**Title**

# Projecting Large Fires in the Western US With an Interpretable and Accurate Hybrid Machine Learning Method



**Journal**
Earth's Future, 12(10)

**ISSN**
2328-4277

**Authors**
Li, Fa
Zhu, Qing
Yuan, Kunxiaojia
et al.

**Publication Date**
2024-10-01

**DOI**
10.1029/2024ef004588

ADVANCING EARTH AND SPACE SCIENCES

# Earth's Future

RESEARCH ARTICLE
10.1029/2024EF004588

Special Collection:
Advancing Interpretable AI/ML Methods for Deeper Insights and Mechanistic Understanding in Earth Sciences: Beyond Predictive Capabilities

# Projecting Large Fires in the Western US With an Interpretable and Accurate Hybrid Machine Learning Method

**Fa Li[1], Qing Zhu[2], Kunxiaojia Yuan[2], Fujiang Ji[1], Arindam Paul[3], Peng Lee[3], Volker C. Radeloff[1], and Min Chen[1]**

[1]Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, Madison, WI, USA, [2]Climate and Ecosystem Sciences Division, Climate Sciences Department, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [3]American Family Insurance, Madison, WI, USA

**Key Points:**

- Substantial structural differences were identified across machine learning (ML) fire models, but existing studies mainly focus on accuracy
- We proposed a fire model with higher accuracy and physical interpretability compared to commonly used ML models
- Our fire model reasonably reflected actual fire risk and revealed complex climate controls on large fires and megafires in the western US

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

F. Li and M. Chen,
fli235@wisc.edu;
mchen392@wisc.edu

**Citation:**

Li, F., Zhu, Q., Yuan, K., Ji, F., Paul, A., Lee, P., et al. (2024). Projecting large fires in the western US with an interpretable and accurate hybrid machine learning method. *Earth's Future*, *12*, e2024EF004588. https://doi.org/10.1029/2024EF004588

Received 28 FEB 2024
Accepted 24 JUL 2024



## Abstract

More frequent and widespread large fires are occurring in the western United States (US), yet reliable methods for predicting these fires, particularly with extended lead times and a high spatial resolution, remain challenging. In this study, we proposed an interpretable and accurate hybrid machine learning (ML) model, that explicitly represented the controls of fuel flammability, fuel availability, and human suppression effects on fires. The model demonstrated notable accuracy with a $F_1$-score of $0.846 \pm 0.012$, surpassing process-driven fire danger indices and four commonly used ML models by up to 40% and 9%, respectively. More importantly, the ML model showed remarkably higher interpretability relative to other ML models. Specifically, by demystifying the "black box" of each ML model using the explainable AI techniques, we identified substantial structural differences across ML fire models, even among those with similar accuracy. The relationships between fires and their drivers, identified by our model, were aligned closer with established fire physical principles. The ML structural discrepancy led to diverse fire predictions and our model predictions exhibited greater consistency with actual fire occurrence. With the highly interpretable and accurate model, we revealed the strong compound effects from multiple climate variables related to evaporative demand, energy release component, temperature, and wind speed, on the dynamics of large fires and megafires in the western US. Our findings highlight the importance of assessing the structural integrity of models in addition to their accuracy. They also underscore the critical need to address the rise in compound climate extremes linked to large wildfires.

## Plain Language Summary

Anthropogenic climate change has caused more frequent and widespread large fires in the western United States (US). Interpretable and accurate methods that can predict large fires in advance and at a high spatial resolution are essential for more targeted fire risk mitigation strategies, but the trade-off between accuracy and interpretability persists in current fire models. For example, machine learning (ML) fire models can be more accurate than process models but operate as "black-boxes" with a poor understanding of fire physics. Process fire models are more physically interpretable but remain less accurate than ML models. This research offers a novel perspective on the tradeoff and quantification of model physical interpretability and accuracy for wildfire predictions in the western US. More importantly, this study proposes a more interpretable and accurate ML fire model that explicitly incorporates fire knowledge while leveraging data-informed modeling advantages. The model predicted fire risk showed high consistency with actually occurred fire events, and revealed strong and multivariate climate controls on large fires and megafires in the western US. Our proposed fire model enables better predictions and understanding of large fires and megafires in the western US.

## 1. Introduction

Large fires in the western United States (US) have become more frequent and widespread, exerting increasingly devastating impacts on ecosystems and human society. The median annual burned area in western US has quadrupled during 2005–2018 relative to 1984–1999 (Iglesias et al., 2022). Notably, the extreme fires (e.g., top one percent in terms of fire size) have remarkedly expanded to a broader geographical extent (Iglesias et al., 2022). These fires have led to pervasively increases in tree mortality rates (Hicke et al., 2016) and have substantially changed ecosystem structures, compositions, as well as their biophysical and biogeochemical properties (Bowman et al., 2009; F. Li et al., 2012; Zhu et al., 2022).

Human activities are a primary cause of these large fires over the western US (Nagy et al., 2018), with more frequent ignitions occurring in the wildland-urban interface (WUI) (Radeloff et al., 2018, 2023). These human-started fires pose significant threats to millions of residential homes and infrastructure, leading to billions of economic lost (Mietkiewicz et al., 2020). Additionally, air pollutants released by these fires, such as particulate matters, are closely linked to serious health issues such as infant mortality (Pullabhotla et al., 2023). The emissions of black carbon along with greenhouse gases like carbon dioxide and methane can exacerbate global warming, creating conditions more favorable for future fires. Over the past few decades, anthropogenic climate change has significantly increased the risk of fires in the western US (Abatzoglou & Williams, 2016; Turco et al., 2023; Zhuang et al., 2021). This trend is anticipated to continue, especially under future high-emission scenarios (Touma et al., 2021). Therefore, accurate predicting of the occurrence of large fires is of utmost importance to manage and mitigate the rising fire risk in the western US.

Despite the notable advancements in fire prediction models, there remains a gap between their accuracy and interpretability. Traditional models, such as empirical process models (Bradshaw, 1984; Jolly et al., 2019) and simplified statistical models (e.g., linear or generalized linear), are interpretable due to their reliance on well established relationships between fire occurrences in the western US and fire weather conditions, such as vapor pressure deficit (VPD), dead fuel moisture, and fire weather indices (FWIs) (Abatzoglou & Williams, 2016; Abatzoglou, Battisti, et al., 2021; Balch et al., 2022; Williams et al., 2019). While effective at capturing inter-annual fire dynamics at a regional scale (Abatzoglou & Williams, 2016; Higuera & Abatzoglou, 2021), these models often fall short in representing sub-regional heterogeneity and intra-annual variations of fire risk at a relatively higher spatial resolution (Kondylatos et al., 2022; J. Li et al., 2020; Wang et al., 2021). The limitations in accuracy may stem from imperfect parameterization of the emergent climate-fire relationships (F. Li et al., 2023; Littell et al., 2016), and inadequate representation of complex interactions between fires and their drivers, such as climate, fuel availability (Parks et al., 2014), topography (Alizadeh et al., 2023), and human activities (Fusco et al., 2016). To address these challenges, recent studies have employed advanced machine learning (ML) models along with critical socio-environmental factors to predict fires in the US (Gray et al., 2018; J. Li et al., 2020; Wang & Wang, 2020; Wang et al., 2021). Despite the improvement in accuracy, most ML models operate as "black boxes" and lack transparency in their decision-making processes (Arrieta et al., 2020; Kondylatos et al., 2022; F. Li et al., 2023). For example, the opacity inherent in neural network or deep-learning models diminishes their interpretability (F. Li et al., 2023; Zhu et al., 2022). Even though some ML models, like decision trees (DTs) and random forest (RF), provide insights into variable importance, this information often remains static across the whole data set, which hinders a comprehensive understanding of the complex, spatio-temporally heterogeneous nature of fire controls (Kondylatos et al., 2022; F. Li et al., 2023; Wang et al., 2021).

To enhance the transparency and interpretability of ML models, *eXplainable AI* (XAI) has emerged as an important tool in Earth science. XAI, as a form of post-hoc interpretability method, offers a deeper understanding of model predictions. It provides detailed insights, such as the magnitude and the sign (positive or negative) of each driver's influence on specific predictions (Arrieta et al., 2020). This capability is especially valuable in discerning the diverse drivers of fire occurrences across space and over time (Kondylatos et al., 2022; Wang et al., 2021). Recent studies have successfully integrated XAI and ML algorithms to predict and analyze fire dynamics in the frequently burned area of continuous US (Wang et al., 2021) and a part of the Eastern Mediterranean (Kondylatos et al., 2022). However, these studies either focus on a coarse spatial resolution (e.g., 0.25°) or limit themselves to short-term predictions (e.g., next-day fire risk prediction). Longer-term fire prediction at a relatively higher spatial resolution allows for more time to implement more spatially-resolved fire risk mitigation and adaptation, but persists as a challenge due to the stochastic characteristic of fires, and the spatiotemporally heterogenous dependencies of fires on underlying drivers.

Additionally, an important yet often overlooked aspect in ML applications to wildfire science is the potential discrepancy between the relationships inferred from ML models and the actual physical or causal relationships between fires and their drivers. For example, even ML models with similar accuracy levels can show substantial differences in the modeled variable importance and the underlying causality importance (Turbé et al., 2023; Yuan, Zhu, Li, et al., 2022). Such discrepancies can lead to divergent model responses to input factors when extrapolating well-trained ML models for spatiotemporal predictions (Yuan, Zhu, Li, et al., 2022; Yuan et al., 2024). The physical or causal interpretability of different ML models remains underexplored in wildfire science, which fuels concerns from practitioners about the model trustworthiness (Jain et al., 2020; Murdoch et al., 2019). While ML models generally outperform process-models regarding accuracy, inconsistency between ML modeled and physical relationships could produce unreliable predictions. This unreliability, in turn, can lower user confidence and hinder the broader adoption of ML models in this field. Therefore, developing ML models that are both highly accurate and interpretable, and that adhere closely to physical processes, is crucial for effectively predicting and understanding wildfires.

In this study, we aim to predict and understand large fires in the western US 1 month in advance at a high spatial resolution (1 km) and in an interpretable manner. Specifically, we developed and employed a hybrid and interpretable ML large fire prediction model to specifically predict the spatiotemporal dynamics of fire risk across grid cells over western US. We assessed the accuracy of the model developed by comparing it with three fire indices, as well as four ML models widely used in fire prediction. Beyond accuracy, we also examined the physical interpretability for each ML model by integrating XAI with well-established principles of fire physics, allowing us to highlight the significant impacts of structural differences in models on their projected fire dynamics, even among those with similar level of accuracy. Finally, we coupled our model and XAI to understand the drivers of large fires, including the megafires in the extreme fire year 2020 (Higuera & Abatzoglou, 2021), in the western US.

## 2. Methods

Following prior studies (Gray et al., 2018; Jain et al., 2020; J. Li et al., 2020; Preisler et al., 2004), we modeled the conditional probability of a fire escalating into a large one over a certain size, given an ignition event. We used a typical threshold of 405 ha (ha) to define "large fires" in the western US, as utilized in previous research (Dennison et al., 2014; Gray et al., 2018; J. Li et al., 2020). This modeling of large fire probability addresses the inherent stochasticity and complexity of ignitions, thus representing the potential risk of large fires without explicitly including ignitions, conceptually similar to that depicted by FWIs (Jolly et al., 2015; Sharples, 2022).

For modeling large fire probability, we employed the hybrid ML model developed by us along with four other widely-used ML models: DT (F. Li et al., 2018; Song & Ying, 2015), RF (Gray et al., 2018; F. Li et al., 2018), Extreme Gradient Boosting (XGBoost) (Michael et al., 2021), and artificial neural network (ANN) (F. Li et al., 2023; Zhu et al., 2022). We also used three fire indices for model comparison. Two of the fire indices were from the United States National Fire-Danger Rating System (USNFDRS) (Bradshaw, 1984), including energy release component (ERC) and burn index (BI). ERC and BI have permeated US firefighting organizations and been used for supporting most of the local wildfire preparedness and management since 1978 (Bradshaw, 1984; Jolly et al., 2019; Schlobohm & Brain, 2002). Another fire index is the Canadian fire weather index (FWI), a widely used fire indices for understanding and estimating fire risk regionally and globally (Jolly et al., 2015; Van Wagner, 1974). We evaluated these models, focusing on both accuracy and physical interpretability in predicting large fires in the western US.

### 2.1. Hybrid Machine Learning Fire Model

We developed the hybrid ML fire model by explicitly considering the controls of fuel availability and human suppression on large fires while representing the complex relationships between fires and their drivers of fuel flammability through the attention mechanism (F. Li et al., 2023; Vaswani et al., 2017). Specifically, as shown in Equation 1 and Figure 1, the large fire probability ($f_{\mathrm{Fire}}$) in the model is determined by three groups of key factors: fuel availability ($f_{\mathrm{FA}}$), fuel flammability ($f_{\mathrm{FF}}$), and human suppression ($f_{\mathrm{HS}}$), based on knowledge-driven and process-based modeling schemes of wildfires (Kelley et al., 2019; F. Li et al., 2012, 2019; Rabin et al., 2017). Similar to prior studies (Kelley et al., 2019; F. Li et al., 2012), $f_{\mathrm{Fire}}$ is represented as the product of $f_{\mathrm{FA}}$, $f_{\mathrm{FF}}$, and $f_{\mathrm{HS}}$ (Equation 1). $f_{\mathrm{FA}}$ and $f_{\mathrm{Fire}}$ non-linearly increase with rising fuel availability (Equation 2), whereas stronger human suppression leads to a curvilinear decrease in both $f_{\mathrm{HS}}$ and $f_{\mathrm{Fire}}$ (Equation 3) (Kelley et al., 2019; F. Li et al., 2012). Fuel flammability, a predominant controlling factor in the temporal dynamics of fires in the western US (Abatzoglou & Williams, 2016; Abatzoglou, Battisti, et al., 2021; Higuera & Abatzoglou, 2021), is influenced by a range of variables including fuel aridity (Abatzoglou & Williams, 2016; Higuera & Abatzoglou, 2021), wind speed (Abatzoglou et al., 2018, 2023; Westerling et al., 2006), topography (Holsinger et al., 2016; Linn et al., 2007), and land cover types (Balch et al., 2022; Calviño-Cancela et al., 2016). The relationship between flammability and these driving-factors can be complex, and often characterized by non-linear (Abatzoglou, Battisti, et al., 2021; Balch et al., 2022), time-lagged (Abolafia-Rosenzweig et al., 2022; F. Li et al., 2023; Littell et al., 2009), and spatially varied dependencies (e.g., the dominant driver can be varied across grid cells or fire

10.1029/2024EF004588

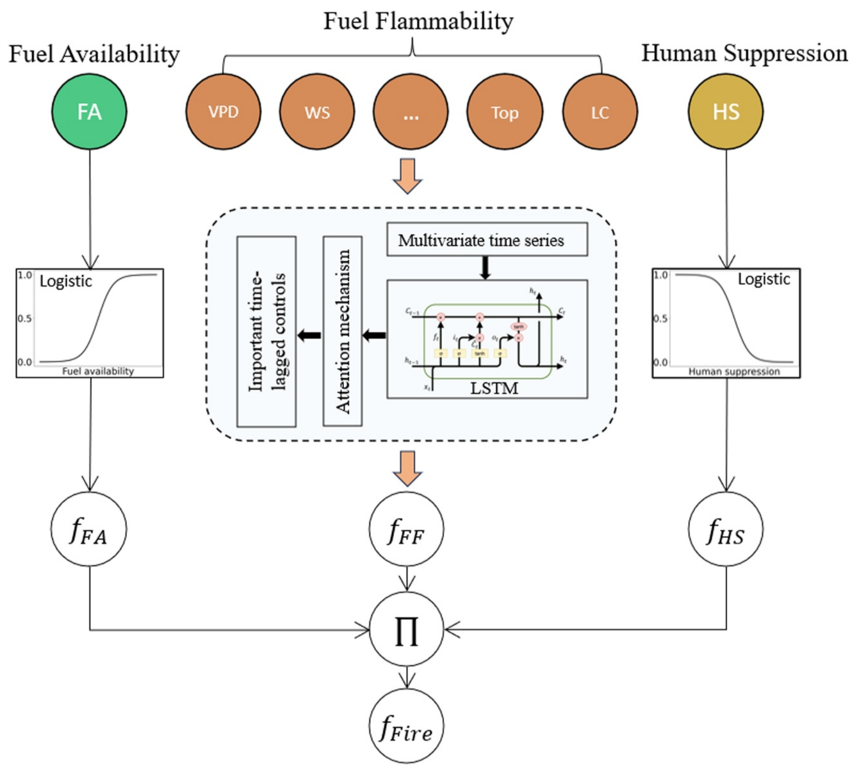


**Figure 1.** The scheme of the hybrid machine learning model proposed for large fire probability prediction. The model output ($f_{\mathrm{Fire}}$) represents the large fire probability if an ignition occurred, and depends on constraints from fuel availability ($f_{\mathrm{FA}}$), fuel flammability ($f_{\mathrm{FF}}$), and human suppression ($f_{\mathrm{HS}}$). The values of $f_{\mathrm{Fire}}$, $f_{\mathrm{FA}}$, $f_{\mathrm{FF}}$, and $f_{\mathrm{HS}}$ range from zero to one with a lower value representing lower probability for a large fire. The $f_{\mathrm{FA}}$ ($f_{\mathrm{HS}}$) is modeled through an empirical logistic function, showing positive (negative) relationship between large fire probability and fuel availability (human suppression). For the $f_{\mathrm{FF}}$, an attention-augmented long short-term memory network model is applied to represent the complex associations among fuel flammability and multi-drivers such as vapor pressure deficit, wind speed, topography (top), and land cover types; all the drivers and their abbreviations used for modeling $f_{\mathrm{FF}}$ are described in Table 1. When $f_{\mathrm{Fire}}$ is larger, the large fire probability is higher; when $f_{\mathrm{FA}}$, $f_{\mathrm{FF}}$, and $f_{\mathrm{HS}}$ are larger, the limitations of fuel availability, fuel flammability, and human suppression on large fires are weaker.

events) (F. Li et al., 2023; Littell et al., 2016). Therefore, to capture these intricate relationships, we applied an attention-augmented long short-term memory network (LSTM) model (F. Li et al., 2023) (Equation 4). It summarized the time series of input variables (Table 1) from the month preceding the fires as hidden state vectors, while the attention mechanism dynamically identified important time steps and variables crucial for fire prediction (Gui et al., 2021; Guo et al., 2019; F. Li et al., 2020; F. Li et al., 2023; Qin et al., 2017). To further alleviate the issue of correlation between different drivers that may bias model training, we adopted the widely used dropout (Srivastava et al., 2014) and weight decay regularization (Zhang et al., 2018) techniques for preventing model overfitting. Details of the input variables and data sets used were listed in Table 1, and details of the attention-based mechanisms for fire modeling were described in F. Li et al. (2023). The four variables in the model ($f_{\mathrm{Fire}}$, $f_{\mathrm{FA}}$, $f_{\mathrm{FF}}$, and $f_{\mathrm{HS}}$) are scaled between 0 and 1. The information of fire occurrence date and fire size was derived from the fire occurrence database used (Table 1). With the occurrence date information, we retrieved the drivers before a large fire used for predictions. In the event of a large fire (i.e., a positive sample), $f_{\mathrm{Fire}}$ equals one, and $f_{\mathrm{Fire}}$ equals zero for small fires (i.e., negative samples). The target is to predict the large fire probability in next month if there is an ignition using the input socio-environmental conditions in Table 1. We employed binary cross entropy (Ruby & Yendapalli, 2020), to quantify the bias between model predictions and observations.

$$f_{\mathrm{Fire}} = f_{\mathrm{FA}} \times f_{\mathrm{FF}} \times f_{\mathrm{HS}} \tag{1}$$

$$f_{\mathrm{FA}} = \frac{1}{1 + e^{-k_1(\mathrm{FA}-\mathrm{FA}_0)}} \tag{2}$$

**Table 1**
*Data Sets Used for Machine Learning Fire Prediction Model Development*

| Variable category | Data sources | Data links | References |
|---|---|---|---|
| Wildfire | Fire Program Analysis fire occurrence database | https://www.fs.usda.gov/rds/archive/Catalog/RDS-2013-0009.5 | Short (2021) |
| Burned index (BI) | Gridded surface meteorological data (gridMET) | https://www.climatologylab.org/gridmet.html | Abatzoglou (2013) |
| Energy release component (ERC) | | | |
| 100-hr dead fuel moisture (FM100) | | | |
| 1,000-hr dead fuel moisture (FM1000) | | | |
| Temperature (T) | | | |
| Relative humidity (RH) | | | |
| Reference Evapotranspiration (ET0) | | | |
| Precipitation (Pr) | | | |
| Vapor pressure deficit (VPD) | | | |
| Wind speed (WS) | | | |
| Topography (Top) | Shuttle Radar Topography Mission (SRTM) data | https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1 | Farr et al. (2007) |
| Land cover types | MODIS annual International Geosphere-Biosphere Program (IGBP) classification data | https://lpdaac.usgs.gov/products/mcd12q1v006/ | Friedl and Sulla-Menashe (2019) |
| Net primary productivity (NPP) | MODIS Net Primary Production data (MYD17A3HGF) | https://lpdaac.usgs.gov/products/myd17a3hv006/ | Running et al. (2015) |
| Population distribution (Popu) | Gridded Population of World Version 4 (GPWv4) | https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-rev11 | Doxsey-Whitfield et al. (2015) |

$$f_{\mathrm{HS}} = \frac{1}{1 + e^{-k_2(\mathrm{HS}-\mathrm{HS}_0)}} \tag{3}$$

$$f_{\mathrm{FF}} = \mathrm{Attention}(x) \tag{4}$$

where FA, and HS represent fuel availability proxied by net primary productivity (Cattau et al., 2020; Ellis et al., 2022), and human suppression proxied by population density (Kelley et al., 2019; F. Li et al., 2012; Zhu et al., 2022), respectively. $\mathrm{FA}_0$ and $\mathrm{HS}_0$ are model parameters representing a 50% limitation of fuel availability (e.g., $f_{\mathrm{FA}}(\mathrm{FA}_0) = 0.5$) and human suppression on large fires, respectively. $k_1$ and $k_2$ are the steepness of the logistic curve; $k_1$ is a positive number representing positive control of fuel availability on large fires; $k_2$ is a negative number indicating negative control of human suppression on large fires. The limitation of fuel flammability ($f_{\mathrm{FF}}$) is modeled through the attention mechanism described by F. Li et al. (2023), and $x$ in Equation 4 represents the drivers of fuel flammability.

## 2.2. Alternative Models for Large Fire Probability Prediction

Four baseline ML models, including DT, RF, XGBoost, and ANN, were compared with the model proposed by us. The DT model resembles a single-tree structure, making it easily interpretable for fire prediction (Coffield et al., 2019). Although DT can outperform other ML models in certain fire prediction cases (Coffield et al., 2019), its single-tree structure generally makes it more prone to overfitting compared to RF and XGBoost. RF deploys an ensemble learning strategy (Breiman, 2001), reducing overfitting by constructing multiple trees, each trained on a randomly selected subset of features and data. In contrast, the XGBoost mitigates overfitting by constructing multiple shallow-depth trees through a gradient boosting algorithm (Ke et al., 2017). In addition, we used the ANN model, which comprises multiple hidden layers and has demonstrated reasonable accuracy in fire prediction

10.1029/2024EF004588


(Joshi & Sukumar, 2021; F. Li et al., 2023). These four ML models predict the probability of large fires in the upcoming month by using various drivers (Table 1) from the preceding month.

Additionally, we used the ERC, BI, and FWI indices for model comparison. ERC and BI were derived from the empirical physics-based USNFDRS (Bradshaw, 1984), and have been most frequently used to guide fire management decisions or predict large fires in US (Gray et al., 2018; Jolly et al., 2019; J. Li et al., 2020). ERC represents the amount of heat released per unit area during flaming, mainly dependent on fuel moisture. BI indicates the flame length, largely determined by both the fuel moisture and the rate of fire spread (Bradshaw, 1984; Jolly et al., 2019). FWI represents fire intensity and fire danger given the meteorological conditions used for the index calculation, including temperature, precipitation, relative humidity, and wind speed (Dowdy et al., 2010; Van Wagner, 1974). The model first calculated three moisture codes, representing moisture levels of fuels in various soil depths at daily, weekly, and monthly scales; then using the calculated codes, two indices that represent fuel available for combustion and fire spread rate, were calculated, respectively; subsequently, the FWI was calculated based the two indices in the second step (Dowdy et al., 2010). The FWI data was downloaded from the Copernicus Emergency Management Service (see the section of Open Research) and re-gridded into the 1 km spatial and monthly temporal resolutions. A higher value of either ERC, BI, or FWI indicates a greater fire risk. The concurrent fire risk during a fire event was denoted as ERC-Cur, BI-Cur, and FWI-Cur, respectively. Moreover, we used ERC, BI, and FWI values from 1 month prior to a fire, labeled as ERC-Pre, BI-Pre, and FWI-Pre, respectively, to assess fire predictability using these time-lagged fire indices.

### 2.3. Model Evaluation

In evaluating fire models, we focused on two key aspects: accuracy and physical interpretability. In terms of accuracy, we used the $F_1$-score (F. Li, Zhu, et al., 2022; Raschka, 2014), which effectively balances precision and recall rates (Equation 5). A higher precision indicates that a greater portion of predicted large fires are correct (Equation 6), whereas a higher recall rate implies that the model successfully detects a larger portion of actual large fires (Equation 7). A larger $F_1$-score value signifies a model that better balances both precision and recall rates.

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

where TP and FP are the number of correctly and falsely predicted large fires, respectively. FN is the number of large fires that model fails to predict.

We assessed the model physical interpretability by combining XAI and fire physical knowledge. For XAI, we used the additive feature attribution (AFA) scheme (Lundberg & Lee, 2017) to estimate each driving factor's contribution to model predictions. AFA assumes that each model prediction can be linearly decomposed into the summed contribution of its driving factors (Equation 8). Specifically, we used the integrated gradients method (Sundararajan et al., 2017) to infer the contribution value ($\varnothing_i$) of each driver in neural-network-based ML fire prediction models. This method, particularly developed for neural-networks, uses the path integral of the gradients for the $i$th dimension as the contribution from the $i$th driver (Sundararajan et al., 2017). Surpassing many XAI methods, the integrated gradients algorithm satisfies two fundamental axioms for ML attribution: sensitivity, meaning that if model predictions differ with one input feature different, that feature should contribute non-zero to the output, and implementation invariance, which asserts that models yielding identical outputs for the same inputs are fundamentally equivalent, despite differences in their implementation. This approach shows superior performance in extracting the underlying rules for neural networks (Sundararajan et al., 2017; Turbé et al., 2023). Additionally, we used the SHapley Additive explanation (SHAP), specifically designed for tree-based ML models and known here as tree-SHAP (Lundberg & Lee, 2017; Lundberg et al., 2018), to interpretate fire prediction models including DT, RF, and XGBoost. Tree-SHAP is computationally efficient and has demonstrated better identification of influential factors than classic attribution methods for tree-based ML models (Lundberg et al., 2018).

$$f_{\mathrm{Fire}} = \sum_{i=0}^{n} \varnothing_i \tag{8}$$

where $\varnothing_i$ represents the contribution of the $i$th driving factor to the large fire probability ($f_{\mathrm{Fire}}$), and $n$ is the number of drivers considered.

The outputs of XAI enable us to explore the dependency of the large fire probability on its drivers. Specifically, the XAI measures the contribution from each driver to the predicted large fire probability (Equation 8). With the information, an XAI-based fire-driver partial dependency plot can be derived with the $X$-axis showing the values of the investigated driver and $Y$-axis showing its contribution to fire predictions (see Section 3.4). Such a plot allows us to visually ascertain whether a driver exerts a positive or negative influence on the predicted large fire probability (Kondylatos et al., 2022; Lundberg et al., 2018; Sundararajan et al., 2017; Wang et al., 2021). In the plot, if increases in the values of a driver increase (decrease) large fire probability, a positive (negative) relationship between the driver and fires is learned by the ML model (Kondylatos et al., 2022; Lundberg & Lee, 2017; Wang et al., 2021). Additionally, we quantitively assessed the sign of a driver's influence on large fire probability using segmented regression (where a positive slope represents positive influence, and vice versa) (Pilgrim, 2021). We then compared these XAI-revealed relationships in the ML models with established physical principles to determine whether a ML model correctly captured the sign (positive or negative) of these relationships. For example, from a mechanistic perspective, if a variable is known to positively influence large fires but the model indicates a negative relationship, we conclude that the model demonstrates poor physical interpretability of that relationship. We compared modeled relationships against all existing well-established physical knowledge (see Section 3.2). The greater the number of physical relationships a model correctly captured, the higher its physical interpretability was.

### 2.4. Experimental Settings

The data sets used to develop the ML fire prediction model are listed in Table 1. For fire modeling, we used the Fire Program Analysis fire occurrence database (Short, 2021), which includes all fire events reported by federal, state, and local fire organizations from 1992 to 2018 (Short, 2021), with each report recording the location, and occurrence date of each fire event. We considered three groups of drivers related to fuel flammability, fuel availability, and human suppression. For fuel flammability, factors such as BI, ERC, 100-hr and 1,000-hr dead fuel moisture (FM100 and FM1000, respectively), air temperature (T), relative humidity (RH), reference evapotranspiration (ET0), precipitation (Pr), and VPD were taken into account. The meteorological data were derived from the gridMET data set, which covers the contiguous US from 1979 to present and has been frequently used for fire modeling in the US (Abatzoglou, 2013; Gray et al., 2018; J. Li et al., 2020). The raw gridMET data, originally with a daily temporal resolution, was aggregated to a monthly scale by averaging the daily values for each variable within each respective month. Additionally, we incorporated the land cover types (500 m spatial resolution) and topography conditions (90 m spatial resolution) into the fire models due to their potential effects on fuel flammability (Alizadeh et al., 2023; Balch et al., 2022; Varner et al., 2015). We further used the NPP product and population distribution data from MODIS and Gridded Population of World Version 4 (GPWV4) (both 1 km spatial resolution) to represent fuel availability and human suppression, respectively. All the data sets were unified into the same spatial (1 km) and temporal (monthly) resolutions. Details of the data sets, including data sources and links, and references, are presented in Table 1.

With the data sets in Table 1, we trained and evaluated the ML fire prediction models. Specifically, data sets jointly covered by all variables in Table 1 were used for model development. During the studied period, 7,019 fires with their sizes larger than 405 ha were regarded as large fires. To avoid a highly imbalanced data, we randomly selected small fires at 1.5 times the number of large fires from small fire samples (Kondylatos et al., 2022). We evaluated fire model performance through both random cross validation and temporal cross validation. For the random cross validation, we randomly selected 80%, 10%, and 10% of the fire events for model training, validation (avoid overfitting during model training, Prechelt, 2002; Yuan, Zhu, Li, et al., 2022), and testing, respectively. Since the random sampling may vary model performance, we repeated the experiments five times, and each of the experiment involved data set sampling, and model training and evaluation. The mean and standard deviation of the model performance across these experiments were used for the final model comparison. For the temporal cross validation, we trained the model using data before 2019 and used the fire samples during 2019–2020 that the model has never seen in the Fire Program Analysis fire occurrence database (Short, 2022) as an independent evaluation data set. Additionally, we analyzed fires and their drivers during 2019–2020 through comparison with the satellite-derived fires in the Monitoring Trends in Burn Severity (MTBS) (Finco et al., 2012) data set. In our model, all the parameters in the three sub-modules, including $f_{\mathrm{FA}}$ (i.e., $k_1$ and $\mathrm{FA}_0$), $f_{\mathrm{FF}}$ (i.e., weights in attention-augmented LSTM), and $f_{\mathrm{HS}}$ (i.e., $k_2$ and $\mathrm{HS}_0$), were learned from the new data sets (Table 1). While the attention-based mechanisms used here were from the F. Li et al. (2023), the parameters were relearned together with the $f_{\mathrm{FA}}$ and $f_{\mathrm{HS}}$ modules rather than directly adopting from the previous study due to their differences in learning tasks (predicting burned area vs. predicting large fire probability) and studied regions (tropics vs. western US). The inputs for the $f_{\mathrm{FA}}$ and $f_{\mathrm{HS}}$ modules were fuel availability and human suppression proxied by NPP and population density, respectively. The inputs for the $f_{\mathrm{FF}}$ module were time series of the environmental factors in Table 1 except NPP and population density. The time steps were set to 8 months (i.e., the months prior to the fire ignition month) to capture the antecedent hydroclimatic conditions from monthly to inter-seasonal scales that may affect fire dynamics in the fire month (Buch et al., 2023; Ermitao et al., 2022; Littell et al., 2016). Details of the hyperparameter settings for the ML models, including the learning rate, batch size, dropout rate and weight decay of our model, are listed in Table S1 in Supporting Information S1, which were determined through the grid search method (F. Li et al., 2020; F. Li et al., 2023; Liashchynskyi & Liashchynskyi, 2019). We also included the code of our model with detailed comments for further reading (see Data Availability Statement section). To further demonstrate the importance of considering the compound climate controls from multiple critical drivers on fire prediction, we conducted ablation experiments. For the ablation experiments, we repeated the aforementioned random cross validation scheme five times, and kept all the experimental settings the same while ablating some important drivers for model training and evaluation. Subsequently, we compared the model performance with and without considering the important drivers (see Section 3.4).

## 3. Results and Discussions

### 3.1. Performance on Large Fire Prediction

We found that our model exhibited higher accuracy in large fire predictions compared to baselines (Figure 2). Among all the models, the $F_1$-score of our model was the highest (0.846 ± 0.012), about 20%–40% higher than that of the index-based models (i.e., BI-Pre, BI-Cur, ERC-Pre, ERC-Cur, FWI-Pre, and FWI-Cur), and approximately 4%–9% higher than those of three ML models (DT, ANN, and XGBoost). The accuracy of our model was slightly higher than that of the RF model (0.842 ± 0.007), although the difference was not statistically significant ($t$-test, $p > 0.05$). This comparative accuracy necessitates further comparison of the two ML models in terms of their structures and interpretability (see Section 3.2). The $F_1$-score of the index-based models ranged from 0.605 ± 0.008 to 0.704 ± 0.011, consistently lower than the ML fire prediction models (Figure 2), which ranged from 0.776 ± 0.009 to 0.846 ± 0.012. The lower performance of the process-driven models highlights the need for further improvement in the empirical relationships between fires and their drivers, and parameterizations of such relationships (F. Li et al., 2019; Zhu et al., 2022). However, the higher performance of ML models does not guarantee more reliable fire predictions, due to the black-box nature of ML models and the risk of overfitting.

### 3.2. Notable Model Structural Discrepancy Relative to Wildfire Physics

We found that the fire-driver relationships captured by our model showed higher consistency with established physical knowledge. We compared the sign of each physical relationship between drivers and fires with those from the ML models (Figure 3). The corresponding signs and references for the fire physics were listed in Table S2 in Supporting Information S1. Compared to other ML models, a larger portion of (10 out of 13) relationships were correctly captured by our model. Among the correctly captured relationships, the positive influences of evaporative demand (ET0), wind speed (WS), and slope, were more consistently captured across most ML models (Figure 3). Although the RF model showed comparable accuracy to our model (Figure 2), its modeled relationships were generally less physically interpretable (Figure 3), similar to the poor interpretability of the DT, ANN, and XGBoost models. These findings suggest that substantial model structural uncertainties could exist across ML models despite their almost equivalent accuracy. The pervasive lack of interpretability for the ML models underscores the importance of trustworthy model development and model evaluation beyond accuracy,

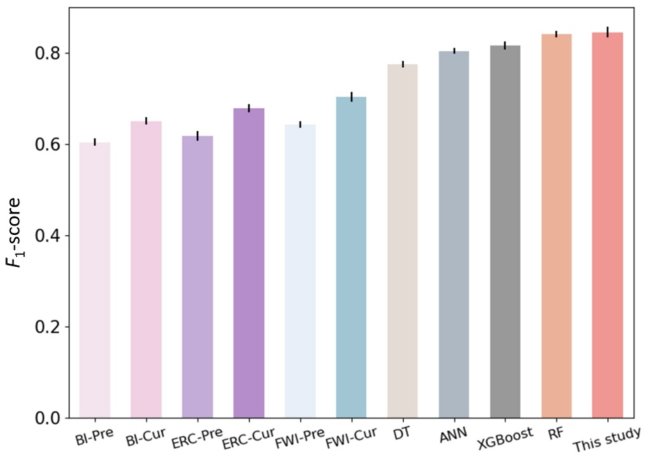


**Figure 2.** The machine learning fire model developed in this study outperforms fire indices and other machine learning fire models on accuracy. Fire indices include different versions of burning index, energy release component, and fire weather index. Machine learning models include decision tree, artificial neural network, extreme gradient boosting, random forest, and the model proposed in this study. The larger $F_1$-score represents better model performance.

such as model physical interpretability or causality structure assessment (Eyring et al., 2019; F. Li, Hao, et al., 2022; F. Li, Zhu, et al., 2022; Yuan, Zhu, Li, et al., 2022; Yuan, Zhu, Riley, et al., 2022).

For the fire physics listed in Figure 3, we acknowledge the complexity of the dependency between fires and climate variables. For example, previous studies have revealed non-linear relationships between fires and climate wetness conditions (Andela & Van Der Werf, 2014; Y. Chen et al., 2011; F. Li et al., 2023). In fuel-limited ecosystems, an increase in precipitation in wet season could amplify fire activities during dry season through enhancing fuel availability (Andela & Van Der Werf, 2014). In contrast, dry-season fires may negatively respond to short-term increasing precipitation due to over-wet fuel moisture (Andela & Van Der Werf, 2014). In our analysis, we considered the drivers present 1 month before fires and explicitly included information on fuel availability for fire predictions. Thus, changes in climate wetness primarily influenced predicted fire probability through their direct impacts on fuel moisture, rather than availability. Given that increases in fuel moisture typically decrease fuel flammability and limit fires, we assume negative relationships between large fire probability and climate wetness variables including FM100, FM1000, Pr, and RH, and positive relationships with BI, ERC, and ET0 (Figure 3). These relationships align with findings from previous studies on the temporal variability of the western US fires during historical period (Abatzoglou & Williams, 2016; Holden et al., 2018; Littell et al., 2016).

### 3.3. Distinct Differences in Model Predicted Fire Patterns

We found distinct differences in the association between actual burned grid cells and large fire probability predicted by different ML models (Figure 4). In grid cells where large fires occurred, we would expect a high large fire probability, as the fire drivers had already met the necessary conditions for the fire occurrence (Jolly et al., 2019; J. Li et al., 2020). While a higher predicted large fire probability of our model generally corresponded to a greater number of burned grid cells, the relationship between large fire probability and the number of burned grid cells exhibited non-monotonic characteristics for the ANN, XGBoost, and RF models (Figure 4). For example, the majority of burned grid cells for RF were distributed with fire probability ranging from 0.68 to 0.94, from 0.66 to 0.92 for XGBoost, and from 0.88 to 0.96 for ANN. Few burned grid cells were observed where the large fire probability reached the highest value in the three models (Figure 4). However, for the DT model, most

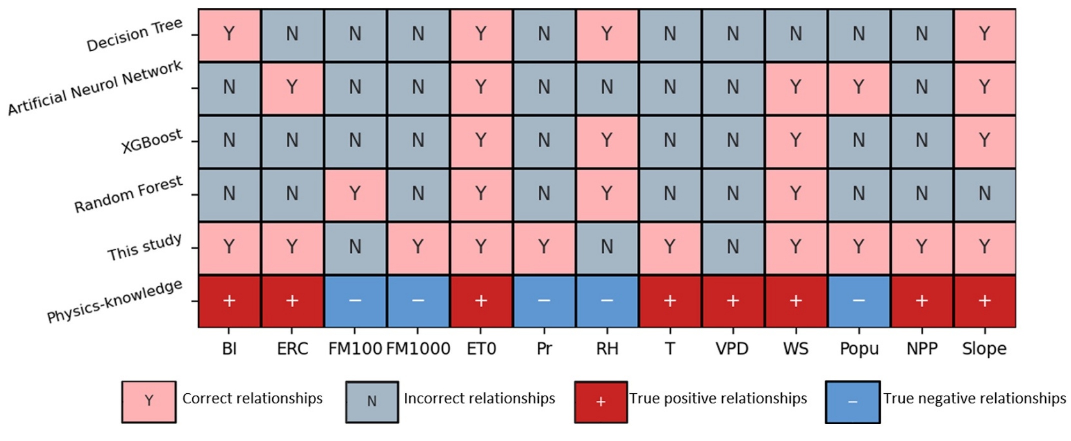


**Figure 3.** Benchmarking machine learning modeled relationships with physical knowledge. The bottom row shows the physics, where "+" and "−" represent positive and negative relationships between fire drivers (full name of the drivers are shown in Table 1) and fires, respectively. The machine learning correctly and incorrectly modeled relationships are shown as "Y" and "N", respectively. When the sign of a modeled relationship between a driver and fires is consistent with that of the physical knowledge, the model correctly captured that relationship here, and vice versa.

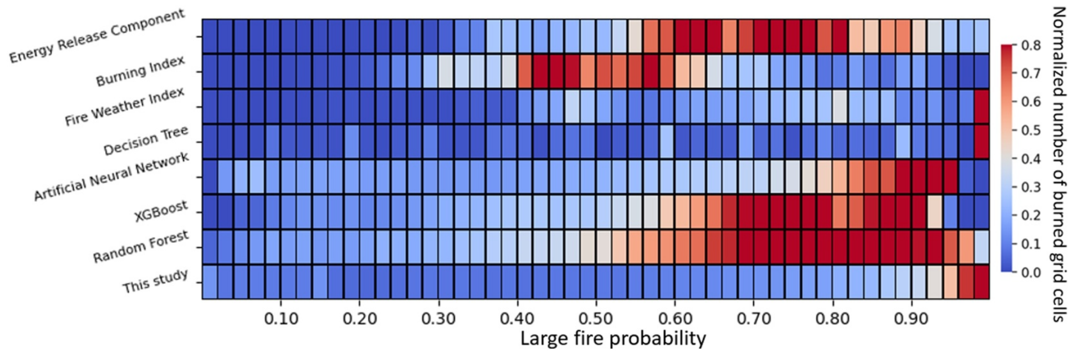


**Figure 4.** Distinct differences in large fire probability over actually burned grid cells predicted by five machine learning fire prediction models and three fire indices. The horizonal axis represents the model predicted large fire probability, and the vertical axis represents models. The deeper red (blue) color represents a larger (smaller) number of burned grid cells falling into the divided bin of the predicted large fire probability on the horizontal axis. The number of burned grid cells in each bin is normalized by the maximum number of burned grid cells across all the horizontal bins for each model.

burned grid cells were concentrated at the highest large fire probability values, despite its lower accuracy and physical interpretability (Figures 2 and 3). These findings suggest that different ML models can yield distinct fire predictions due to structural uncertainties (Figure 3), despite having comparable prediction accuracies (e.g., our model vs. RF). For the three fire indices, while most of the burned grid cells occurred with the highest FWI values (Figure 4), moderate values of BI and ERC coincided with more burned grid cells than their highest values. Similar patterns of BI and ERC compared with actual fires have also been revealed by Jolly et al. (2019), suggesting the need for more advanced fire indices that can more effectively reflect actual fire activities and aid fire management (Bradshaw, 1984; Jolly et al., 2019; J. Li et al., 2020; Van Wagner, 1987). The tendency observed in our model, where a higher large fire probability corresponds to more burned grid cells, indicates its potential utility for early warning of large fires 1 month in advance.

### 3.4. Strong and Compound Climate Controls on Large Fires

Using the high accurate and interpretable fire model proposed, we found that climate variables, including ET0, ERC, T, and WS, predominantly controlled fuel flammability and thus large fires (Figure 5a) during 2002–2018. Specifically, increases in these dominant drivers were associated with increased fuel flammability (Figure 5b), consistent with existing fire physical knowledge (Abatzoglou et al., 2023; Holden et al., 2018; Jolly et al., 2019; McEvoy et al., 2020). ET0 emerged as the largest contributor to fuel flammability, followed by ERC (Figure 5a). ET0 is strongly associated with drought and wildfire risk in the western US (Abatzoglou & Williams, 2016; Littell et al., 2016; McEvoy et al., 2016, 2020), with a higher value reflecting a larger evaporative demand over water-unlimited surface. Similarly, a higher ERC value represents a drier fuel moisture condition and thus higher fire danger (Brown et al., 2004; Finney et al., 2011; Jolly et al., 2019). After the two variables, the impacts of T and WS were the most prominent (Figure 5a). High temperature can dry out dead fuels (Holden et al., 2018; Williams et al., 2012) and high wind speed can accelerate fire spread, contributing to a considerable portion of large fires in the western US (Abatzoglou et al., 2018, 2023; Westerling et al., 2006). While ET0 and ERC were the most important two variables (Figure 5a), our model performance significantly (one-tailed $t$-test, $P < 0.05$) decreased by 7.5%, and 5.6% if we only used ET0 or ERC as the hydroclimatic drivers for model training and evaluation (i.e., ablating other hydroclimatic drivers), respectively. These findings indicate the importance of considering the compound effects of multiple key drivers, rather than focusing solely on one specific driver, when predicting large fires in the western US (Richardson et al., 2022).

We further predicted and analyzed the top five megafires in 2020 (Figure 6). We found that the large fire probability for the top five megafires in 2020 was anomalously high compared to the multi-year (2002–2020) average fire probability over the same grid cells and in the same month (Figure 6a). While about one fifth of grid cells in western US exhibited high large fire probability (fire probability greater than 0.9) (Figure S1 in Supporting Information S1), the majority of grid cells burned overlapped with high large fire probability, especially for the top four megafires (Figure 6b). The fire model predictions were able to reflect the interannual variations in fire occurrences. For example, compared to the smaller fire occurrences in 2019 (Figure S2 in Supporting

10.1029/2024EF004588

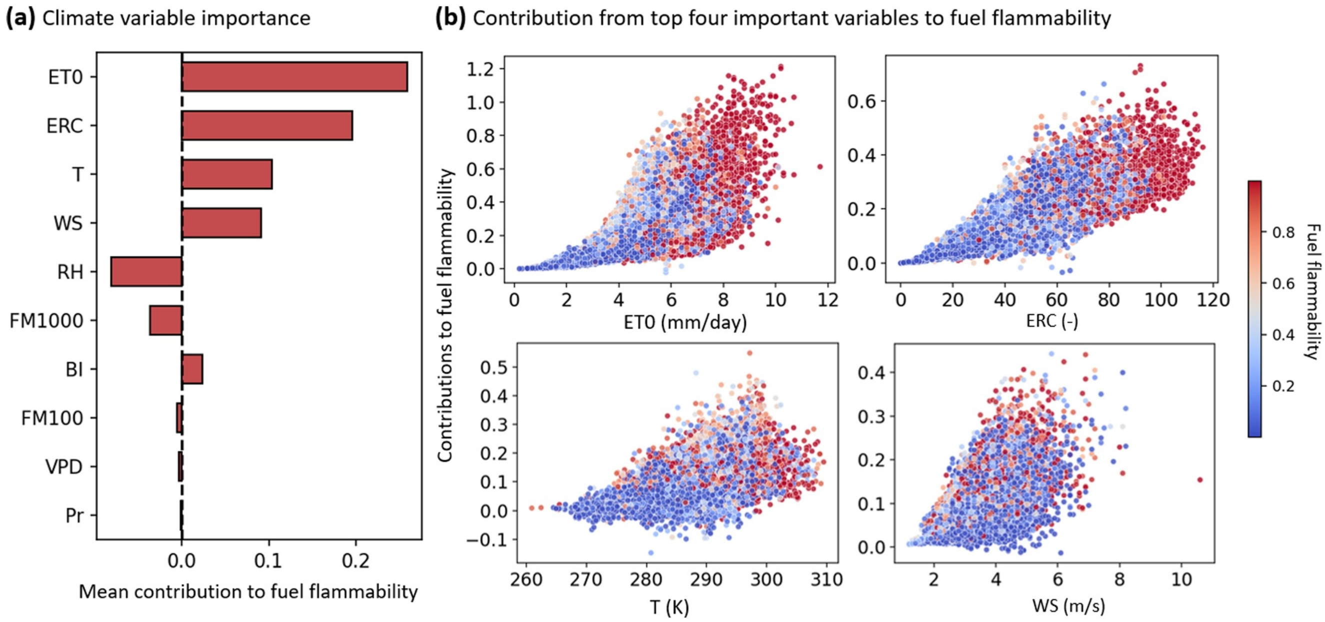


**Figure 5.** Dominant climate variables and their quantitative contributions to fuel flammability. The mean contribution from each driver to quantified fuel flammability (a) where positive (negative) values represent positive (negative) contributions, and a larger absolute value represents a larger contribution. The contributions from the top four dominant variables to fuel flammability (b) where the deeper red (blue) color represents a higher (lower) value of fuel flammability predicted by the machine learning fire model.

Information S1) and the multi-year average, the large fire probability in 2020 was notably higher for several megafires, including August Complex, SCU Lightening Complex, North Complex, and Hennessey (Figure 6c), which burned approximately 4,325, 1,642, 1,281, and 1,268 $km^2$, respectively. In contrast, the Cameron Peak megafire exhibited a moderately large fire probability (Figures 6b and 6c). The effective prediction of those megafires with 1-month lead time allows for more targeted fire management strategies, such as prescribed burning and fuel treatment (Kolden, 2019; Stephens & Ruth, 2005).

Further attribution analysis confirmed the prominent impacts of aforementioned multiple dominant climate variables on the five megafires (Figures 6d–6h). Interestingly, although the key variables remained consistent, their ranking and relative importance varied across the five megafires (Figure 6). For example, ERC was the most important variable for the August Complex fire (Figure 6d), while ET0 was more important for the other megafires (Figures 6e–6h). Additionally, RH was more important for the August Complex and North Complex fires (Figures 6d and 6f). These results suggest the spatially varied controls on wildfires, which cannot be differentiated using static variable importance information across the entire data set. Existing studies suggest that anthropogenic climate change is likely to lead to more intense and frequent multi-variate-controlled climate extremes and compound hazards, including drought related to wildfires (Abatzoglou, Rupp, et al., 2021; AghaKouchak et al., 2020; Hao et al., 2022). Therefore, a reliable approach to modeling such spatiotemporally-varied and compound climate effects on wildfires is essential for effectively mitigating and adapting the adverse impacts of climate change.

### 3.5. Perspectives

The fire model that we developed in this study well balanced accuracy and physical interpretability (Figure 7) and could aid a more trustworthy solution for fire predictions. Similar to prior studies (Arrieta et al., 2020; Hu et al., 2023), we revealed the tradeoff between model accuracy and interpretability for both ML models and process-driven fire indices (Figure 7). Both highly interpretable but less accurate fire indices, and highly accurate but less interpretable ML models, could reduce confidence in fire predictions. While we acknowledge that the fire model proposed has limitations, particularly being less interpretable than process-driven models, it is much more interpretable than the other ML models, and shows substantially higher accuracy compared to process-driven index-based models (Figure 7). Our model exhibits high accuracy relative to the commonly used ML models

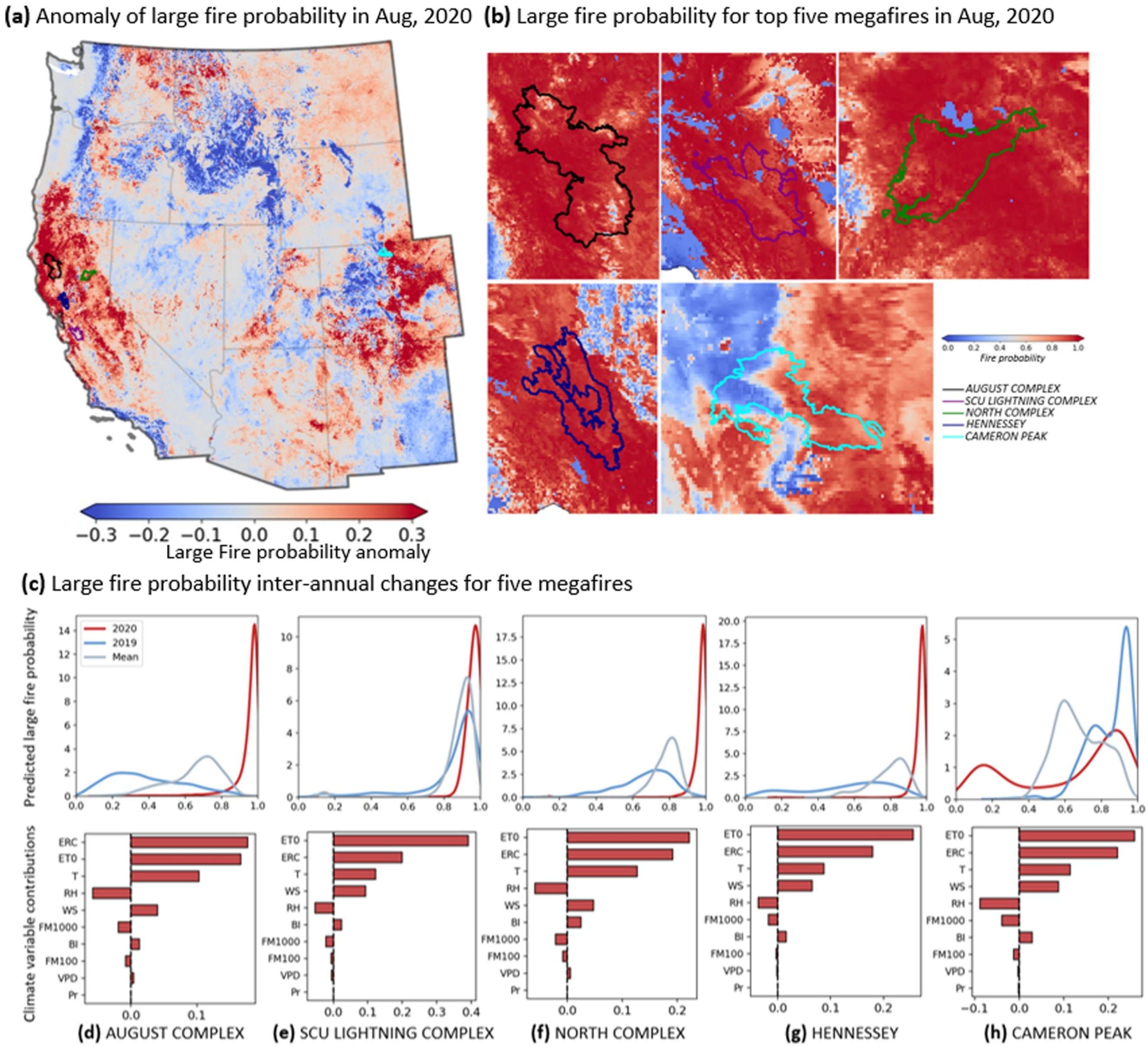


**Figure 6.** Anomaly high large fire probability driven by compound fire weather overlapped with the top five largest megafires in 2020. The large fire probability anomaly (a) predicted by the model proposed 1 month in advance and their overlap with the top five megafires (b). For each grid cell in (a), deeper red (blue) color represents anomaly higher (lower) fire probability compared to that of the grid-level multi-year (2002–2020) average in the same month. The colored boundaries in (a) and (b) represent the area burned depicted by the Monitoring Trends in Burn Severity data. Comparison of the 2020 large fire probability with that of 2019 and multi-year average (c). The contributions from climate variables to the top five megafires (d–h).

in terms of $F_1$-score through the random-cross validation (Figures 2 and 7), and also shows reasonable accuracy through the temporal-cross validation scheme ($F_1$-score = 0.839 ± 0.006). Additionally, we emphasize the great importance of physical knowledge in the process-driven models, despite their structural and parametrization uncertainties (Ji et al., 2024; F. Li, Hao, et al., 2022; F. Li, Hao, et al., 2024; Wu et al., 2020; Yuan et al., 2021; Yuan, Zhu, Riley, et al., 2022). Such knowledge serves as a foundation for designing more trustworthy ML models and is essential for validating these models beyond just their accuracy.

We underscore that high interpretability of ML with reasonable accuracy is essential for more trustworthy ML development, however, building a trustworthy ML model could be beyond interpretability and accuracy. First, a ML model with lower or less acceptable accuracy can indicate likely missing or biased critical drivers of the

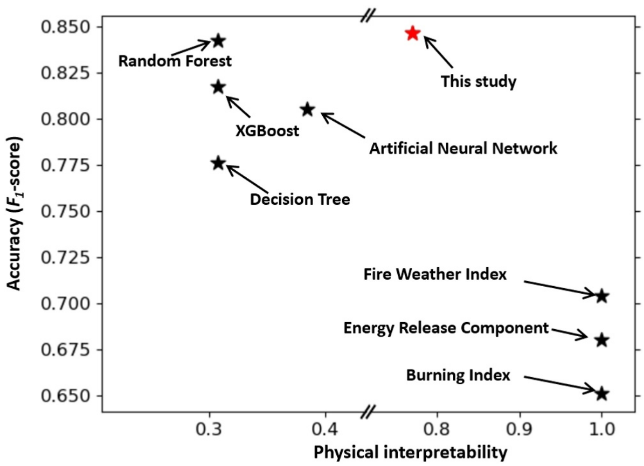


**Figure 7.** Comparison of fire model performance from two dimensions: accuracy and physical interpretability. The physical interpretability here is measured as the percentage of correctly modeled relationships between fires and their drivers relative to fire physics.

predicted variable or defective model structures or parameterization (F. Li et al., 2023; Yuan, Zhu, Li, et al., 2022; Yuan, Zhu, Riley, et al., 2022). If so, the relationships learned by ML models can be less trustworthy (Murdoch et al., 2019). Second, if a model shows reasonable accuracy but its modeled relationships are less interpretable or violate principal physics, further investigations are required to diagnose the causes (e.g., model structure issues) (F. Li, Zhu, et al., 2022; Runge et al., 2019, 2023; Yuan et al., 2021; Zeng et al., 2023). Maximizing both interpretability and accuracy ideally aids trustworthiness (Murdoch et al., 2019), however, a tradeoff between them is practically common (e.g., Figure 7) (Arrieta et al., 2020). A ML model that well balances interpretability and accuracy is, therefore, useful in real-world applications such as wildfire prediction (Jain et al., 2020; F. Li et al., 2023; Marcinkevičs & Vogt, 2020; Murdoch et al., 2019). Finally, while the concept of trustworthiness can be varied in different disciplines (Toreini et al., 2020), trustworthy ML in computer science includes other critical aspects apart from interpretability and accuracy, such as fairness, auditability, privacy, and safety (Eshete, 2021; Murdoch et al., 2019; Toreini et al., 2020). The concept of trustworthy ML and how to quantify trustworthiness in Earth science (e.g., wildfire science) warrant further exploration before we can better evaluate and compare trustworthiness across Earth ML models.

We highlight the importance of evaluating the interpretability of ML models, particularly with reference to well-known physics. With XAI, we found substantial uncertainty in ML learned relationships (Figure 3). Specifically, integrated gradients and tree-SHAP were used to infer the fire-driver relationships for neural-network and tree-based ML models, respectively. Since the two XAI methods used are not mathematically identical, we did not directly compare the sign of the XAI outputs (i.e., Equation 8). Instead, similar to previous studies (Kondylatos et al., 2022; Lundberg & Lee, 2017; Wang et al., 2021), we derived the dependency plot between each driver and its contribution to fires (e.g., Figure 5b), and then checked whether increases in a driver positively or negatively (i.e., the signs) influence the large fire probability. By doing so, the inferred signs reflect the trend of the association between each driver and fires rather than the signs of the original XAI outputs. Nevertheless, we acknowledge that differences in XAI methods may still influence the inferred relationships across different ML models, and more efforts are required to develop more advanced XAI methods that are computational efficient and appliable to all kinds of ML models (Saeed & Omlin, 2023; Turbé et al., 2023). We further demonstrated that the ML structural uncertainties can bias model predictions (Figure 4), indicating the critical need for improving model structures rather than only focusing on accuracy. Our proposed interpretability evaluation framework integrates XAI and physics, aiding better understanding and diagnosing the black-box ML models in the Earth science.

This study mainly focuses on next-month large fire probability prediction using prior fire conditions. The potential physics that support such a modeling scheme include (a) pre-fire conditions can exert important effects on fuel flammability (e.g., fuel moisture) and availability (e.g., fuel accumulation) and thus fires in the following month (Abolafia-Rosenzweig et al., 2022; Ermitao et al., 2022); (b) hydroclimatic conditions may show certain persistence characteristics (e.g., persistent drought) and thus prior fire conditions can reflect certain fire conditions in the next month (Littell et al., 2016; Robertson & Vitart, 2018; Tuel & Martius, 2023). The reasonable high accuracy (Figure 2) indicates that most of the large fires in the western US can be successfully predicted by our modeling scheme 1 month in advance, and the lead time enables fire management before fire occurrence. For megafires that can spread for multiple months such as the August Complex megafire which started from August and ended in early November (Varga et al., 2022), our model reasonably captured the persistently high fire risk in August, September, and October over the footprint of the megafire (Figure S3 in Supporting Information S1). However, our model can only iteratively predict the next-month fire risk when the latest data set of fire drivers is available, rather than directly predicting the fire risk with lead time of few months. Furthermore, we acknowledge that when fire conditions sharply change between the predicted month and its prior months (e.g., pre-fire conditions can be substantially different from the next-month fire occurrence conditions), the ML model may fail to predict fires in the next month. While including the hydroclimatic conditions in the fire-occurred month can be an optional prediction scheme (Gray et al., 2018; J. Li et al., 2020; Wang et al., 2021), the scheme itself requires predicted hydroclimatic information from weather/climate forecast models which remain challenging from sub-seasonal to seasonal scales (Doblas-Reyes et al., 2013; He et al., 2021; F. Li, Zhu, et al., 2022; Robertson & Vitart, 2018). Additionally, in our proposed model, we followed the previous studies (Balch et al., 2017; B. Chen et al., 2021; Ellis et al., 2022; Hantson et al., 2015) and considered NPP as a proxy of fuel availability. Further investigation is required by considering improved spatiotemporal representation of fuel availability (e.g., using fuel products upscaled by integrating field measurements, remote sensing, and ML (D'Este et al., 2021; Morais et al., 2021; Walker et al., 2020)), possibly varied fire responses to fuel availability across land cover types (Ermitao et al., 2022), and other critical aspects of fuel characteristics, such as fuel size, types, spatial arrangement and continuity, and chemical composition of fuels (Gale et al., 2021).

## 4. Conclusions

Predicting large fires with a long-term lead time and at a high spatial resolution is essential for early fire warning and fire management. This study developed a fire model, which demonstrated higher accuracy in large fire prediction compared to other commonly used ML models and fire indices. With explainable AI, we revealed substantial model structural uncertainties among ML fire prediction models, which affected their predictions despite their comparable accuracy. Our model exhibited higher interpretability and was more consistent with established fire physics. With the highly accurate and interpretable fire model, we revealed strong compound effects of climate variables related to evaporative demand, ERC, temperature, and wind speed on controlling large fires and top megafires in the western US. Our developed fire model could facilitate early fire warning and enable targeted fire management. Our results highlight the urgent need for high accurate, interpretable, and trustworthy AI development, underscore the necessity for moving model evaluation beyond accuracy, and demonstrate the importance of comprehensively modeling the compound climate controls on large fires.

## Data Availability Statement

All the dataset used and the hybrid fire model code are available in F. Li, Zhu, et al. (2024).

**Acknowledgments**
Support for this research was provided by American Family Insurance through a research partnership with the University of Wisconsin–Madison's Data Science Institute. Q.Z. is supported by the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Scientific Focus Area Project, sponsored by the Earth and Environmental Systems Modeling (EESM) Program under the Office of Biological and Environmental Research of the US Department of Energy Office of Science. K.Y. is supported by an Early Career Development Grant (ECDG), which was sponsored by the Earth and Environmental Sciences Area of Lawrence Berkeley National Laboratory.

## References

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, *33*(1), 121–131. https://doi.org/10.1002/joc.3413

Abatzoglou, J. T., Balch, J. K., Bradley, B. A., & Kolden, C. A. (2018). Human-related ignitions concurrent with high winds promote large wildfires across the USA. *International Journal of Wildland Fire*, *27*(6), 377–386. https://doi.org/10.1071/wf17149

Abatzoglou, J. T., Battisti, D. S., Williams, A. P., Hansen, W. D., Harvey, B. J., & Kolden, C. A. (2021). Projected increases in Western US forest fire despite growing fuel constraints. *Communications Earth & Environment*, *2*, 1–8. https://doi.org/10.1038/s43247-021-00299-0

Abatzoglou, J. T., Kolden, C., Williams, A., Sadegh, M., Balch, J., & Hall, A. (2023). Downslope wind-driven fires in the Western United States. *Earth's Future*, *11*(5), e2022EF003471. https://doi.org/10.1029/2022ef003471

Abatzoglou, J. T., Rupp, D. E., O'Neill, L. W., & Sadegh, M. (2021). Compound extremes drive the Western Oregon wildfires of September 2020. *Geophysical Research Letters*, *48*(8), e2021GL092520. https://doi.org/10.1029/2021gl092520

Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across Western US forests. *Proceedings of the National Academy of Sciences*, *113*(42), 11770–11775. https://doi.org/10.1073/pnas.1607171113

Abolafia-Rosenzweig, R., He, C., & Chen, F. (2022). Winter and spring climate explains a large portion of interannual variability and trend in Western US summer fire burned area. *Environmental Research Letters*, *17*(5), 054030. https://doi.org/10.1088/1748-9326/ac6886

AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasni, O., et al. (2020). Climate extremes and compound hazards in a warming world. *Annual Review of Earth and Planetary Sciences*, *48*(1), 519–548. https://doi.org/10.1146/annurev-earth-071719-055228

Alizadeh, M. R., Abatzoglou, J. T., Adamowski, J., Modaresi Rad, A., AghaKouchak, A., Pausata, F. S., & Sadegh, M. (2023). Elevation-dependent intensification of fire danger in the Western United States. *Nature Communications*, *14*(1), 1773. https://doi.org/10.1038/s41467-023-37311-4

Andela, N., & Van Der Werf, G. R. (2014). Recent trends in African fires driven by cropland expansion and El Niño to La Niña transition. *Nature Climate Change*, *4*(9), 791–795. https://doi.org/10.1038/nclimate2313

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Balch, J. K., Abatzoglou, J. T., Joseph, M. B., Koontz, M. J., Mahood, A. L., McGlinchy, J., et al. (2022). Warming weakens the night-time barrier to global fire. *Nature*, *602*(7897), 442–448. https://doi.org/10.1038/s41586-021-04325-1

Balch, J. K., Bradley, B. A., Abatzoglou, J. T., Nagy, R. C., Fusco, E. J., & Mahood, A. L. (2017). Human-started wildfires expand the fire niche across the United States. *Proceedings of the National Academy of Sciences*, *114*(11), 2946–2951. https://doi.org/10.1073/pnas.1617394114

Bowman, D. M., Balch, J. K., Artaxo, P., Bond, W. J., Carlson, J. M., Cochrane, M. A., et al. (2009). Fire in the Earth system. *Science*, *324*(5926), 481–484. https://doi.org/10.1126/science.1163886

Bradshaw, L. S. (1984). *The 1978 national fire-danger rating system: Technical documentation*. US Department of Agriculture, Forest Service, Intermountain Forest and Range.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Brown, T. J., Hall, B. L., & Westerling, A. L. (2004). The impact of twenty-first century climate change on wildland fire danger in the Western United States: An applications perspective. *Climatic Change*, *62*(1–3), 365–388. https://doi.org/10.1023/b:clim.0000013680.07783.de

Buch, J., Williams, A. P., Juang, C. S., Hansen, W. D., & Gentine, P. (2023). SMLFire1.0: A stochastic machine learning (SML) model for wildfire activity in the Western United States. *Geoscientific Model Development*, *16*(12), 3407–3433. https://doi.org/10.5194/gmd-16-3407-2023

Calviño-Cancela, M., Chas-Amil, M. L., García-Martínez, E. D., & Touza, J. (2016). Wildfire risk associated with different vegetation types within and outside wildland-urban interfaces. *Forest Ecology and Management*, *372*, 1–9. https://doi.org/10.1016/j.foreco.2016.04.002

Cattau, M. E., Wessman, C., Mahood, A., & Balch, J. K. (2020). Anthropogenic and lightning-started fires are becoming larger and more frequent over a longer season length in the USA. *Global Ecology and Biogeography*, *29*(4), 668–681. https://doi.org/10.1111/geb.13058

Chen, B., Jin, Y., Scaduto, E., Moritz, M. A., Goulden, M. L., & Randerson, J. T. (2021). Climate, fuel, and land use shaped the spatial pattern of wildfire in California's Sierra Nevada. *Journal of Geophysical Research: Biogeosciences*, *126*(2), e2020JG005786. https://doi.org/10.1029/2020jg005786

Chen, Y., Randerson, J. T., Morton, D. C., DeFries, R. S., Collatz, G. J., Kasibhatla, P. S., et al. (2011). Forecasting fire season severity in South America using sea surface temperature anomalies. *Science*, *334*(6057), 787–791. https://doi.org/10.1126/science.1209472

Coffield, S. R., Graff, C. A., Chen, Y., Smyth, P., Foufoula-Georgiou, E., & Randerson, J. T. (2019). Machine learning to predict final fire size at the time of ignition. *International Journal of Wildland Fire*, *28*(11), 861–873. https://doi.org/10.1071/wf19023

Dennison, P. E., Brewer, S. C., Arnold, J. D., & Moritz, M. A. (2014). Large wildfire trends in the Western United States, 1984–2011. *Geophysical Research Letters*, *41*(8), 2928–2933. https://doi.org/10.1002/2014gl059576

D'Este, M., Elia, M., Giannico, V., Spano, G., Lafortezza, R., & Sanesi, G. (2021). Machine learning techniques for fine dead fuel load estimation using multi-source remote sensing data. *Remote Sensing*, *13*(9), 1658. https://doi.org/10.3390/rs13091658

Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., & Rodrigues, L. R. (2013). Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, *4*, 245–268. https://doi.org/10.1002/wcc.217

Dowdy, A. J., Mills, G. A., Finkele, K., & de Groot, W. (2010). Index sensitivity analysis applied to the Canadian forest fire weather index and the McArthur forest fire danger index. *Meteorological Applications*, *17*(3), 298–312. https://doi.org/10.1002/met.170

Doxsey-Whitfield, E., MacManus, K., Adamo, S. B., Pistolesi, L., Squires, J., Borkovska, O., & Baptista, S. R. (2015). Taking advantage of the improved availability of census data: A first look at the gridded population of the world, version 4. *Papers in Applied Geography*, *1*(3), 226–234. https://doi.org/10.1080/23754931.2015.1014272

Ellis, T. M., Bowman, D. M., Jain, P., Flannigan, M. D., & Williamson, G. J. (2022). Global increase in wildfire risk due to climate-driven declines in fuel moisture. *Global Change Biology*, *28*(4), 1544–1559. https://doi.org/10.1111/gcb.16006

Ermitao, T., Gouveia, C. M., Bastos, A., & Russo, A. C. (2022). Interactions between hot and dry fuel conditions and vegetation dynamics in the 2017 fire season in Portugal. *Environmental Research Letters*, *17*(9), 095009. https://doi.org/10.1088/1748-9326/ac8be4

Eshete, B. (2021). Making machine learning trustworthy. *Science*, *373*(6556), 743–744. https://doi.org/10.1126/science.abi5052

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, *45*(2), RG2004. https://doi.org/10.1029/2005rg000183

Finco, M., Quayle, B., Zhang, Y., Lecker, J., Megown, K. A., & Brewer, C. K. (2012). Monitoring trends and burn severity (MTBS): Monitoring wildfire activity for the past quarter century using Landsat data. In *Moving from status to trends: Forest inventory and analysis (FIA) symposium* (pp. 4–6).

Finney, M. A., McHugh, C. W., Grenfell, I. C., Riley, K. L., & Short, K. C. (2011). A simulation of probabilistic wildfire risk components for the continental United States. *Stochastic Environmental Research and Risk Assessment*, *25*(7), 973–1000. https://doi.org/10.1007/s00477-011-0462-z

Friedl, M., & Sulla-Menashe, D. (2019). MCD12Q1 MODIS/Terra+Aqua land cover type yearly L3 global 500 m SIN grid V006. In *NASA EOSDIS land processes distributed active archive center*.

Fusco, E. J., Abatzoglou, J. T., Balch, J. K., Finn, J. T., & Bradley, B. A. (2016). Quantifying the human influence on fire ignition across the Western USA. *Ecological Applications*, *26*(8), 2390–2401. https://doi.org/10.1002/eap.1395

Gale, M. G., Cary, G. J., Van Dijk, A. I., & Yebra, M. (2021). Forest fire fuel through the lens of remote sensing: Review of approaches, challenges and future directions in the remote sensing of biotic determinants of fire behaviour. *Remote Sensing of Environment*, *255*, 112282. https://doi.org/10.1016/j.rse.2020.112282

Gray, M. E., Zachmann, L. J., & Dickson, B. G. (2018). A weekly, continually updated dataset of the probability of large wildfires across Western US forests and woodlands. *Earth System Science Data*, *10*(3), 1715–1727. https://doi.org/10.5194/essd-10-1715-2018

Gui, Z., Sun, Y., Yang, L., Peng, D., Li, F., Wu, H., et al. (2021). LSI-LSTM: An attention-aware LSTM for real-time driving destination prediction by considering location semantics and location importance of trajectory points. *Neurocomputing*, *440*, 72–88. https://doi.org/10.1016/j.neucom.2021.01.067

Guo, T., Lin, T., & Antulov-Fantulin, N. (2019). Exploring interpretable LSTM neural networks over multi-variable data. In *International conference on machine learning* (pp. 2494–2504). PMLR.

Hantson, S., Lasslop, G., Kloster, S., & Chuvieco, E. (2015). Anthropogenic effects on global mean fire size. *International Journal of Wildland Fire*, *24*(5), 589–596. https://doi.org/10.1071/wf14208

Hao, Z., Hao, F., Xia, Y., Feng, S., Sun, C., Zhang, X., et al. (2022). Compound droughts and hot extremes: Characteristics, drivers, changes, and impacts. *Earth-Science Reviews*, *235*, 104241. https://doi.org/10.1016/j.earscirev.2022.104241

He, S., Li, X., DelSole, T., Ravikumar, P., & Banerjee, A. (2021). Sub-seasonal climate forecasting via machine learning: Challenges, analysis, and advances. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 169–177).

Hicke, J. A., Meddens, A. J., & Kolden, C. A. (2016). Recent tree mortality in the Western United States from bark beetles and forest fires. *Forest Science*, *62*(2), 141–153. https://doi.org/10.5849/forsci.15-086

Higuera, P. E., & Abatzoglou, J. T. (2021). Record-setting climate enabled the extraordinary 2020 fire season in the Western United States. *Global Change Biology*, *27*(1), 1–2. https://doi.org/10.1111/gcb.15388

Holden, Z. A., Swanson, A., Luce, C. H., Jolly, W. M., Maneta, M., Oyler, J. W., et al. (2018). Decreasing fire season precipitation increased recent Western US forest wildfire activity. *Proceedings of the National Academy of Sciences*, *115*(36), E8349–E8357. https://doi.org/10.1073/pnas.1802316115

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

Hu, T., Zhang, X., Bohrer, G., Liu, Y., Zhou, Y., Martin, J., et al. (2023). Crop yield prediction via explainable AI and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield. *Agricultural and Forest Meteorology*, *336*, 109458. https://doi.org/10.1016/j.agrformet.2023.109458

Iglesias, V., Balch, J. K., & Travis, W. R. (2022). US fires became larger, more frequent, and more widespread in the 2000s. *Science Advances*, *8*(11), eabc0020. https://doi.org/10.1126/sciadv.abc0020

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

Ji, F., Li, F., Hao, D., Shiklomanov, A. N., Yang, X., Townsend, P. A., et al. (2024). Unveiling the transferability of PLSR models for leaf trait estimation: Lessons from a comprehensive analysis with a novel global dataset. *New Phytologist*, *243*(1), 111–131. https://doi.org/10.1111/nph.19807

Jolly, W. M., Cochrane, M. A., Freeborn, P. H., Holden, Z. A., Brown, T. J., Williamson, G. J., & Bowman, D. M. (2015). Climate-induced variations in global wildfire danger from 1979 to 2013. *Nature Communications*, *6*, 1–11. https://doi.org/10.1038/ncomms8537

Jolly, W. M., Freeborn, P. H., Page, W. G., & Butler, B. W. (2019). Severe fire danger index: A forecastable metric to inform firefighter and community wildfire risk management. *Fire*, *2*(3), 47. https://doi.org/10.3390/fire2030047

Joshi, J., & Sukumar, R. (2021). Improving prediction and assessment of global fires using multilayer neural networks. *Scientific Reports*, *11*(1), 3295. https://doi.org/10.1038/s41598-021-81233-4

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems* (Vol. 30).

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

Kolden, C. A. (2019). We're not doing enough prescribed fire in the Western United States to mitigate wildfire risk. *Fire*, *2*, 30. https://doi.org/10.3390/fire2020030

Kondylatos, S., Prapas, I., Ronco, M., Papoutsis, I., Camps-Valls, G., Piles, M., et al. (2022). Wildfire danger prediction and understanding with deep learning. *Geophysical Research Letters*, *49*(17), e2022GL099368. https://doi.org/10.1029/2022gl099368

Li, F., Gui, Z., Wu, H., Gong, J., Wang, Y., Tian, S., & Zhang, J. (2018). Big enterprise registration data imputation: Supporting spatiotemporal analysis of industries in China. *Computers, Environment and Urban Systems*, *70*, 9–23. https://doi.org/10.1016/j.compenvurbsys.2018.01.010

Li, F., Gui, Z., Zhang, Z., Peng, D., Tian, S., Yuan, K., et al. (2020). A hierarchical temporal attention-based LSTM encoder-decoder model for individual mobility prediction. *Neurocomputing*, *403*, 153–166. https://doi.org/10.1016/j.neucom.2020.03.080

Li, F., Hao, D., Zhu, Q., Yuan, K., Braghiere, R. K., He, L., et al. (2022). Vegetation clumping modulates global photosynthesis through adjusting canopy light environment. *Global Change Biology*, *29*(3), 731–746. https://doi.org/10.1111/gcb.16503

Li, F., Hao, D., Zhu, Q., Yuan, K., Braghiere, R. K., He, L., et al. (2024). Global impacts of vegetation clumping on regulating land surface heat fluxes. *Agricultural and Forest Meteorology*, *345*, 109820. https://doi.org/10.1016/j.agrformet.2023.109820

Li, F., Val Martin, M., Hantson, S., Andreae, M. O., Arneth, A., Lasslop, G., et al. (2019). Historical (1700–2012) global multi-model estimates of the fire emissions from the fire modeling intercomparison project (FireMIP).

Li, F., Zeng, X., & Levis, S. (2012). A process-based fire parameterization of intermediate complexity in a dynamic global vegetation model. *Biogeosciences*, *9*(7), 2761–2780. https://doi.org/10.5194/bg-9-2761-2012

Li, F., Zhu, Q., Riley, W. J., Yuan, K., Wu, H., & Gui, Z. (2022). Wetter California projected by CMIP6 models with observational constraints under a high GHG emission scenario. *Earth's Future*, *10*(4), e2022EF002694. https://doi.org/10.1029/2022ef002694

Li, F., Zhu, Q., Riley, W. J., Zhao, L., Xu, L., Yuan, K., et al. (2023). AttentionFire_v1.0: Interpretable machine learning fire model for burned-area predictions over tropics. *Geoscientific Model Development*, *16*(3), 869–884. https://doi.org/10.5194/gmd-16-869-2023

Li, F., Zhu, Q., Yuan, K., Ji, F., Paul, A., Lee, P., et al. (2024). Projecting large fires in the Western US with an interpretable and accurate hybrid machine learning method [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.12112363

Li, J., Albert, A., White, B. L., & Mudigonda, M. (2020). *A random forest model for the probability of large wildfires in California*. ICLR.

Liashchynskyi, P., & Liashchynskyi, P. (2019). Grid search, random search, genetic algorithm: A big comparison for NAS. arXiv preprint arXiv:1912.06059.

Linn, R., Winterkamp, J., Edminster, C., Colman, J. J., & Smith, W. S. (2007). Coupled influences of topography and wind on wildland fire behaviour. *International Journal of Wildland Fire*, *16*(2), 183–195. https://doi.org/10.1071/wf06078

Littell, J. S., McKenzie, D., Peterson, D. L., & Westerling, A. L. (2009). Climate and wildfire area burned in Western US ecoprovinces, 1916–2003. *Ecological Applications*, *19*(4), 1003–1021. https://doi.org/10.1890/07-1183.1

Littell, J. S., Peterson, D. L., Riley, K. L., Liu, Y., & Luce, C. H. (2016). A review of the relationships between drought and forest fire in the United States. *Global Change Biology*, *22*(7), 2353–2369. https://doi.org/10.1111/gcb.13275

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (Vol. 30).

Marcinkevičs, R., & Vogt, J. E. (2020). Interpretability and explain ability: A machine learning zoo mini-tour. arXiv preprint arXiv:2012.01805.

McEvoy, D. J., Huntington, J. L., Hobbins, M. T., Wood, A., Morton, C., Anderson, M., & Hain, C. (2016). The evaporative demand drought index. Part II: CONUS-wide assessment against common drought indicators. *Journal of Hydrometeorology*, *17*(6), 1763–1779. https://doi.org/10.1175/jhm-d-15-0122.1

McEvoy, D. J., Pierce, D. W., Kalansky, J. F., Cayan, D. R., & Abatzoglou, J. T. (2020). Projected changes in reference evapotranspiration in California and Nevada: Implications for drought and wildland fire danger. *Earth's Future*, *8*(11), e2020EF001736. https://doi.org/10.1029/2020ef001736

Michael, Y., Helman, D., Glickman, O., Gabay, D., Brenner, S., & Lensky, I. M. (2021). Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time-series. *Science of the Total Environment*, *764*, 142844. https://doi.org/10.1016/j.scitotenv.2020.142844

Mietkiewicz, N., Balch, J. K., Schoennagel, T., Leyk, S., St. Denis, L. A., & Bradley, B. A. (2020). In the line of fire: Consequences of human-ignited wildfires to homes in the US (1992–2015). *Fire*, *3*, 50. https://doi.org/10.3390/fire3030050

Morais, T. G., Teixeira, R. F., Figueiredo, M., & Domingos, T. (2021). The use of machine learning methods to estimate aboveground biomass of grasslands: A review. *Ecological Indicators*, *130*, 108081. https://doi.org/10.1016/j.ecolind.2021.108081

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, *116*(44), 22071–22080. https://doi.org/10.1073/pnas.1900654116

Nagy, R. C., Fusco, E., Bradley, B., Abatzoglou, J. T., & Balch, J. (2018). Human-related ignitions increase the number of large wildfires across US ecoregions. *Fire*, *1*(4), 4. https://doi.org/10.3390/fire1010004

Parks, S. A., Parisien, M.-A., Miller, C., & Dobrowski, S. Z. (2014). Fire activity and severity in the Western US vary along proxy gradients representing fuel amount and fuel moisture. *PLoS One*, *9*(6), e99699. https://doi.org/10.1371/journal.pone.0099699

Pilgrim, C. (2021). Piecewise-regression (aka segmented regression) in Python. *Journal of Open Source Software*, *6*(68), 3859. https://doi.org/10.21105/joss.03859

Prechelt, L. (2002). Early stopping-but when? In *Neural networks: Tricks of the trade* (pp. 55–69). Springer.

Preisler, H. K., Brillinger, D. R., Burgan, R. E., & Benoit, J. (2004). Probability based models for estimation of wildfire risk. *International Journal of Wildland Fire*, *13*(2), 133–142. https://doi.org/10.1071/wf02061

Pullabhotla, H. K., Zahid, M., Heft-Neal, S., Rathi, V., & Burke, M. (2023). Global biomass fires and infant mortality. *Proceedings of the National Academy of Sciences*, *120*(23), e2218210120. https://doi.org/10.1073/pnas.2218210120

Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. (2017). A dual-stage attention-based recurrent neural network for time series prediction. arXiv preprint arXiv:1704.02971.

Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., et al. (2017). The fire modeling Intercomparison Project (FireMIP), phase 1: Experimental and analytical protocols with detailed model descriptions. *Geoscientific Model Development*, *10*(3), 1175–1197. https://doi.org/10.5194/gmd-10-1175-2017

Radeloff, V. C., Helmers, D. P., Kramer, H. A., Mockrin, M. H., Alexandre, P. M., Bar-Massada, A., et al. (2018). Rapid growth of the US wildland-urban interface raises wildfire risk. *Proceedings of the National Academy of Sciences*, *115*(13), 3314–3319. https://doi.org/10.1073/pnas.1718850115

Radeloff, V. C., Mockrin, M. H., Helmers, D., Carlson, A., Hawbaker, T. J., Martinuzzi, S., et al. (2023). Rising wildfire risk to houses in the United States, especially in grasslands and shrublands. *Science*, *382*(6671), 702–707. https://doi.org/10.1126/science.ade9223

Raschka, S. (2014). An overview of general performance metrics of binary classifier systems. arXiv preprint arXiv:1410.5330.

Richardson, D., Black, A. S., Irving, D., Matear, R. J., Monselesan, D. P., Risbey, J. S., et al. (2022). Global increase in wildfire potential from compound fire weather and drought. *NPJ climate and atmospheric science*, *5*(1), 23. https://doi.org/10.1038/s41612-022-00248-4

Robertson, A., & Vitart, F. (2018). *Sub-seasonal to seasonal prediction: The gap between weather and climate forecasting*. Elsevier.

Ruby, U., & Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(4), 5393–5397. https://doi.org/10.30534/ijatcse/2020/175942020

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., et al. (2019). Inferring causation from time series in Earth system sciences. *Nature Communications*, *10*(1), 2553. https://doi.org/10.1038/s41467-019-10105-3

Runge, J., Gerhardus, A., Varando, G., Eyring, V., & Camps-Valls, G. (2023). Causal inference for time series. *Nature Reviews Earth & Environment*, *4*(7), 487–505. https://doi.org/10.1038/s43017-023-00431-y

Running, S., Mu, Q., & Zhao, M. (2015). MYD17A3H MODIS/aqua net primary production yearly L4 global 500 m SIN grid V006. In *NASA EOSDIS land processes distributed active archive Center*.

Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, *263*, 110273. https://doi.org/10.1016/j.knosys.2023.110273

Schlobohm, P., & Brain, J. (2002). Gaining an understanding of the national fire danger rating system. In *National wildfire coordinating group, PMS, 932*.

Sharples, J. J. (2022). A note on fire weather indices. *International Journal of Wildland Fire*, *31*(7), 728–734. https://doi.org/10.1071/wf21134

Short, K. C. (2021). Spatial wildfire occurrence data for the United States, 1992–2018 [FPA_FOD_20210617].

Short, K. C. (2022). Spatial wildfire occurrence data for the United States, 1992–2020 [FPA_FOD_20221014].

Song, Y.-Y., & Ying, L. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*, 130.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.

Stephens, S. L., & Ruth, L. W. (2005). Federal forest-fire policy in the United States. *Ecological Applications*, *15*(2), 532–542. https://doi.org/10.1890/04-0545

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *International conference on machine learning* (pp. 3319–3328). PMLR.

Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & Van Moorsel, A. (2020). The relationship between trust in AI and trustworthy machine learning technologies. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 272–283).

Touma, D., Stevenson, S., Lehner, F., & Coats, S. (2021). Human-driven greenhouse gas and aerosol emissions cause distinct regional impacts on extreme fire weather. *Nature Communications*, *12*(1), 212. https://doi.org/10.1038/s41467-020-20570-w

Tuel, A., & Martius, O. (2023). Weather persistence on sub-seasonal to seasonal timescales: A methodological review. *Earth System Dynamics*, *14*(5), 955–987. https://doi.org/10.5194/esd-14-955-2023

Turbé, H., Bjelogrlic, M., Lovis, C., & Mengaldo, G. (2023). Evaluation of post-hoc interpretability methods in time-series classification. *Nature Machine Intelligence*, *5*(3), 250–260. https://doi.org/10.1038/s42256-023-00620-w

Turco, M., Abatzoglou, J. T., Herrera, S., Zhuang, Y., Jerez, S., Lucas, D. D., et al. (2023). Anthropogenic climate change impacts exacerbate summer forest fires in California. *Proceedings of the National Academy of Sciences*, *120*(25), e2213815120. https://doi.org/10.1073/pnas.2213815120

Van Wagner, C. E. (1974). *Structure of the Canadian forest fire weather index*. Environment Canada, Forestry Service Ottawa.

Van Wagner, C. E. (1987). Development and structure of the Canadian forest fire weather index system.

Varga, K., Jones, C., Trugman, A., Carvalho, L. M., McLoughlin, N., Seto, D., et al. (2022). Megafires in a warming world: What wildfire risk factors led to California's largest recorded wildfire. *Fire*, *5*(1), 16. https://doi.org/10.3390/fire5010016

Varner, J. M., Kane, J. M., Kreye, J. K., & Engber, E. (2015). The flammability of forest and woodland litter: A synthesis. *Current Forestry Reports*, *1*(2), 91–99. https://doi.org/10.1007/s40725-015-0012-x

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *Advances in neural information processing systems* (Vol. 30).

Walker, X. J., Rogers, B. M., Veraverbeke, S., Johnstone, J. F., Baltzer, J. L., Barrett, K., et al. (2020). Fuel availability not fire weather controls boreal wildfire severity and carbon emissions. *Nature Climate Change*, *10*(12), 1130–1136. https://doi.org/10.1038/s41558-020-00920-8

Wang, S. S. C., Qian, Y., Leung, L. R., & Zhang, Y. (2021). Identifying key drivers of wildfires in the contiguous US using machine learning and game theory interpretation. *Earth's Future*, *9*(6), e2020EF001910. https://doi.org/10.1029/2020ef001910

10.1029/2024EF004588

Wang, S. S. C., & Wang, Y. (2020). 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*(18), 11065–11087. https://doi.org/10.5194/acp-20-11065-2020

Westerling, A. L., Hidalgo, H. G., Cayan, D. R., & Swetnam, T. W. (2006). Warming and earlier spring increase Western US forest wildfire activity. *Science*, *313*(5789), 940–943. https://doi.org/10.1126/science.1128834

Williams, A. P., Abatzoglou, J. T., Gershunov, A., Guzman-Morales, J., Bishop, D. A., Balch, J. K., & Lettenmaier, D. P. (2019). Observed impacts of anthropogenic climate change on wildfire in California. *Earth's Future*, *7*(8), 892–910. https://doi.org/10.1029/2019ef001210

Williams, A. P., Allen, C. D., Macalady, A. K., Griffin, D., Woodhouse, C., Meko, D. M., et al. (2012). Temperature as a potent driver of regional forest drought stress and tree mortality.

Wu, G., Cai, X., Keenan, T. F., Li, S., Luo, X., Fisher, J. B., et al. (2020). Evaluating three evapotranspiration estimates from model of different complexity over China using the ILAMB benchmarking system. *Journal of Hydrology*, *590*, 125553. https://doi.org/10.1016/j.jhydrol.2020.125553

Yuan, K., Li, F., McNicol, G., Chen, M., Hoyt, A., Knox, S., et al. (2024). Boreal–Arctic wetland methane emissions modulated by warming and vegetation activity. *Nature Climate Change*, *14*(3), 1–7. https://doi.org/10.1038/s41558-024-01933-3

Yuan, K., Zhu, Q., Li, F., Riley, W. J., Torn, M., Chu, H., et al. (2022). Causality guided machine learning model on wetland $CH_4$ emissions across global wetlands. *Agricultural and Forest Meteorology*, *324*, 109115. https://doi.org/10.1016/j.agrformet.2022.109115

Yuan, K., Zhu, Q., Riley, W. J., Li, F., & Wu, H. (2022). Understanding and reducing the uncertainties of land surface energy flux partitioning within CMIP6 land models. *Agricultural and Forest Meteorology*, *319*, 108920. https://doi.org/10.1016/j.agrformet.2022.108920

Yuan, K., Zhu, Q., Zheng, S., Zhao, L., Chen, M., Riley, W. J., et al. (2021). Deforestation reshapes land-surface energy-flux partitioning. *Environmental Research Letters*, *16*(2), 024014. https://doi.org/10.1088/1748-9326/abd8f9

Zeng, Y., Hao, D., Park, T., Zhu, P., Huete, A., Myneni, R., et al. (2023). Structural complexity biases vegetation greenness measures. *Nature Ecology & Evolution*, *7*(11), 1790–1798. https://doi.org/10.1038/s41559-023-02187-6

Zhang, G., Wang, C., Xu, B., & Grosse, R. (2018). Three mechanisms of weight decay regularization. arXiv preprint arXiv:1810.12281.

Zhu, Q., Li, F., Riley, W. J., Xu, L., Zhao, L., Yuan, K., et al. (2022). Building a machine learning surrogate model for wildfire activities within a global Earth system model. *Geoscientific Model Development*, *15*(5), 1899–1911. https://doi.org/10.5194/gmd-15-1899-2022

Zhuang, Y., Fu, R., Santer, B. D., Dickinson, R. E., & Hall, A. (2021). 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*(45), e2111875118. https://doi.org/10.1073/pnas.2111875118

## References From the Supporting Information

Jain, P., Castellanos-Acuna, D., Coogan, S. C., Abatzoglou, J. T., & Flannigan, M. D. (2022). Observed increases in extreme fire weather driven by atmospheric humidity and temperature. *Nature Climate Change*, *12*(1), 63–70. https://doi.org/10.1038/s41558-021-01224-1

Keeley, J. E., & Syphard, A. D. (2019). Twenty-first century California, USA, wildfires: Fuel-dominated vs. wind-dominated fires. *Fire Ecology*, *15*, 1–15. https://doi.org/10.1186/s42408-019-0041-0

Knorr, W., Kaminski, T., Arneth, A., & Weber, U. (2014). Impact of human population density on fire frequency at the global scale. *Biogeosciences*, *11*(4), 1085–1102. https://doi.org/10.5194/bg-11-1085-2014

Nolan, R. H., Boer, M. M., Resco de Dios, V., Caccamo, G., & Bradstock, R. A. (2016). Large-scale, dynamic transformations in fuel moisture drive wildfire activity across southeastern Australia. *Geophysical Research Letters*, *43*(9), 4229–4238. https://doi.org/10.1002/2016gl068614

Nolan, R. H., de Dios, V. R., Boer, M. M., Caccamo, G., Goulden, M. L., & Bradstock, R. A. (2016). Predicting dead fine fuel moisture at regional scales using vapour pressure deficit from MODIS and gridded weather data. *Remote Sensing of Environment*, *174*, 100–108. https://doi.org/10.1016/j.rse.2015.12.010

Pimont, F., Dupuy, J.-L., & Linn, R. (2012). Coupled slope and wind effects on fire spread with influences of fire size: A numerical study using FIRETEC. *International Journal of Wildland Fire*, *21*(7), 828–842. https://doi.org/10.1071/wf11122

Van Der Werf, G. R., Randerson, J. T., Giglio, L., Gobron, N., & Dolman, A. (2008). Climate controls on the variability of fires in the tropics and subtropics. *Global Biogeochemical Cycles*, *22*(3), GB3028. https://doi.org/10.1029/2007gb003122

Weise, D. R., & Biging, G. S. (1997). A qualitative comparison of fire spread models incorporating wind and slope effects. *Forest Science*, *43*(2), 170–180. https://doi.org/10.1093/forestscience/43.2.170