimate  
604 much larger emissions than other products, which can be explained by the fact that the emission coefficients used  
605 to obtain emissions are constrained by MODIS AOD and the uncertainties within FRP (Pan et al., 2020). Despite  
606 the known discrepancy between GFED and other data products, the GFED data still shows bimodal peaks in spring  
607 and fall over the southeastern US, while most FireMIP models fail to reproduce the first peak (Fig. 7g in the  
608 manuscript). For the western US, GFED and FINN are generally consistent regarding the magnitude and variability  
609 of fire emissions (Urbanski et al., 2018). As stated above, different fire emission inventories have uncertainties.  
610 Future works are required to include other fire emission datasets for model evaluation.

611 To summarize, we utilize the ML model with SHAP importance to diagnose the fire emissions simulated  
612 by process-based models and attributed model biases to several factors. First, the sensitivities of fire emissions to  
613 meteorology in the models are stronger than the observed, leading to overestimations. Second, the large-scale  
614 meteorological patterns conducive to fires are not fully considered in the process-based models, which are  
615 important contributors of large fire emissions in western US and southeastern US. Third, the spatial resolutions of  
616 models and/or the atmospheric forcing they used may be too coarse to resolve the effects of regional weather  
617 phenomenon such as Santa Ana winds. Fourth, biases in the atmospheric forcing data may result in biases of fire  
618 emission predictions. Last but not least, human activities are a critical component shaping fire regimes but the

619 human effects on fire activities in the FireMIP models may not reflect the human-fire relationships in the real-  
620 world. This is also an issue in the ML model as the human-related predictors in the ML model may be too simple  
621 to represent the human influences. The underrepresentation of human effects in both types of models may cause  
622 additional uncertainties in projecting future fire activities and their impacts on climate. By training the ML model  
623 using the GFED emissions, the ML model is able to better explain fire emissions in the US, which makes it a useful  
624 tool for diagnosing processes or relationships that may be missing or not well represented in the process-based  
625 models to guide future development for improving their performance. Besides its use in diagnosing process-based  
626 models, the ML model *with SHAP* provides a different and novel approach to simulate fire emissions more  
627 accurately and identify the important predictor variables. While the ML model generally has higher accuracy than  
628 the FireMIP models, the feedbacks between fire emissions and climate are not *included*, which could potentially  
629 affect the reliability of ML-based models in fire emission prediction under future climate change scenario (Zou et  
630 al., 2020). Lastly, due to limited training data, the ML model cannot predict fires in regions with longer fire return  
631 intervals, posing additional uncertainties in their use for making future projections.

632  
633  
634 *Code availability.* Model code is available upon request to the first author.

635  
636 *Data availability.* The ML prediction and predictor dataset used in this study are publicly accessible online at  
637 <https://zenodo.org/record/5076646#.YOZ14zZKjOQ>.

638  
639 *Author contributions.* SW, YQ, RL conceived the research ideas. SW wrote the initial draft of the paper, performed  
640 analyses, and model development. All authors contributed to the interpretation of the results and the preparation of  
641 the manuscript.

642  
643 *Acknowledgements.*

644 This research was performed at PNNL and funded under Assistance Agreement No. RD835871 by the U.S.  
645 Environmental Protection Agency to Yale University through the SEARCH (Solutions for Energy, AiR, Climate,  
646 and Health) Center. It has not been formally reviewed by EPA. The views expressed in this document are solely  
647 those of the SEARCH Center and do not necessarily reflect those of the Agency. EPA does not endorse any products  
648 or commercial services mentioned in this publication.

## 649 **References**

650 Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., Crounse, J. D., and Wennberg, P. O.:  
651 Emission factors for open and domestic biomass burning for use in atmospheric models, 11, 4039–4072,  
652 <https://doi.org/10.5194/acp-11-4039-2011>, 2011.

653 Abatzoglou, J. T. and Kolden, C. A.: Relationships between climate and macroscale area burned in the western United States,  
654 *Int. J. Wildland Fire*, 22, 1003–1020, <https://doi.org/10.1071/WF13019>, 2013.

655 Abatzoglou, J. T. and Williams, A. P.: Impact of anthropogenic climate change on wildfire across western US forests, *PNAS*,  
656 113, 11770–11775, <https://doi.org/10.1073/pnas.1607171113>, 2016.

657 Abatzoglou, J. T., Kolden, C. A., Balch, J. K., and Bradley, B. A.: Controls on interannual variability in lightning-caused fire  
658 activity in the western US, *Environ. Res. Lett.*, 11, 045005, <https://doi.org/10.1088/1748-9326/11/4/045005>, 2016.

- 659 Aldersley, A., Murray, S. J., and Cornell, S. E.: Global and regional analysis of climate and human drivers of wildfire, *Science*  
660 of The Total Environment, 409, 3472–3481, <https://doi.org/10.1016/j.scitotenv.2011.05.032>, 2011.
- 661 Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D.,  
662 Benjamins, R., Chatila, R., and Herrera, F.: Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and  
663 challenges toward responsible AI, *Information Fusion*, 58, 82–115, <https://doi.org/10.1016/j.inffus.2019.12.012>, 2020.
- 664 Balch, J. K., Bradley, B. A., Abatzoglou, J. T., Nagy, R. C., Fusco, E. J., and Mahood, A. L.: Human-started wildfires expand  
665 the fire niche across the United States, *PNAS*, 114, 2946–2951, <https://doi.org/10.1073/pnas.1617394114>, 2017.
- 666 Bartier, P. M. and Keller, C. P.: Multivariate interpolation to incorporate thematic surface data using inverse distance weighting  
667 (IDW), *Computers & Geosciences*, 22, 795–799, [https://doi.org/10.1016/0098-3004\(96\)00021-0](https://doi.org/10.1016/0098-3004(96)00021-0), 1996.
- 668 Bedia, J., Herrera, S., and Gutiérrez, J. M.: Assessing the predictability of fire occurrence and area burned across phytoclimatic  
669 regions in Spain, 14, 53–66, <https://doi.org/10.5194/nhess-14-53-2014>, 2014.
- 670 Birch, D. S., Morgan, P., Kolden, C. A., Abatzoglou, J. T., Dillon, G. K., Hudak, A. T., and Smith, A. M. S.: Vegetation,  
671 topography and daily weather influenced burn severity in central Idaho and western Montana forests, 6, art17,  
672 <https://doi.org/10.1890/ES14-00213.1>, 2015.
- 673 Brey, S. J., Barnes, E. A., Pierce, J. R., Wiedinmyer, C., and Fischer, E. V.: Environmental Conditions, Ignition Type, and Air  
674 Quality Impacts of Wildfires in the Southeastern and Western United States, 6, 1442–1456,  
675 <https://doi.org/10.1029/2018EF000972>, 2018.
- 676 Carter, T. S., Heald, C. L., Jimenez, J. L., Campuzano-Jost, P., Kondo, Y., Moteki, N., Schwarz, J. P., Wiedinmyer, C.,  
677 Darmenov, A. S., da Silva, A. M., and Kaiser, J. W.: How emissions uncertainty influences the distribution and radiative  
678 impacts of smoke from fires in North America, 20, 2073–2097, <https://doi.org/10.5194/acp-20-2073-2020>, 2020.
- 679 Carvalho, A., Flannigan, M. D., Logan, K., Miranda, A. I., and Borrego, C.: Fire activity in Portugal and its relationship to  
680 weather and the Canadian Fire Weather Index System, 17, 328–338, <https://doi.org/10.1071/WF07014>, 2008.
- 681 Center For International Earth Science Information Network-CIESIN-Columbia University: Gridded Population of the World,  
682 Version 4 (GPWv4): Population Density, Revision 11, <https://doi.org/10.7927/H49C6VHW>, 2017.
- 683 Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD*  
684 *International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 785–794,  
685 <https://doi.org/10.1145/2939672.2939785>, 2016.
- 686 Coffield, S. R., Graff, C. A., Chen, Y., Smyth, P., Foufloula-Georgiou, E., and Randerson, J. T.: Machine learning to predict  
687 final fire size at the time of ignition, 28, 861–873, 2019.
- 688 Cortez, P. and Morais, A.: A Data Mining Approach to Predict Forest Fires using Meteorological Data, *Proceedings of the*  
689 *13th Portuguese Conference on Artificial Intelligence*, Portugal, 512–523, 2007.
- 690 Crevoisier, C., Shevliakova, E., Gloor, M., Wirth, C., and Pacala, S.: Drivers of fire in the boreal forests: Data constrained  
691 design of a prognostic model of burned area for use in dynamic global vegetation models, 112,  
692 <https://doi.org/10.1029/2006JD008372>, 2007.
- 693 Crimmins, M. A.: Synoptic climatology of extreme fire-weather conditions across the southwest United States, 26, 1001–1016,  
694 <https://doi.org/10.1002/joc.1300>, 2006.

- 695 Cummins, K. L. and Murphy, M. J.: An Overview of Lightning Locating Systems: History, Techniques, and Data Uses, With  
696 an In-Depth Look at the U.S. NLDN, 51, 499–518, <https://doi.org/10.1109/TEM.2009.2023450>, 2009.
- 697 Dillon, G. K., Holden, Z. A., Morgan, P., Crimmins, M. A., Heyerdahl, E. K., and Luce, C. H.: Both topography and climate  
698 affected forest and woodland burn severity in two regions of the western US, 1984 to 2006, 2, art130,  
699 <https://doi.org/10.1890/ES11-00271.1>, 2011.
- 700 Dong, L., Leung, L. R., Qian, Y., Zou, Y., Song, F., and Chen, X.: Meteorological Environments Associated With California  
701 Wildfires and Their Potential Roles in Wildfire Changes During 1984–2017, 126, e2020JD033180,  
702 <https://doi.org/10.1029/2020JD033180>, 2021.
- 703 Ford, B., Martin, M. V., Zelasky, S. E., Fischer, E. V., Anenberg, S. C., Heald, C. L., and Pierce, J. R.: Future Fire Impacts on  
704 Smoke Concentrations, Visibility, and Health in the Contiguous United States, 2, 229–247,  
705 <https://doi.org/10.1029/2018GH000144>, 2018.
- 706 Forkel, M., Andela, N., Harrison, S. P., Lasslop, G., van Marle, M., Chuvieco, E., Dorigo, W., Forrest, M., Hantson, S., Heil,  
707 A., Li, F., Melton, J., Sitch, S., Yue, C., and Arneeth, A.: Emergent relationships with respect to burned area in global satellite  
708 observations and fire-enabled vegetation models, 16, 57–76, <https://doi.org/10.5194/bg-16-57-2019>, 2019.
- 709 Friedl, M. and Sulla-Menashe, D.: MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V006,  
710 <https://doi.org/10.5067/MODIS/MCD12C1.006>, 2015.
- 711 Giglio, L., Randerson, J. T., and Werf, G. R. van der: Analysis of daily, monthly, and annual burned area using the fourth-  
712 generation global fire emissions database (GFED4), 118, 317–328, <https://doi.org/10.1002/jgrg.20042>, 2013.
- 713 Gunning, D.: Explainable artificial intelligence (xAI), Defense Advanced Research Projects Agency (DARPA), 2017.
- 714 Hantson, S., Kelley, D. I., Arneeth, A., Harrison, S. P., Archibald, S., Bachelet, D., Forrest, M., Hickler, T., Lasslop, G., Li, F.,  
715 Mangeon, S., Melton, J. R., Nieradzick, L., Rabin, S. S., Prentice, I. C., Sheehan, T., Sitch, S., Teckentrup, L., Voulgarakis, A.,  
716 and Yue, C.: Quantitative assessment of fire and vegetation properties in simulations with fire-enabled vegetation models from  
717 the Fire Model Intercomparison Project, 13, 3299–3318, <https://doi.org/10.5194/gmd-13-3299-2020>, 2020.
- 718 Holden, Z. A., Swanson, A., Luce, C. H., Jolly, W. M., Maneta, M., Oyler, J. W., Warren, D. A., Parsons, R., and Affleck, D.:  
719 Decreasing fire season precipitation increased recent western US forest wildfire activity, PNAS, 115, E8349–E8357,  
720 <https://doi.org/10.1073/pnas.1802316115>, 2018.
- 721 Hooker, G.: Diagnostics and Extrapolation in Machine Learning, Stanford University, 2004.
- 722 Hu, X., Yu, C., Tian, D., Ruminski, M., Robertson, K., Waller, L. A., and Liu, Y.: Comparison of the Hazard Mapping System  
723 (HMS) fire product to ground-based fire records in Georgia, USA, 121, 2901–2910, <https://doi.org/10.1002/2015JD024448>,  
724 2016.
- 725 Hughes, M. and Hall, A.: Local and synoptic mechanisms causing Southern California’s Santa Ana winds, Clim Dyn, 34, 847–  
726 857, <https://doi.org/10.1007/s00382-009-0650-4>, 2010.
- 727 Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., Fisk, J. P., Hibbard, K., Houghton, R. A., Janetos,  
728 A., Jones, C. D., Kindermann, G., Kinoshita, T., Klein Goldewijk, K., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E.,  
729 Thomson, A., Thornton, P., van Vuuren, D. P., and Wang, Y. P.: Harmonization of land-use scenarios for the period 1500–  
730 2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands, Climatic Change,  
731 109, 117, <https://doi.org/10.1007/s10584-011-0153-2>, 2011.

- 732 Johnston, F. H., Henderson, S. B., Chen, Y., Randerson, J. T., Marlier, M., Defries, R. S., Kinney, P., Bowman, D. M. J. S.,  
733 and Brauer, M.: Estimated global mortality attributable to smoke from landscape fires, *Environ Health Perspect*, 120, 695–  
734 701, <https://doi.org/10.1289/ehp.1104422>, 2012.
- 735 Kane, V. R., Cansler, C. A., Povak, N. A., Kane, J. T., McGaughey, R. J., Lutz, J. A., Churchill, D. J., and North, M. P.: Mixed  
736 severity fire effects within the Rim fire: Relative importance of local climate, fire weather, topography, and forest structure,  
737 *Forest Ecology and Management*, 358, 62–79, <https://doi.org/10.1016/j.foreco.2015.09.001>, 2015.
- 738 Kaulfus, A. S., Nair, U., Jaffe, D., Christopher, S. A., and Goodrick, S.: Biomass Burning Smoke Climatology of the United  
739 States: Implications for Particulate Matter Air Quality, *Environ. Sci. Technol.*, 51, 11731–11741,  
740 <https://doi.org/10.1021/acs.est.7b03292>, 2017.
- 741 Keane, R. E., Reinhardt, E. D., Scott, J., Gray, K., and Reardon, J.: Estimating forest canopy bulk density using six indirect  
742 methods, 724–739, 2005.
- 743 Keeley, J. E., Safford, H., Fotheringham, C. J., Franklin, J., and Moritz, M.: The 2007 Southern California Wildfires: Lessons  
744 in Complexity, *Journal of Forestry*, 107, 287–296, <https://doi.org/10.1093/jof/107.6.287>, 2009.
- 745 Klein Goldewijk, K., Beusen, A., and Janssen, P.: Long-term dynamic modeling of global population and built-up area in a  
746 spatially explicit way: HYDE 3.1, *The Holocene*, 20, 565–573, <https://doi.org/10.1177/0959683609356587>, 2010.
- 747 Kloster, S., Mahowald, N. M., Randerson, J. T., Thornton, P. E., Hoffman, F. M., Levis, S., Lawrence, P. J., Feddema, J. J.,  
748 Oleson, K. W., and Lawrence, D. M.: Fire dynamics during the 20th century simulated by the Community Land Model, 7,  
749 1877–1902, <https://doi.org/10.5194/bg-7-1877-2010>, 2010.
- 750 Knorr, W., Jiang, L., and Arneth, A.: Climate, CO<sub>2</sub> and human population impacts on global wildfire emissions, 13, 267–282,  
751 <https://doi.org/10.5194/bg-13-267-2016>, 2016.
- 752 Koplitz, S. N., Nolte, C. G., Pouliot, G. A., Vukovich, J. M., and Beidler, J.: Influence of uncertainties in burned area estimates  
753 on modeled wildland fire PM<sub>2.5</sub> and ozone pollution in the contiguous U.S., *Atmos Environ* (1994), 191, 328–339,  
754 <https://doi.org/10.1016/j.atmosenv.2018.08.020>, 2018.
- 755 Kummu, M., Taka, M., and Guillaume, J. H. A.: Gridded global datasets for Gross Domestic Product and Human Development  
756 Index over 1990–2015, *Sci Data*, 5, 180004, <https://doi.org/10.1038/sdata.2018.4>, 2018.
- 757 Lasslop, G. and Kloster, S.: Impact of fuel variability on wildfire emission estimates, *Atmospheric Environment*, 121, 93–102,  
758 <https://doi.org/10.1016/j.atmosenv.2015.05.040>, 2015.
- 759 Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B.,  
760 Kampenhout, L. van, Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi,  
761 M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., Broeke, M. van den, Brunke, M. A., Burns,  
762 S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman,  
763 F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D.,  
764 Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Martin, M.  
765 V., and Zeng, X.: The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing  
766 Uncertainty, 11, 4245–4287, <https://doi.org/10.1029/2018MS001583>, 2019.
- 767 Le Quéré, C., Peters, G. P., Andres, R. J., Andrew, R. M., Boden, T. A., Ciais, P., Friedlingstein, P., Houghton, R. A., Marland,  
768 G., Moriarty, R., Sitch, S., Tans, P., Arneth, A., Arvanitis, A., Bakker, D. C. E., Bopp, L., Canadell, J. G., Chini, L. P., Doney,  
769 S. C., Harper, A., Harris, I., House, J. I., Jain, A. K., Jones, S. D., Kato, E., Keeling, R. F., Klein Goldewijk, K., Körtzinger,

- 770 A., Koven, C., Lefèvre, N., Maignan, F., Omar, A., Ono, T., Park, G.-H., Pfeil, B., Poulter, B., Raupach, M. R., Regnier, P.,  
771 Rödenbeck, C., Saito, S., Schwinger, J., Segschneider, J., Stocker, B. D., Takahashi, T., Tilbrook, B., van Heuven, S., Viovy,  
772 N., Wanninkhof, R., Wiltshire, A., and Zaehle, S.: Global carbon budget 2013, 6, 235–263, [https://doi.org/10.5194/essd-6-](https://doi.org/10.5194/essd-6-235-2014)  
773 235-2014, 2014.
- 774 Li, F., Levis, S., and Ward, D. S.: Quantifying the role of fire in the Earth system &ndash; Part 1: Improved global fire modeling  
775 in the Community Earth System Model (CESM1), 10, 2293–2314, <https://doi.org/10.5194/bg-10-2293-2013>, 2013.
- 776 Li, F., Val Martin, M., Andreae, M. O., Arneth, A., Hantson, S., Kaiser, J. W., Lasslop, G., Yue, C., Bachelet, D., Forrest, M.,  
777 Kluzek, E., Liu, X., Mangeon, S., Melton, J. R., Ward, D. S., Darmenov, A., Hickler, T., Ichoku, C., Magi, B. I., Sitch, S., van  
778 der Werf, G. R., Wiedinmyer, C., and Rabin, S. S.: Historical (1700–2012) global multi-model estimates of the fire emissions  
779 from the Fire Modeling Intercomparison Project (FireMIP), 19, 12545–12567, <https://doi.org/10.5194/acp-19-12545-2019>,  
780 2019.
- 781 Littell, J. S., McKenzie, D., Peterson, D. L., and Westerling, A. L.: Climate and wildfire area burned in western U.S.  
782 ecoprovinces, 1916–2003, 19, 1003–1021, 2009.
- 783 Liu, J. C., Mickley, L. J., Sulprizio, M. P., Dominici, F., Yue, X., Ebisu, K., Anderson, G. B., Khan, R. F. A., Bravo, M. A.,  
784 and Bell, M. L.: Particulate Air Pollution from Wildfires in the Western US under Climate Change, *Clim Change*, 138, 655–  
785 666, <https://doi.org/10.1007/s10584-016-1762-6>, 2016.
- 786 Liu, T., Mickley, L. J., Marlier, M. E., DeFries, R. S., Khan, M. F., Latif, M. T., and Karambelas, A.: Diagnosing spatial biases  
787 and uncertainties in global fire emissions inventories: Indonesia as regional case study, *Remote Sensing of Environment*, 237,  
788 111557, <https://doi.org/10.1016/j.rse.2019.111557>, 2020.
- 789 Liu, Y., Zhang, K., Qian, Y., Wang, Y., Zou, Y., Song, Y., Wan, H., Liu, X., and Yang, X.-Q.: Investigation of short-term  
790 effective radiative forcing of fire aerosols over North America using nudged hindcast ensembles, 18, 31–47,  
791 <https://doi.org/10.5194/acp-18-31-2018>, 2018.
- 792 Liu, Z. and Wimberly, M. C.: Climatic and Landscape Influences on Fire Regimes from 1984 to 2010 in the Western United  
793 States, *PLOS ONE*, 10, e0140839, <https://doi.org/10.1371/journal.pone.0140839>, 2015.
- 794 Lundberg, S. and Lee, S.: A Unified Approach to Interpreting Model Predictions, *Neural Information Processing Systems*  
795 (NIPS 2017), Long Beach, CA, USA, 2017.
- 796 Lundberg, S. M., Erion, G. G., and Lee, S.-I.: Consistent Individualized Feature Attribution for Tree Ensembles, 2019.
- 797 Mann, M. L., Battlori, E., Moritz, M. A., Waller, E. K., Berck, P., Flint, A. L., Flint, L. E., and Dolfi, E.: Incorporating  
798 Anthropogenic Influences into Fire Probability Models: Effects of Human Activity and Climate Change on Fire Activity in  
799 California, *PLOS ONE*, 11, e0153589, <https://doi.org/10.1371/journal.pone.0153589>, 2016.
- 800 Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery,  
801 E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W.: North American  
802 Regional Reanalysis, *Bull. Amer. Meteor. Soc.*, 87, 343–360, <https://doi.org/10.1175/BAMS-87-3-343>, 2006.
- 803 Morton, D. C., Collatz, G. J., Wang, D., Randerson, J. T., Giglio, L., and Chen, Y.: Satellite-based assessment of climate  
804 controls on US burned area, 10, 247–260, <https://doi.org/10.5194/bg-10-247-2013>, 2013.
- 805 Myneni, R., Knyazikhin, Y., and Park, T.: MCD15A2H MODIS/Terra+Aqua Leaf Area Index/FPAR 8-day L4 Global 500m  
806 SIN Grid V0006, <https://doi.org/10.5067/MODIS/MCD15A2H.006>, 2015.

- 807 NIFC: Total Wildland Fires and Acres (1983-2020), National Interagency Fire Center, 2020.
- 808 NOAA: Severe Weather Data Inventory (SWDI), <https://www1.ncdc.noaa.gov/pub/data/swdi>, 2006.
- 809 NOAA: State of the Climate: Wildfires for Annual 2011, National Centers for Environmental Information, 2012.
- 810 NOAA: State of the Climate: Wildfires for Annual 2012, National Centers for Environmental Information, 2013.
- 811 Nowell, H. K., Holmes, C. D., Robertson, K., Teske, C., and Hiers, J. K.: A New Picture of Fire Extent, Variability, and  
812 Drought Interaction in Prescribed Fire Landscapes: Insights From Florida Government Records, 45, 7874–7884,  
813 <https://doi.org/10.1029/2018GL078679>, 2018.
- 814 Pan, X., Ichoku, C., Chin, M., Bian, H., Darmenov, A., Colarco, P., Ellison, L., Kucsera, T., da Silva, A., Wang, J., Oda, T.,  
815 and Cui, G.: Six global biomass burning emission datasets: intercomparison and application in one global aerosol model, 20,  
816 969–994, <https://doi.org/10.5194/acp-20-969-2020>, 2020.
- 817 Parisien, M.-A., Miller, C., Parks, S. A., DeLancey, E. R., Robinne, F.-N., and Flannigan, M. D.: The spatially varying  
818 influence of humans on fire probability in North America, *Environ. Res. Lett.*, 11, 075005, <https://doi.org/10.1088/1748-9326/11/7/075005>, 2016.
- 820 Parks, S. A., Parisien, M.-A., and Miller, C.: Spatial bottom-up controls on fire likelihood vary across western North America,  
821 3, art12, <https://doi.org/10.1890/ES11-00298.1>, 2012.
- 822 Paton-Walsh, C., Emmons, L. K., and Wiedinmyer, C.: Australia’s Black Saturday fires – Comparison of techniques for  
823 estimating emissions from vegetation fires, *Atmospheric Environment*, 60, 262–270,  
824 <https://doi.org/10.1016/j.atmosenv.2012.06.066>, 2012.
- 825 Pechony, O. and Shindell, D. T.: Fire parameterization on a global scale, 114, <https://doi.org/10.1029/2009JD011927>, 2009.
- 826 Pfeiffer, M., Spessa, A., and Kaplan, J. O.: A model for global biomass burning in preindustrial time: LPJ-LMfire (v1.0), 6,  
827 643–685, <https://doi.org/10.5194/gmd-6-643-2013>, 2013.
- 828 Pollina, J. B., Colle, B. A., and Charney, J. J.: Climatology and Meteorological Evolution of Major Wildfire Events over the  
829 Northeast United States, *Wea. Forecasting*, 28, 175–193, <https://doi.org/10.1175/WAF-D-12-00009.1>, 2013.
- 830 Pyne, S. J.: Introduction to wildland fire. Fire management in the United States., John Wiley & Sons, New York, NY, USA,  
831 1984.
- 832 Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., Kaplan, J. O., Li, F., Mangeon, S., Ward, D.  
833 S., Yue, C., Arora, V. K., Hickler, T., Kloster, S., Knorr, W., Nieradzik, L., Spessa, A., Folberth, G. A., Sheehan, T.,  
834 Voulgarakis, A., Kelley, D. I., Prentice, I. C., Sitch, S., Harrison, S., and Arneeth, A.: The Fire Modeling Intercomparison  
835 Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions, 10, 1175–1197,  
836 <https://doi.org/10.5194/gmd-10-1175-2017>, 2017.
- 837 Radeloff, V. C., Helmers, D. P., Kramer, H. A., Mockrin, M. H., Alexandre, P. M., Bar-Massada, A., Butsic, V., Hawbaker,  
838 T. J., Martinuzzi, S., Syphard, A. D., and Stewart, S. I.: Rapid growth of the US wildland-urban interface raises wildfire risk,  
839 *PNAS*, 115, 3314–3319, <https://doi.org/10.1073/pnas.1718850115>, 2018.
- 840 Randerson, J. T., Chen, Y., Werf, G. R. van der, Rogers, B. M., and Morton, D. C.: Global burned area and biomass burning  
841 emissions from small fires, 117, <https://doi.org/10.1029/2012JG002128>, 2012.

- 842 Rap, A., Scott, C. E., Spracklen, D. V., Bellouin, N., Forster, P. M., Carslaw, K. S., Schmidt, A., and Mann, G.: Natural aerosol  
843 direct and indirect radiative effects, 40, 3297–3301, <https://doi.org/10.1002/grl.50441>, 2013.
- 844 Seiler, W. and Crutzen, P. J.: Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from  
845 biomass burning, *Climatic Change*, 2, 207–247, <https://doi.org/10.1007/BF00137988>, 1980.
- 846 Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data, in: Proceedings of the 1968 23rd ACM  
847 national conference, New York, NY, USA, 517–524, <https://doi.org/10.1145/800186.810616>, 1968.
- 848 Spracklen, D. V., Mickley, L. J., Logan, J. A., Hudman, R. C., Yevich, R., Flannigan, M. D., and Westerling, A. L.: Impacts  
849 of climate change from 2000 to 2050 on wildfire activity and carbonaceous aerosol concentrations in the western United States,  
850 114, <https://doi.org/10.1029/2008JD010966>, 2009.
- 851 Stephens, S. L.: Forest fire causes and extent on United States Forest Service lands, *Int. J. Wildland Fire*, 14, 213,  
852 <https://doi.org/10.1071/WF04006>, 2005.
- 853 Stowell, J. D., Geng, G., Saikawa, E., Chang, H. H., Fu, J., Yang, C.-E., Zhu, Q., Liu, Y., and Strickland, M. J.: Associations  
854 of wildfire smoke PM<sub>2.5</sub> exposure with cardiorespiratory events in Colorado 2011–2014, *Environment International*, 133,  
855 105151, <https://doi.org/10.1016/j.envint.2019.105151>, 2019.
- 856 Syphard, A. D., Radeloff, V. C., Keuler, N. S., Taylor, R. S., Hawbaker, T. J., Stewart, S. I., Clayton, M. K., Syphard, A. D.,  
857 Radeloff, V. C., Keuler, N. S., Taylor, R. S., Hawbaker, T. J., Stewart, S. I., and Clayton, M. K.: Predicting spatial patterns of  
858 fire on a southern California landscape, *Int. J. Wildland Fire*, 17, 602–613, <https://doi.org/10.1071/WF07087>, 2008.
- 859 Thomas, D. S., Butry, D. T., Gilbert, S. W., Webb, D. H., and Fung, J. F.: The Costs and Losses of Wildfires, 2017.
- 860 Thonicke, K., Spessa, A., Prentice, I. C., Harrison, S. P., Dong, L., and Carmona-Moreno, C.: The influence of vegetation, fire  
861 spread and fire behaviour on biomass burning and trace gas emissions: results from a process-based model, 7, 1991–2011,  
862 <https://doi.org/10.5194/bg-7-1991-2010>, 2010.
- 863 Trouet, V., Taylor, A. H., Carleton, A. M., and Skinner, C. N.: Interannual variations in fire weather, fire extent, and synoptic-  
864 scale circulation patterns in northern California and Oregon, 95, 349–360, <https://doi.org/10.1007/s00704-008-0012-x>, 2009.
- 865 Tsubaki, M. and Mizoguchi, T.: Quantum Deep Field: Data-Driven Wave Function, Electron Density Generation, and  
866 Atomization Energy Prediction and Extrapolation with Machine Learning, *Phys. Rev. Lett.*, 125, 206401,  
867 <https://doi.org/10.1103/PhysRevLett.125.206401>, 2020.
- 868 Urbanski, S. P., Reeves, M. C., Corley, R. E., Silverstein, R. P., and Hao, W. M.: Contiguous United States wildland fire  
869 emission estimates during 2003–2015, 10, 2241–2274, <https://doi.org/10.5194/essd-10-2241-2018>, 2018.
- 870 Urbieto, I. R., Zavala, G., Bedia, J., Gutiérrez, J. M., Miguel-Ayanz, J. S., Camia, A., Keeley, J. E., and Moreno, J. M.: Fire  
871 activity as a function of fire–weather seasonal severity and antecedent climate across spatial scales in southern Europe and  
872 Pacific western USA, *Environ. Res. Lett.*, 10, 114013, <https://doi.org/10.1088/1748-9326/10/11/114013>, 2015.
- 873 Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A Multiscalar Drought Index Sensitive to Global Warming:  
874 The Standardized Precipitation Evapotranspiration Index, 23, 1696–1718, <https://doi.org/10.1175/2009JCLI2909.1>, 2010.
- 875 Wang, S. S.-C. and Wang, Y.: Quantifying the effects of environmental factors on wildfire burned area in the south central US  
876 using integrated machine learning techniques, 20, 11065–11087, <https://doi.org/10.5194/acp-20-11065-2020>, 2020.

- 877 Wang, S. S.-C., Qian, Y., Leung, L. R., and Zhang, Y.: Identifying Key Drivers of Wildfires in the Contiguous US Using  
878 Machine Learning and Game Theory Interpretation, 9, e2020EF001910, <https://doi.org/10.1029/2020EF001910>, 2021.
- 879 Wang, S.-C., Wang, Y., Estes, M., Lei, R., Talbot, R., Zhu, L., and Hou, P.: Transport of Central American Fire Emissions to  
880 the U.S. Gulf Coast: Climatological Pathways and Impacts on Ozone and PM<sub>2.5</sub>, 123, 8344–8361,  
881 <https://doi.org/10.1029/2018JD028684>, 2018.
- 882 Wang, Y., Xie, Y., Cai, L., Dong, W., Zhang, Q., and Zhang, L.: Impact of the 2011 Southern U.S. Drought on Ground-Level  
883 Fine Aerosol Concentration in Summertime, 72, 1075–1093, <https://doi.org/10.1175/JAS-D-14-0197.1>, 2015.
- 884 Ward, D. S., Kloster, S., Mahowald, N. M., Rogers, B. M., Randerson, J. T., and Hess, P. G.: The changing radiative forcing  
885 of fires: global model estimates for past, present and future, 12, 10857–10886, <https://doi.org/10.5194/acp-12-10857-2012>,  
886 2012.
- 887 Wear, D. N. and Greis, J. G.: The Southern Forest Futures Project: technical report, 178, 1–542, 2013.
- 888 Wei, Y., Liu, S., Huntzinger, D. N., Michalak, A. M., Viovy, N., Post, W. M., Schwalm, C. R., Schaefer, K., Jacobson, A. R.,  
889 Lu, C., Tian, H., Ricciuto, D. M., Cook, R. B., Mao, J., and Shi, X.: The North American Carbon Program Multi-scale Synthesis  
890 and Terrestrial Model Intercomparison Project – Part 2: Environmental driver data, 7, 2875–2893,  
891 <https://doi.org/10.5194/gmd-7-2875-2014>, 2014.
- 892 van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E.,  
893 Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997–2016, 9,  
894 697–720, <https://doi.org/10.5194/essd-9-697-2017>, 2017.
- 895 Williams, A. P., Seager, R., Macalady, A. K., Berkelhammer, M., Crimmins, M. A., Swetnam, T. W., Trugman, A. T.,  
896 Buening, N., Noone, D., McDowell, N. G., Hryniw, N., Mora, C. I., and Rahn, T.: Correlations between components of the  
897 water balance and burned area reveal new insights for predicting forest fire area in the southwest United States, *Int. J. Wildland*  
898 *Fire*, 24, 14–26, <https://doi.org/10.1071/WF14023>, 2015.
- 899 Williams, A. P., Abatzoglou, J. T., Gershunov, A., Guzman-Morales, J., Bishop, D. A., Balch, J. K., and Lettenmaier, D. P.:  
900 Observed Impacts of Anthropogenic Climate Change on Wildfire in California, 7, 892–910,  
901 <https://doi.org/10.1029/2019EF001210>, 2019.
- 902 Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B.,  
903 Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., and Mocko, D.: Continental-scale water and energy flux analysis and  
904 validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and  
905 application of model products, 117, <https://doi.org/10.1029/2011JD016048>, 2012.
- 906 Yue, X., Mickley, L. J., Logan, J. A., and Kaplan, J. O.: Ensemble projections of wildfire activity and carbonaceous aerosol  
907 concentrations over the western United States in the mid-21st century, *Atmospheric Environment*, 77, 767–780,  
908 <https://doi.org/10.1016/j.atmosenv.2013.06.003>, 2013.
- 909 Yue, X., Mickley, L. J., and Logan, J. A.: Projection of wildfire activity in southern California in the mid-twenty-first century,  
910 *Clim Dyn*, 43, 1973–1991, <https://doi.org/10.1007/s00382-013-2022-3>, 2014.
- 911 Zhong, S., Yu, L., Heilman, W. E., Bian, X., and Fromm, H.: Synoptic weather patterns for large wildfires in the northwestern  
912 United States—a climatological analysis using three classification methods, 76, <https://doi.org/10.1007/s00704-020-03235-y>,  
913 2020.

914 Zou, Y., Wang, Y., Ke, Z., Tian, H., Yang, J., and Liu, Y.: Development of a REgion-Specific Ecosystem Feedback Fire  
 915 (RESFire) Model in the Community Earth System Model, 11, 417–445, <https://doi.org/10.1029/2018MS001368>, 2019.

916 Zou, Y., Wang, Y., Qian, Y., Tian, H., Yang, J., and Alvarado, E.: Using CESM-RESFire to understand climate–fire–  
 917 ecosystem interactions and the implications for decadal climate variability, 20, 995–1020, [https://doi.org/10.5194/acp-20-995-](https://doi.org/10.5194/acp-20-995-2020)  
 918 2020, 2020.

919

920

921

922

923 **Table 1.** Advantages and limitations of different types of fire models

	<b>Representative method</b>	<b>Advantages</b>	<b>Limitations</b>
<b>Data-driven model</b>	Multiple Linear regression (MLR)	1. Computationally efficient 2. Simple model 3. It is easy to interpret	1. It cannot capture the non-linear relationships between fires and predictors 2. It assumes that the predictor variables are independent 3. It is sensitive to outliers
	Machine learning method (e.g., neural network, decision tree etc.)	1. Computationally cheap 2. The performance improves when more training data are included 3. It can easily handle multi-dimensional data	1. It requires a lot of training data 2. It is relatively hard to interpret 3. The interactions between fires and vegetation/atmosphere cannot be updated to the model
<b>Process-based model</b>	Dynamic global vegetation model (DGVM)	1. Physics-driven 2. The simulations can include feedbacks between fires and climate or vegetation	1. Computationally expensive 2. The same parameterization may not be applied to all regions 3. It only parameterizes the known processes or phenomena

924

925

926

927 **Table 2.** Predictor variables used in the ML model

<b>Variables</b>	<b>Abbreviation</b>	<b>Categories</b>	<b>Temporal resolution</b>	<b>Spatial resolution</b>	<b>Data Source</b>	<b>References</b>
Monthly mean surface temperature	temp	Local meteorology	monthly	32 km	North American Reanalysis (NARR)	Mesinger et al. (2006)
Monthly mean relative humidity	RH	Local meteorology	monthly	32 km	North American Reanalysis (NARR)	Mesinger et al. (2006)
Monthly mean of daily precipitation	precip	Local meteorology	monthly	32 km	North American Reanalysis (NARR)	Mesinger et al. (2006)
Monthly mean zonal component of wind speed	U	Local meteorology	monthly	32 km	North American Reanalysis (NARR)	Mesinger et al. (2006)
Monthly mean meridional component of wind speed	V	Local meteorology	monthly	32 km	North American Reanalysis (NARR)	Mesinger et al. (2006)
Monthly Standardized Precipitation Evapotranspiration Index	SPEI	Local meteorology	monthly	0.5°×0.5°	SPEI	Vicente-Serrano et al. (2010)
Monthly mean 1000-hour dead fuel moisture	FM1000	Local meteorology	daily	4 km	gridMET	Abatzoglou (2013)
Monthly mean energy release component	ERC	Local meteorology	daily	4 km	gridMET	Abatzoglou (2013)
Monthly mean vapor pressure deficit	VPD	Local meteorology	daily	4 km	gridMET	Abatzoglou (2013)
Monthly lightning flashes density	lightning	Local meteorology	daily	0.1°×0.1°	SWDI/NLDN	NOAA (2006); Cummins and Murphy (2009)
Monthly standard deviation of daily SVDs for northern California	SVD1_NCA and SVD2_NCA	Large-scale meteorological patterns	monthly	Regional	North American Reanalysis (NARR)	Wang et al. (2021)
Monthly standard deviation of daily SVDs for southern Rocky Mountain	SVD1_SRM and SVD2_SRM	Large-scale meteorological patterns	monthly	Regional	North American Reanalysis (NARR)	Wang et al. (2021)

Monthly standard deviation of daily SVDs for southeastern US (with 2-month lag)	SVD1_SElag2 and SVD2_SElag2	Large-scale meteorological patterns	monthly	Regional	North American Reanalysis (NARR)	Wang et al. (2021)
Monthly mean evapotranspiration	ET	Land-surface properties	monthly	0.25°×0.25°	North American Land Data Assimilation System (NLDAS-2)	Xia et al. (2012)
Monthly mean surface soil moisture	soilm	Land-surface properties	monthly	0.25°×0.25°	Global Land Data Assimilation System (GLDAS-2)	Xia et al. (2012)
Monthly mean vegetation fraction	Veg_frac	Land-surface properties	monthly	0.25°×0.25°	Global Land Data Assimilation System (GLDAS-2)	Xia et al. (2012)
Monthly mean Leaf Area Index	LAI	Land-surface properties	8 days	500 m	MODerate resolution Imaging Spectroradiometer (MODIS); LAI classification scheme	Myneni et al. (2015)
Monthly fuel load/normalized fuel load	fuel_load/fuel_load_nor	Land-surface properties	monthly	0.9°×1.25°	Community Land Model (CLM)	Lawrence et al. (2019)
Land cover percentage	p_	Land-surface properties	Yearly	0.05°×0.05°	MODerate resolution Imaging Spectroradiometer (MODIS); LAI classification scheme	Friedl (2015)
Median Topography (slope and elevation)	Slope and elevation	Land-surface properties	Not change by time	100 km		Amatulli et al. (2018)
Gross domestic product	GDP	Socioeconomic and coordinate variables	Yearly	5 arc		Kummu et al. (2018)
Population density	Pop	Socioeconomic and coordinate variables	Yearly	30 arc	Gridded Population of the World data collection (GPW v4)	CIESIN-Columbia University (2017)

928

929

930

931 **Table 3.** The ML model performance for different regions: western forest area, Mediterranean California, southwestern US,  
 932 and southeastern US

	Western forest area	Mediterranean California	Southwestern US	Southeastern US	Whole US
<b>Grid scale (individual grid)</b>					
RMSE (km <sup>2</sup> )	0.29	0.32	0.10	0.02	0.16
Correlation (r)	0.79	0.51	0.76	0.84	0.76
IoA	0.86	0.60	0.85	0.90	0.84
Percentage of grids with correlation > 0.4 (%)	68	47	52	80	74
<b>Regional scale (summation over the region)</b>					
RMSE (km <sup>2</sup> )	37.80	13.94	2.76	3.37	49.98
Correlation (r)	0.98	0.81	0.94	0.97	0.97
IoA	0.98	0.81	0.95	0.98	0.97

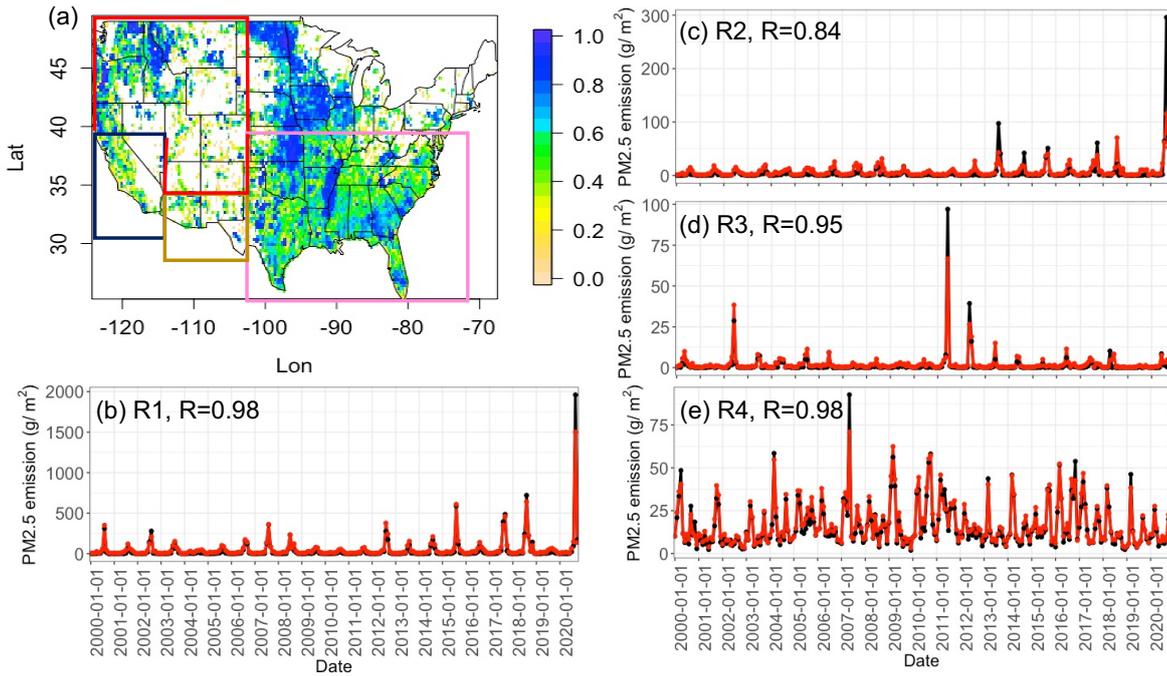
933

934

935 **Table 4.** The model performance for the ML model and FireMIP models

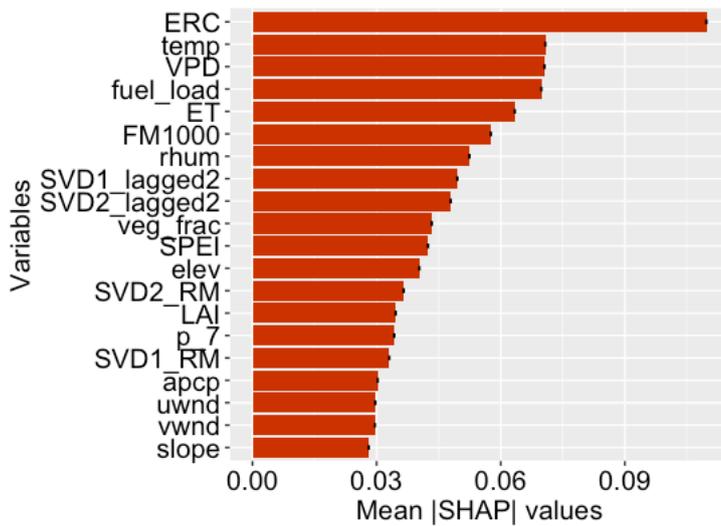
	ML model	CLM	CTEM	JSBACH	LPJ-SPI	LPJ- Glob	LPJ-SIM	ORCHIDEE	JULES	
<b>Total amounts of fire PM2.5 emissions (Tg=10<sup>12</sup> g) (GFED: 4.89 Tg)</b>										
Total fire PM2.5 emissions over 2000- 2012 (Tg)	8.33	16.54	41.50	19.92	16.23	79.49	35.38	2.43	33.43	
<b>Correlation of interannual/seasonal variability for the CONUS</b>										
Correlation (interannual/seasonal)	0.87/0.98	0.71/0.92	0.28/0.87	0.15/0.89	0.15/0.92	0.02/-	0.23/0.65	0.03/0.91	0.55/0.93	
<b>Correlation of interannual/seasonal variability for the selected regions</b>										
Western forest area	0.93/0.98	0.70/0.93	0.33/0.88	0.21/0.88	0.38/0.79	0.51/-	0.46/0.98	0.05/0.94	0.60/0.92	
Mediterranean California	0.72/0.97	-0.01/0.87	0.05/0.94	-0.30/0.89	-0.07/0.90	-0.14/-	-0.19/0.83	0.25/0.13	-0.21/0.87	
Southwestern US	0.95/0.99	0.14/0.85	-0.26/0.62	-0.28/0.45	0.34/0.42	0.30/-	0.40/0.94	0.45/0.72	-0.07/0.69	
Southeastern US	0.96/0.99	0.57/-0.27	-0.16/0.09	0.72/-0.14	0.08/0.35	0.39/-	0.18/0.68	0.16/0.13	0.36/0.01	

936



938

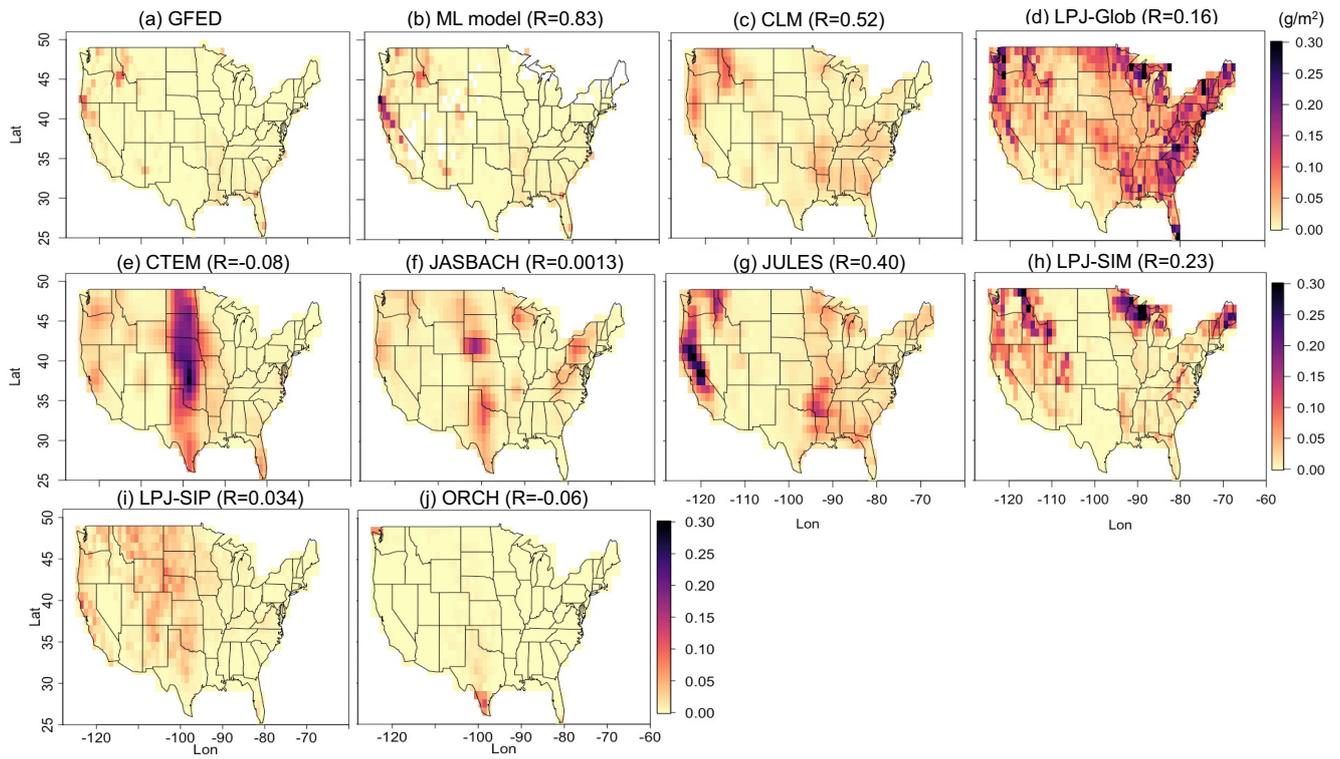
939 **Fig. 1.** (a) The map of temporal correlation between observed and predicted PM<sub>2.5</sub> fire emission for each grid. Time series of  
 940 observed (black) and predicted (red) PM<sub>2.5</sub> fire emission average across (b) western forest area (red box in 1a), (c)  
 941 Mediterranean California (blue box), (d) southwestern US (dusty box), (e) southeastern US (pink box).



942

943 **Fig. 2.** Top 20 variables for the model based on the mean absolute SHAP value with the 95% confidence intervals.

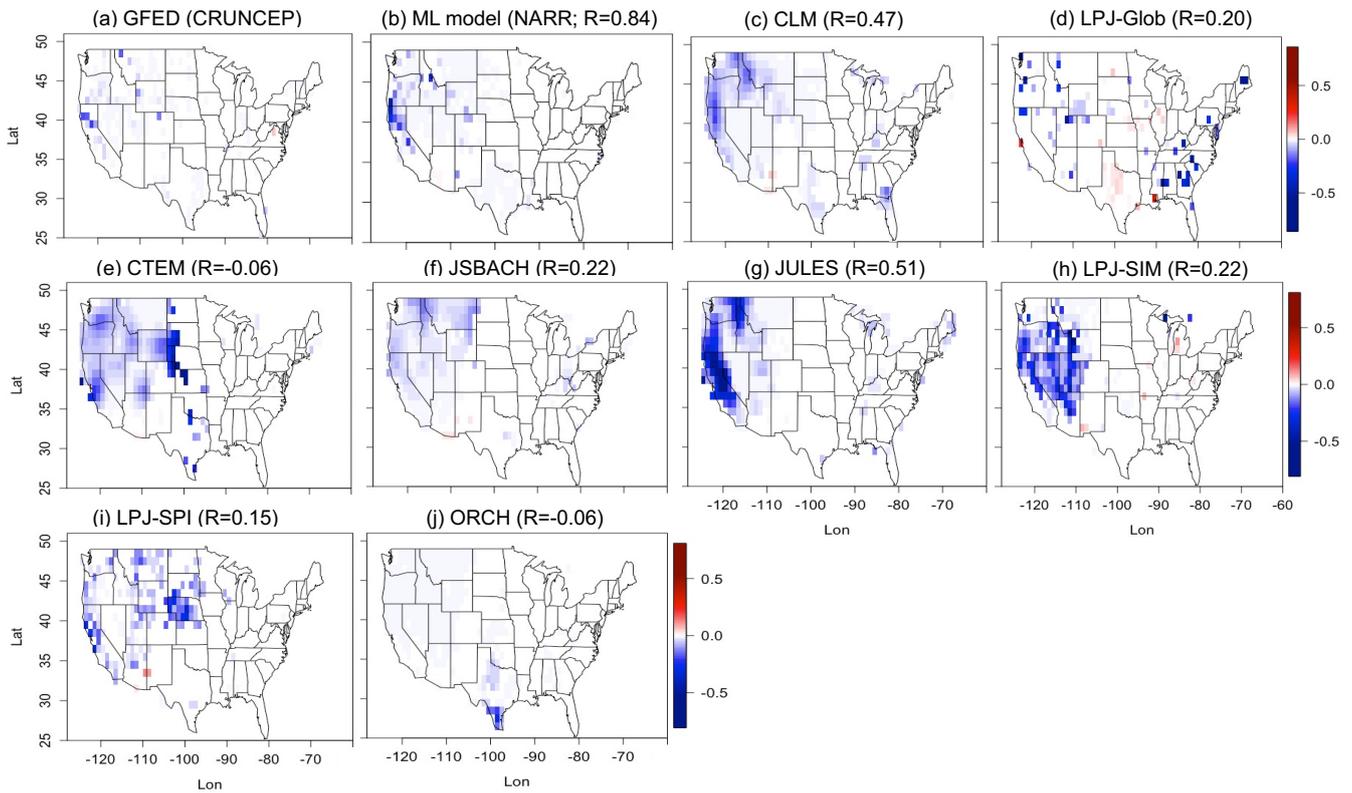
944



945

946 **Fig. 3.** Spatial distributions of the monthly mean  $PM_{2.5}$  fire emission ( $g/m^2/month$ ) over 2000-2012.

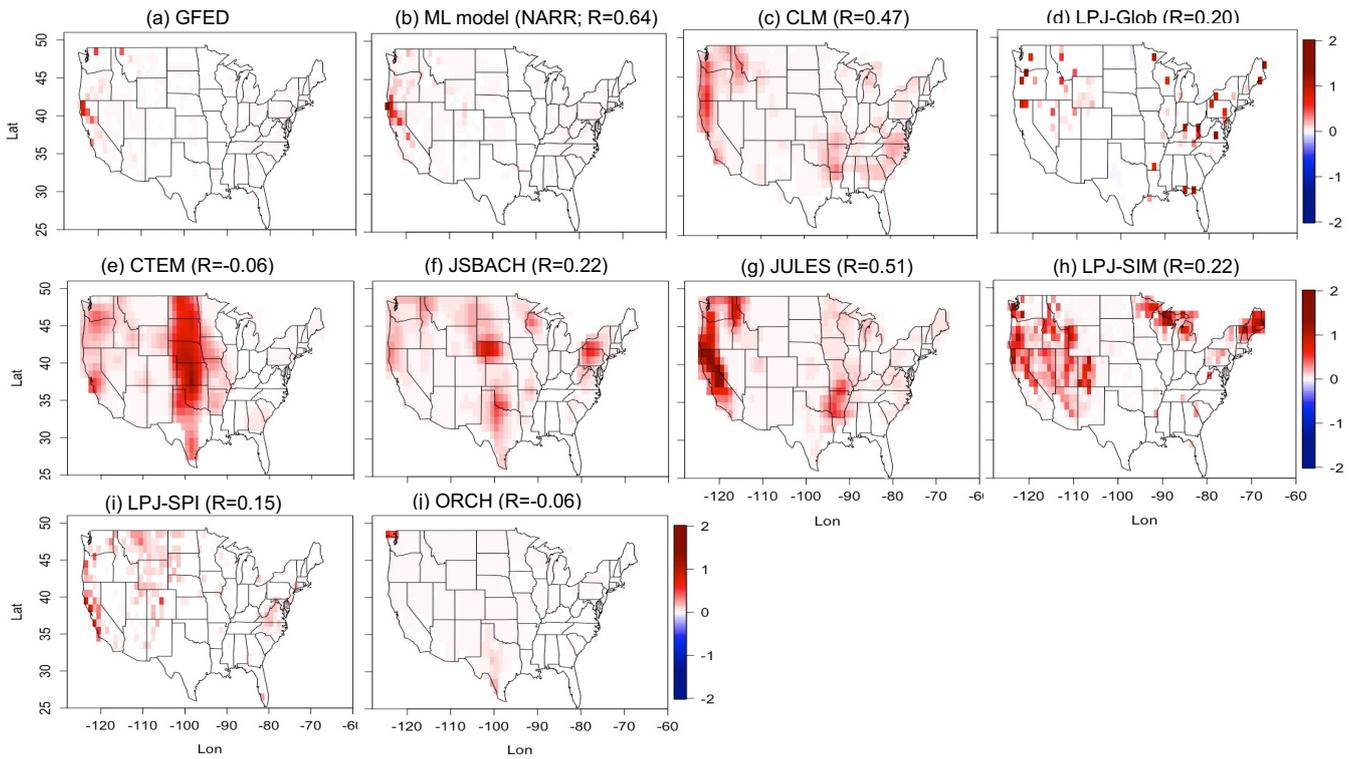
947



948

949 **Fig. 4.** Spatial distributions of the linear regression slope for the dependence of annual mean  $PM_{2.5}$  fire emissions on annual  
 950 mean RH. Only the grids with slopes that are statistically significant ( $p < 0.05$ ) are shown.

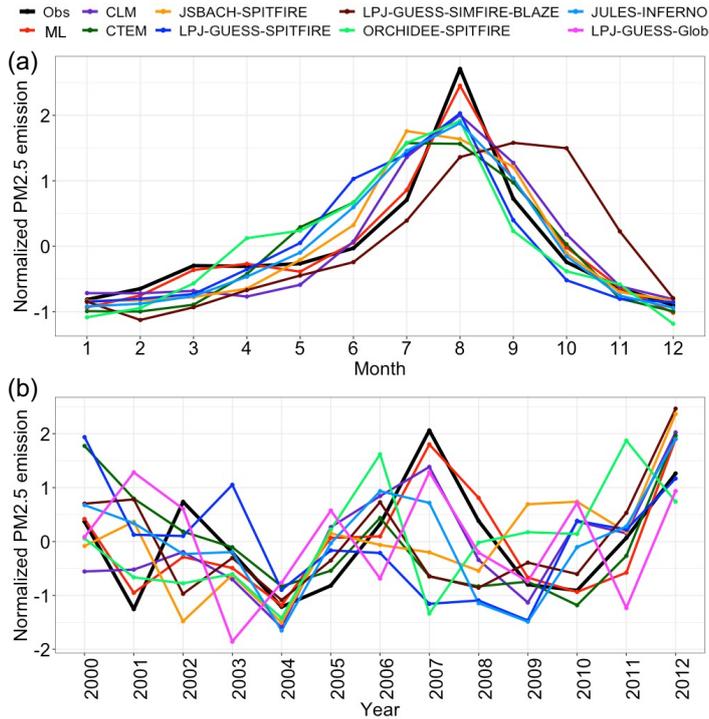
951



952

953 **Fig. 5.** Spatial distributions of the linear regression slope for the dependence of annual mean  $PM_{2.5}$  fire emissions on annual  
 954 mean temperature. Only the grids with slopes that are statistically significant ( $p < 0.05$ ) are shown.

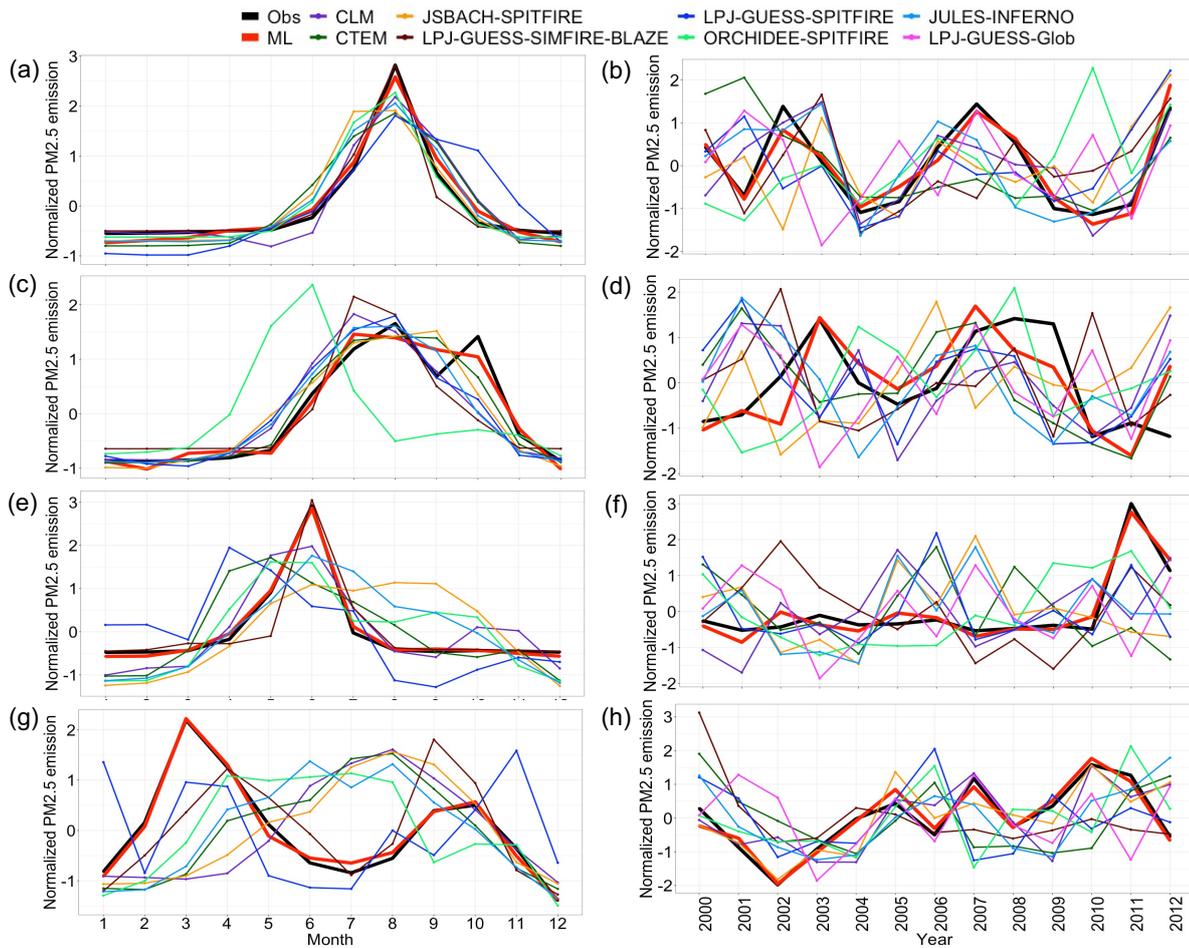
955



956

957 **Fig. 6.** (a) Seasonality and (b) interannual variability of the normalized averaged PM<sub>2.5</sub> fire emission from the GFED (black  
 958 line), ML model (red line), and the FireMIP models (color lines). The PM<sub>2.5</sub> fire emissions are first averaged over CONUS  
 959 and normalized by the monthly (annual) mean and standard deviation for seasonality (interannual variability) plots.

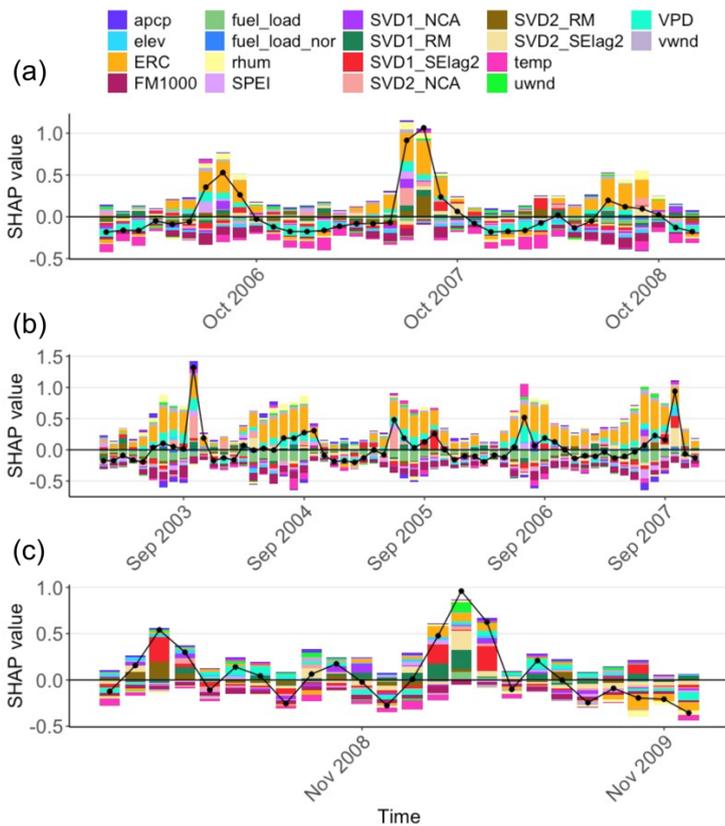
960



961

962 **Fig. 7.** Seasonality and interannual variability of the PM<sub>2.5</sub> fire emission from the GFED (black line), ML model (red line),  
 963 and the FireMIP models (color lines) for (a, b) western forest area, (c, d) Mediterranean California, (e, f) southwestern US,  
 964 and (g, h) southeastern US.

965

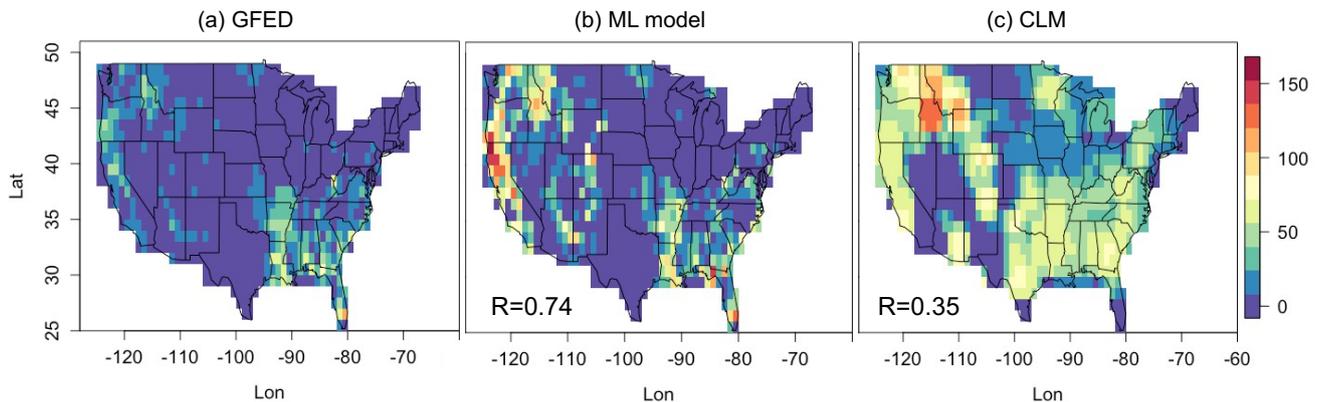


966

967 **Fig. 8.** Time series of the average SHAP values (bar) and predicted normalized PM<sub>2.5</sub> fire emission (line) for (a) western forest  
 968 area from 2006 to 2008, (b) Mediterranean California from 2003 to 2007, and (c) southeastern US from 2008 to 2009. The  
 969 SHAP values indicate the contribution of the predictors to the prediction of normalized fire emission.

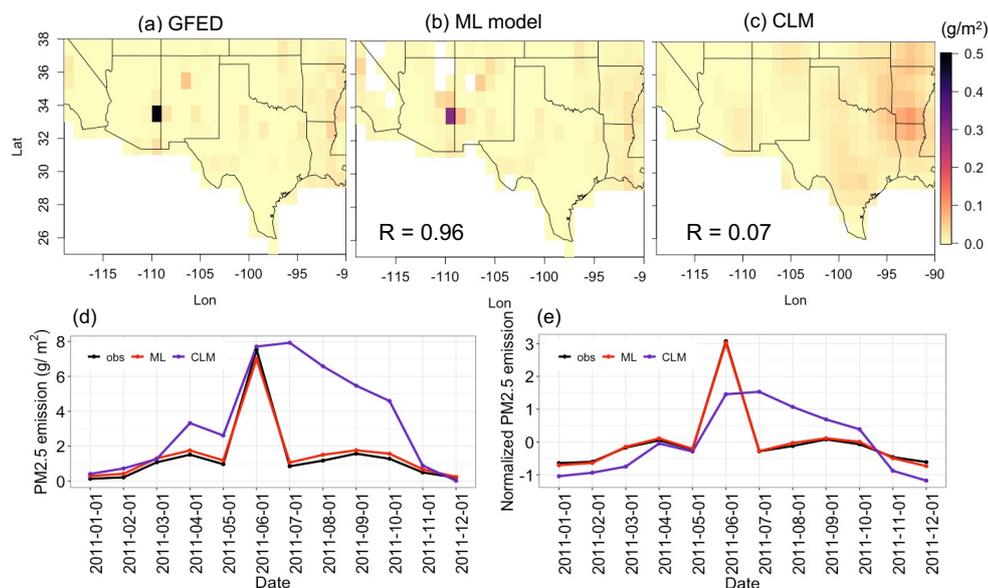
970

971



972

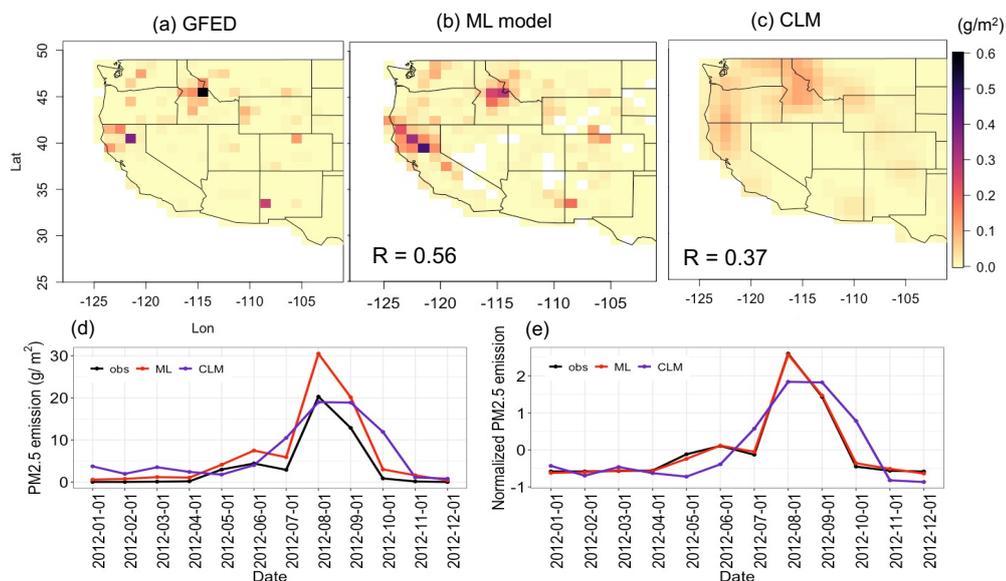
973 **Fig. 9.** Spatial distributions of number of months with large fire emissions (> 95<sup>th</sup> percentiles of PM<sub>2.5</sub> fire emission over all  
 974 the grids in 2000-2012) for (a) GFED, (b) ML model, and (c) CLM.



976

977 **Fig. 10.** Top panel: Spatial distributions of the annual mean PM<sub>2.5</sub> fire emission in 2011 for (a) GFED, (b) ML model, and (c)  
 978 CLM. Bottom panel: Time series of the (d) total PM<sub>2.5</sub> fire emissions and (e) normalized PM<sub>2.5</sub> fire emission over southern US  
 979 domain during 2011.

980



981

982 **Fig. 11.** Top panel: Spatial distributions of the annual mean PM<sub>2.5</sub> fire emission in 2012 for (a) GFED, (b) ML model, and (c)  
 983 CLM. Bottom panel: Time series of the (d) total PM<sub>2.5</sub> fire emissions and (e) normalized PM<sub>2.5</sub> fire emission over western US  
 984 domain during 2012.