SMLFire1.0: a stochastic machine learning (SML) model for wildfire activity in the western United States

Jatan Buch[1], A. Park Williams[2], Caroline Juang[1, 3], Winslow D. Hansen[4], and Pierre Gentine[5]

[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA
[2]Department of Geography, University of California, Los Angeles, CA, USA
[3]Department of Earth and Environmental Sciences, Columbia University, New York, NY, USA
[4]Cary Institute of Ecosystem Studies, Millbrook, NY, USA
[5]Department of Earth and Environmental Engineering, Columbia University, New York, NY, USA

Correspondence: Jatan Buch (jb4625@columbia.edu)

Abstract. The annual area burned due to wildfires in the western United States (WUS) increased by more than 300% between 1984 and 2020. However, accounting for the nonlinear, spatially heterogeneous interactions between climate, vegetation, and human predictors driving the trends in fire frequency and sizes at different spatial scales remains a challenging problem for statistical fire models. Here we introduce a novel stochastic machine learning (SML) framework, SMLFire1.0, to model observed fire frequencies and sizes in $12\,\mathrm{km}\times12\,\mathrm{km}$ grid cells across the WUS. This framework is implemented using Mixture Density Networks trained on a wide suite of input predictors. The modeled WUS fire frequency matches observations at both monthly ($r = 0.94$) and annual ($r = 0.85$) timescales, as do the monthly ($r = 0.90$) and annual ($r = 0.88$) area burned. Moreover, the modeled annual time series of both fire variables exhibit strong correlations ($r \geq 0.6$) with observations in 16 out of 18 ecoregions. Our ML model captures the interannual variability and the distinct multidecade increases in annual area burned for both forested and non-forested ecoregions. Evaluating predictor importance with Shapley additive explanations, we find that fire month vapor pressure deficit (VPD) is the dominant driver of fire frequencies and sizes across the WUS, followed by 1000-hour dead fuel moisture (FM1000), total monthly precipitation (Prec), mean daily maximum temperature (Tmax), and fraction of grassland cover in a grid cell. Our findings serve as a promising use case of ML techniques for wildfire prediction in particular and extreme event modeling more broadly. They also highlight the power of ML driven parameterizations for potential implementation in fire modules of Dynamic Global Vegetation Models (DGVMs) and Earth System Models (ESMs).

1 Introduction

Wildfire is an important biophysical process that structures natural and anthropogenic systems, and is, in turn, affected by climate, vegetation, and humans (Bowman et al., 2009; Krawchuk et al., 2009). The relative strength of each driver and the interactions between them, however, vary across multiple spatial and temporal scales. For instance, sediment charcoal records indicate that while global biomass burning, a proxy for total area burned, responded strongly to warming and drought in the past, these relationships weakened beginning in the late 1800s in many regions due to changes in land use as well as more active fire management (Marlon et al., 2008). Modern satellite observations between 1998 and 2015 (Giglio et al., 2013), on the other

hand, indicate divergent trends along tree cover gradients (Andela et al., 2017); although the decreased fire activity in grasslands and shrublands contributed to the overall decline in global burned area, forest area burned increased across the globe (Zheng et al., 2021). In fact, for regions like the western United States (WUS), there was a $\gtrsim 300\%$ increase in the total area burned between 1984 and 2020, promoted by high flammability of fuels induced by more frequent hot temperature extremes, rising atmospheric aridity, and prolonged drought-like conditions (Dennison et al., 2014; Abatzoglou and Williams, 2016; Zhuang et al., 2021; Kuhn-Régnier et al., 2021). The effect of recent warming and drought on area burned is also exacerbated due to the fuels accumulated in many areas as a result of century-long fire suppression efforts (Marlon et al., 2012; Parks et al., 2015). Incidences of large and severe fires often result in severe environmental and social impacts, such as: poor air quality (O'Dell et al., 2019; Xie et al., 2022), negative health effects from smoke exposure (Burke et al., 2022), enhanced streamflow (Williams et al., 2022), increased flood and debris risk (Jong-Levinger et al., 2022), major vegetation shifts in ecosystems (Coop et al., 2020), and mass displacement of human populations (Jia et al., 2020). Moreover, to manage these fires, federal firefighting expenditures in the United States soared from $\sim$ \$0.5 billion in the late 1980s to an average of $\sim$ \$3 billion between 2016 and 2021 (source: https://www.nifc.gov/fire-information/statistics/suppression-costs). Thus, understanding the complex, multiscale interactions between climate, vegetation, and human drivers of wildfire activity is of vital scientific and societal importance.

Individual wildfire events in the WUS are caused by the coincidence of fire conducive hot and arid weather in presence of adequate vegetation and sources of ignition (Parisien and Moritz, 2009; Williams and Abatzoglou, 2016). However, the influence of specific climatic conditions such as high temperatures and low precipitation may vary spatially due to the fuel moisture content, biomass distribution, and local topography in flammability-limited regions such as forests (Westerling, 2016), and temporally through the response of vegetation growth to antecedent conditions in fuel-limited regions such as grasslands and shrublands (Swetnam and Betancourt, 1998). Meanwhile, the larger WUS fires typically burn over a period of several weeks or more, so the climatic effect on total area burned is regulated by short-term fire weather conditions such as prolonged temperature and aridity extremes (Gutierrez et al., 2021; Juang et al., 2022), sustained intense wind events over multiple days (Potter and McEvoy, 2021), or even the continuity provided by fuels within a landscape's heterogeneous vegetation structure (Rollins et al., 2002). Although difficult to model precisely, fire regimes across the WUS are also affected by the spatial variability of lightning strikes (Romps et al., 2014; Kalashnikov et al., 2022) and stochastic human ignition patterns (Balch et al., 2017; Keeley and Syphard, 2018; Keeley et al., 2021). When aggregated over multiple wildfire events, the observed trends in fire frequency and total area burned carry imprints of the nonlinear, spatially heterogeneous, temporally integrated interactions between climate, vegetation, human, and topographic variables. Physical models of wildfire activity in the WUS, consequently, require a wide suite of input predictors over multiple spatiotemporal scales to accurately represent the various dynamical processes that promote or inhibit fire ignitions and growth.

Here we focus on statistical models for two important fire variables, frequency and area burned. Broadly, these models infer the empirical relationships between observed wildfire activity at a given spatiotemporal scale and its various climate, vegetation, and human drivers. To account for the multiple degrees of freedom characteristic to the problem, regression based models tend to study the mean state relationship between wildfire activity and its drivers by averaging all variables along spatial (Abatzoglou and Williams, 2016) or temporal dimensions (Parisien and Moritz, 2009; Parisien et al., 2012). Despite

being instrumental in clarifying the role of different fire drivers on large spatiotemporal scales, these, and similar, analyses are unable to model fire activity at smaller scales that are important for allocating fire suppression and rescue resources or

60   identifying regions for preventive fuel treatment. On the other hand, other efforts based on classical (Westerling et al., 2011) and Bayesian (Joseph et al., 2019) statistical methods as well as machine learning (ML) approaches (Coffield et al., 2019; Jain et al., 2020; Wang and Wang, 2020; Wang et al., 2021; Joshi and Sukumar, 2021; Kuhn-Régnier et al., 2021; Kondylatos et al., 2022; Richards et al., 2022) have modeled grid-scale fire activity across various spatial extents. Besides representation of finer-scale processes, another key advantage of the grid-scale analyses over the mean state approach is their ability to determine

65   the hierarchy of important wildfire drivers at various spatiotemporal scales.

In this paper, we introduce a stochastic ML (SML) model, SMLFire1.0, to estimate the probability distributions of monthly fire frequencies and sizes in $12\,\mathrm{km} \times 12\,\mathrm{km}$ grid cells across the WUS based on data from 1984 to 2020. SMLFire1.0 consists of a pair of mixture density networks (MDNs) constructed by appending a custom loss function (Ebert-Uphoff et al., 2021) to a neural network. We adopt the MDNs to determine the parametric distributions of fire frequencies and sizes using a combination

70   of static and dynamic climate, vegetation, human, and topographic predictors. We then simulate fire frequencies for each grid cell as well as sizes for grid cells with non-zero frequencies. Our results are visualized and discussed at broader spatial scales of ecoregions for ease of comparison with results from previous analyses.

Our modeling approach for SMLFire1.0 builds upon and extends previous methods in four important ways: *a)* unlike other ML methods based on gradient boosted trees or quantile regression, our use of parametric distributions in SMLFire1.0, es-

75   pecially for individual fire sizes, provides a straightforward way to implement uncertainty quantification for our predictions; *b)* we account for the spatiotemporal variability of the predictors and their nonlinear interactions; *c)* our model includes fire frequencies and locations while simulating the total area burned, thus enabling projections of total area burned for different idealized future scenarios of fuel flammability, human ignition patterns, and fuel treatment; *d)* the combined frequency and size ML framework serves as a single model across the entire WUS and does not require separate training for predicting fire

80   activity in each constituent region. While we do not explore the scenario in detail here, the flexibility and efficiency of our ML framework also makes it an ideal subgrid-scale parameterization scheme for the fire modules of regional scale Dynamic Vegetation Models (DGVMs) (Li et al., 2012; Rabin et al., 2017) as well as Earth System Models (ESMs) (Zou et al., 2020).

2   Data

2.1   Study Ecoregions and Divisions

85   Our study region consists of all $12\,\mathrm{km} \times 12\,\mathrm{km}$ grid cells in the continental US west of $103°$ W longitude. We visualized the results of our analysis at the Bailey's Level III (L3) ecoregion (Bailey, 1996) scale for clarity and ease of comparison, especially in terms of interannual variability, with prior results in the literature (Abatzoglou et al., 2017; Williams et al., 2019; Joseph et al., 2019). Moreover, in several analyses (Littell et al., 2009; Parisien and Moritz, 2009; Dennison et al., 2014), organizing the study region in terms of ecoregions or ecoprovinces has been useful in identifying the broad contours of climate-fire

90   relationships. We define an "Ecoregion" to be constituted by one or more similar L3 ecoregions to ensure sufficient statistics
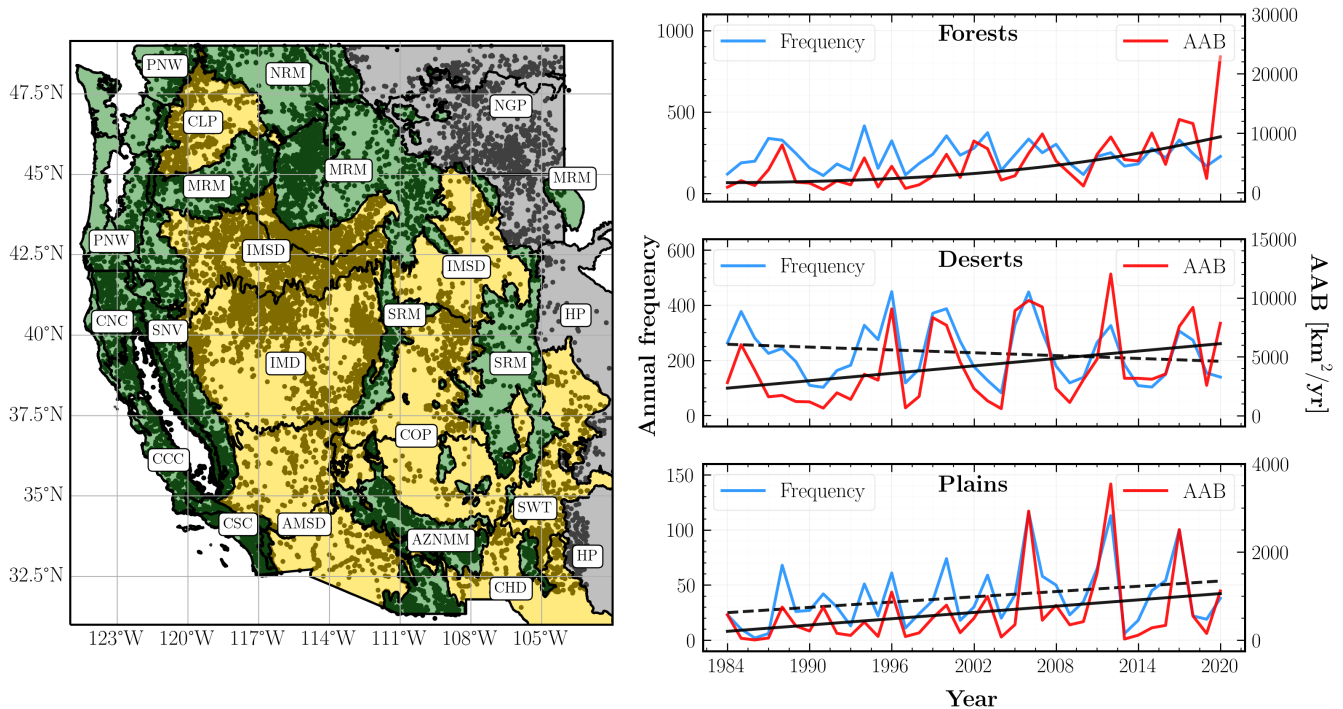
Figure 1. Wildfire activity in the western United States (WUS) from 1984 to 2020. *Left:* locations of fire centroids (black dots) across the WUS with the spatial extent of three ecological Divisions characterized by their primary vegetation type – Forests (green), Deserts (yellow), and Plains (gray). Individual Ecoregions are delineated with thick black lines and referenced with abbreviated names in rounded boxes. A full list of Ecoregion names is given in Table 2. More details about Level III ecoregions, including high resolution maps, can be found here: https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states *Right:* Observed annual fire frequencies (blue) and annual area burned (AAB) (red) for each Division. The black curves indicate the statistically significant ($p < 0.05$) trends for each AAB (solid) and annual frequency (dashed) time series.

(refer to Table S1 for more details), considering a total of 18 Ecoregions across the western United States for this study. Further, we follow (Brey et al., 2018) and organize our Ecoregions in terms of three broad ecological "Divisions" that are characterized by their primary vegetation types, namely Forests, Deserts, and Plains. Note that all three Division types consist of a combination of both forested and non-forested areas albeit in different proportions.

2.2 Wildfire Activity

We focus on two primary fire variables in this analysis: occurrences and sizes. Both these variables are available in the Western US MTBS-Interagency (WUMI) wildfire dataset (Juang et al., 2022) that contains 18646 fire locations and burned areas from 1984 to 2020. The recently released WUMI dataset (accessed Sep 12, 2022) is a collection of unique fires $\geq 4\,\mathrm{km}^2$ from the Monitoring Trends in Burn Severity (MTBS) program (Eidenshink et al., 2007) and fires $\geq 1\,\mathrm{km}^2$ from the following

federal agencies: California Department of Forestry and Fire Protection (CalFire), US Fish and Wildlife Service (FWS), US Forest Service (FS), Bureau of Indian Affairs (BIA), Bureau of Land Management (BLM), Bureau of Reclamation (BOR), and the National Park Service (NPS). Notably, the WUMI dataset underrepresents fires $\leq 4\,\mathrm{km}^2$ from 2018–2020, especially in non-forested areas outside of California because of missing post-2017 data from the BLM, BIA, BOR, and NPS (source: https://famit.nwcg.gov/applications/FireAndWeatherData/ZipFiles; accessed Sep 24, 2022). Although fires smaller than $4\,\mathrm{km}^2$ have a negligible contribution to the total area burned, they constitute $\sim 50\%$ of all fires in our study domain. Thus, the artificially low frequency of smaller fires in 2018-2020 as represented in the current version of the WUMI database likely hinders with accuracy with which our current modeling effort can simulate the probability of small fires.

In Fig. 1, we map all WUMI fire locations as well as plot the annual frequency and annual area burned (AAB) time series for the Forests, Deserts, and Plains Divisions. We also indicate all statistically significant ($p < 0.05$) trends, which were determined using the Student's t-test. The AAB trends are evaluated using a least squares linear regression fit to the log transformed area burned time series as in Williams et al. (2019).

2.3 Input Predictors

We consider four broad classes of input predictors – three dynamic plus one static – aggregated to the $12\,\mathrm{km} \times 12\,\mathrm{km}$ grid scale: climate and fire weather, vegetation, human-related (henceforth human), and topographic. At this spatial scale, a vast majority of fires ($\sim 97\%$) have sizes smaller than the size of the grid cell. Choosing a finer resolution would require explicitly modeling the spatial autocorrelation between the burned area in neighboring grid cells, whereas a coarser resolution results in lower accuracy while correlating fire properties to its environmental variables.

We select six primary climate and fire weather predictors: temperature, precipitation, vapor pressure deficit (VPD), snow water equivalent (SWE), wind speed, and lightning. Monthly climate grids for mean daily maximum temperature (Tmax), mean daily minimum temperature (Tmin), and precipitation total (Prec) are taken from the National Oceanic and Atmospheric Administration's (NOAA) Climgrid dataset (Vose et al., 2014); additionally, gridded dew point temperatures for computing VPD are adapted from PRISM (Daly et al., 2004). We consider two additional fire danger predictors which have been shown to significantly correlate with fire activity (Abatzoglou and Kolden, 2013): 1,000-hr dead fuel moisture (FM1000), and the Fosberg Fire Weather Index (FFWI). Monthly mean FM1000 values, an indicator of climate-derived moisture balance, were adapted from gridMET (Abatzoglou, 2013). The FFWI, which is calculated using temperature, humidity, and wind speed (Fosberg, 1978), has been shown to be an important correlate of dry, windy conditions associated with fire weather (Moritz et al., 2010). Since wind speed in gridMET is derived using a spatial interpolation of the National Atmospheric Regional Reanalysis (NARR) data from a coarser ($32\,\mathrm{km} \times 32\,\mathrm{km}$) resolution, we instead use high ($9\,\mathrm{km} \times 9\,\mathrm{km}$) resolution temperature, humidity, and wind speed predictors from the dynamically downscaled UCLA ERA5-WRF reanalysis (Rahimi et al., 2022) to calculate the monthly mean FFWI. Furthermore, we use daily scale data from the UCLA ERA5-WRF reanalysis to calculate the monthly maximum X-day running average of daily maximum and minimum temperature (Tmax$^{\mathrm{maxX}}$, Tmin$^{\mathrm{maxX}}$), where $X \in \{3, 5, 7\}$. Similar X-day extreme predictors are also derived for VPD, FFWI, and wind speed. The X-day running average variables are included as predictors to improve the model's sensitivity to synoptic scale extreme weather caused by events such as heatwaves.

The monthly mean and maximum daily SWE variables come from the gridded National Snow and Ice Data Center (NSIDC) dataset (Zeng et al., 2018, 2019). Antecedent conditions often exhibit significant correlations with fire activity through drying of soils and fuels as well as promoting fuel growth over multiple months (Westerling et al., 2006; Wang and Wang, 2020; Abolafia-Rosenzweig et al., 2022). Thus, for a given month m with potential fire activity (henceforth fire month), we include temperature, precipitation, VPD, and SWE based predictors that are running averages of monthly mean values from month $m-1$ to $m-t$, where $t \in \{2,3,4\}$. We also include the mean annual precipitation for each of the two years prior to the fire year ($\mathrm{AntPrec_{lag1}}$ and $\mathrm{AntPrec_{lag2}}$) as additional predictors to probe long drought legacy effects (Bastos et al., 2020; Wu et al., 2022). An important source of ignitions over a large area of the WUS is lightning, most frequently as part of summer thunderstorms. We use the Vaisala National Lighning Detection Network (NLDN) lightning strike density data (Wacker and Orville, 1999; Orville and Huffines, 2001) aggregated to monthly scale with coverage from 1987 to 2020. For all months between December 1983 and January 1987, we assume monthly climatological means for the missing lightning data.

We leverage land type data from the National Land Cover Dataset (NLCD) (Yang et al., 2018) for deriving annual scale vegetation predictors. Since the NLCD classifies land cover type (e.g., evergreen forest, cropland etc.) across the US at a 30m spatial resolution, we calculate the fraction of each $12\,\mathrm{km} \times 12\,\mathrm{km}$ grid cell occupied by a given NLCD land-cover classification. The NLCD is not an annual product and provides maps of landcover classification for: 1992, 2001, 2004, 2006, 2008, 2011, 2013, 2016, and 2019. For years between two NLCD years, the landcover in each grid cell is linearly interpolated between the NLCD years, whereas for years before or after 1992-2019, landcover is assumed to be the same as the nearest NLCD year. We adopt three predictors: Grassland, Shrubland, and Forest, each of which represents the fraction of landcover in a grid cell covered by the respective vegetation type. Besides the fraction of landcover, we include a more direct representation of fuel abundance through the aboveground biomass map from (Spawn et al., 2020). Although the biomass map (Biomass) is available for only one year, 2010, we justify its inclusion by positing that the spatial variability of vegetation across the WUS is more dominant than the temporal variability of the vegetation in a majority of grid cells. Thus, for all modeling purposes, we treat Biomass as a static predictor. In future work, it will be ideal to include simulated vegetation biomass maps (e.g., (Hansen et al., 2022)) in a coupled framework within the wildfire model.

Combining the following NLCD land cover types that reflect the presence of urban areas: "Developed High", "Developed Low", "Developed Medium", and "Developed Open" we construct a single, annual scale human predictor, Urban fraction. For more granular information of human settlements, we include the following predictors: distance from the nearest area with population density greater than 10 people per square kilometer (Pop10_dist), mean population density (Popdensity), and mean housing density (Housedensity). These predictors are adapted to annual timescales using data for three years: 1990, 2000, and 2010 from the SILVIS dataset (Radeloff et al., 2005) following the same interpolation procedure that we used for NLCD predictors. Other static human predictors include: mean number of camp grounds (Camp_num), mean distance to nearest camp grounds (Camp_dist), and the distance to nearest highway (Road_dist) derived from publicly available datasets (source: https://www.naturalearthdata.com/downloads/10m-cultural-vectors/roads/; http://www.uscampgrounds.info/takeit.html). These predictors serve as potential correlates of human ignitions for fire occurrences as well as proxies for access to fire suppression or

containment resources. Some predictors such as Popdensity could play a dual role through both increasing the likelihood of ignitions while also providing easier access for fire suppression.

170 Lastly, to incorporate the effect of topography on fire activity (Holsinger et al., 2016; Harris and Taylor, 2017), we include two static variables: mean slope (Slope) and mean south-facing degree of slope (Southness). In the Northern Hemisphere, Southness is associated with higher insolation which results in drier conditions and low fuel moisture relative to other slope directions (Rollins et al., 2002; Dillon et al., 2011). Altogether we include a total of 51 potential predictors, and summarize their names, identifiers, spatial resolution, temporal scale, and sources in Table S2. A short summary of the predictors' physical

175 meaning as well as their qualitative effect on fire frequency and size is provided in Table S3.

Before analyzing the data with a statistical model, we perform an additional preprocessing step. To account for spatiotemporal heterogenity of the WUS ecological landscape, we "standardize", *i.e* subtract the mean and divide by the standard deviation, all input predictors. Dynamic predictors, including all climate and most vegetation variables, at each location are standardized in time, whereas the static predictors are standardized across the entire spatial domain.

3 Model Description

3.1 Theory

Our main goal is to develop a statistical model for fire frequency and sizes as a function of input predictors described in the previous section. Specifically, we want our model to: *a)* capture the nonlinear, spatially heterogeneous interactions among the climate, vegetation, human, and topographic variables that influence wildfire activity; *b)* rely on physical variables and be

185 independent of location and time of year; *c)* be based on parametric distributions that could be sampled using Monte Carlo simulations for estimating the mean and parametric model uncertainty of modeled fire frequency and sizes. While tree-based ML approaches using xGBoost have shown high performance in predicting area burned across the continental US (Wang et al., 2021), we adopt a neural network based architecture here because it combines the flexibility of machine learning techniques with the robustness of parameterized distribution based methods traditionally used in statistical fire modeling (Westerling et al.,

190 2011; Joseph et al., 2019). Recent work (Levin et al., 2022) has also shown that neural network models are more powerful at learning feature representations than gradient-boosted trees. However, generalizing the learned relationships between input predictors and fires to out-of-sample data from future climate states or different fire regimes remains a challenging problem for most ML approaches, including neural network based models (Rasp et al., 2018; Yuval and O'Gorman, 2020).

In SMLFire1.0, we use two mixture density networks (MDNs) to separately model the conditional probability (henceforth

195 conditional for brevity) distributions for fire frequency and sizes on a monthly time scale. A MDN is a fully connected, feedforward neural network whose output layer consists of parameters of a mixture model (Bishop, 1994). In other words, we use a neural network with multiple hidden layers, illustrated in Fig. 2, to map the nonlinear functional relationship between different predictor variables and output data onto the parameters of a mixture of standard statistical distributions. A mixture distribution is a useful tool for representing the probability distribution of outputs with multiple modes or peaks. Thus, given

200 observed data Y, we learn the functional mapping between input predictors X and output parameters ψ by minimizing a loss
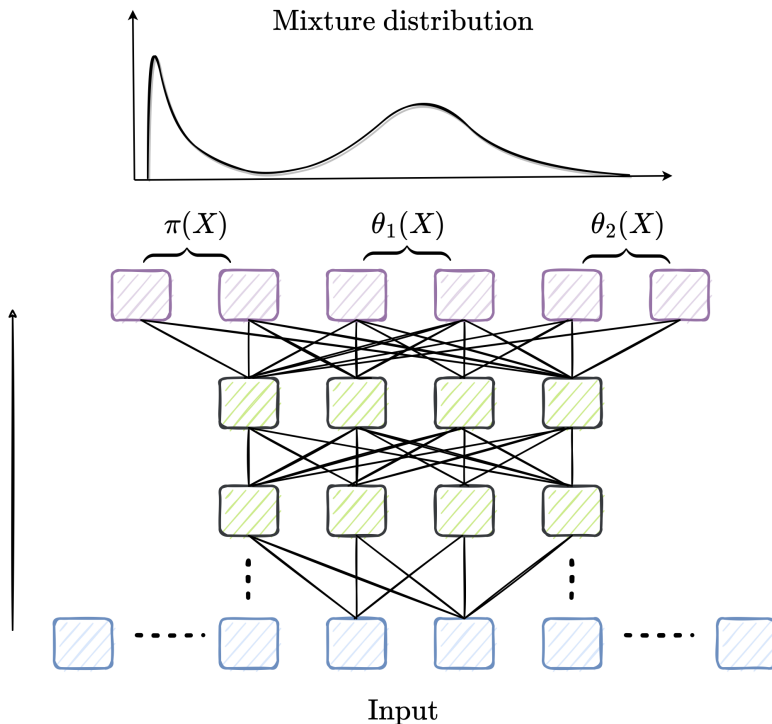
Figure 2. Schematic diagram illustrating the input (blue), hidden (green), and output (purple) layers of a Mixture Density Network (MDN) model within the SMLFire1.0 framework. While a fully connected neural network is implemented in practice, only a partial connected one is shown here for clarity; the solid black line on the left denotes the direction from the input to the output layer, whereas the dotted black lines represent additional nodes and layers in the network. Also shown above the output layer are the parameters for a two component mixture distribution of the form given in Eq. 1.

function of the general form,

$$\mathcal{L}(Y|X; \boldsymbol{\psi} \in \{\pi, \theta\}) = \prod_{n=1}^{N} \left(\sum_{m=1}^{M} \pi_m(X_n) p_m(Y_n | \theta_m(X_n)) \right); \quad \sum_{m=1}^{M} \pi_m = 1, \tag{1}$$

where M and N denote the number of mixture components and data points respectively, and each mixture component consists of a conditional distribution $p_m(\theta_m)$ as well as a weight parameter π_m. To ensure that the resultant mixture distribution is normalized, we constrain the sum of all individual weight parameters to be 1.

We use the monthly fire counts (including zeros) in each grid cell across the WUS as the data for our fire frequency model. In total, we consider data in about ~ 10 million grid cells out of which only 17489 correspond to observed fires. Common choices of parametric distributions for representing count data include: binomial, Poisson, and negative binomial distributions. All of these distributions are often also used in conjunction with another distribution, as part of a zero-inflated mixture model, that accounts for the additional zeros in the data coming from an independent process such as fire suppression. In this analysis,

for each space-time grid cell (henceforth grid cell), n, we use a zero-inflated Poisson lognormal distribution (ZIPD) to model the observed fire frequencies f_n as a function of the input predictors X_n,

$$\mathcal{L}(f_n|X_n;\pi,\mu,\delta) = \begin{cases} \pi(X_n) + (1 - \pi(X_n))\operatorname{Pois}(f_n|\mu(X_n),\delta(X_n)), & , f_n = 0; \\ (1 - \pi(X_n))\operatorname{Pois}(f_n|\mu(X_n),\delta(X_n)) & , f_n > 0, \end{cases} \tag{2}$$

where π is the probability of an independent process that generates zeros, and the rate parameter of the Poisson (Pois) distribution is drawn from a lognormal distribution with mean μ and variance δ^2. There are two major challenges to this approach for modeling fire frequencies: *a)* a large proportion of grid cells contain no fires, resulting in a significant data imbalance problem, *b)* minimizing the risk of missing fires in our predictions, we tend to overpredict fires in grid cells that saw no fires, leading to a high false positive rate. To address point *a)*, we experiment with both downsampling, *i.e.* considering only a random subset of all available grid cells with no fires similar in size to the number of observed fires, and upsampling, *i.e.* generating multiple duplicate samples of the observed fires to match the size of grid cells with no fires, to address this imbalance in our analysis. On the other hand, to fix the effects of a high false positive rate due to the large number of non-fire grid cells, we use a spline regression model for calibrating the mean and variance of the predicted frequencies to those of the observed data at the Ecoregional scale. Finally, we aggregate the predicted fire frequencies across all grid cells within an Ecoregion and compute the mean and variance of the fire frequency, F, for a given month, l, as follows,

$$\mathbb{E}[F^l|X] \equiv \sum_{n \in \mathrm{L3}} \mathbb{E}_{\mathcal{L}_n(\hat{\pi},\hat{\mu},\hat{\delta})}[f_n^l|X_n^l],$$

$$\operatorname{Var}[F^l|X] \equiv \sum_{n \in \mathrm{L3}} \operatorname{Var}_{\mathcal{L}_n(\hat{\pi},\hat{\mu},\hat{\delta})}[f_n^l|X_n^l], \tag{3}$$

where $\mathbb{E}_{\mathcal{L}_n(\cdot)}[\cdot]$ and $\operatorname{Var}_{\mathcal{L}_n(\cdot)}[\cdot]$ indicate the expected value, or mean, and variance with respect to the conditional frequency distribution given by Eq. 2, and the hats denote the distribution parameters fixed to their optimal values determined by training the MDN. The expected value and variance are evaluated using Monte Carlo (MC) simulations of the frequency distribution at monthly and annual timescales. We treat the variance as an estimate of the parametric model uncertainty, or equivalently the uncertainty in modeled frequency due to different realizations of a parametric model.

Meanwhile, as shown previously (Schoenberg et al., 2003; Littell et al., 2009; Li and Banerjee, 2021), wildfire sizes across several spatial scales follow a probability distribution with large tails, or equivalently an extreme value distribution, such that a majority of fires are small but the total area burned is dominated by a small number of large fires. In this analysis, we consider three extreme value mixture distributions for fire sizes: Generalized Pareto distribution (GPD), Lognormal distribution, and a composite Lognormal-Generalized Pareto distribution (Lognormal-GPD) (Scollnik, 2007). The Lognormal-GPD is included to account for the possibility that the fire sizes follow a hybrid distribution with a pronounced hump as well as a significant tail, which are the features of the Lognormal distribution and GPD respectively. We consider 9953 fires from the WUMI dataset with sizes greater than a threshold of $4\,\mathrm{km}^2$ for the GPD and Lognormal-GPD fire size models, whereas we consider all fires for the Lognormal case as it does not require a threshold for extreme events. Due to our fixed choice of the grid for input predictors, a majority of the fires have burned areas that are spread across two or more grid cells. This raises the question:

which values of the input variables should we consider as predictors for our model? Approximating each fire as a circle with area equal to its size, we consider the "effective" input predictors for a given fire to be the average of inputs over all grid cells intersected by the fire's perimeter weighted by the fraction of burned area in each grid cell. In future work, we will use burned area polygons from MTBS for large fires instead of the circular approximation while deriving the effective input predictors.

We model each fire j as independent draws from a conditional mixture distribution (Carreau and Bengio, 2007) of the form,

$$\mathcal{L}(A_j|X_j;\pi,\theta_1,\theta_2) = w(A_j) \sum_{m=1}^{M} \pi_m(X_j) p_m(A_j|\theta_{m,1}(X_n),\theta_{m,2}(X_j)), \tag{4}$$

where (θ_1,θ_2) are the parameters of a heavy tailed distribution determined by the MDN for each fire. For the GPD, (θ_1,θ_2) represent the scale and concentration parameters, whereas for the Lognormal case they represent the mean and standard deviation of the distribution's natural logarithm. We include a weight factor $w(A_j)$, that is inversely proportional to the frequency of size A_j in the training data, to account for data imbalance due to the relative disparity in the number of small and large fires.

The conditional distributions of monthly and annual area burned (MAB and AAB respectively) are obtained by aggregating the distribution of fire sizes for each grid cell in an Ecoregion with a fire. We compute the mean and variance for the fire size distributions using MC simulations and formulas similar to the ones described in Eq. 3.

The expressions for MAB and AAB can be schematically interpreted as follows: assuming that the mean size of all fires at a given spatiotemporal scale l are identical and denoted by A, the mean total area burned A_l is simply given by $\mathbb{E}[A_l|X_l] = \mathbb{E}[F_l|X_l] \times A$. Phrased differently, the expected area burned at a given spatiotemporal scale is linearly proportional to the mean fire frequency, $\mathbb{E}[F_l|X_l]$ with a constant coefficient A. We note, in practice, that since the mean fire size of the GPD model is similar for most fires, the mean fire frequency plays an important role in determining the mean MAB or AAB.

Following Iglesias et al. (2022), we consider the total size distribution to be time-dependent or nonstationary, in general. We allow for an enhancement or weakening in the response of fire sizes to one or more predictors by including a nonstationary response in the size model. In particular, we use the entire training dataset to construct two models for consecutive time periods: a reweighted GPD loss function (GPD-Ext) for the time period with larger fire sizes and an unweighted loss function (GPD) for the remaining years. We stitch the two models together at a breakpoint year to construct the combined GPD MDN model.

In order to isolate the role of frequency in the total area burned, we first derive the AAB using the combined GPD MDN model evaluated with observed fire frequencies and values of input predictors corresponding to the observed locations of fires given in the WUMI dataset. We also explore three further variations for the WUS AAB time series: one, using modeled frequencies for each Ecoregion from the frequency MDN model with observed fire locations; second, with observed frequencies but input predictors corresponding to model fire locations predicted by the frequency MDN; and third, with both fire frequencies and locations drawn from the frequency MDN model. Since our modeled frequencies also include smaller fires, we apply an additional time-dependent scaling factor to account for the relative abundance of large fires ($\geq 4\,\mathrm{km}^2$) while deriving the area burned with modeled frequencies.

Lastly, to obtain percentiles of the burned area distribution we require the full probability density function defined over all grid cells with fires,

$$p(A^l|X, N^l) = p_{A_1^l|X_1^l} * \cdots * p_{A_{N^l}^l|X_{N^l}^l}; \quad p_{A_j^l|X_j^l} = \mathcal{L}(A_j^l|X_j^l, f_j^l > 0; \hat{\pi}_j^l, \hat{\theta}_{1;j}^l, \hat{\theta}_{2;j}^l). \tag{5}$$

275 where $*$ denotes the convolution operator, and N^l are the number of fires at a given spatiotemporal scale. Rather than solving this expression analytically (Nadarajah et al., 2018), we sample it with MC simulations and report the 0.5^{th}, 50th and 99.5^{th} percentiles of the monthly and annual area burned at the WUS and Divisional scales.

3.2 Implementation

We implement our SMLFire1.0 framework using the `Keras` interface for `TensorFlow` library version 2.7.0. During training,
280 we allow our neural networks to have the following tunable hyperparameters: number of hidden layers, n_l; number of neurons per layer, n_{ne}; and the number of mixture components, n_c. We use the relevant distributions available in the `TensorFlow Probability` library version 0.15.0 for designing custom loss functions for both the frequency and size data.

For model training, we hold out 1 contiguous year (which we take to be 2020, unless specified otherwise) of fires and input predictors as test data, and split the monthly data from the remaining 36 years 70% to 30%, chosen randomly, between training
285 and validation data. The out-of-sample validation data is useful for evaluating model performance, which we measure through several metrics defined in the section below.

We train our model for up to 500 epochs using the Adam optimizer with learning rate, `lr_rate` $= 10^{-4}$, to learn the optimal distribution parameters for each model. Since a typical MDN with several hidden layers consists of $\sim \mathcal{O}(1000)$ hidden weights and biases, it is plausible that the model overfits the training data; we address this issue by applying three regularization
290 steps: early stopping of the training process when the model performance does not improve for 10 epochs; L2 regularization, which constrains the squared sum of all the network weights; and a Dropout layer which randomly sets a fraction, `dr_frac`, of inputs to 0 to improve co-learning of the remaining weights. Based on numerical experiments in the pre-training phase, we fix the L2 regularization rate to be `reg_rate` $= 10^{-3}$, and the dropout fraction to be `dr_frac` $= 0.4$. Altogether, the regularization procedure reduces overfitting and improves generalization for the MDN.

295 ## 3.3 Metrics

We define metrics for two broad purposes: enabling model selection and measuring model performance. For the former, a straightforward choice is the value of the loss function, given in Eq. 1, averaged across all batches and epochs: reducing the loss improves the model. As our model size is small, we ignore the effects of model complexity that may be included through metrics such as the Akaike Information Criterion (AIC). Since we are predicting a mixture distribution rather than
300 a point estimate of the output, we use a variation of the Kullback-Leibler (KL) divergence to measure the distance between distributions,

$$D'_{\text{KL}}(C_\psi||\hat{C}) \equiv \sum_{n=1}^{N} \log\left(\frac{C_\psi(y_n|x_n)}{\hat{C}(y_n|x_n)}\right); \quad \text{Accuracy} = 100 \times (1 - D'_{\text{KL}}(C_\psi||\hat{C})), \tag{6}$$

Fire variable	MDN type	(n_l, n_{ne}, n_c)	Loss	Accuracy	LOO-CV score
Frequency	ZIPD (Upsampled)	(2, 16, 2)	3.29	87.53 %	75.75 %
	ZIPD (Downsampled)	(2, 16, 2)	**0.85**	**95.67**%	94.18 %
Size	GPD	(3, 8, 2)	**4.22**	**91.71**%	90.60 %
	Lognormal distribution	(3, 8, 2)	4.45	91.39 %	86.92 %
	Lognormal-GPD	(2, 16, 4)	4.36	87.15 %	85.49 %

Table 1. Summary of the mixture density network (MDN) architecture and performance metrics used for modeling fire frequencies and sizes in SMLFire1.0. n_l, n_{ne}, and n_c refer to the number of hidden layers, neurons per layer, and the number of mixture components respectively. The loss is dimensionless for the frequency MDN whereas the loss for size MDN has units of inverse area burned, or km^{-2}. Both the loss and accuracy metrics are reported for validation data, whereas the leave-one-out cross validation (LOO-CV) score is the validation accuracy averaged over different subsets of fire years, where each subset is obtained by excluding all fires in a randomly chosen year. Lower values of loss as well as higher values of accuracy and LOO-CV score indicate better model performance.

where C_ψ is the cumulative distribution function (CDF) of fire frequencies or sizes predicted by the respective MDN, and $\hat{C}$ is the empirical CDF estimated from data. Unlike the ratios of the probability density functions used in calculating the classic
305 KL divergence, we adopt a ratio of the CDFs which is an equivalent estimator in the asymptotic limit (Perez-Cruz, 2008). We also perform leave-one-out cross validation (LOO-CV) to approximate the model's generalization error. Specifically, we create subsets of fires by excluding the fires that occurred in 1 contiguous year for all years in our study period and report the validation accuracy averaged over all such subsets as the LOO-CV score.

Having selected a model, we characterize its performance by measuring, statistically, how well or poorly the modeled time
310 series fits the observed data. We use the Pearson's correlation coefficient, r, to gauge the proportion of variance in the observed data explained by the predicted time series. Moreover, to account for the point-wise uncertainty that we obtain from our MC simulations, we use the chi-squared statistic,

$$\chi^2 = \sum_{n \in \mathcal{S}_T} \frac{(y_n - \hat{y}_n)^2}{\hat{\sigma}_n^2}, \tag{7}$$

as a measure of the goodness-of-fit for all frequencies or area burned at a particular spatiotemporal scale, $\mathcal{S}_T$. To ensure
315 uniformity of scale, we report the reduced chi-squared statistic, χ_r^2, which is the chi-squared value defined above divided by the degrees of freedom. In our case, the degrees of freedom are simply given by the number of years with non-zero values minus the number of parameters.

Finally, we determine the optimal number of predictor variables by iteratively dropping all predictors that do not improve overall model performance and are highly correlated ($r \geq 0.5$) with other predictors. We include all fire month and static
320 predictors in this step, using the iterative procedure to identify the most important antecedent and extreme weather permutations of each relevant climate variable. After this step we are able to narrow our predictor basis from 51 to 28 variables. Successive removal of each additional variables drastically reduces model performance.

3.4 Predictor importance

We estimate the sensitivity of model output to input predictors using the SHapley Additive exPlanation (SHAP) technique (Lundberg and Lee, 2017). SHAP values are a recent approach (see Wang et al. (2021) for an application to fire modeling) to characterize the marginal contribution of each input predictor on local predictions by using a game-theoretic approach to account for the contributions of all possible coalitions of the remaining predictors. This is in contrast to traditional predictor importance techniques which only rely on a fixed coalition of predictors to assess the importance of any predictor. We implement the SHAP technique by adapting the `KernelExplainer` method from the `shap` package (source: https://github.com/slundberg/shap). We note that a drawback of using the `KernelExplainer` method is that its assumption of predictor independence could lead, in practice, to a biased estimation of predictor importance in the presence of two or more strongly correlated features.

A SHAP value s for an input predictor p can be interpreted as follows: p contributes s additional units to the model output determined by combining the mean baseline value along with the contributions of all other predictors. To estimate the sensitivity of our model outputs to various predictors in a particular Ecoregion, we first randomly choose a subset of grid cells within that Ecoregion with no observed frequencies, or background points, to compute our mean baseline SHAP value. Then combining a fraction of the background points with grid cells that have observed fires in a fixed ratio to create a pool of test points, we evaluate the SHAP values for all predictors relative to the mean baseline value at each test point. The choice of the ratio does not affect our results as long as the number of background points constitute a minority fraction of the test points. For the results shown in the following section, we use a 1:3 ratio of background to test points for each Ecoregion to ensure sufficient statistics. In total, we evaluate the SHAP values for all input predictors at $\sim 20,000$ test points across the WUS.

We visualize our results using two types of plots: a summary plot that shows the SHAP values of the leading input predictors at each test point colored by the predictor value alongside the partial dependence plots of two important predictors, and a global feature importance plot of the leading predictors ordered according to their mean absolute SHAP values, or S, for each Ecoregion. We also assess the interaction effect between predictors by applying a color gradient of one predictor's values to all test points in the partial dependence plots of the other. However, since all test points do not explicitly include information about vegetation transitions under climatic perturbations in the respective grid cell or the effect of repeated fire burns, the indicated predictor importance and partial dependence plots are valid only under the assumption of a stationary relationship between the input predictors and fire sizes.

4 Results

4.1 Model Selection

An important step in our model selection process is determining the optimal hyperparameter configuration for each loss function. We train the frequency and size MDNs on a subset ($\sim 40\%$) of the overall training data over a grid of hyperparamater configurations. In particular, for both the frequency MDNs as well as the Lognormal-GPD size MDN, we fix the number of components, $n_c = 2$, while varying the number of hidden layers, n_l, the number of neurons per layer, n_{ne}; for the rest of the

cases, we vary n_c as well. The model performance is evaluated using the three metrics defined above: average loss, maximum accuracy, and the LOO-CV score computed over all the epochs. The optimal configuration for a MDN type is defined as the one with the lowest loss and the highest validation accuracy. Since our choice of validation data as a random subset of the training data (as opposed to selecting data for consecutive years) already serves as a form of cross-validation, the LOO-CV score may be interpreted as a measure of the model's mean performance across different initial conditions as it is computed with different subsets of validation data.

The optimal hyperparameters and performance metrics for each MDN are outlined in Table 1. For the fire frequency model, we indicate results for the ZIPD MDN trained with upsampled and downsampled data separately. We find that the downsampled ZIPD MDN performs slightly better than its upsampled counterpart, despite the latter using more data (see Chatterji et al. (2022) for a discussion of an analogous problem in the ML literature). Among the size models, we find that the GPD MDN has the best performance and prefers only two components, while the optimal Lognormal distribution MDN configuration contains four components albeit with a higher loss and lower validation accuracy. The optimal Lognormal-GPD MDN has an intermediate performance relative to those of the GPD and Lognormal distributions. We also calculated the LOO-CV score with 3 contiguous years of data held out and find that there is only a marginal decline in model performance, highlighting the robustness of our ML models. In the following sections, unless stated otherwise, we refer to the downsampled ZIPD and GPD MDNs as the frequency MDN and size MDN respectively.

4.2 Fire Frequency

We use a MDN trained on downsampled training data to determine the parameters of the ZIPD for fire frequencies in each grid cell across the WUS from 1984 to 2020. MC simulations of the parametric frequency distributions for all grid cells are aggregated to compute the mean fire frequency and its 2σ uncertainty intervals over monthly and annual time scales. These are plotted for the entire study domain in Fig. 3 as well as individual Ecoregions selected on the basis of total fire counts and their quality of fit in Figs. 4 and 5. The frequencies are plotted for both monthly and annual timescales and are contrasted with the observed values at the respective spatial scale. We note that the observed fire frequency between 2018 and 2020 could increase after including several missing smaller fires in the WUMI dataset, which may also potentially affect our modeled frequencies. We evaluated the goodness-of-fit between our predictions and observations at the annual timescale through the Pearson's correlation coefficient and the reduced chi-squared statistic shown in Table 2 for all the WUS Ecoregions.

An upshot of our likelihood-based MDN model in SMLFire1.0 is the availability of uncertainty estimates (Riley and Thompson, 2016) for the predicted fire frequencies. Since our frequencies are modeled as a Poisson distribution, we expect their standard deviation to scale as $\sim \sqrt{N}$ for N fire counts. Thus, the relatively narrow 95% error band shown for both the WUS and regional frequency plots comes with an important caveat: we only estimate the statistical uncertainty for our results while ignoring the (possibly dominant) contribution to the model uncertainty from sources such as climate-vegetation linkages (Kitzberger et al., 2017; Zhou et al., 2019; Bastos et al., 2020; Tschumi et al., 2022).

At the WUS level, our mean modeled frequencies are in good agreement with the total number of observed fires, exhibiting high correlations at both monthly ($r = 0.94$) and annual ($r = 0.85$) timescales. Our model also successfully captures the
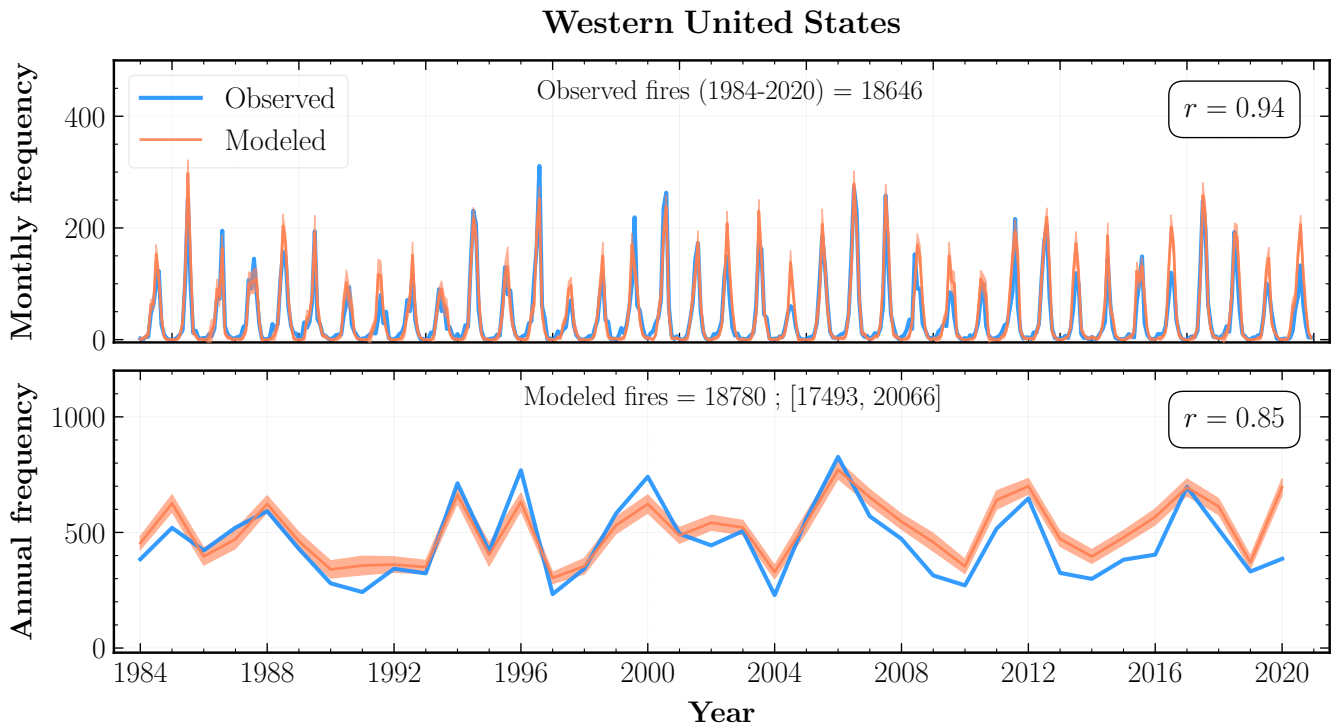
Figure 3. Observed (blue) and modeled (orange) fire frequencies across the western United States at monthly (top) and annual (bottom) time scales from 1984 to 2020. Orange shaded regions represent 2σ uncertainty intervals for the mean fire frequency aggregated over the Monte Carlo simulations for all grid cells. The mean number of modeled fires over the study period as well as its 2σ uncertainty interval are indicated at the top of the bottom panel. Also shown is the Pearson correlation (r) between the observed and modeled time series.

interannual variability and monthly extremes across most of the Ecoregions. In particular, the modeled annual frequencies for Sierra Nevada, Pacific Northwest Mountains, Northern, Middle, and Southern Rockies among the Forests Division; American Semidesert, Intermountain (IM) Desert, Columbia and Colorado Plateaus among the Deserts Division; and the Northern Great Plains are strong correlated ($r \geq 0.7$) with the observed frequencies.

Broadly, the trends in fire frequencies can be characterized as a competition between three independent drivers. One, an increasing trend in climate dryness (Seager et al., 2015; Abatzoglou and Williams, 2016), which is correlated with regional water balance and hence fuel flammability; two, better communication of fire risk resulting in fewer accidental ignitions (Keeley and Syphard, 2018) and enhanced preparedness levels; and three, increased human fire suppression efforts through improvements in fire prevention and containment techniques. While the WUMI dataset indicates moderate increases in the annual number of wildfires $> 1\,\mathrm{km}^2$ in areas defined as forest by NLCD, this is almost fully compensated by a reduction in frequency of fires in non-forested areas. As a result, there is no clear trend in the observed fire frequency for the WUS. On the other hand, our modeled frequencies indicate a mildly increasing trend at the overall WUS scale as well as for several Ecoregions such as PNW Mountains, Columbia Plateau, and IM Desert. A potential contributing factor to this variability could be the high sensitivity
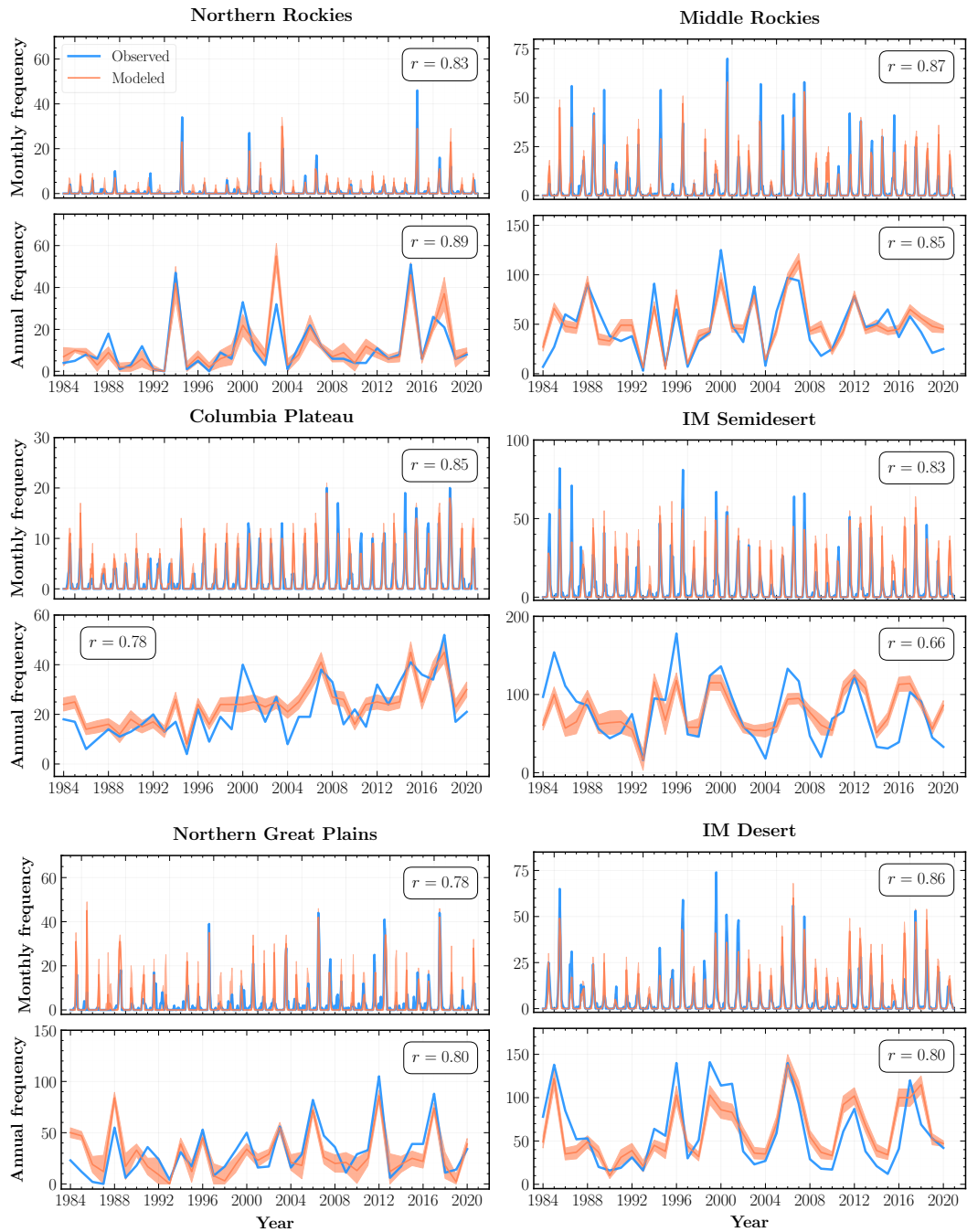
Figure 4. Observed (blue) and modeled (orange) fire frequencies at monthly and annual scales from 1984 to 2020 for Ecoregions selected based on the total number of fires and goodness-of-fit metrics. The orange shaded regions within each subplot indicate 2σ uncertainty intervals for the mean regional fire frequency aggregated over the Monte Carlo (MC) simulations for all constituent grid cells. Also shown is the Pearson correlation coefficient (r) between the observed and modeled fire frequency time series for each Ecoregion.
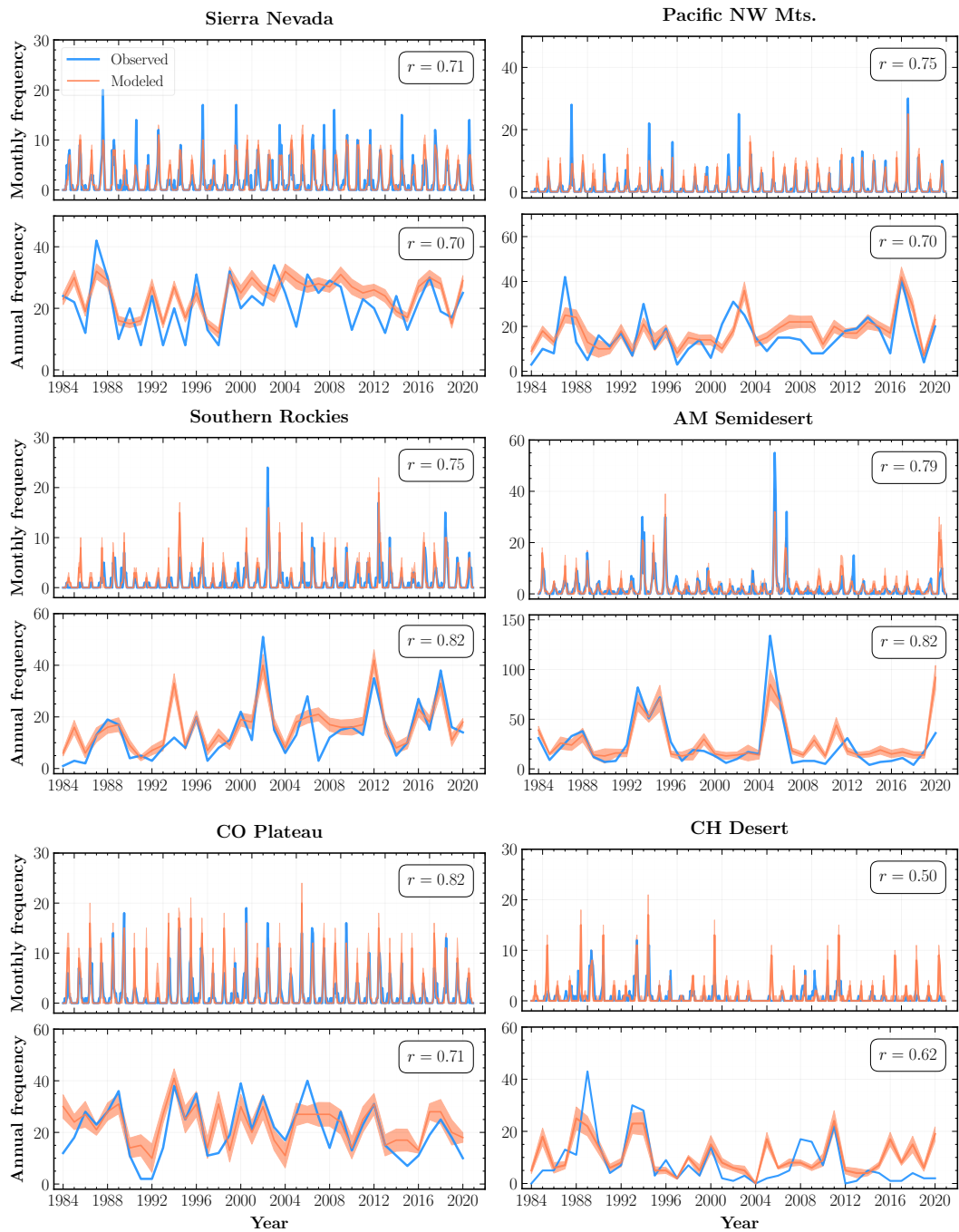
Figure 5. As in Fig. 4, but with different WUS Ecoregions.

		Frequency		Size	
Division	Ecoregion	r	χ_r^2	r	χ_r^2
Forests	Sierra Nevada [SNV]	0.70	20.98	0.60	0.84
	California (CA) North Coast [CNC]	0.68	107.75	0.58	1.52
	CA Central Coast [CCC]	0.46	75.64	0.60	1.17
	CA South Coast [CSC]	0.61	40.91	0.66	1.70
	Pacific Northwest Mountains [PNW]	0.70	9.63	0.81	0.66
	Northern Rockies [NRM]	0.89	11.28	0.93	0.34
	Middle Rockies [MRM]	0.85	25.69	0.84	1.30
	Southern Rockies [SRM]	0.82	10.20	0.79	0.81
	Arizona/New Mexico Mountains [AZNMM]	0.72	45.57	0.63	1.18
Deserts	American (AM) Semidesert [AMSD]	0.88	15.80	0.95	0.77
	Intermountain (IM) Semidesert [IMSD]	0.66	36.52	0.90	0.43
	IM Desert [IMD]	0.80	34.99	0.92	0.39
	Chihuahuan (CH) Desert [CHD]	0.62	68.93	0.91	0.52
	Columbia Plateau [CLP]	0.78	14.79	0.77	0.56
	Colorado Plateau [COP]	0.71	12.10	0.72	0.39
	Southwestern (SW) Tablelands [SWT]	0.67	15.88	0.94	0.71
Plains	Northern Great Plains [NGP]	0.87	5.75	0.93	0.43
	High Plains [HP]	0.65	8.73	0.94	0.57

Table 2. Goodness-of-fit metrics in terms of Pearson's correlation (r) and the reduced chi-squared statistic (χ_r^2) between the observed and modeled time series for both frequencies and area burned at an annual time scale. Both r and χ_r^2 are dimensionless metrics; higher values of r and lower values of χ_r^2 indicate a better fit. Results are shown for each Ecoregion organized by their ecological Division. The abbreviated Ecoregion names shown in Fig. 1 are given in square brackets.

of fire frequencies to hot and dry conditions in our model combined with the inadequate representation of human action. The latter is important especially since human predictors such as population and housing density can have a dual effect on fire frequencies: proximity to urban settlements increases the probability of ignitions and access to suppression resources, whereas
405 reductions in fuel continuity due to development and land management drastically reduce the probability that an individual ignition grows into a large wildfire (Knorr et al., 2014; Andela et al., 2017).

Since our model is trained on data over the whole WUS, its performance, on average, is better over Ecoregions with larger number of fires, such as the Middle Rockies and IM Semidesert. On the other hand, our model performs quite poorly for regions with a low number of total fires, where it is more likely to exhibit large interannual variations over a baseline of very
410 few to no fires. This behavior is evident in the plot for CH Desert in Fig. 5 as well as the low r and high χ_r^2 values in Table 2 for Ecoregions such as CA North Coast, SW Tablelands, CH Desert, and High Plains.

The SHAP values for individual predictors of the frequency MDN as well as the partial dependence plots for two important predictors, VPD ($S = 0.042$) and FM1000 ($S = 0.030$), with a color gradient corresponding to Tmax values are plotted for the WUS in Fig. 6. Here, S denotes the mean absolute SHAP values of each predictor. Similar plots at the Divisional level for Forests, Deserts, and Plains are shown in Fig. 7. We also include the mean SHAP plots for each of the 18 Ecoregions considered in our analysis in Figs. S2 and S3. The main drivers of fire frequencies at all spatial scales are climate predictors correlated with hot and arid fire weather conditions. Thus, high VPD is the leading predictor for most Ecoregions, most frequently combined with low fuel moisture in large diameter fuels, FM1000. Other important predictors include Tmax, Prec, and Slope, such that high daily maximum temperatures, lower fire month precipitation total, and higher mean slope all contribute to higher fire frequency on average.

Among the other subdominant predictors, antecedent climate conditions play a varying role across different Divisions. Antecedent snowpack estimated using 3 month average SWE, $\text{AvgSWE}_{3\text{mo}}$, is an important predictor especially in the Rocky Mountains Ecoregions ($S = 0.023$), with lower SWE in 3-4 antecedent months leading to higher predicted frequencies in the fire month. Increases in modeled fire frequencies for several Ecoregions in Deserts and Plains are also driven by antecedent conditions at both the seasonal and annual timescales through lower values of $\text{AntPrec}_{3\text{mo}}$ and higher values of $\text{AntPrec}_{1\text{yr}}$ predictors. The latter result corroborates previous analyses (Crimmins et al., 2004; Abatzoglou et al., 2017) which have highlighted the role of high prior year precipitation in promoting fuel growth within arid regions where vegetation is often too limiting to allow for large fires. The spatial variability in vegetation predictors, however, is of low importance for most Ecoregions. Interestingly, for similar dryness levels, our model simulates more fires for sites with lower values of Biomass relative to
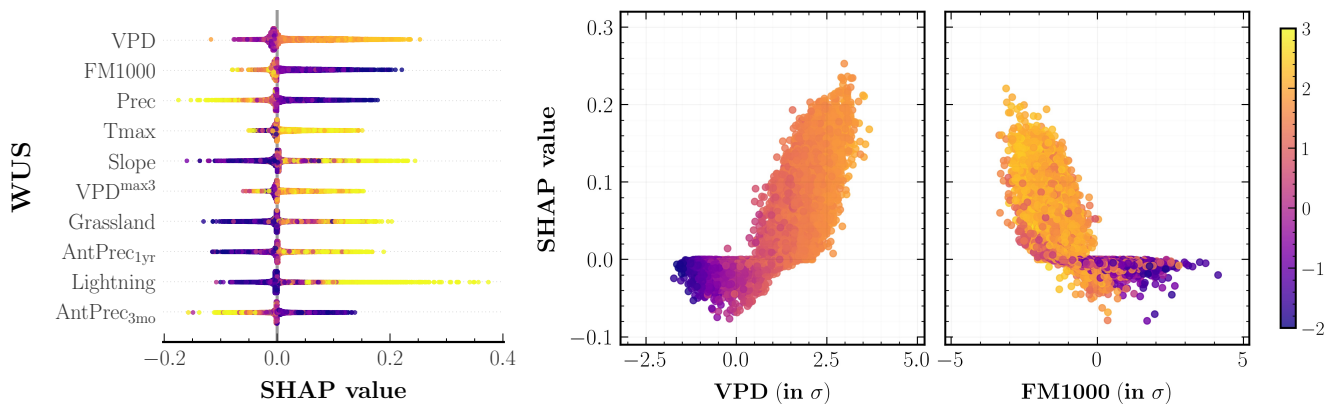
Figure 6. SHapley Additive exPlanation (SHAP) analysis of the fire frequency MDN model outputs across the western United States. *Left:* Input predictors sorted in descending order of their mean SHAP values aggregated over the entire study period. Each colored point along the x-axis represents an individual prediction with the color corresponding to high (yellow) or low (indigo) values of the respective input predictor. *Middle* and *Right:* Partial dependence plots for two important predictors shown on the x-axis, colored corresponding to high (yellow) or low (indigo) values of the mean daily maximum temperature, Tmax. The colorbar (far right) is normalized in terms of standard deviations (σs) for all relevant values across the three panels.
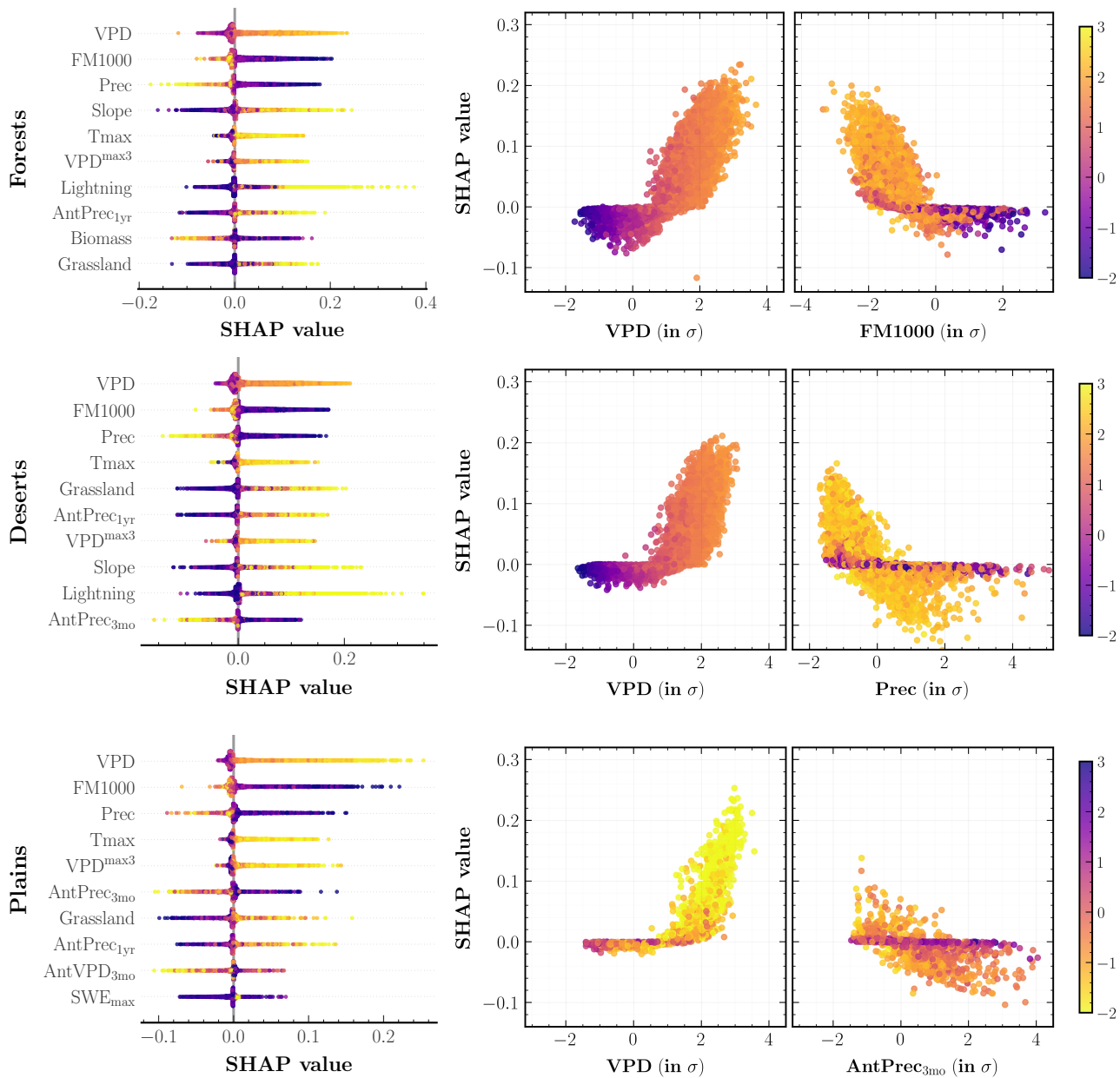
Figure 7. SHapley Additive exPlanation (SHAP) analysis of the fire frequency MDN model outputs for different western United States' Divisions: (top) Forests, (middle) Deserts, and (bottom) Plains. *Left column:* Input predictors sorted in descending order of their mean SHAP values aggregated over the entire study period. Each colored point along the x-axis represents an individual prediction with the color corresponding to high (yellow) or low (indigo) values of the respective input predictor. *Middle*, *Right:* Partial dependence plots for two important predictors shown on the x-axis, colored corresponding to high (yellow) or low (indigo) values of the mean daily maximum temperature, Tmax (top and middle panel), and high (indigo) or low (yellow) values of the 1000-hr dead fuel moisture, FM1000 (bottom panel). The colorbar (far right) is normalized in terms of standard deviations (σs) for all relevant values across the three panels.

430   sites with higher Biomass in Forests; meanwhile, higher fraction of grassland results in a higher fire frequency across all three

Divisions. Our model considers lightning strike density as an important predictor across several Ecoregions, most notably over

the CA North ($S = 0.021$) and South ($S = 0.024$) Coasts, PNW Mountains ($S = 0.021$), and Middle Rockies ($S = 0.019$).

Human predictors, on the other hand, are not among the top 10 predictors for any Ecoregion. Given that a large fraction of fires

in parts of the WUS, especially Mediterranean California and coastal PNW, are human ignited (Balch et al., 2017), this result

435   could stem from a skewed sampling of fires while training SMLFire1.0 as well as the lack of correlation between our chosen

human predictors and fire occurrences.

     We visualize the response of fire frequencies to individual predictors through the partial dependence plots in Figs. 6 and 7.

SHAP values for all four variables shown in the plots: VPD, FM1000, Prec, and $\mathrm{AntPrec_{3mo}}$ exhibit near-linear relationships

above a threshold with their respective predictor values. The color gradient of all test points in Forests and Deserts shows that

440   VPD and FM1000 are strongly correlated with Tmax, while also highlighting the interaction effect between Tmax and Prec.

In Plains, instead of Tmax, we consider the interaction effect of FM1000 on VPD and $\mathrm{AntPrec_{3mo}}$ to explore the influence

of antecedent and fire month predictors on fuel moisture. FM1000 values exhibit a strong interaction effect with antecedent

precipitation in our model, but not with fire month VPD since fuel moisture shows significant correlations with atmospheric

aridity. In other words, SHAP values for our frequency model vary with Prec and $\mathrm{AntPrec_{3mo}}$ predictors only for sites with

445   high values of Tmax and low values of FM1000.

4.3   Fire Size

We use MDNs trained on fires $\geq 4\ \mathrm{km}^2$ to determine the parameters of the combined GPD and GPD-Ext (henceforth combined

GPD) distribution of individual fire sizes. MC simulations of the parametric size distributions for all observed fires from

1984 to 2020 are aggregated to compute the mean of the monthly and annual area burned (MAB and AAB respectively) and

450   their 1σ uncertainty intervals. The total MAB and AAB simulated using the combined GPD model with a breakpoint after

2004 are plotted for the entire WUS in Fig. 8 and separately at the Ecoregion level in Figs. 9 and 10. The goodness-of-fit

metrics, namely Pearson's correlation and reduced chi-squared statistic, between the predicted and observed sizes for each

Ecoregion are summarized in Table 2. We plot the SHAP values for individual predictors at test points across the WUS in

Fig. 12, alongside partial dependence plots for two important fire size predictors, VPD and Grassland, with a color gradient

455   corresponding to Tmax values. These plots are constructed using a procedure similar to the one described in the previous

section for fire frequencies. We also show the partial dependence plots at the Division level are in Fig. 13 and plot the mean

SHAP values for individual Ecoregions in Figs. S7 and S8.

     The introduction of a time-dependent response and a breakpoint after 2004 in our modeling is justified through the following

analysis. As indicated by Fig. 8, there is a rising trend in the AAB for the WUS with significantly more large MAB months

460   ~ 2000 onward than in 1984-1999. Moreover, as shown by Juang et al. (2022), this increase in AAB is driven by the exponential

response of fire size to atmospheric aridity and not due to increasing fire frequency. We confirm their result by noting that the

complementary cumulative distribution function (CCDF) of observed fire sizes between 1984-2004 is markedly different from

the CCDF for sizes observed between 2005-2020 as plotted in Figs. S4 and S5. In fact, as shown in Fig. S4 (right panel), the

Western United States
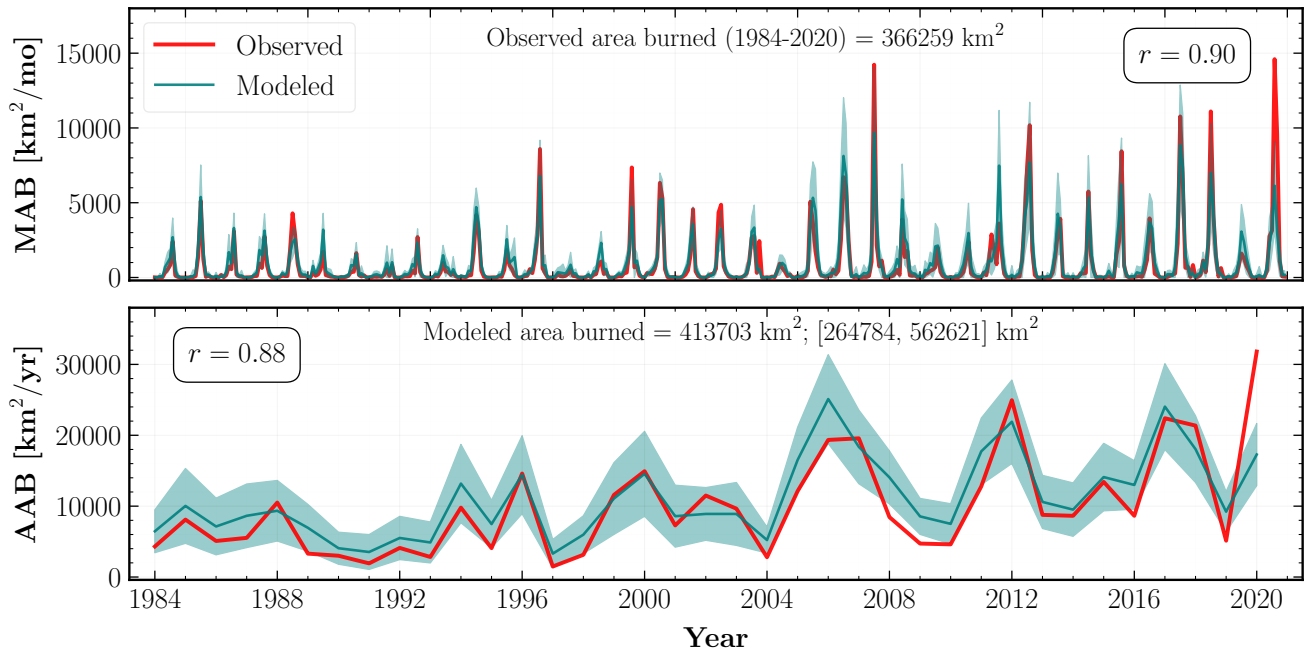
Figure 8. Observed (red) and mean modeled (teal) area burned across the western United States at monthly (MAB) and annual (AAB) time scales from 1984 to 2020. The teal shaded regions indicate 1σ uncertainty intervals for the mean area burned aggregated over the Monte Carlo (MC) simulations from the combined GPD model for all observed fires. The mean total area burned over the study period as well as its 1σ uncertainty interval are indicated at the top of the lower panel. Also shown within each subplot is the Pearson correlation coefficient between the observed and modeled burned area time series.

CCDF of the combined GPD model is in much better agreement with the observed CCDF than the CCDFs of either model
465 individually (Fig. S4; left and middle panels). After varying the breakpoint for different years between 2000 and 2006, we find that a breakpoint after 2004 results in the best fit to the observed area burned. Thus, we successfully model the shift to larger fire sizes observed after 2004 by including an additional GPD distribution with fatter tails.

We emphasize that the improved agreement between the CCDFs of observed and modeled sizes is not merely an artifact of the breakpoint procedure; in fact the choice of the distribution plays a critical role. Specifically, we verify this by repeating our
470 analysis with the Lognormal distribution, which has thinner tails than the GPD. In Fig. S5, we demonstrate that the CCDF of the reweighted Lognormal MDN (Lognorm-Ext) underestimates large portions of the observed sizes' CCDF while being able to account for only the most extreme fires. Consequently, the combined distribution (Lognorm-Comb) is an inadequate model for the observed fire sizes over the study period.

The choice of fire frequencies – either observed or modeled – and the stochasticity in fire locations affects both the inter-
475 annual variability and total area burned of the modeled AAB time series. Contrasting the results shown in Fig. 8 and Fig.
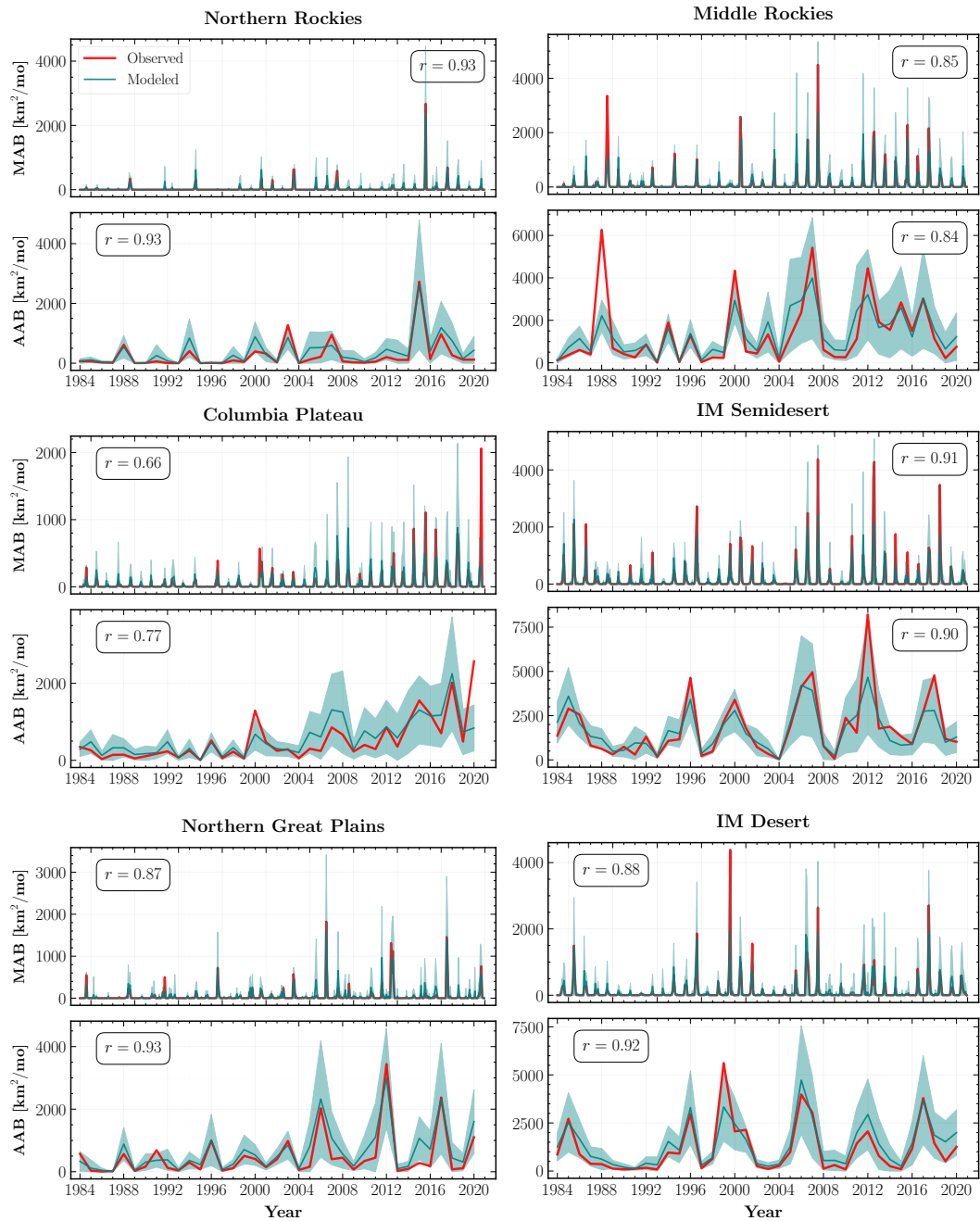
Figure 9. Observed (red) and mean modeled (teal) monthly (MAB) and annual area burned (AAB) from 1984 to 2020 for the Ecoregions shown in Fig. 4. The teal shaded regions within each subplot indicate 1σ uncertainty intervals for the mean regional area burned aggregated over the Monte Carlo (MC) simulations from the combined GPD model for all observed fires. Also shown is the Pearson correlation coefficient (r) between the observed and modeled area burned time series for each Ecoregion.
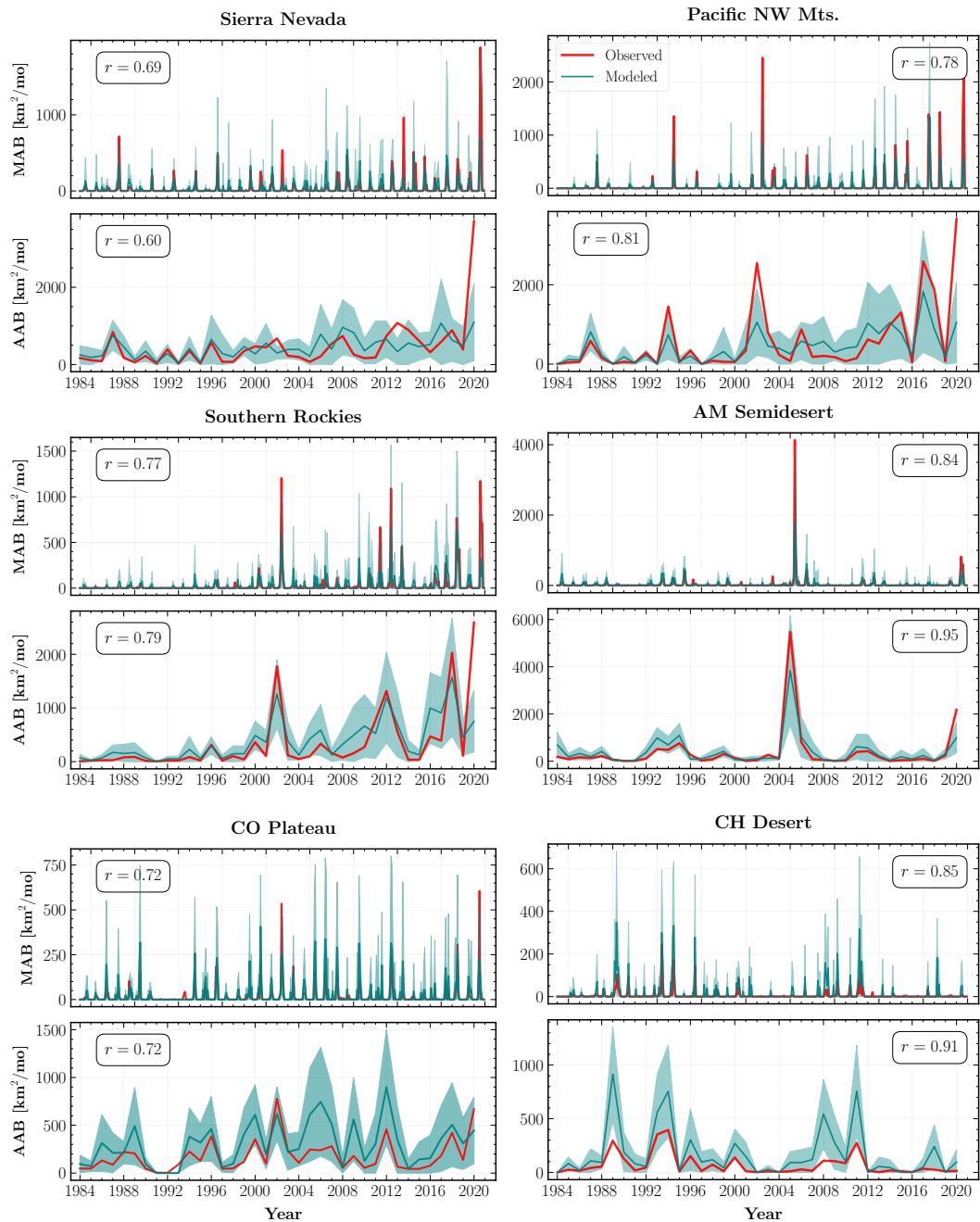
Figure 10. As in Fig. 9, but with different WUS Ecoregions.

S6, we note that the AAB using modeled frequencies and observed locations results in a moderate decrease in the total area burned along with a marginal improvement in the correlation with the observed AAB. Further, simulating fire sizes with model locations leads to a significant rise in the total area burned irrespective of the frequency source, although the AAB time series with observed frequencies has a notably weaker correlation ($r = 0.77$) compared to the case with modeled frequencies ($r = 0.91$). We explain this behavior with respect to both fire locations and frequencies. First, the frequency MDN tends to locate fires in grid cells with extreme values of input predictors such as VPD and Prec, leading to simulated fire sizes larger than those at observed fire locations; important fire predictors at the latter, especially before 2004, having high (or low), but not extreme, values. Second, as shown in Fig. 3, the modeled frequencies are consistently lower than the observed ones between 1993 and 2001 leading to a lower simulated area burned relative to the case with observed frequencies for the same source of fire locations. At the same time, the anomalously high modeled frequency for 2020 leads to improved correlations for the modeled AAB time series irrespective of the choice of fire locations. As a cross-check, we compute the correlation coefficient for the two AAB time series simulated with observed locations and the observed AAB between 1984 and 2019, and find that the simulation with observed frequencies has a higher correlation ($r = 0.93$) than the one with modeled frequencies ($r = 0.89$). We do not find a similar improvement in the correlations of the two AAB time series simulated with model locations. These results serve as an important lesson for modeling fire activity: *improved correlations of model predictions with observed data may not always be a good indicator of improved model accuracy.*

Using the combined GPD MDN model, our modeled MAB ($r = 0.90$) and AAB ($r = 0.88$) time series are very good fits to the observed area burned, within the 1σ uncertainty interval, when considering the entire WUS. Our model performs well across most of the WUS with AAB predictions for 15 out of the 18 Ecoregions exhibiting strong correlations ($r \geq 0.7$) with the observed area burned. Moreover, as shown in Figs. 9 and 10, the trends in the modeled AAB time series successfully emulate the distinct multidecade increases in observed AAB over both forested and non-forested Ecoregions.

Our model has mixed skill in predicting large MAB and AAB during the study period. For example, our model is able to simulate the full range of AAB variability in the Northern Rockies, Northern Great Plains (Fig. 9; top and middle left panels), and American Semidesert (Fig. 10; middle right panel), but it fails to capture the largest AAB between 1984 and 2019 in Middle Rockies, IM Semidesert (Fig. 9; top and middle right panels), and PNW Mountains (Fig. 10; top right panel). This tendency holds even after reweighting the size distribution for post-2004 fires. In particular, the total AAB during the year 2020 deserves further scrutiny. The modeled AAB significantly underestimates the observed 2020 AAB in Fig. 8, predicting only about half of the observed value. We see similar behavior at the Ecoregional level in Columbia Plateau (Fig. 9; middle left panel), Sierra Nevada, and Southern Rockies (Fig. 10; top and middle left panels).

With these results, we can make a stronger assessment about our modeling framework: first, for almost all years in our study period, the mean of the aggregate area burned distribution is a good approximation for the observed time series, so the only challenging part is determining the time dependence of the mean sizes of individual fires; and second, while the discrepancy between modeled and observed area burned in 2020 highlights a clear limitation of our model, can we use still use it to make meaningful predictions for anomalous extreme fire years?
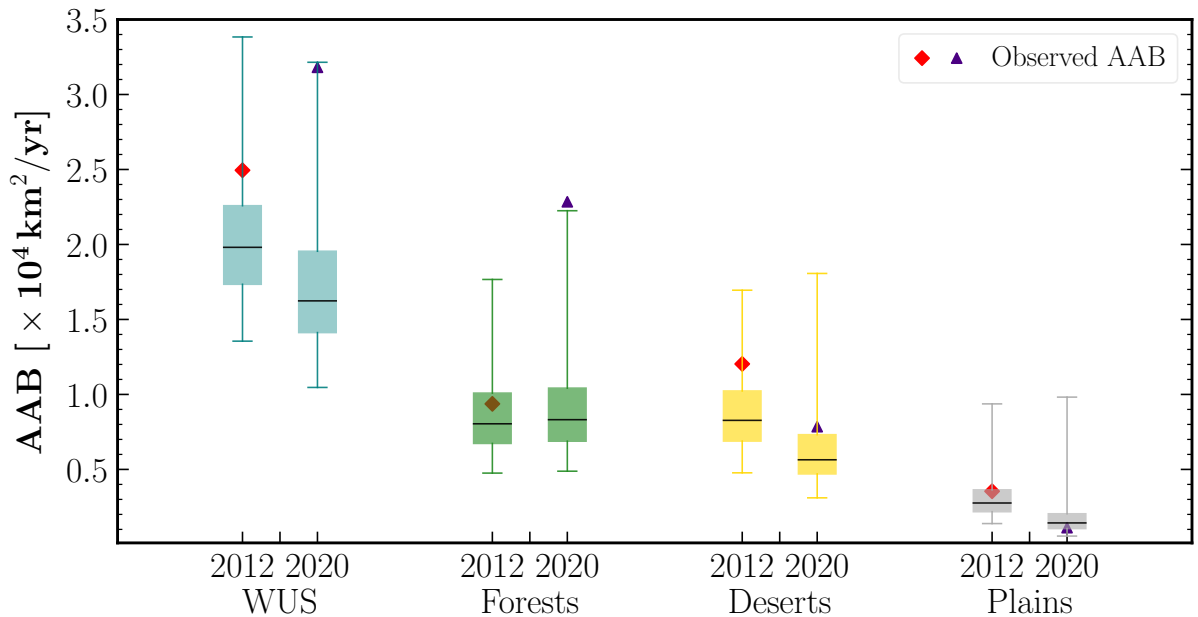
Figure 11. Boxplots of predicted annual area burned (AAB) for two extreme fire years, 2012 and 2020, for the entire western United States (WUS) (teal) and three Divisions organized by their primary vegetation types: Forests (green), Deserts (yellow), and Plains (gray). The lower and upper whiskers of each boxplot indicate the 0.5^{th} and 99.5^{th} percentile of the predicted AAB distribution, whereas the black line represents its median value. Also shown for reference are the observed AAB for both 2012 (red diamond) and 2020 (indigo traingle).

510   In Fig. 11, we show the total modeled AAB for two extreme fire years with the largest area burned, 2012 and 2020, at both the WUS and Divisional scales. We find that the observed AAB for 2012 is in the $\sim 80^{\text{th}}$ percentile of the predicted WUS AAB, driven primarily by the large AAB in Deserts. However, for 2020, the observed $\gtrsim 99.5^{\text{th}}$ percentile Forests AAB resulted in the total WUS AAB to be in the 99.5^{th} percentile of our model predictions. Contrasted with previous extreme fire years, 2012 and 2017 (see Fig. S9), the observed 2020 AAB is: *a)* in very high percentiles of the modeled AAB simulated

515 with observed frequencies,implying that the observed fire sizes for 2020 were much greater than those in the 1984-2019 period that our model is trained on; *b)* dominated by anomalously high AAB in the Forests Division, and is a striking example of an extreme fire year driven by large fires in flammability-limited areas rather than fuel-limited ones.

    Among the 10 input predictors of fire size shown in descending order of importance in Fig. 12, the SHAP technique selects VPD ($S = 3.95$), Grassland ($S = 2.60$), and FM1000 ($S = 2.34$) as the three most important predictors at the WUS level.

520 Again, S refers to the mean absolute SHAP value for each predictor. These are also among the top predictors at the Divisional scale as shown in Fig. 13, with Grassland being more important than FM1000 in Deserts ($S = 2.75$) and Plains ($S = 3.33$). Broadly, SHAP values for all predictors besides FM1000, Prec, and $\text{AntPrec}_{\text{3mo}}$, have a positive relationship with higher predictor values. Another important predictor at the WUS and Forests level is Slope ($S = 2.43$): its SHAP values indicate that fire size is promoted by large topographic slope, which is consistent with previous findings (Andrews, 2018). Assessing the

predictor importance at the Ecoregion level, as illustrated in Fig. S7 and S8, we find that climate and fire weather predictors are dominant drivers across Forests, whereas Grassland plays a larger role in Deserts and Plains. The importance of grassland cover could also signal the role of invasive grass species (Knapp, 1998; Balch et al., 2013) in driving large area burned within our model. Thus, vegetation plays a much more important role in simulating area burned for the size model as compared to the frequency model. This is also true for most Ecoregions in Forests Division, where the spatial distribution of aboveground biomass serves as an important secondary predictor. The mean absolute SHAP values suggest that weekly scale extreme weather predictors such as $\mathrm{FFWI^{max3}}$ are also important predictors in several Ecoregions. We interpret the response of fire sizes simulated by our model to the climate at different temporal scales as follows: monthly to seasonal scale hot and arid weather create favorable conditions for fire spread, while the growth of large fires is facilitated by weekly scale extreme fire weather. The SHAP plot for Pop10_dist in Deserts ($S = 1.54$) and Plains ($S = 1.32$) indicates that higher predictor values result in simulation of larger fire sizes. This could be because increased distance from populated areas is correlated with a decrease in accessibility for fire suppression efforts, and therefore higher occurrences of larger fires in the observed record.

We also retrain our fire size model with two different variations of the input predictors selected for the main analysis: first with relative humidity (RH), average RH over 3 antecedent months ($\mathrm{AvgRH_{3mo}}$), and 3 day minimum RH ($\mathrm{RH^{min3}}$) instead of VPD and its corresponding permutations; and second including all VPD and RH predictors. There are no significant differences in the correlations between the simulated and observed WUS AAB time series in either case, but the SHAP summary plots shown in Fig. S10 provide valuable insights. In the first case where VPD is not considered as a potential predictor, RH ($S = 1.67$) replaces VPD as the leading fire size driver across the WUS, and $\mathrm{AvgRH_{3mo}}$ ($S = 1.00$) is more important relative to other antecedent climate predictor. When allowing both VPD and RH to serve as predictors, VPD ($S = 3.58$) has higher predictive power than RH ($S = 1.95$) in our model at both the WUS and Forests Divisional scales. From the perspective of predictor importance, there might actually be an advantage to using RH instead of VPD: the correlation between VPD and Tmax leads to a small but spurious trend in SHAP value for Tmax in Forests as shown in Fig. 13, whereas using only RH and
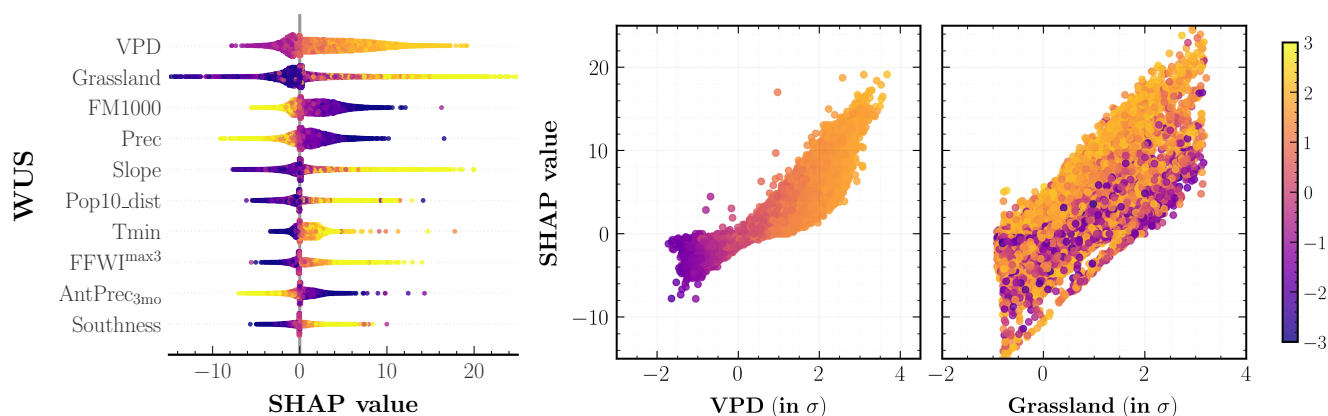
Figure 12. As in Fig. 6, but for the fire size MDN model.
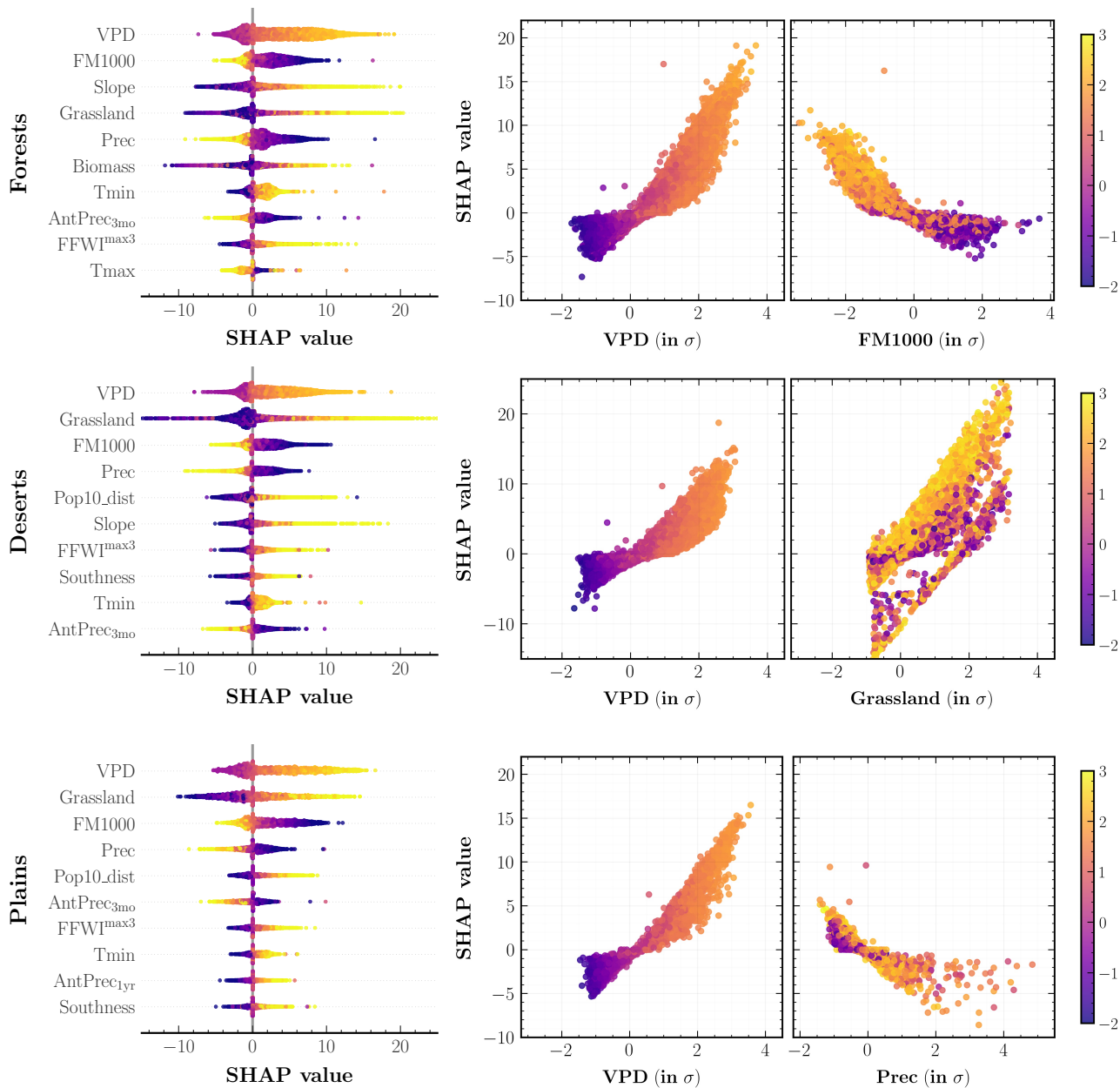
Figure 13. SHapley Additive exPlanation (SHAP) analysis of the fire size MDN model outputs for different western United States' Divisions: (top) Forests, (middle) Deserts, and (bottom) Plains. *Left column:* Input predictors sorted in descending order of their mean SHAP values aggregated over the entire study period. Each colored point along the x-axis represents an individual prediction with the color corresponding to high (yellow) or low (indigo) values of the respective input predictor. *Middle*, *Right:* Partial dependence plots for two important predictors shown on the x-axis, colored corresponding to high (yellow) or low (indigo) values of the mean daily maximum temperature, Tmax (all panels). The colorbar (far right) is normalized in terms of standard deviations (σs) for all relevant values across the three panels.

Tmax yields the correct Tmax effect on fire size (see lower left panel of Fig. S10). On the other hand, from the perspective of future climate-fire relationships, the fact that the decision of including VPD, RH, or both does not substantially affect model performance does not mean that this decision is unimportant. As Brey et al. (2021) point out, VPD is projected to continue rising dramatically while projected RH decreases are more moderate. In this paper we prioritize the model that uses VPD, because VPD is more directly representative of the atmosphere's evaporative demand (Anderson, 1936; Monteith, 1965).

Lastly, we show the responses of fire sizes to individual predictor values for all test points at the WUS and Divisional level in Figs. 12 and 13 respectively. We find that fire sizes simulated by our model respond, above a threshold, exponentially to increases in VPD and decreases in FM1000 at all spatial scales, although the response is notably stronger in Forests. This result is consistent with the findings of Juang et al. (2022) who showed that the exponential response of fire sizes to increasing aridity appears to arise from the fact that large fires have much greater capacity for area growth than smaller fires. Meanwhile, we do not find any significant interactions between Tmax and VPD as well as FM1000; Grassland shows a weak interaction effect with Tmax, such that sites with the same fraction of grassland cover yield larger sizes at higher values of Tmax.

5 Discussion

We have developed a novel stochastic ML framework, SMLFire1.0, for modeling fire activity across different WUS Ecoregions. Although the fire frequency and area burned time series simulated using this framework are in good agreement with observations at multiple spatial and temporal scales, there are several areas of improvement across three interconnected themes: modeling approach and architecture, vegetation, and other potential predictors. We discuss each one of these themes in detail below.

- *Modeling:* A limitation of the frequency model is that we are, effectively, estimating a joint distribution between ignitions and fire likelihood. In other words, we are using data for observed fires, which occur randomly, to learn the relationships between different predictors that contribute to fire conducive conditions. However, such an approach may introduce a bias in ignition-limited regions that could have large fire-prone areas with no fire occurrences (Parisien and Moritz, 2009). One way to improve our framework would be to model ignitions using spatial stochastic processes, or to compute fire probabilities using a presence-only approach (Chen et al., 2021). Alternatively, we could leverage the seasonal differences between human and lightning started fires to account for potential selection biases in training data for SMLFire1.0. We also expect that further improvements to the WUMI dataset, especially for smaller fires, would improve the accuracy of our modeled frequency time series. For the fire size model, a major limitation of our current approach is its reliance on climate predictors whose spatial and temporal scales are coarse relative to the physical scales involved in fire front propagation (Bakhshaii and Johnson, 2019). Bridging these two regimes (Sullivan, 2009) is an important focus of our ongoing work to improve predictability of extreme fire behavior. Moreover, we combined two GPD distributions with a breakpoint after 2004 to obtain a distribution that best fit the cumulative distribution of observed fire sizes. Rather than introduce a breakpoint by hand, in future work we intend to explore and model the mechanisms that may have led to such a distribution shift. At the computational level, we plan on incorporating a recurrent neural network component to

our ML architecture in order to model the nonstationary fire size responses over longer timescales. An extension would be to expand the parametric size distribution by including a smooth, differentiable form of the Lognormal-GPD with an arbitrary threshold parameter. Using such a hybrid distribution would ensure that our model has more flexibility in simulating a large number of small fires from a distribution with finite mean and variance as well as a small number of larger fires from a distribution with finite mean but infinite variance (Cohen and Xu, 2015). To improve model interpretability and avoid post hoc evaluations of feature importance such as SHAP (Hooker et al., 2021), a more robust alternative would be to build an interpretable ML model from the bottom up as outlined in Alvarez-Melis and Jaakkola (2018).

– *Vegetation:* A major area of desired improvement for SMLFire1.0 is the representation as well as the dynamic structure of vegetation predictors. This could be done in several different ways. First, by including finer scale vegetation characteristics through a combination of integrated data products, such as Effective Vegetation Type (EVT) (Rollins, 2009) or Normalized Difference Vegetation Index (NDVI) (Didan, 2015), and outputs from physically parameterized models (Hansen et al., 2022). These predictors would be helpful in informing the model about the type and spatial distribution of different live and dead fuels. Second, for predictions of future fire activity over longer time scales, it would be important to account for the nonstationarity of the climate-vegetation relationship, a pivotal factor in determining the spatially heterogeneous shifts in vegetation patterns (Higuera et al., 2009; Bradstock, 2010; Hansen et al., 2018). We may already be seeing evidence of this effect in our analysis: recent increases in aridity coupled with transitions in vegetation patterns could have precipitated a shift in the fire regimes across the WUS and promoted larger and more severe fires in the past two decades. Third, besides climate induced shifts, vegetation patterns are also affected by human and natural disturbances such as changes in land use (Klein Goldewijk and Ramankutty, 2004), tree mortality (Williams et al., 2013), insect range expansions as well as infestations (Pureswaran et al., 2018), and fire itself (Parks et al., 2018). Importantly, multiple studies have shown that fire-induced fuel limitations are expected to slow, but not abate the continued heat- and drought-induced increases in annual area burned across the WUS over the next few decades (Hurteau et al., 2019; Abatzoglou et al., 2021a). Thus, coupling the current stochastic ML model framework with a DGVM for a variety of vegetation types and different human intervention scenarios is a promising research direction.

– *Other predictors:* Several potentially relevant land-surface predictors were not considered here since their records are not available over the full duration of our study period. For instance, recent work has highlighted the role of remotely sensed soil moisture (Rigden et al., 2020) and the sensitivity of live fuel moisture content to atmospheric aridity (Rao et al., 2022) in regulating wildfire ignitions and area burned respectively. Reliable measurements over the WUS for both these predictors are only available after 2015. Meanwhile, human influence on individual fire sizes could be affected by synchronous fire activity over several regions. Abatzoglou et al. (2021b) approximate this effect by concurrent fire danger days, a metric that measures the strain on available resources for suppressing new ignitions as well as containment of ongoing fires. We will explore the role of additional land-surface and human action predictors in forthcoming analyses.

6 Conclusions

Disentangling the various climate, vegetation, and human drivers of wildfire frequency and sizes in the western United States is critical for developing accurate seasonal as well as longer term forecasts of fire activity. In this paper, we introduced a novel stochastic ML framework, SMLFire1.0, for estimating the parametric distributions of observed fire frequencies and sizes in $12\text{km} \times 12\text{km}$ grid cells observed on a monthly time scale. These distributions were sampled using Monte Carlo simulations to obtain the mean and variance for fire frequency as well as area burned at monthly and annual scales in several WUS Ecoregions. We improve upon previous regression-based and ML approaches in several concrete ways. In particular, our model: relies only on the spatiotemporal variability of dynamic predictors, the spatial variability of static predictors, and not on any predictors related to location and time such as latitude or calendar month; captures the nonlinear interactions among different predictors at multiple spatial scales; and provides robust parametric model uncertainty estimates for our results.

Our main results are as follows: *a)* the time series for both modeled frequencies and area burned are in good statistical agreement with the observed data over monthly and annual timescales at spatial scales from Ecoregions to the whole WUS; *b)* the modeled area burned successfully accounts for the interannual variability and multidecadal trends in the observed area burned in both forested and nonforested regions; *c)* for anomalous extreme fire years such as 2020, the stochastic model is useful for estimating the upper percentiles, *i.e.* 95th, 99th..., of the total annual area burned distribution; *d)* the cumulative observed fire size distribution is best fit by a combined GPD model with finite mean but infinite variance, which has important consequences for how resources are allocated for fuel treatment and fire containment.

We used the SHAP technique to evaluate the predictor importance for the frequency and size models at the Ecoregional, Divisional, and WUS scales. While VPD is the leading predictor at both smaller and larger scales, the order of subleading fire month predictors – precipitation total, mean daily maximum and minimum temperatures, moisture in large diameter dead fuels – as well as the fraction of grassland cover, aboveground biomass, and topography varies across Ecoregions, indicating that our model is able to generalize well across spatially heterogeneous climate, vegetation, and human gradients. Furthermore, we visualized the different functional relationships between predictor values and wildfire activity with potential interaction effects through partial dependence plots for several important predictors. Besides fire month variables, we find that increased fire frequencies in our model are driven by a set of antecedent predictors acting at two distinct timescales across Forests, Deserts, and Plains: a seasonal (3-4 months) scale associated with snow or precipitation drought, and a cumulative longer term (1-2 years) scale correlated with wetter conditions that promote fuel growth. Modeled fire sizes, on the other hand, are mostly sensitive to seasonal scale antecedent conditions.

Future research directions will focus on expanding the SMLFire1.0 model framework to include: a stochastic model for human ignitions, nonstationary relationships between predictors and fire activity, fire-fuel feedback over different climate and vegetation gradients, as well as additional finer scale moisture and human action predictors. Moreover, we intend to incorporate SMLFire1.0 as a subgrid-scale parameterization scheme for the fire modules of a regional scale DGVM and ESM while also benchmarking it against existing parameterizations.

Code and data availability. The code for training and validating the SMLFire1.0 model as well as reproducing the figures shown in this paper is available at: https://doi.org/10.5281/zenodo.7277980. The WUMI dataset (Juang et al., 2022) contains all the fire occurrences and sizes analyzed in this paper and is hosted at: https://doi.org/10.5061/dryad.sf7m0cg72.

Author contributions. J.B., A.P.W., and P.G. conceived the study. A.P.W., C.J., and W.D.H. handled the data collection and processing. J.B. performed the analysis and wrote the paper. All authors discussed the results and contributed to the final version of the draft.

Financial support. This work is funded by the Zegar Family Foundation and the Amazon Reseach program. C.J. is supported by a Future Investigators in NASA Earth and Space Science and Technology (FINESST) Grant 80NSSC20K1617. W.D.H. acknowledges support from The Gordon and Betty Moore Foundation (GBMF10763), the Environmental Defense Fund, and the Three Cairns Group. P.G.'s research is funded by the NSF LEAP Science and Technology Center # 2019625 and the USMILE European Research Council grant.

Competing interests. The authors declare no competing interests.

Acknowledgements. We acknowledge initial work from Kenza Amara during an internship at Columbia University on machine learning and wildfires. We are also grateful for the critical feedback from three anonymous reviewers that improved the presentation of our results.

References

Abatzoglou, J. T.: Development of gridded surface meteorological data for ecological applications and modelling, International Journal of Climatology, 33, 121–131, https://doi.org/https://doi.org/10.1002/joc.3413, 2013.

Abatzoglou, J. T. and Kolden, C. A.: Relationships between climate and macroscale area burned in the western United States, International Journal of Wildland Fire, 22, 1003–1020, https://doi.org/10.1071/WF13019, 2013.

Abatzoglou, J. T. and Williams, A. P.: Impact of anthropogenic climate change on wildfire across western US forests, Proceedings of the National Academy of Sciences, 113, 11 770–11 775, https://doi.org/10.1073/pnas.1607171113, 2016.

Abatzoglou, J. T., Kolden, C. A., Williams, A. P., Lutz, J. A., and Smith, A. M. S.: Climatic influences on interannual variability in regional burn severity across western US forests, International Journal of Wildland Fire, 26, 269–275, https://doi.org/10.1071/WF16165, 2017.

Abatzoglou, J. T., Battisti, D. S., Williams, A. P., Hansen, W. D., Harvey, B. J., and Kolden, C. A.: Projected increases in western US forest fire despite growing fuel constraints, Communications Earth & Environment, 2, https://doi.org/10.1038/s43247-021-00299-0, 2021a.

Abatzoglou, J. T., Juang, C. S., Williams, A. P., Kolden, C. A., and Westerling, A. L.: Increasing Synchronous Fire Danger in Forests of the Western United States, Geophysical Research Letters, 48, e2020GL091 377, https://doi.org/https://doi.org/10.1029/2020GL091377, 2021b.

Abolafia-Rosenzweig, R., He, C., and Chen, F.: Winter and spring climate explains a large portion of interannual variability and trend in western U.S. summer fire burned area, Environmental Research Letters, 17, https://doi.org/10.1088/1748-9326/ac6886, 2022.

Alvarez-Melis, D. and Jaakkola, T. S.: Towards Robust Interpretability with Self-Explaining Neural Networks, arXiv e-prints, 2018.

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

Anderson, D. B.: Relative Humidity or Vapor Pressure Deficit, Ecology, 17, 277–282, http://www.jstor.org/stable/1931468, 1936.

Andrews, P. L.: The Rothermel surface fire spread model and associated developments: A comprehensive explanation, Gen. Tech. Rep. RMRS-GTR-371. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. 121 p., 371, 2018.

Bailey, R. G.: Ecoregions of the United States, in: Ecosystem Geography, pp. 83–104, Springer New York, New York, NY, https://doi.org/10.1007/978-1-4612-2358-0_7, 1996.

Bakhshaii, A. and Johnson, E.: A review of a new generation of wildfire–atmosphere modeling, Canadian Journal of Forest Research, 49, 565–574, https://doi.org/10.1139/cjfr-2018-0138, 2019.

Balch, J. K., Bradley, B. A., D'Antonio, C. M., and Gómez-Dans, J.: Introduced annual grass increases regional fire activity across the arid western USA (1980-2009), Global Change Biology, 19, 173–183, https://doi.org/10.1111/gcb.12046, 2013.

Balch, J. K., Bradley, B. A., Abatzoglou, J. T., Nagy, R. C., Fusco, E. J., and Mahood, A. L.: Human-started wildfires expand the fire niche across the United States, Proceedings of the National Academy of Sciences, 114, 2946–2951, https://doi.org/10.1073/pnas.1617394114, 2017.

Bastos, A., Ciais, P., Friedlingstein, P., Sitch, S., Pongratz, J., Fan, L., Wigneron, J. P., Weber, U., Reichstein, M., Fu, Z., Anthoni, P., Arneth, A., Haverd, V., Jain, A. K., Joetzjer, E., Knauer, J., Lienert, S., Loughran, T., McGuire, P. C., Tian, H., Viovy, N., and Zaehle, S.: Direct and seasonal legacy effects of the 2018 heat wave and drought on European ecosystem productivity, Science Advances, 6, https://doi.org/10.1126/sciadv.aba2724, 2020.

Bishop, C.: Mixture density networks, Working paper, Aston University, 1994.

Bowman, D. M. J. S., Balch, J. K., Artaxo, P., Bond, W. J., Carlson, J. M., Cochrane, M. A., D'Antonio, C. M., DeFries, R. S., Doyle, J. C., Harrison, S. P., Johnston, F. H., Keeley, J. E., Krawchuk, M. A., Kull, C. A., Marston, J. B., Moritz, M. A., Prentice, I. C., Roos, C. I., Scott, A. C., Swetnam, T. W., van der Werf, G. R., and Pyne, S. J.: Fire in the Earth System, Science, 324, 481–484, https://doi.org/10.1126/science.1163886, 2009.

Bradstock, R. A.: A biogeographic model of fire regimes in Australia: current and future implications, Global Ecology and Biogeography, 19, 145–158, https://doi.org/https://doi.org/10.1111/j.1466-8238.2009.00512.x, 2010.

Brey, S. J., Barnes, E. A., Pierce, J. R., Wiedinmyer, C., and Fischer, E. V.: Environmental Conditions, Ignition Type, and Air Quality Impacts of Wildfires in the Southeastern and Western United States, Earth's Future, 6, 1442–1456, https://doi.org/10.1029/2018EF000972, 2018.

Brey, S. J., Barnes, E. A., Pierce, J. R., Swann, A. L. S., and Fischer, E. V.: Past Variance and Future Projections of the Environmental Conditions Driving Western U.S. Summertime Wildfire Burn Area, Earth's Future, 9, https://doi.org/https://doi.org/10.1029/2020EF001645, 2021.

Burke, M., Heft-Neal, S., Li, J., Driscoll, A., Baylis, P., Stigler, M., Weill, J. A., Burney, J. A., Wen, J., Childs, M. L., and Gould, C. F.: Exposures and behavioural responses to wildfire smoke, Nature Human Behaviour, https://doi.org/10.1038/s41562-022-01396-6, 2022.

Carreau, J. and Bengio, Y.: A Hybrid Pareto Model for Conditional Density Estimation of Asymmetric Fat-Tail Data, in: Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics, edited by Meila, M. and Shen, X., vol. 2 of *Proceedings of Machine Learning Research*, pp. 51–58, PMLR, San Juan, Puerto Rico, https://proceedings.mlr.press/v2/carreau07a.html, 2007.

Chatterji, N. S., Haque, S., and Hashimoto, T.: Undersampling is a Minimax Optimal Robustness Intervention in Nonparametric Classification, arXiv e-prints, 2022.

Chen, B., Jin, Y., Scaduto, E., Moritz, M. A., Goulden, M. L., and Randerson, J. T.: Climate, Fuel, and Land Use Shaped the Spatial Pattern of Wildfire in California's Sierra Nevada, Journal of Geophysical Research: Biogeosciences, 126, https://doi.org/10.1029/2020JG005786, 2021.

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

Cohen, J. E. and Xu, M.: Random sampling of skewed distributions implies Taylor's power law of fluctuation scaling, Proceedings of the National Academy of Sciences, 112, 7749–7754, https://doi.org/10.1073/pnas.1503824112, 2015.

Coop, J. D., Parks, S. A., Stevens-Rumann, C. S., Crausbay, S. D., Higuera, P. E., Hurteau, M. D., Tepley, A., Whitman, E., Assal, T., Collins, B. M., Davis, K. T., Dobrowski, S., Falk, D. A., Fornwalt, P. J., Fulé, P. Z., Harvey, B. J., Kane, V. R., Littlefield, C. E., Margolis, E. Q., North, M., Parisien, M.-A., Prichard, S., and Rodman, K. C.: Wildfire-Driven Forest Conversion in Western North American Landscapes, BioScience, 70, 659–673, https://doi.org/10.1093/biosci/biaa061, 2020.

Crimmins, M. A., Comrie, A. C., Crimmins, M. A., and Comrie, A. C.: Interactions between antecedent climate and wildfire variability across south-eastern Arizona, International Journal of Wildland Fire, 13, 455–466, https://doi.org/10.1071/WF03064, 2004.

Daly, C., Gibson, W., Doggett, M., Smith, J., and Taylor, G.: Up-to-date monthly climate maps for the conterminous United States, Proc., 14th AMS Conf. on Applied Climatology, 84th AMS Annual Meeting Combined Preprints, 2004.

Dennison, P. E., Brewer, S. C., Arnold, J. D., and Moritz, M. A.: Large wildfire trends in the western United States, 1984–2011, Geophysical Research Letters, 41, 2928–2933, https://doi.org/10.1002/2014GL059576, 2014.

Didan, K.: MOD13Q1 MODIS/Terra vegetation indices 16-day L3 global 250m SIN grid V006, NASA EOSDIS Land Processes DAAC, 10, 415, 2015.

Dillon, G. K., Holden, Z. A., Morgan, P., Crimmins, M. A., Heyerdahl, E. K., and Luce, C. H.: Both topography and climate affected forest and woodland burn severity in two regions of the western US, 1984 to 2006, Ecosphere, 2, https://doi.org/https://doi.org/10.1890/ES11-00271.1, 2011.

Ebert-Uphoff, I., Lagerquist, R., Hilburn, K., Lee, Y., Haynes, K., Stock, J., Kumler, C., and Stewart, J. Q.: CIRA Guide to Custom Loss Functions for Neural Networks in Environmental Sciences – Version 1, https://arxiv.org/abs/2106.09757, 2021.

Eidenshink, J. C., Schwind, B., Brewer, K., Zhu, Z.-L., Quayle, B., and Howard, S. M.: A project for monitoring trends in burn severity, Fire Ecology, 3, 3–21, https://doi.org/10.4996/fireecology.0301003, 2007.

Fosberg, M. A.: Weather in wildland fire management: The fire-weather index, Paper presented at the Conference on Sierra Nevada Meteorology, *Am. Meteorol. Soc.*, South Lake Tahoe, California, 1978.

Giglio, L., Randerson, J. T., and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), Journal of Geophysical Research: Biogeosciences, 118, 317–328, https://doi.org/https://doi.org/10.1002/jgrg.20042, 2013.

Gutierrez, A. A., Hantson, S., Langenbrunner, B., Chen, B., Jin, Y., Goulden, M. L., and Randerson, J. T.: Wildfire response to changing daily temperature extremes in California's Sierra Nevada, Science Advances, 7, https://doi.org/10.1126/sciadv.abe6417, 2021.

Hansen, W. D., Braziunas, K. H., Rammer, W., Seidl, R., and Turner, M. G.: It takes a few to tango: changing climate and fire regimes can cause regeneration failure of two subalpine conifers, Ecology, 99, 966–977, https://doi.org/https://doi.org/10.1002/ecy.2181, 2018.

Hansen, W. D., Krawchuk, M. A., Trugman, A. T., and Williams, A. P.: The Dynamic Temperate and Boreal Fire and Forest-Ecosystem Simulator (DYNAFFOREST): Development and evaluation, Environmental Modelling & Software, 156, 105 473, https://doi.org/https://doi.org/10.1016/j.envsoft.2022.105473, 2022.

Harris, L. and Taylor, A. H.: Previous burns and topography limit and reinforce fire severity in a large wildfire, Ecosphere, 8, https://doi.org/https://doi.org/10.1002/ecs2.2019, 2017.

Higuera, P. E., Brubaker, L. B., Anderson, P. M., Hu, F. S., and Brown, T. A.: Vegetation mediated the impacts of postglacial climate change on fire regimes in the south-central Brooks Range, Alaska, Ecological Monographs, 79, 201–219, https://doi.org/https://doi.org/10.1890/07-2019.1, 2009.

Holsinger, L., Parks, S. A., and Miller, C.: Weather, fuels, and topography impede wildland fire spread in western US landscapes, Forest Ecology and Management, 380, 59–69, https://doi.org/https://doi.org/10.1016/j.foreco.2016.08.035, 2016.

Hooker, G., Mentch, L., and Zhou, S.: Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance, Statistics and Computing, 31, 82, https://doi.org/10.1007/s11222-021-10057-z, 2021.

Hurteau, M. D., Liang, S., Westerling, A. L., and Wiedinmyer, C.: Vegetation-fire feedback reduces projected area burned under climate change, Scientific Reports, 9, 2838, https://doi.org/10.1038/s41598-019-39284-1, 2019.

Iglesias, V., Balch, J. K., and Travis, W. R.: U.S. fires became larger, more frequent, and more widespread in the 2000s, Science Advances, 8, https://doi.org/10.1126/sciadv.abc0020, 2022.

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

Jia, S., Kim, S. H., Nghiem, S. V., Doherty, P., and Kafatos, M. C.: Patterns of population displacement during mega-fires in California detected using Facebook Disaster Maps, Environmental Research Letters, 15, https://doi.org/10.1088/1748-9326/ab8847, 2020.

Jong-Levinger, A., Banerjee, T., Houston, D., and Sanders, B. F.: Compound Post-Fire Flood Hazards Considering Infrastructure Sedimentation, Earth's Future, 10, https://doi.org/https://doi.org/10.1029/2022EF002670, 2022.

Joseph, M. B., Rossi, M. W., Mietkiewicz, N. P., Mahood, A. L., Cattau, M. E., Denis, L. A. S., Nagy, R. C., Iglesias, V., Abatzoglou, J. T., and Balch, J. K.: Spatiotemporal prediction of wildfire size extremes with Bayesian finite sample maxima, Ecological Applications, 29, https://doi.org/10.1002/eap.1898, 2019.

Joshi, J. and Sukumar, R.: Improving prediction and assessment of global fires using multilayer neural networks, Scientific Reports, 11, 3295, https://doi.org/10.1038/s41598-021-81233-4, 2021.

Juang, C. S., Williams, A. P., Abatzoglou, J. T., Balch, J. K., Hurteau, M. D., and Moritz, M. A.: Rapid Growth of Large Forest Fires Drives the Exponential Response of Annual Forest-Fire Area to Aridity in the Western United States, Geophysical Research Letters, 49, https://doi.org/10.1029/2021GL097131, 2022.

Kalashnikov, D. A., Abatzoglou, J. T., Nauslar, N. J., Swain, D. L., Touma, D., and Singh, D.: Meteorological and geographical factors associated with dry lightning in central and northern California, Environmental Research: Climate, 1, https://doi.org/10.1088/2752-5295/ac84a0, 2022.

Keeley, J. E. and Syphard, A. D.: Historical patterns of wildfire ignition sources in California ecosystems, International Journal of Wildland Fire, 27, 781–799, https://doi.org/10.1071/WF18026, 2018.

Keeley, J. E., Guzman-Morales, J., Gershunov, A., Syphard, A. D., Cayan, D., Pierce, D. W., Flannigan, M., and Brown, T. J.: Ignitions explain more than temperature or precipitation in driving Santa Ana wind fires, Science Advances, 7, https://doi.org/10.1126/sciadv.abh2262, 2021.

Kitzberger, T., Falk, D. A., Westerling, A. L., and Swetnam, T. W.: Direct and indirect climate controls predict heterogeneous early-mid 21st century wildfire burned area across western and boreal North America, PLOS ONE, 12, https://doi.org/10.1371/journal.pone.0188486, 2017.

Klein Goldewijk, K. and Ramankutty, N.: Land cover change over the last three centuries due to human activities: The availability of new global data sets, GeoJournal, 61, 335–344, https://doi.org/10.1007/s10708-004-5050-z, 2004.

Knapp, P. A.: Spatio-Temporal Patterns of Large Grassland Fires in the Intermountain West, U.S.A., Global Ecology and Biogeography Letters, 7, 259, https://doi.org/10.2307/2997600, 1998.

Knorr, W., Kaminski, T., Arneth, A., and Weber, U.: Impact of human population density on fire frequency at the global scale, Biogeosciences, 11, 1085–1102, https://doi.org/10.5194/bg-11-1085-2014, 2014.

Kondylatos, S., Prapas, I., Ronco, M., Papoutsis, I., Camps-Valls, G., Piles, M., Fernandez-Torres, M.-A., and Carvalhais, N.: Wildfire Danger Prediction and Understanding With Deep Learning, Geophysical Research Letters, 49, https://doi.org/https://doi.org/10.1029/2022GL099368, 2022.

Krawchuk, M. A., Moritz, M. A., Parisien, M.-A., Van Dorn, J., and Hayhoe, K.: Global Pyrogeography: the Current and Future Distribution of Wildfire, PLoS ONE, 4, https://doi.org/10.1371/journal.pone.0005102, 2009.

Kuhn-Régnier, A., Voulgarakis, A., Nowack, P., Forkel, M., Prentice, I. C., and Harrison, S. P.: The importance of antecedent vegetation and drought conditions as global drivers of burnt area, Biogeosciences, 18, 3861–3879, https://doi.org/10.5194/bg-18-3861-2021, 2021.

Levin, R., Cherepanova, V., Schwarzschild, A., Bansal, A., Bruss, C. B., Goldstein, T., Wilson, A. G., and Goldblum, M.: Transfer Learning with Deep Tabular Models, arXiv preprint arXiv:2206.15306, 2022.

Li, F., Zeng, X. D., and Levis, S.: A process-based fire parameterization of intermediate complexity in a Dynamic Global Vegetation Model, Biogeosciences, 9, 2761–2780, https://doi.org/10.5194/bg-9-2761-2012, 2012.

Li, S. and Banerjee, T.: Spatial and temporal pattern of wildfires in California from 2000 to 2019, Scientific Reports, 11, https://doi.org/10.1038/s41598-021-88131-9, 2021.

Littell, J. S., McKenzie, D., Peterson, D. L., and Westerling, A. L.: Climate and wildfire area burned in western U.S. ecoprovinces, 1916–2003, Ecological Applications, 19, 1003–1021, https://doi.org/https://doi.org/10.1890/07-1183.1, 2009.

810 Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems 30, edited by Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., pp. 4765–4774, http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf, 2017.

Marlon, J. R., Bartlein, P. J., Carcaillet, C., Gavin, D. G., Harrison, S. P., Higuera, P. E., Joos, F., Power, M. J., and Prentice, I. C.: Climate and human influences on global biomass burning over the past two millennia, Nature Geoscience, 1, 697–702, https://doi.org/10.1038/ngeo313, 815 2008.

Marlon, J. R., Bartlein, P. J., Gavin, D. G., Long, C. J., Anderson, R. S., Briles, C. E., Brown, K. J., Colombaroli, D., Hallett, D. J., Power, M. J., Scharf, E. A., and Walsh, M. K.: Long-term perspective on wildfires in the western USA, Proceedings of the National Academy of Sciences, 109, https://doi.org/10.1073/pnas.1112839109, 2012.

Monteith, J. L.: Evaporation and environment, in: Symposia of the society for experimental biology, vol. 19, pp. 205–234, Cambridge 820 University Press (CUP), 1965.

Moritz, M. A., Moody, T. J., Krawchuk, M. A., Hughes, M., and Hall, A.: Spatial variation in extreme winds predicts large wildfire locations in chaparral ecosystems, Geophysical Research Letters, 37, https://doi.org/https://doi.org/10.1029/2009GL041735, 2010.

Nadarajah, S., Zhang, Y., and Pogány, T. K.: On sums of independent Generalized Pareto random variables with applications to Insurance and CAT bonds, Probability in the Engineering and Informational Sciences, 32, 296–305, https://doi.org/10.1017/S0269964817000055, 825 2018.

Orville, R. E. and Huffines, G. R.: Cloud-to-Ground Lightning in the United States: NLDN Results in the First Decade, 1989–98, Monthly Weather Review, 129, 1179 – 1193, https://doi.org/10.1175/1520-0493(2001)129<1179:CTGLIT>2.0.CO;2, 2001.

O'Dell, K., Ford, B., Fischer, E. V., and Pierce, J. R.: Contribution of Wildland-Fire Smoke to US PM2.5 and Its Influence on Recent Trends, Environmental Science & Technology, 53, 1797–1804, https://doi.org/10.1021/acs.est.8b05430, 2019.

830 Parisien, M.-A. and Moritz, M. A.: Environmental controls on the distribution of wildfire at multiple spatial scales, Ecological Monographs, 79, 127–154, https://doi.org/10.1890/07-1289.1, 2009.

Parisien, M.-A., Snetsinger, S., Greenberg, J. A., Nelson, C. R., Schoennagel, T., Dobrowski, S. Z., and Moritz, M. A.: Spatial variability in wildfire probability across the western United States, International Journal of Wildland Fire, 21, 313, https://doi.org/10.1071/WF11044, 2012.

835 Parks, S. A., Miller, C., Parisien, M.-A., Holsinger, L. M., Dobrowski, S. Z., and Abatzoglou, J.: Wildland fire deficit and surplus in the western United States, 1984–2012, Ecosphere, 6, 1–13, https://doi.org/https://doi.org/10.1890/ES15-00294.1, 2015.

Parks, S. A., Parisien, M.-A., Miller, C., Holsinger, L. M., and Baggett, L. S.: Fine-scale spatial climate variation and drought mediate the likelihood of reburning, Ecological Applications, 28, 573–586, https://doi.org/https://doi.org/10.1002/eap.1671, 2018.

Perez-Cruz, F.: Kullback-Leibler divergence estimation of continuous distributions, in: 2008 IEEE International Symposium on Information 840 Theory, pp. 1666–1670, https://doi.org/10.1109/ISIT.2008.4595271, 2008.

Potter, B. E. and McEvoy, D.: Weather Factors Associated with Extremely Large Fires and Fire Growth Days, Earth Interactions, 25, 160 – 176, https://doi.org/10.1175/EI-D-21-0008.1, 2021.

Pureswaran, D. S., Roques, A., and Battisti, A.: Forest Insects and Climate Change, Current Forestry Reports, 4, 35–50, https://doi.org/10.1007/s40725-018-0075-6, 2018.

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

850 Radeloff, V. C., Hammer, R. B., Stewart, S. I., Fried, J. S., Holcomb, S. S., and McKeefry, J. F.: The Wildland-Urban Interface in the United States, Ecological Applications, 15, 799–805, https://doi.org/https://doi.org/10.1890/04-1413, 2005.

Rahimi, S., Krantz, W., Lin, Y., Bass, B., Goldenson, N., Hall, A., Jebo, Z., and Norris, J.: Evaluation of a Reanalysis-Driven Configuration of WRF4 Over the Western United States From 1980-2020., Journal of Geophysical Research: Atmospheres, 127, https://doi.org/10.1029/2021JD035699, 2022.

855 Rao, K., Williams, A. P., Diffenbaugh, N. S., Yebra, M., and Konings, A. G.: Plant-water sensitivity regulates wildfire vulnerability, Nature Ecology & Evolution, 6, 332–339, https://doi.org/10.1038/s41559-021-01654-2, 2022.

Rasp, S., Pritchard, M. S., and Gentine, P.: Deep learning to represent subgrid processes in climate models, Proceedings of the National Academy of Sciences, 115, 9684–9689, https://doi.org/10.1073/pnas.1810286115, 2018.

Richards, J., Huser, R., Bevacqua, E., and Zscheischler, J.: Insights into the drivers and spatio-temporal trends of extreme Mediterranean 860 wildfires with statistical deep-learning, arXiv preprint arXiv:2212.01796, 2022.

Rigden, A. J., Powell, R. S., Trevino, A., McColl, K. A., and Huybers, P.: Microwave Retrievals of Soil Moisture Improve Grassland Wildfire Predictions, Geophysical Research Letters, 47, e2020GL091 410, https://doi.org/10.1029/2020GL091410, 2020.

Riley, K. and Thompson, M.: An Uncertainty Analysis of Wildfire Modeling, chap. 13, pp. 191–213, American Geophysical Union (AGU), https://doi.org/https://doi.org/10.1002/9781119028116.ch13, 2016.

865 Rollins, M. G.: LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment, International Journal of Wildland Fire, 18, 235–249, https://doi.org/10.1071/WF08088, 2009.

Rollins, M. G., Morgan, P., and Swetnam, T.: Landscape-scale controls over 20th century fire occurrence in two large Rocky Mountain (USA) wilderness areas, Landscape Ecology, 17, 539–557, https://doi.org/10.1023/A:1021584519109, 2002.

Romps, D. M., Seeley, J. T., Vollaro, D., and Molinari, J.: Projected increase in lightning strikes in the United States due to global warming, 870 Science, 346, 851–854, https://doi.org/10.1126/science.1259100, 2014.

Schoenberg, F. P., Peng, R., and Woods, J.: On the distribution of wildfire sizes, Environmetrics, 14, 583–592, https://doi.org/10.1002/env.605, 2003.

Scollnik, D. P. M.: On composite lognormal-Pareto models, Scandinavian Actuarial Journal, pp. 20–33, https://doi.org/10.1080/03461230601110447, 2007.

875 Seager, R., Hooks, A., Williams, A. P., Cook, B., Nakamura, J., and Henderson, N.: Climatology, Variability, and Trends in the U.S. Vapor Pressure Deficit, an Important Fire-Related Meteorological Quantity, Journal of Applied Meteorology and Climatology, 54, 1121 – 1141, https://doi.org/10.1175/JAMC-D-14-0321.1, 2015.

Spawn, S. A., Sullivan, C. C., Lark, T. J., and Gibbs, H. K.: Harmonized global maps of above and belowground biomass carbon density in the year 2010, Scientific Data, 7, 112, https://doi.org/10.1038/s41597-020-0444-4, 2020.

880 Sullivan, A. L.: Wildland surface fire spread modelling, 1990–2007. 3: Simulation and mathematical analogue models, International Journal of Wildland Fire, 18, 387–403, 2009.

Swetnam, T. W. and Betancourt, J. L.: Mesoscale Disturbance and Ecological Response to Decadal Climatic Variability in the American Southwest, Journal of Climate, 11, 3128 – 3147, https://doi.org/10.1175/1520-0442(1998)011<3128:MDAERT>2.0.CO;2, 1998.

Tschumi, E., Lienert, S., van der Wiel, K., Joos, F., and Zscheischler, J.: The effects of varying drought-heat signatures on terrestrial carbon dynamics and vegetation composition, Biogeosciences, 19, 1979–1993, https://doi.org/10.5194/bg-19-1979-2022, publisher: Copernicus GmbH, 2022.

Vose, R., Applequist, S., Squires, M., Durre, I., Menne, M., Williams, C., Fenimore, C., Gleason, K., and Arndt, D.: Improved Historical Temperature and Precipitation Time Series for U.S. Climate Divisions, Journal of Applied Meteorology and Climatology, 53, 1232–1251, https://doi.org/10.1175/JAMC-D-13-0248.1, 2014.

Wacker, R. S. and Orville, R. E.: Changes in measured lightning flash count and return stroke peak current after the 1994 U.S. National Lightning Detection Network upgrade: 1. Observations, Journal of Geophysical Research: Atmospheres, 104, 2151–2157, https://doi.org/https://doi.org/10.1029/1998JD200060, 1999.

Wang, S. S.-C. and Wang, Y.: Quantifying the effects of environmental factors on wildfire burned area in the south central US using integrated machine learning techniques, Atmospheric Chemistry and Physics, 20, 11 065–11 087, https://doi.org/10.5194/acp-20-11065-2020, 2020.

Wang, S. S.-C., Qian, Y., Leung, L. R., and Zhang, Y.: Identifying Key Drivers of Wildfires in the Contiguous US Using Machine Learning and Game Theory Interpretation, Earth's Future, 9, https://doi.org/10.1029/2020EF001910, 2021.

Westerling, A. L.: Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring, Philosophical Transactions of the Royal Society B: Biological Sciences, 371, https://doi.org/10.1098/rstb.2015.0178, 2016.

Westerling, A. L., Hidalgo, H. G., Cayan, D. R., and Swetnam, T. W.: Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity, Science, 313, 940–943, https://doi.org/10.1126/science.1128834, 2006.

Westerling, A. L., Turner, M. G., Smithwick, E. A. H., Romme, W. H., and Ryan, M. G.: Continued warming could transform Greater Yellowstone fire regimes by mid-21st century, Proceedings of the National Academy of Sciences, 108, 13 165–13 170, https://doi.org/10.1073/pnas.1110199108, 2011.

Williams, A. P. and Abatzoglou, J. T.: Recent Advances and Remaining Uncertainties in Resolving Past and Future Climate Effects on Global Fire Activity, Current Climate Change Reports, 2, 1–14, https://doi.org/10.1007/s40641-016-0031-0, 2016.

Williams, A. P., Allen, C. D., Macalady, A. K., Griffin, D., Woodhouse, C. A., Meko, D. M., Swetnam, T. W., Rauscher, S. A., Seager, R., Grissino-Mayer, H. D., Dean, J. S., Cook, E. R., Gangodagamage, C., Cai, M., and McDowell, N. G.: Temperature as a potent driver of regional forest drought stress and tree mortality, Nature Climate Change, 3, 292–297, https://doi.org/10.1038/nclimate1693, 2013.

Williams, A. P., Abatzoglou, J. T., Gershunov, A., Guzman-Morales, J., Bishop, D. A., Balch, J. K., and Lettenmaier, D. P.: Observed Impacts of Anthropogenic Climate Change on Wildfire in California, Earth's Future, 7, 892–910, https://doi.org/10.1029/2019EF001210, 2019.

Williams, A. P., Livneh, B., McKinnon, K. A., Hansen, W. D., Mankin, J. S., Cook, B. I., Smerdon, J. E., Varuolo-Clarke, A. M., Bjarke, N. R., Juang, C. S., and Lettenmaier, D. P.: Growing impact of wildfire on western US water supply, Proceedings of the National Academy of Sciences, 119, https://doi.org/10.1073/pnas.2114069119, 2022.

Wu, X., Liu, H., Hartmann, H., Ciais, P., Kimball, J. S., Schwalm, C. R., Camarero, J. J., Chen, A., Gentine, P., Yang, Y., Zhang, S., Li, X., Xu, C., Zhang, W., Li, Z., and Chen, D.: Timing and Order of Extreme Drought and Wetness Determine Bioclimatic Sensitivity of Tree Growth, Earth's Future, 10, e2021EF002 530, https://doi.org/https://doi.org/10.1029/2021EF002530, 2022.

Xie, Y., Lin, M., Decharme, B., Delire, C., Horowitz, L. W., Lawrence, D. M., Li, F., and Séférian, R.: Tripling of western US particulate pollution from wildfires in a warming climate, Proceedings of the National Academy of Sciences, 119, https://doi.org/10.1073/pnas.2111372119, 2022.

920   Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G. C., Rigge, M., and Xian, G.: A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies, ISPRS Journal of Photogrammetry and Remote Sensing, 146, 108–123, https://doi.org/https://doi.org/10.1016/j.isprsjprs.2018.09.006, 2018.

Yuval, J. and O'Gorman, P. A.: Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions,
925   Nature Communications, 11, 3295, https://doi.org/10.1038/s41467-020-17142-3, 2020.

Zeng, X., Broxton, P., and Dawson, N.: Snowpack Change From 1982 to 2016 Over Conterminous United States, Geophysical Research Letters, 45, 12,940–12,947, https://doi.org/https://doi.org/10.1029/2018GL079621, 2018.

Zeng, X., Broxton, P., and Dawson, N.: Daily 4 km Gridded SWE and Snow Depth from Assimilated In-Situ and Modeled Data over the Conterminous US, Version 1, NASA National Snow and Ice Data Center Distributed Active Archive Center,
930   https://doi.org/10.5067/0GGPB220EX6A, 2019.

Zheng, B., Ciais, P., Chevallier, F., Chuvieco, E., Chen, Y., and Yang, H.: Increasing forest fire emissions despite the decline in global burned area, Science Advances, 7, https://doi.org/10.1126/sciadv.abh2646, 2021.

Zhou, S., Williams, A. P., Berg, A. M., Cook, B. I., Zhang, Y., Hagemann, S., Lorenz, R., Seneviratne, S. I., and Gentine, P.: Land-atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity, Proceedings of the National Academy of Sciences, 116, 18 848–
935   18 853, https://doi.org/10.1073/pnas.1904955116, 2019.

Zhuang, Y., Fu, R., Santer, B. D., Dickinson, R. E., and Hall, A.: Quantifying contributions of natural variability and anthropogenic forcings on increased fire weather risk over the western United States, Proceedings of the National Academy of Sciences, 118, https://doi.org/10.1073/pnas.2111875118, 2021.

Zou, Y., Wang, Y., Qian, Y., Tian, H., Yang, J., and Alvarado, E.: Using CESM-RESFire to understand climate–fire–ecosystem interactions
940   and the implications for decadal climate variability, Atmospheric Chemistry and Physics, 20, 995–1020, https://doi.org/10.5194/acp-20-995-2020, 2020.