Building a machine learning surrogate model for wildfire activities within a global earth system model

Qing Zhu1,*, Fa Li1,2, William J. Riley1, Li Xu3, Lei Zhao4, Kunxiaojia Yuan1,2, Huayi Wu2, Jianya Gong5, James Randerson3

1 Climate and Ecosystem Sciences Division, Climate Sciences Department, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
2 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China
3 Department of Earth System Science, University of California Irvine, Irvine, CA, USA
4 Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, Champaign, IL, USA
5 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

*Correspondence to Qing Zhu (qzhu@lbl.gov)
Abstract
Wildfire is an important ecosystem process, influencing land biogeophysical and biogeochemical dynamics and atmospheric composition. Fire-driven loss of vegetation cover, for example, directly modifies the surface energy budget as a consequence of changing albedo, surface roughness, and partitioning of sensible and latent heat fluxes. Carbon dioxide and methane emitted by fires contribute to a positive atmospheric forcing, whereas emissions of carbonaceous aerosols may contribute to surface cooling. Process-based modeling of wildfires in earth system land models is challenging due to limited understanding of human, climate, and ecosystem controls on fire number, fire size, and burned area. Integration of mechanistic wildfire models within Earth system models requires careful parameter calibration, which is computationally expensive and subject to equifinality. To explore alternative approaches, we present a deep neural network (DNN) scheme that surrogates the process-based wildfire model within the Energy Exascale Earth System Model (E3SM). The DNN wildfire model accurately simulates observed burned area with over 90% higher accuracy with a large reduction in parameterization time compared with the current process-based wildfire model. The surrogate wildfire model successfully captured global dynamics of wildfire burned area between years 2011 and 2015 ($R^2 = 0.93$). Since the DNN wildfire model has the same input and output requirements as the E3SM process-based wildfire model, our results demonstrate the applicability of machine learning for high accuracy and efficient large-scale land model development and predictions.
1. Introduction

Wildfires burn ~500 million hectares of vegetated land surface each year, which significantly modifies the physical properties and biogeochemical cycles of terrestrial ecosystems [Andela et al., 2017; Bond-Lamberty et al., 2007; Pellegrini et al., 2018; Randerson et al., 2006]. Living vegetation biomass, surface litter, and coarse woody debris are directly combusted and removed by wildfire [Harden et al., 2006; Walker et al., 2019]. It has been suggested that global forest would double if fire were eliminated [Bond et al., 2005]. Fire has multiple important consequences for the climate system, including directly releasing greenhouse gases (e.g., CO₂, CH₄) [Kasischke and Bruhwiler, 2002; Ross et al., 2013] and aerosols [Jiang et al., 2020]; changing land surface albedo and energy budgets [French et al., 2016; Rother and De Sales, 2020] and land-atmosphere exchanges of heat, mass, and momentum [Chambers and Chapin, 2002]; limiting plant transpiration and regional water recycling [Brando et al., 2020; Holden et al., 2018]; and reshaping forest composition [Mekonnen et al., 2019]. In addition, biomass burning emits a large amount of fine particulate matter that contributes to about 30% of cloud condensation nuclei globally [Day, 2004]. Soil organic matter decomposition, nitrogen mineralization, and the richness and diversity of soil fungal communities [Oliver et al., 2015] could also be influenced by wildfire through modifying litter substrate supply and degraded enzymatic activities [Bowd et al., 2019; Holden et al., 2018; Pellegrini et al., 2018; Pellegrini et al., 2020].

Climate change and land use activities have jointly affected fire spatial distribution, frequency, and intensity [Andela et al., 2017; Kelley et al., 2019; Xu et al., 2020] since the pre-industrial era. For example, warmer and drier climate conditions enhance fuel aridity and favor fire occurrence in forest ecosystems where fuels have built up over a period of decades and centuries [Abatzoglou and Williams, 2016; Williams et al., 2019]. Even if annual precipitation does not decline, redistribution of precipitation towards wet season extreme rainfall events could contribute to longer dry periods and thus more severe fire activity [Xu et al., 2020]. Human activities often shape wildfire activity through regulating patterns of ignition and fire occurrence (e.g., powerline ignition) [Keeley and Syphard, 2018] and suppressing wildfire activity by means of land fragmentation, fire management, and livestock grazing [Andela et al., 2017]. In California, fire density is highly associated with population density and the distance to the wildland urban interface (WUI) [Syphard et al., 2007]. At the global scale, along gradients of...
increasing population density, fire frequency initially increases by up to 20% and then gradually declines in more densely populated areas [Knorr et al., 2014]. Although global wildfire burned area has declined over the recent two decades [Andela et al., 2017], many vulnerable ecosystems and geographic regions have experienced significant increases in wildfire activity [Abatzoglou and Williams, 2016; Walker et al., 2019] resulting in large losses of natural resources and economic assets [Papakosta et al., 2017; Stephenson et al., 2013]. Over western U.S. forests, wildfire has dramatically increased, costing billions of dollars each year and gaining wide public attention. This regional wildfire increase is mainly driven by concurrent increases of spring temperature and declining snowpack [Westerling et al., 2006], mid-summer increases in vapor pressure deficit [Williams et al., 2019], and increases in drought stress during fall [Goss et al., 2020]. The enhancement of wet and dry oscillations favors initial vegetation growth and subsequent wildfire activity [Heyerdahl et al., 2002; Saha et al., 2019].

Wildfire models have played an important role in many aspects of wildfire research, including monitoring fire spread [Finney, 1998; Radke et al., 2019], analyzing controllers of wildfire short-term and long-term variability [Kelley et al., 2019], predicting severity of the upcoming fire seasons [Preisler and Westerling, 2007] and climate-scale fire variability [Girardin and Mudelsee, 2008; Yue et al., 2013], and understanding the complex climate-wildfire-ecosystem feedbacks [Clark et al., 2004; Mekonnen et al., 2019; Zou et al., 2020]. Two types of wildfire models are widely used: process-based models and data-driven statistical models. Process-based wildfire models consider detail processes related to natural fire ignition [Prentice and Mackerras, 1977], anthropogenic ignition [Venevsky et al., 2002], fire spread and duration [Thonicke et al., 2010], fire suppression [Lenihan and Bachelet, 2015], and fire mass and heat fluxes [Li et al., 2012]. Process-based wildfire models have been widely used in dynamic vegetation models and coupled earth system models (ESMs) with various complexities of parameterization [Li et al., 2019]. As more and more detailed fire processes are considered and parameterized, structural and parametric uncertainties may increase due to incomplete representation of individual processes and imperfect mathematical formulation [Riley and Thompson, 2017].

Although explicit processes are simulated, the accuracy of process-based wildfire models are highly dependent on parameterization, which is computationally expensive [Teckentrup et al., 2018; Zhu and Zhuang, 2014]. Data-driven models, however, directly link the driving
variables (e.g., climate factors) to the fire activity using simple statistical models or more sophisticated machine learning techniques, ignoring the explicit processes and feedbacks associated with wildfire [Ganapathi Subramanian and Crowley, 2018; Radke et al., 2019; Tonini et al., 2020]. Through training and validation, statistical representations of wildfire dynamics are learned by models using principles from machine learning. Data-driven wildfire models are diverse in terms of driving variables and model structure. For example, many current machine learning wildfire models rely on remote oceanic dynamics (e.g., sea surface temperature variability) and atmospheric teleconnections to simulate land surface fire activities [Chen et al., 2020; Chen et al., 2011; Yu et al., 2020]. Another group of data-driven wildfire models draws more heavily upon regional climate, plant functional type, and human infrastructure driver variables [Coffield et al., 2019; Sayad et al., 2019].

In this study, we develop a machine learning wildfire model using the process representation of wildfire in the Energy Exascale Earth System Model (E3SM) land model (ELMv1) [Zhu et al., 2019] the observationally-inferred Global Fire Emissions Database v4 (GFEDv4), and a deep neural network approach [Goodfellow et al., 2016]. We implemented a deep learning model that can better capture the complex and non-linear interactions between controlling factors and wildfire activity. The objectives of this study are to surrogate the wildfire parameterization in ELMv1 with the deep neural network and improve the model simulated wildfire burned area across various fire regions [Giglio et al., 2013].

2. Methodology

2.1 ELMv1 wildfire model

The process-based wildfire model in ELMv1 originates from the Community Land Model (CLM4.5) [Li et al., 2012]; we take this wildfire model as the baseline (hereafter refer to as BASE-Fire). BASE-Fire combines information regarding ignition, fuel conditions, surface climate, and anthropogenic suppression to simulate total burned area based on the fire counts and spread area of each fire (Figure 1). The fire count in BASE-Fire is modeled as the sum of anthropogenic ignition and natural ignition, where the latter is proportional to lightning density [Prentice and Mackerras, 1977] and the former is determined by population density [Venevsky et al., 2002]. Human activity may also intentionally suppress wildfire occurrence if the fire is detected at early stage. For example, developed regions with high population density and gross
domestic product are less likely to use fire to remove surface biomass. On the other hand, developed regions more likely suppress fire given more effective fire management policy and suppression capability. Fire count is also affected by surface fuel availability (aboveground biomass) and fuel combustibility (relative humidity, topsoil temperature and moisture). The fire spread area in BASE-Fire is modeled as an elliptical shaped region controlled by wind speed and fuel wetness (using topsoil (0 – 15 cm) moisture as a proxy). The fire duration is set to be one day based on a study that reported years 2001-2004 mean global fire persistence [Giglio et al., 2006a]. BASE-Fire also does not explicitly consider roads, rivers, and firefighting activity [Arora and Boer, 2005].

Figure 1. Schematic representation of the ELMv1 process-based BASE-Fire model and the components to be surrogated with the Deep Neural Network (DNN) model (dark grey).

2.2 Deep neural network wildfire surrogate model

We developed the new fire model in two steps: (1) surrogating BASE-Fire with a deep neural network (DNN) approach and (2) improving that surrogate model using the Global Fire Emissions Database v4 (GFEDv4 [Giglio et al., 2013]). First, we surrogated BASE-Fire with a DNN approach (hereafter refer to as DNN-Fire) that uses the same input and output variables as BASE-Fire but treats the explicit intermediate processes (e.g., ignition, fire spread) as latent variables coded by hidden layers in the DNN (Figure 1). DNN-Fire was developed with five
hidden layers and five neurons in each layer for burned area simulation. The DNN approach uses a fully-connected feedforward neural network [Schmidhuber, 2015] that comprises input, hidden, and output layers:

\[
\begin{align*}
    h_1 &= f_1(W_1 I + b_1) \\
    h_2 &= f_2(W_2 h_1 + b_2) \\
    h_3 &= f_3(W_3 h_2 + b_3) \\
    h_4 &= f_4(W_4 h_3 + b_4) \\
    h_5 &= f_5(W_5 h_4 + b_5) \\
    O &= f_6(W_6 h_5 + b_6)
\end{align*}
\]

where \( I \) denotes the input layer (e.g., climate factors) with 11 neurons, each corresponding to an input variable listed in Table 1. \( h_1, h_2, h_3, h_4, \) and \( h_5 \) are five hidden vectors that are calculated with two steps. First is a linear combination of previous layers’ input vector \( h \) and the trainable weight parameter matrix \([W_1, W_2, W_3, W_4, W_5, W_6]\), considering biases \( b_1, b_2, b_3, b_4, b_5, \) and \( b_6 \). Then, nonlinear activation functions \( f_1, f_2, f_3, f_4, f_5, \) and \( f_6 \) are applied to the output from the previous step. In this study we used softplus as the activation function [Zheng et al., 2015] that is a non-linear transformation of input signals. \( O \) denotes the output layer that summarize the latent variables from the last hidden layer \( (h_5) \) and calculate burned area.

Table 1. Input and output variables of ELMv1 BASE-Fire and surrogate DNN-Fire models

<table>
<thead>
<tr>
<th>Variable category</th>
<th>Variable name</th>
<th>Data source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tree coverage</td>
<td>LUH2</td>
<td>[Hurtt et al., 2020]</td>
<td></td>
</tr>
<tr>
<td>Fuel load</td>
<td>ELMv1 total biomass</td>
<td>[Zhu and Riley, 2015; Zhu et al., 2019]</td>
<td></td>
</tr>
<tr>
<td>Fuel conditions</td>
<td>Fuel wetness</td>
<td>ELMv1 topsoil moisture</td>
<td>[Zhu and Riley, 2015; Zhu et al., 2019]</td>
</tr>
<tr>
<td>Fuel temperature</td>
<td>ELMv1 topsoil temperature</td>
<td>[Zhu and Riley, 2015; Zhu et al., 2019]</td>
<td></td>
</tr>
<tr>
<td>Climate factors</td>
<td>Precipitation</td>
<td>GSWP3</td>
<td>[Dirmeyer et al., 2006]</td>
</tr>
</tbody>
</table>
Second, we improved the surrogate DNN-Fire by fine-tuning the weight parameters using observations (hereafter refer to DNN-Fire-GFED). Between 2001 and 2010, we initialized DNN-Fire-GFED’s weight parameters \(W_1, W_2, W_3, W_4, W_5, \text{ and } W_6\) using results from DNN-Fire, replaced the BASE-Fire burned area by GFEDv4 burned area [Giglio et al., 2013], and adjusted weight parameters until the model best reproduced the observed burned area. This two-step approach will also allow rapid parameterization of the Fire model as new fire data and baseline fire model results become available. DNN-Fire-GFED can be more easily generalized since BASE-Fire provides explicit physical guidance and a larger-than-observation input and output feature space for development of the machine learning fire model.

### 2.3 Model setup and simulation protocol

We ran ELMv1 with BASE-Fire at 1.9° by 2.5° spatial resolution [Zhu et al., 2020; Zhu et al., 2016] to generate training and testing datasets for the DNN wildfire model. BASE-Fire was first spun up for 600 years with accelerated soil decomposition followed by 200 years regular spinup with regular soil decomposition [Koven et al., 2013]. The spinup simulations were forced with constant atmospheric CO\(_2\) concentration (285 ppmv) and 1901-1920 repeated climate forcing from GSWP3 (Global Soil Wetness Project) [Dirmeyer et al., 2006]. The purpose of the spinup was to initialize ecosystem carbon pools and stabilize plant and soil carbon and water fluxes. Restarting from the spunup conditions, a transient simulation was then conducted from 1901 to 2015 with GSWP3 transient climate forcing, atmospheric CO\(_2\) concentrations, and
nitrogen and phosphorus deposition [Lamarque et al., 2005; Mahowald et al., 2008]. Wildfire
associated variables were selected for output with a monthly temporal resolution (Table 1).
BASE-Fire output from years 1981 to 2010 were used to train, test, and fine-tune
DNN-Fire. We developed 14 region-specific models, corresponding to 14 widely used GFEDv4
regions. For each region, all GFEDv4 land gridcells (comprising no fire history, infrequent fire,
and repeated fire) were concatenated into one data matrix (where rows consist of the number of
samples and columns of the number of variables). 80% of the data matrix was randomly sampled
for the training dataset and the remaining 20% of the data were reserved for testing. All training
and testing datasets were normalized to the range [0, 1] with the following scaler:

$$ X = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} $$

where $X$ is the variable vector of interest and $X_{\text{min}}$ and $X_{\text{max}}$ are minimum and maximum values of
$X$, respectively. During the training stage, we randomly initialized the weighting parameters (Eq.
1-6) and optimized them using the Adaptive Moment Estimation method [Kingma and Ba,
2014], which is a variant of the gradient descent optimization method but considers adaptive
learning rate and momentum-like exponentially decaying gradients. The parameter optimization
aimed to minimize a mean squared error cost function:

$$ J = \frac{1}{n} \sum_{i=1}^{n} (Y_{i}^{\text{DNN}} - Y_{i}^{\text{BASE}})^2 $$

where $y_{i}^{\text{DNN}}$ and $y_{i}^{\text{BASE}}$ are DNN-Fire and BASE-Fire generated burned area, respectively. Cost
function $J$ summarizes the overall magnitude of the error between the surrogate DNN-Fire and
BASE-Fire. We then evaluated model performance using metrics of mean absolute error
($\frac{1}{n} \sum_{i=1}^{n} |Y_{i}^{\text{DNN}} - Y_{i}^{\text{BASE}}|$), Pearson correlation ($\frac{\text{covariance}(y^{\text{DNN}}, y^{\text{BASE}})}{\text{variance}(y^{\text{DNN}})\text{variance}(y^{\text{BASE}})}$), and coefficient of
determination ($R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_{i}^{\text{DNN}} - Y_{i}^{\text{BASE}})^2}{\sum_{i=1}^{n} (Y_{i}^{\text{BASE}} - \bar{Y}_{\text{mean}})^2}$).

3. Results and discussion
3.1 Evaluation of wildfire surrogate model
BASE-Fire performed reasonably well for total global burned area (508 ± 53 Mha yr\(^{-1}\)
(million hectare per year) between years 2001 and 2010 compared with the GFEDv4 value of 469
± 35 Mha yr\(^{-1}\); Figure 2). BASE-Fire also captured the global declining trend of wildfire burned
area over this time period, attributed to a decrease in tropical fires [Andela et al., 2017]. At the
regional scale, however, BASE-Fire underestimated tropical (S23.5° - N23.5°) burned area and overestimated temperate (N23.5° - N67.5°) and boreal (N67.5 above) burned area (Figure 2). Large spatial heterogeneity existed for BASE-Fire regional bias. For example, over tropical GFEDv4 regions, BASE-Fire overestimated wildfire burned area over Southern Hemisphere South America (SHSA), but underestimated wildfire burned area over both Southern and Northern Hemisphere Africa regions (SHAF and NHAF), despite an overall underestimation over the tropical region (Figure 3). In contrast, consistent overestimation occurred over all temperate GFEDv4 regions. For example, wildfire burned was overestimated by about a factor of 16 (~1 versus 16 Mha yr⁻¹) over the Europe GFEDv4 region (EURO) (Figure 3). Although there is room to improve BASE-Fire performance, the parameterization would involve large ensemble simulations and computational resources. Instead, we first use BASE-Fire generated data as training and validation datasets to parameterize DNN-Fire against observed burned area.

Figure 2. BASE-Fire simulated and GFEDv4 observationally inferred burned area at (a) global scale; (b) Tropical (S23.5° -N23.5°); (c) Temperate (N23.5° - N 67.5°); and (d) Boreal (north of N 67.5°) regions.
Figure 3. A comparison of wildfire burned area between estimates from the ELMv1 process-based model (BASE-Fire) and GFEDv4 observations over 14 regions.

Next we compare DNN-Fire and BASE-Fire outputs of burned area. Using BASE-Fire generated 1.9° × 2.5° resolution datasets of surface fuel conditions (fuel load (vegetation biomass), fuel temperature (topsoil temperature), and fuel wetness (topsoil moisture)) with gridded climate forcing (GSWP3) [Dirmeyer et al., 2006], land use (LUH2 dataset) [Hurtt et al., 2020], and social economic [Dobson et al., 2000; van Vuuren et al., 2007] factors, DNN-Fire captured the spatial pattern of BASE-fire predicted wildfire activity (Figure 4). Across all GFEDv4 regions, mean absolute error of DNN-Fire was 4.4 Mha yr⁻¹ (<1% of total burn area), with median and maximum errors of 1.8 and 13.0 Mha yr⁻¹, respectively (Figure 5). Equatorial Asia (EQAS), Northern Hemisphere South America (NHSA), Central America (CEAS), and Europe (EURO) regions had the lowest DNN-Fire errors (< 1.0 Mha yr⁻¹), while Southern Hemisphere Africa (SHAF), and Boreal Asia (BOAS) had the largest errors (10-13 Mha yr⁻¹).

Overall, the correlation coefficient between BASE-Fire and DNN-Fire simulated burned area was 0.91 ($p$ value < 0.01) and the coefficient of determination ($R^2$) was 0.79. Across seasons, DNN-Fire also reasonably captured the BASE-Fire peak fire months (June to October), which were dominated by Southern Hemisphere Africa and Southern Hemisphere South America (Figure 6, orange and blue lines).

By surrogating BASE-Fire, DNN-Fire is expected to have similar biases and uncertainties. The deficiency of BASE-Fire model will propagate to DNN-Fire. In our future work we will overcome such limitation by training multiple DNN-Fire models with ensemble simulations of BASE-Fire models that differ in critical parameters and vary in model structures.
Figure 4. The performance of the Deep Neural Network wildfire model (DNN-Fire), compared with the original ELMv1 process-based wildfire model (BASE-Fire) over 14 GFEDv4 regions between years 2001 and 2010.

Figure 5. A comparison of wildfire burned area between the original ELMv1 process-based wildfire model (BASE-Fire) and Deep Neural Network wildfire model (DNN-Fire) over 14 GFEDv4 regions. Error bars represent temporal (2001-2010) standard deviation for each GFEDv4 region.
Figure 6. A comparison of wildfire burned area among Deep Neural Network wildfire model (DNN-Fire), Deep Neural Network wildfire model fine-tuned with GFEDv4 (DNN-Fire-GFED), and observations over 14 GFEDv4 regions. Error bars represent temporal (2001-2010) standard deviation for each GFEDv4 region.

3.2 Calibrating the wildfire surrogate model using GFEDv4

Although the global pattern was reasonably captured, BASE-Fire had relatively large biases in several GFEDv4 regions, as discussed above. Since DNN-Fire was trained and validated only with BASE-Fire generated inputs (e.g., fuel conditions) and outputs (burned area), we expect that, at best, DNN-Fire would have comparable biases as BASE-Fire. Starting from DNN-Fire, we further calibrated the model weighting parameters using observed burned area from GFEDv4 between years 2001 and 2010 to create DNN-Fire-GFED. We note that the conventional approach to calibrate a process-based wildfire model requires many ensemble simulations requiring large computational resources and time. Since DNN-Fire was guided by BASE-Fire, we found that parameterization time could be substantially reduced (several minutes for the global calculation).

Dramatic improvements were found in most of the 14 GFEDv4 regions simulated by DNN-Fire-GFED (Figure 6). Overall, DNN-Fire-GFED increased simulated burned area compared to DNN-Fire by 26 Mha yr$^{-1}$ (73% in terms of mean absolute error averaged across all GFEDv4 regions). Pearson correlation coefficient between the DNN-Fire-GFED simulated and GFEDv4 burned area was 0.98 ($p$ value < 0.001) with an $R^2$ of 0.97. Bias reduction was disproportionally distributed across the GFEDv4 regions (Figure 6). For example, severely burned regions, including Southern and Northern Hemisphere Africa (SHAF and NHAF) and...
Southern Hemisphere South America (SHSA) greatly benefited from the tuning and their regional biases were reduced by 94, 64, and 44 Mha yr\(^{-1}\) (or 90%, 83%, 95% reduction), respectively. Although Temperate Northern America (TENA) and Europe (EURO) wildfire burned area is relatively small (1-3 Mha yr\(^{-1}\)), the impacts of wildfire activity were significant due to their high population densities. DNN-Fire tended to overestimate the burned area in TENA and EURO by 42 and 14 Mha yr\(^{-1}\), while DNN-Fire-GFED significantly reduced biases in both regions to less than 1 Mha yr\(^{-1}\) (a 98% reduction).

BASE-Fire tended to overestimate inter-annual variability (IAV) and had opposite burned area anomalies between years 2001 and 2005. DNN-Fire dampened BASE-Fire’s IAV, but had systematic overestimation of burned area. DNN-Fire-GFED agreed well with the GFEDv4 observed IAV between years 2001 and 2010 (Figure 7). The seasonal cycle was also improved in DNN-Fire-GFED in terms of reducing BASE-Fire’s overestimation of burned area during peak fire seasons (Figure 7), although we note that DNN-Fire-GFED is biased high during low fire seasons (March and April).

![Figure 7. Inter-annual variation of burned area from years 2001 to 2010 (a) and the averaged seasonal cycle (b) of burned area estimated by the ELMv1 process-based wildfire model (BASE-Fire), Deep Neural Network wildfire model (DNN-Fire), Deep Neural Network wildfire model fine-tuned with GFEDv4 (DNN-Fire-GFED), and GFEDv4 observations.](https://doi.org/10.5194/gmd-2021-83)

3.3 Prognostic simulation and limitations

We next evaluated the DNN-Fire-GFED model against GFEDv4 for the period 2011 to 2015, using data which were not used to train and validate the model. Averaged latitudinal distribution of simulated burned area during this period showed that global wildfire activity...
peaked around $S10^\circ- S15^\circ$ and $N5^\circ- N10^\circ$, together accounting for burning 12-16% of the land surface (Figure 8). These two peaks were dominated by large burned area over Southern (SHAF) and Northern Hemisphere Africa (NHAF) fire regions. Over this period, GFEDv4 IAV was relatively larger over the Southern Hemisphere than that over the Northern Hemisphere (Figure 8 shaded area). Compared to GFEDv4, DNN-Fire-GFED simulated reasonable burned area IAV over the Northern Hemisphere but lower IAV over the Southern Hemisphere. Overall, DNN-Fire-GFED simulated $411\pm14$ Mha yr$^{-1}$ global burned area, compared with GFEDv4 observed $419\pm40$ Mha yr$^{-1}$. DNN-Fire-GFED overestimated NHAF and Central Asia (CEAS) annual burned area by 22 and 15 Mha yr$^{-1}$, respectively, while it underestimated the SHAF and Australia and New Zealand (AUST) annual burned area by 20 and 15 Mha yr$^{-1}$ (Figure 8, left panel). In summary, the prognostic DNN-Fire-GFED simulation is reasonably accurate and: (1) improved the simulated wildfire spatial and temporal distributions in ELMv1; (2) enabled effective and efficient parameterization of global fire.

**Figure 8.** Prognostic simulation of annual wildfire burned area with the Deep Neural Network wildfire model fine-tuned with GFEDv4 (DNN-Fire-GFED) compared with GFEDv4 observations averaged over 2011-2015 for (a) 14 GFEDv4 regions and (b) latitudinal distribution.

We acknowledge several challenges and limitations in our modeling framework. First, the original ELMv1 wildfire model has a unified mathematical representation of how fuel, climate, and social-economic conditions control wildfire burned area [Li et al., 2012]. However, training one single DNN wildfire model across the globe will produce a model dominated by gridcells that have high burned area (e.g., Africa). The performance of the trained DNN model, therefore, will likely have larger biases over the low fire gridcells although the globally
aggregated burned area could be reasonable. We partly overcame this challenge by applying the widely used 14 GFEDv4 fire regions that assume unique and relatively uniform dynamics over each region [Giglio et al., 2006b]. Although the regionally specific wildfire model introduces additional complexity, it better represents distinct characteristics of wildfire activity over different climate regimes and biomes [Zheng and Zhu, 2013; Zou et al., 2019] and allows for future analyses of how the relevant controllers vary across the globe.

4. Conclusions

In this study, we first surrogated the baseline ELMv1 wildfire model with a Deep Neural Network (DNN) approach (Pearson correlation coefficient = 0.91 ($p$ value < 0.01), $R^2 = 0.79$). The development was based on inputs and outputs from the baseline ELMv1 wildfire simulation, which is process-based and reasonably simulates global burned area, although regional biases existed. We then calibrated the neural network weights using the years 2001-2010 GFEDv4 observationally inferred burned area. The final calibrated DNN wildfire model (DNN-Fire-GFED) was shown to be very accurate over the 14 GFED regions. For example, reductions in absolute error over Africa, South America, and Europe were over 90%. More importantly, the DNN-Fire-GFED model global parameters could be calibrated within minutes, compared with traditional ELMv1 parameterization ensemble simulations that consume a large amount of computational time. The improved DNN-Fire-GFED model also accurately prognosted global and regional burned area in the five-year period following the training period from 2011 to 2015 (modeled 411±14 versus observed 419±40 Mha yr$^{-1}$). We conclude that the improved surrogate wildfire model (DNN-Fire-GFED) developed in this study can serve as an effective alternative to the process-based fire model currently used in ELMv1. More broadly, we conclude that machine learning techniques can facilitate earth system model development, parameterization, and uncertainty reduction with high efficiency and accuracy.

Acknowledgements

This research was supported by Energy Exascale Earth System Modeling (E3SM, https://e3sm.org/) Project and the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Scientific Focus Area, Office of Biological and Environmental Research of the U.S. Department of Energy Office of Science, Lawrence
Berkeley National Laboratory (LBNL) is managed by the University of California for the U.S. Department of Energy under contract DE-AC02-05CH11231.

Author contribution

Code availability
https://github.com/qzhu-lbl/ANN_wildfire

Data availability

Reference


Giglio, L., I. Csiszar, and C. O. Justice (2006a), Global distribution and seasonality of active fires as observed with the Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) sensors, *Journal of geophysical research: Biogeosciences*, 111(G2).


Giglio, L., G. Van der Werf, J. Randerson, G. Collatz, and P. Kasibhatla (2006b), Global estimation of burned area using MODIS active fire observations.


