BuRNN (v1.0): A Data-Driven Fire Model

Seppe Lampe[1], Lukas Gudmundsson[2], Basil Kraft[2], Stijn Hantson[3], Douglas Kelley[4], Vincent Humphrey[5], Bertrand Le Saux[6], Emilio Chuvieco[7], and Wim Thiery[1]

[1]Department of Water and Climate, Vrije Universiteit Brussel, Brussels, Belgium
[2]Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland
[3]School of Sciences and Engineering, Universidad del Rosario, Bogotá, Colombia
[4]Centre for Ecology and Hydrology, Wallingford, UK
[5]Federal Office of Meteorology and Climatology MeteoSwiss, Zürich, Switzerland
[6]Φ-lab, European Space Agency, Frascati, Italy
[7]Environmental Remote Sensing Research Group, Universidad de Alcalá, Alcalá de Henares, Spain

Correspondence: Seppe Lampe (seppe.lampe@vub.be)

Abstract. Fires play an important role in the Earth system but remain complex phenomena that are challenging to model numerically. Here, we present the first version of BuRNN, a data-driven model simulating burned area on a global $0.5° \times 0.5°$ grid with a monthly time resolution. We trained Long Short-Term Memory networks to predict satellite-based burned area (GFED5) from a range of climatic, vegetation and socio-economic parameters. We employed a region-based cross-validation strategy to account for the high spatial autocorrelation in our data. BuRNN outperforms the process-based fire models participating in ISIMIP3a on a global scale across a wide range of metrics. Regionally, BuRNN outperforms almost all models across a set of benchmarking metrics in all regions. ~~However, in the African savannah regions and Australia burned area is underestimated, leading to a global underestimation of total area burned.~~ Through eXplainable AI (XAI) we unravel the difference in regional drivers of burned area in our models, showing that the presence/absence of bare ground and C4 grasses along with the fire weather index have the largest effects on our predictions of burned area. Lastly, we used BuRNN to reconstruct global burned area for 1901-2019 and compare the simulations against independent long-term historical fire observation databases in five countries and the EU. Our approach highlights the potential of machine learning to improve burned area simulations and our understanding of past fire behaviour.

1 Introduction

Fire plays an important role in the Earth system by influencing ecosystem dynamics, biogeochemical cycles and atmospheric composition (Bowman et al., 2020). Fires drive ecosystem dynamics by affecting plant evolution (Simon et al., 2009), vegetation species composition and the physical, chemical and biological properties of soils (McLauchlan et al., 2020). Many of these ecosystem characteristics in turn also shape fire behaviour (Archibald et al., 2018). Emissions from vegetation fires affect the radiative balance of the Earth as the gases (H_2O, CO_2) trap energy through the greenhouse effect, while the aerosols reduce the amount of solar radiation that reaches Earth's surface (Bowman et al., 2009; Ward et al., 2012). Smoke of fires affects a wide range of systems including the radiative balance (Hodzic et al., 2007; Chakrabarty et al., 2023), plant fertilization (Fritze et al., 1994; Bauters et al., 2021), albedo (Beck et al., 2011; Veraverbeke et al., 2012) and air quality (Carvalho et al., 2011; Chen et al., 2017). Fires act as a big natural hazard and can also precondition post-fire hazards such as floods, landslides and large-scale erosion (Zscheischler et al., 2020; Jacobs et al., 2016; Girona-García et al., 2021; Brogan et al., 2017; Shakesby, 2011). Global observations of fire activity are typically provided by satellite products. However, these observations contain substantial uncertainties due to their spatial resolution, cloud cover and temporal resolution affecting their ability to detect small and short-lived fires. Moreover, smoke, rapid regrowth and obscuration by unburned vegetation further complicates satellite-based fire detection. Nonetheless, satellites provide the most reliable estimates of global fire activity to date. Vegetation fires burn approximately 3.5-4.5 million km^2 of surface area per year (Giglio et al., 2018; Lizundia-Loiola et al., 2020) and emit between 1.8 and 3.0 Pg Cyr^{-1} (Lizundia-Loiola et al., 2020; van der Werf et al., 2017). More recent estimates from Global Fire Emissions Database version 5 (GFED5) however suggest the amount of surface area burned per year to be around 6.5-9.5 million km^2 (Chen et al., 2023b) with an emission of 2.9-3.7 Pg Cyr^{-1} (Chen et al., 2023b), comparable to around 20-30% of the annual emissions from anthropogenic greenhouse gases (Friedlingstein et al., 2025). Fires thus play an active role in our Earth system. Yet, despite their key role, it is not fully understood and quantified how socio-economical development and climate change have affected fire occurrence in the past, and how these will affect future fire dynamics. Moreover, satellite observations suffer uncertainties due to (i) cloud cover, (ii) limited spatial resolution, which affects the detection of small fires, (iii) rapid regrowth and (iv) obscuration by unburned vegetation (Chen et al., 2023b). All of these uncertainties are propagated further into modelling efforts.

To understand how climate change and socio-econmic conditions affect vegetation fires, researchers typically model fire activity with fire-coupled Dynamic Global Vegetation Models (DGVMs) (e.g., Burton et al., 2024; Park et al., 2024). These process-based fire models simulate vegetation fires as a function of vegetation characteristics, weather, socio-economic conditions, lightning and land use (Hantson et al., 2016). Vegetation dynamics are typically supplied by the DGVM, while the other factors are provided as inputs derived from climate and integrated assessment models (Frieler et al., 2024). From these drivers, most fire models simulate ignitions (natural + anthropogenic), fuel (dry vegetation), fire spread and fire suppression, which are then transformed to fire characteristics such as burned area, fire intensity and fire emissions (Rabin et al., 2017; Li et al., 2019; Hantson et al., 2020). However, this extensive processing chain requires fine-tuning many parameterizations and formulae, each of which has the potential to alter the outcome substantially. As a result, current state-of-the-art process-based fire models are not always able to reproduce observed fire events (Burton et al., 2024; Park et al., 2024), and their projections contain substantial spread (Teckentrup et al., 2019; Lange et al., 2020; Thiery et al., 2021; Grant et al., 2025). Moreover, (sub)national fire databases are often incomplete and inconsistent (Bowman, 2018; Gincheva et al., 2024)

Machine learning algorithms have the advantage of being able to fit (non-linear) functions to data rather than prescribing them manually. In complex tasks, such as fire modelling, where the real world relations and interactions are hard or near-impossible to pin down mathematically, machine learning can provide a valuable solution (Qi and Majda, 2020; Bracco et al., 2025). At the same time, machine learning often lacks interpretability (Rudin, 2019; Yang et al., 2024; Bracco et al., 2025), which can be a disadvantage compared to process-based models when process understanding or fine-grained control is the primary objective. Thus, machine learning can serve as a complementary rather than a substitutive approach to process-based fire modelling.

Here we present a data-driven fire model "BUrned area modelling by Recurrent Neural Networks (BuRNN)". BuRNN combines traditional fire model inputs and intermediary DGVM outputs such as Gross Primary Production (GPP) with machine learning to predict burned area. We first describe the architecture and training process of the model. Then, we evaluate the skill of BuRNN against satellite data, using state-of-the-art process-based wildfire models as benchmark. Next, we attempt to understand the inner workings of BuRNN through XAI methods. Finally, we apply BuRNN to generate a monthly gridded burned area reconstruction from 1901 to 2019 at 0.5° × 0.5° spatial resolution and evaluate this new dataset against regional wildfire records.

2 Materials & Methods

2.1 Data

To train BuRNN, we make use of ~~six~~ five different data sources. BuRNN is trained on a monthly timescale and receives ~~27~~ 24 features as input, each providing information on (i) climate, (ii) land or vegetation properties or (iii) socio-economic conditions (Table 1). Climate-related variables are: (i) monthly mean of the daily maximum temperature, mean monthly precipitation and mean monthly wind speed from the daily NOAA-CIRES-DOE 20th Century Reanalysis version 3 homogenized to W5E5 (20CRv3-W5E5) product (Compo et al., 2011; Slivinski et al., 2021; Lange, 2019; Lange et al., 2021), (ii) monthly mean Fire Weather Index (FWI) calculated from 20CRv3-W5E5 ~~,~~ and (iii) ~~Standardised Precipitation-Evapotranspiration Index (SPEI) with a 1, 3 and 6 month time lag calculated through the FAO-56 Penman-Monteith estimation for potential evapotranspiration (Beguería et al., 2014) and (iv)~~ lightning density. The land and vegetation characteristics are (i) land cover from the Community Land Model (CLM), which are generated based on Land Use Harmonization phase 2 (LUH2; Hurtt et al., 2020), (ii) land use provided by Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) also based on LUH2 and (iii) intermediate DGVM outputs (ensemble mean) from the ISIMIP biome sector for GPP (n=7), Carbon Mass in Vegetation (cVeg) (n=3) and Leaf Area Index (LAI) (n=5) Table A1. Lastly, socio-economic conditions are provided by ISIMIP in terms of population densities and Gross Domestic Product (GDP) (Table 1). The LUH2 derived data from ISIMIP and CLM was linearly interpolated from a yearly to monthly timescale. Moreover, we removed and grouped a number of related land use/land cover classes in order to bring the total number of features down. ~~Additionally, earlier research has pointed out that the driving factors for savannah fires are different on different continents (Lehmann et al., 2014; Alvarado et al., 2020; Simpson et al., 2022). Therefore, we also split the original C4 grasses variable into three additional variables based on their location i.e., (i) North-America and South-America (C4 grasses Americas), (ii) Africa (C4 grasses Africa) and (iii) Europe, Asia and Oceania (C4 grasses Eurasia_Oceania).~~ We chose these input variables as all are available on a monthly timescale from 1901 onwards at a 0.5° × 0.5° spatial resolution (or higher) and represent many drivers, or proxies thereof, of fire behaviour. To train BuRNN, we use GFED5 as target data (Chen et al., 2023b), we remapped the original 0.25° × 0.25° grid to 0.5° × 0.5° using area-weighted regridding from the Python Package *Iris - SciTools*. GFED5 derives burned area estimates for 2001–2020 from the Moderate

Resolution Imaging Spectroradiometer (MODIS) MCD64A1 product (Giglio et al., 2018), applying region-, land cover-, and tree cover-specific corrections for commission and omission errors based on spatiotemporally aligned Landsat and Sentinel-2 burned area observations. Burned area in croplands, peatlands, and deforestation regions is separately estimated using MODIS active fire detections (Giglio et al., 2016). To extend the record back to 1997, active fire data from the Along-Track Scanning Radiometer (ATSR) and the Visible and Infrared Scanner (VIRS) were used, which carry higher uncertainties (Chen et al., 2023b). Although GFED5 almost doubles the observed burned area compared to other satellite products, we consider it most suitable for ground truth as it matches high-resolution burned area observations for Africa (Chuvieco et al., 2022). Moreover, literature suggests that 'traditional' burned area products, such as FireCCI51 severely underestimate actual burned area (Zhu et al., 2017; Franquesa et al., 2022; Khairoun et al., 2024), supporting our choice for GFED5 as target dataset.

2.2 Model Description

We aim to design a machine learning model that is able to learn the lagged and cumulative effects of climate variability, land use and socio-economic conditions on fire dynamics. Unlike traditional machine learning algorithms, which often treat each observation independently, Long Short-Term Memorys (LSTMs) are capable to capture non-linear temporal dependencies in sequential data (Hochreiter and Schmidhuber, 1997), making them ideal for our use case. Although LSTMs were originally designed for natural language processing (Gers et al., 2000), LSTMs have also successfully been applied in a number of climate related applications such as modelling vegetation dynamics (Reddy and Prasad, 2018), predicting river streamflow (Hunt et al., 2022), weather forecasting (Karevan and Suykens, 2020) and even detection of forest fires (Cao et al., 2019). Therefore, we chose the LSTM as main component of BuRNN. The LSTM maintains its own hidden states acting as *memory*, which is updated dynamically in interaction with the input features. The hidden state at each time step is mapped to ~~a single output~~ three outputs using a dense neural layer~~, yielding the predicted burnt area fraction.~~. The first of the outputs is be used as a binary classifier, determining whether it burns or not. The second and third represent parameters (mean and variance) of the modelled burned area distribution. Predicted burned area is constructed via Eq. (2), assuming a normal distribution. Despite the simple model architecture, a couple of hyperparameters have to be chosen. To automate the search for optimal hyperparameters, we used the *Optuna* framework (Akiba et al., 2019). We used the Tree-structured Parzen Estimator (TPE) sampler inside the framework to find appropriate values for the learning rate, number of LSTM layers, hidden size of the LSTM layer(s), activation functions, number of dense neural layers, size of the dense neural layers and dropout fraction (Bergstra et al., 2011). Currently, BuRNN is a single layered LSTM with a hidden size of 64 connected to a dense neural layer with Rectified Linear Unit (ReLU) activation function given in Eq. 1 (see also Fig. 1):

$$\mathrm{ReLU}(x) = \max(0, x) \tag{1}$$

$$\hat{y} = \mathbf{1}\{\hat{p} > 0.5\} \cdot (\mu + \tfrac{1}{2}\sigma^2) \tag{2}$$

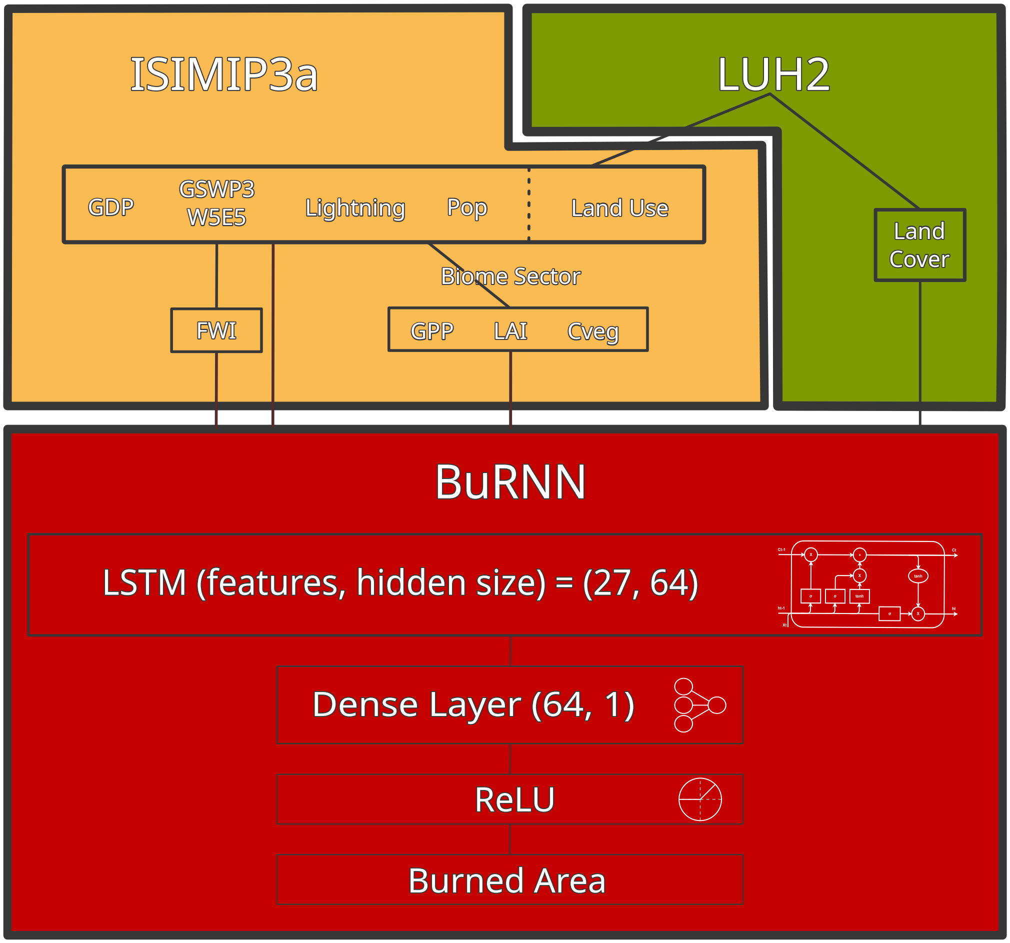

Figure 1. Structure of BuRNN. The top row denotes the origin of all the features supplied to our model, split into the ~~three~~ main sources~~(ISIMIP3, LUH2, and SPEIbase)~~. The red rectangle reflects the architecture of BuRNN.

Given the nature of our data, our input variables (and targets) contain a high degree of spatial autocorrelation. Applying a traditional random train-test split or random train-test folds would likely lead to an overestimation of performance and poor predictive power (Diniz-Filho et al., 2008; Le Rest et al., 2014; Meyer et al., 2019). Therefore, we trained our LSTM networks according to a region-based cross-validation. We split our data according to 43 Intergovernmental Panel on Climate Change (IPCC) land regions (we removed the two Antarctic regions and Greenland) and manually grouped these regions into 11 folds (Fig. A1), whereby we made sure that the 3-4 regions in each fold represent different continents and biomes (Iturbide et al., 2020). For each fold, we use two different folds as validation set and the remaining 8 folds as training set. We repeat this five times for each fold, each time with two different folds as validation set. For example, when fold 1 is chosen as test fold, we first select folds 2 and 3 as validation set and folds 4-11 as training set. Then we choose folds 4 and 5 as validation set and folds 2-3 and 6-11 as training set, etc. This results in a total of 55 (11 times 5) models. Then, when we make predictions with our model for an IPCC region, it is the mean estimate of five LSTMs which have never seen data for that IPCC region before. The validation folds are used to monitor model convergence and overfitting by using the early stopping algorithm; As soon as model performance on these validation folds started to decrease after a given set of training iterations, training was stopped and the best model was restored. This model was then used to make predictions on the independent test set.

Before training, we combine the data from all different sources, convert the time dimension to have identical units and split them into the 11 pre-defined folds. We normalize the training data and use the mean and standard deviation of the training set to normalize the validation folds (and the test fold during prediction). Each time we change the training folds, we undo the normalization operation, and redo it based on the mean and standard deviation of the new training set. Additionally, we log-transform the target variable (GFED5 percentage burned area) as the original data is strongly right-skewed ~~.~~ We do this by applying the natural logarithm of one plus the target (log1p). After this, we normalise the targets by subtracting the mean and dividing by the standard deviation. Pre-processing of our data happens through a combination of *Xarray* and *NumPy*. We use *PyTorch* and *PyTorch Lightning* to build our model architecture and to handle training and validation (Paszke et al., 2019; Falcon and The PyTorch Lightning team, 2019). In the training phase the LSTM layer is followed by batch normalization. During training, we provide the samples in batches of 32, an error is calculated based on the cumulative error of the predictions for these 32 samples (see further) after which the model is updated/improved. Batch normalization normalizes the features of each batch (based on the batch's mean and standard deviation) and results in faster and more stable training (Santurkar et al., 2018). This layer is followed by a dropout layer for which the optimal dropout fraction was found to be 0.2. This randomly ignores, on average, 20% of the connections between the LSTM and dense layer, which has been proposed to improve generalisation and reduce overfitting (Srivastava et al., 2014). During training we ignore the first 36 predictions (3 years) to allow the LSTM's memory state to spin up and then evaluate the predictions of the following 3 years using ~~Mean Squared Error (MSE) as~~ a custom loss function. In each epoch, we pass each ~~pixel/location~~ gridcell in the training set once and randomly select a 6 year time slice. The spinup period of 3 years was chosen based on fire-process understanding and the prediction length of 3 years was chosen in function of model convergence speed. The loss function expects three outputs from the model, the first is used as a binary classifier (will it burn or not) and is scored through binary cross entropy (Eq. (3)), the second represents the mean (log-transformed and normalised) burned area and the third is the (log-transformed and normalised) variance. This mean and variance are used in a Gaussian negative log likelihood loss (Eq. (4)). Then, the Gaussian negative log likelihood loss is multiplied by 1000 so it reaches a similar magnitude as the binary cross entropy loss, after which both loss terms are added up. After training, the normalisation and log-transform can be inverted to obtain predictions in fraction of burned area per cell again.

$$\mathcal{L}_{\mathrm{BCE}}(y,\hat{p}) = -\left[y\log(\hat{p}) + (1-y)\log(1-\hat{p})\right]. \tag{3}$$

Here, $y \in \{0,1\}$ is the binary target (fire occurrence) and $\hat{p} \in (0,1)$ is the predicted probability of fire occurrence.

$$\mathcal{L}_{\mathrm{NLL}}(y,\mu,\sigma^2) = \frac{1}{2}\left[\log(\sigma^2) + \frac{(y-\mu)^2}{\sigma^2}\right]. \tag{4}$$

Here, y is the observed value, μ is the predicted mean, and σ^2 is the predicted variance of a Gaussian distribution.

2.3 Model Evaluation

We evaluate our predictions for ~~1997-2019~~2003-2019, the common period between the ~~observational GFED5 product~~ full availability of Terra/Aqua in MODIS and the ISIMIP fire sector simulations (forced with the GSWP3-W5E5 reanalysis). We evaluate our 3D (time, latitude, longitude) data cubes for several metrics in different dimensions (spatial, temporal and spatio-temporal). By calculating the Root Mean Squared Error (RMSE) between the modelled and observed 3D cubes, we obtain an error expressed in % burned area. Similarly, by calculating the Pearson correlation we obtain a metric that informs on spatial and temporal patterns, ignoring the mean and scale bias the process-based models and BuRNN have (Hantson et al., 2020; Burton et al., 2024). The spatial pattern is evaluated by computing the mean over time, resulting in a 2D data cube (latitude, longitude), and we calculate both spatial RMSE and correlation. Similarly, by taking the sum over the spatial domain (latitude and longitude), we arrive at a monthly and yearly time series of global burned area. We calculate yearly correlation, which assesses the interannual variability, and monthly correlation, which represents seasonality.

2.4 Driver Analysis

To better understand the inner workings of BuRNN, which is *in se* a black box model, we employ an explainable AI method. ~~SHapley Additive exPlanations (SHAP) is a method from game theory that investigates the effects features have on the prediction outcomes of machine learning models (Lundberg and Lee, 2017). SHAP assigns each feature a score that represents its contribution to the difference between the actual prediction and the average prediction over a dataset. It provides a consistent way to quantify how much each feature pushes the prediction higher or lower. We randomly select 50 samples for each GFED region for each model and calculate SHAP values over 24 months (~~Integrated Gradients (IG) is an attribution method for differentiable models, like LSTMs, that quantifies the contribution of each input feature to a specific prediction. IG compares the prediction at an input x to the prediction with a reference baseline input x_0 and integrates the model's gradients along a straight-line path between them (Sundararajan et al., 2017b). Here, we applied the global mean for each feature as baseline. Thus, the IG results need to be interpreted as 'How strong does each feature affect burned area in this region compared to the global mean of this feature'. A caveat of this approach is that when a feature in a region tends to be close to the global mean, then attribution for that feature will be low as the integration between sample and baseline will be performed over a short path. Moreover, our approach does not inform on the direction of influence as the direction can vary based on the timing of the feature. For example, precipitation a year before a fire can actually increase burned area by stimulating vegetation growth and increasing future fuel loads, but precipitation right before a fire typically negatively affects burned area. As to not average these two effects out, we take the absolute value of each attribution and thus only look at how important each feature is, not at the actual effect (positive or negative) of each feature. Lastly, highly correlated features will have their attributed importance spread across each other and hence be lower than if only a single of these features was provided. For each of the 55 LSTMs, we pass it the test data of 2002-2008 and attribute the predictions of 2005-2008 (using 2002-2004 as spinup period; see Section 2.2) ~~. We use GradientExplainer from the aptly named SHAP Python package to calculate the SHAP values. GradientExplainer calculates the gradients of the input features (how much does the output of~~ and store this per GFED region. The total number of attributions is the multiplication of the number of models per gridcell (n=5) by the number of land gridcells in the dataset (n=65797) by each predicted timestep (n=48, since we don’t attribute the 3 years spinup i.e., 2002-2004) by the number of features (n=24) and by each considered timestep in the attribution. For the latter, we consider the previous 3 years and the features of the ~~model change when each input is changed slightly) and approximates the SHAP values by integrating the gradients over the input features, moving from the average input (mean of each feature in the dataset)to an actual sample in the dataset. The GradientExplainer approach is an extension of integrated gradients proposed originally in Sundararajan et al. (2017a)~~predicted timestep itself (n=37). Resulting in a total of 14 billion attributions, or around 580 million attributions per feature. We take the absolute value of these attributions and average this per region per feature. We note here upfront that ~~SHAP~~ IG does not provide causal insights into the real-world processes underlying the data. ~~Rather, SHAP values offer~~ Instead, IG offers a post hoc explanation of the model~~'~~’s internal logic by attributing contributions to input features in a way that reflects the model~~'~~’s learned associations. When applied to structured or interdependent data, ~~SHAP~~ IG values can be ~~particularly~~ difficult to interpret ~~and may not yield a complete picture~~ because feature dependencies may obscure how importance is distributed, and the method may not capture the full complexity of how the model uses such inputs. ~~Nonetheless, they can still offer~~ Nevertheless, IG can still provide a useful high-level ~~perspective on~~ view of the patterns and dependencies the model has ~~captured. This analysis therefore serves the purpose of understanding the statistical associations the model has learned, rather than uncovering mechanistic relationships inherent to the system under study~~learned. Thus, our analysis aims to characterize the statistical associations encoded by the model rather than to infer mechanistic relationships in the underlying system.

2.5 Burned Area Reconstruction

After training, we use the models to simulate burned area for the period 1901-2019. During training we employed a 3 year spinup period (see Section 2.2). Therefore, we add 1901-1903 in front of the dataset so this can be used as spinup. We analyse this full reconstruction per region and also compare it against a 1997-2019 run to verify the stability of the model.

Moreover, we compare the reconstruction to the FireCCiLT11 product, which is based on Advanced Very-High-Resolution Radiometer (AVHRR; Otón et al., 2021). FireCCiLT11 is available from 1982-2018, with the exception of 1994. We calculate the regional 1982-1993 correlations for annual burned area between BuRNN and FireCCiLT11 and compare those to the 1997-2018 correlations between BuRNN and FireCCiLT11 and between GFED5 and FireCCiLT11. Ideally, the latter values are high, indicating both observational products are in agreement. If this is the case, then a good reconstruction (1982-1993) should have a similar correlation (to FireCCiLT11) for both periods.

Additionally, we compare the reconstruction to regional datasets where available (see subsection 3.3). For Canada, we assess the National Burned Area Composite (NBAC) and National Fire Database (NFDB) datasets. NBAC is fire polygon database from Landsat (30m) starting from 1972 and contains data on $\sim$35000 fires (Canadian Forest Service, 2024). NFDB combines data from various Canadian agencies and contains data for over 700 000 fires between 1959 and 2022 (Hanes et al., 2019). For the Unites States, we compare our reconstruction to Monitoring Trends in Burn Severity (MTBS) and Fire Occurrence Database (FOD). MTBS estimates burned area from Landsat and provides data on fires >2km^2 since 1984 (Picotte et al., 2020). FOD encompasses fire records from several US agencies for 1992 to 2020 and excludes prescribed burning (Short, 2022). For Brazil we use data from the MapBiomas project, which produces gridded burned area over Brazil from 1985 to 2023 based on Landsat (Souza Jr et al., 2020). For Chile, the database is managed by Chilean Forest Service (CONAF) and is also based on Landsat, it contains information on over 200 000 fires from 1985 to 2021. We obtained the Chilean data from Gincheva et al. (2024). European Forest Fire Information System (EFFIS) provides us with country-level data on non-agricultural fires for 21 countries in the EU (excluding Austria, Belgium, Denmark, Ireland, Luxembourg and Malta). The data comes from the individual EU countries and is available for different time periods for each country, the earliest is 1980 for Portugal. Lastly, we also asses fires over Australia, making use of data from over 75% of the Australian surface area. Data was provided by different state and territory agencies and was combined by Gincheva et al. (2024) and is available from 1950 to 2021. All these datasets come with a number of caveats, especially in the earlier periods. They are (i) often incomplete, (ii) use different protocols between products, but also for different time periods within a dataset and (iii) they report different things (some exclude agricultural and/or managed fires, others exclude small fires) (Gincheva et al., 2024). Nonetheless, they are the best independent reference data we have available.

Table 1. List of the ~~27~~ 24 features provided to BuRNN along with their origin.

Type	Source	Description	Number of Features
Climate	20CRv3-W5E5	We aggregate the daily values for daily maximum temperature (tasmax; in K), total precipitation (pr; in kg m^{-2} s^{-1}) and near-surface wind speed (sfcWind; in m s^{-1}) to monthly means.	3
		Canadian FWI calculated on a daily timescale from tasmax, pr, tasmax and near-surface relative humidity (hurs; in %) (van Wagner, 1987). These daily values are then aggregated to monthly means through CDO.	1
	~~SPEIbase 1, 3 and 6-month lagged SPEI from Beguería et al. (2014) with potential evapotranspiration calculated via FAO-56 Penman-Monteith estimation. 3 -~~ HistLight & WGLC	Lightning density provided by combining HistLight (1901-2009) and WGLC (2010-2019) (Kaplan and Lau, 2022a, b).	1
Land & vegetation	CLM	Land cover maps originating from LUH2 (Hurtt et al., 2020) and processed for use as input to the Community Land Model (CLM, Lawrence and Chase, 2007; Lawrence et al., 2019). We regrouped the original 17 land cover types into 11 groups (all represented as fraction of grid cell area): Urban, Lake, Crop, Bare Ground, Needleleaf tree, Broadleaf evergreen tree, Broadleaf deciduous tree, Broadleaf shrub - temperate, Broadleaf deciduous shrub - boreal, C3 grass and C4 grass. ~~Then we added three additional variables for C4 grasses i.e., C4 grasses Americas, C4 grasses Africa and C4 grasses Europe-Asia-Oceania.~~	~~14~~ 11
	ISIMIP	Land use maps originating from LUH2 and processed for use in ISIMIP (Volkholz and Ostberg, 2022). Given the similarity between the land cover and land use datasets, only the grid cell fractions managed pastures and rangeland were added to the feature list.	2
		ISIMIP ensemble mean of Leaf Area Index (LAI; n=5), Gross Primary Production (GPP; n=7) and Carbon stored in Vegetation (CVeg; n=3).	3
Socio-economic	ISIMIP	Rural and urban population along with GDP from ISIMIP3a (Volkholz et al., 2024; Sauer et al., 2024) .	3

3 Results

3.1 Model Evaluation

Our global-scale evaluation results highlight that BuRNN outperforms all process-based fire models on each of the skill metrics we consider (~~see Section 2.3, except for interannual variability, where only CLASSIC outperforms BuRNN; Fig. 2)~~Fig. 2) with respect to GFED5 and in all but one metric with respect to FireCCI51 (Fig. A7). Moreover, for spatial RMSE, spatial correlation and monthly correlation its performance falls in the inter-observational uncertainty. This implies that BuRNN's performance for these metrics is indiscernible from observational products and that further improvement is meaningless until inter-observational uncertainty is decreased. BuRNN has a RMSE of 1.59, while the process-based fire models fall between ~~2.02 and 3.07~~1.92 and 3.00 and inter-observational uncertainty is 1.18. Similarly, the correlation factor is 0.7~~and~~ between 0.01 and 0.51 for the FireMIP models and 0.85 between GFED5 and FireCCI51. The spatial RMSE of BuRNN is 0.49, ~~while~~ the process-based models fall between 0.82 and 1.32 and the inter-observational spatial RMSE is 0.49. The spatial correlation is 0.88 for BuRNN~~and~~, between -0.01 and ~~0.69~~ 0.67 for the FireMIP models and 0.90 for GFED5 and FireCCI51. Monthly correlation, representing seasonality, is 0.86 for BuRNN~~and between -0.13 and 0.66~~, between -0.12 and 0.73 for the FireMIP models and 0.87 for FireCCI51 and GFED5. BuRNN's yearly correlation, representing interannual variability, is 0.87, while it is between ~~-0.34 and 0.80~~ -0.37 and 0.76 for the process-based models (Fig. 2) and 0.94. Three example maps of burned area prediction by BuRNN are shown alongside those of GFED5 and the two best-performing process-based models in Figs. A4 to A6. Hence ~~, with one exception (yearly correlation of the CLASSIC model)~~ BuRNN scores better than any other fire model for each considered performance metric (total of 54 model-metric combinations). These evaluation results thus overall indicate that at the global scale, BuRNN largely outperforms state-of-the-art global wildfire models. Fig. 4 depicts the mean monthly burned area from GFED5 (upper left), BuRNN (upper right) and the nine FireMIP models. In general, the spatial pattern of BuRNN matches closely the pattern of GFED5. ~~However, there is a clear underestimation of burned area in the Northern Australian savannahs and mainland Southeast Asia.~~ This is made further clear in Fig. 5, which shows the difference in mean monthly burned area between BuRNN and GFED5 (upper right) and between the FireMIP models ~~between the FireMIP models~~ and GFED5. The density plot in the upper left depicts the distribution of the error over all land pixels for BuRNN and the FireMIP models, where the difference between GFED5 and each of the models is considered the error. The distribution of BuRNN falls more closely around zero than any of the FireMIP models, indicating again better spatial performance.

~~Global evaluation scores of BuRNN and the FireMIP models. Colour scaling has been done based on the normalized values (value - row mean)/(row standard deviation) with the minimum and maximum values set to -2 and 2, respectively. Better scores (lower for RMSE and higher for Pearson correlation) are marked in blue, while worse performance is in red.~~

We also evaluate our results across 14 fire regions defined by Giglio et al. (2010) in Fig. 6 ~~. We find a general tendency to underestimate the annual burned area in most regions.~~ and **??.** Interannual variability is relatively well modelled, although the amplitude is lower than observed for ~~most~~ some regions. However, there are differences in performance across regions. Regions

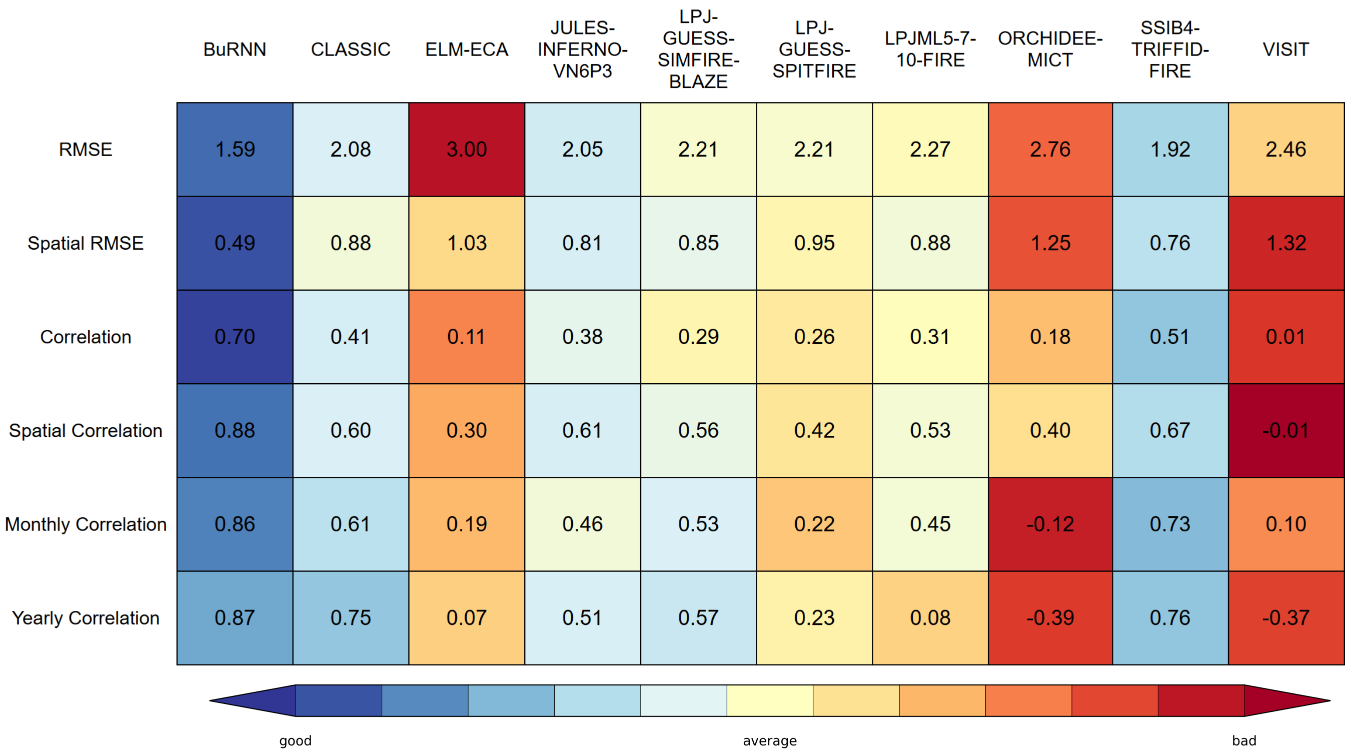

Figure 2. Global evaluation scores of BuRNN and the FireMIP models for 2003-2019. Colour scaling has been done based on the normalized values (value - row mean)/(row standard deviation) with the minimum and maximum values set to -2 and 2, respectively. Better scores (lower for RMSE and higher for Pearson correlation) are marked in blue, while worse performance is in red.

~~BuRNN~~

~~CLASSICELM-ECAJULES-INFERNO-VN6P3LPJ-GUESS-SIMFIRE-BLAZELPJ-GUESS-SPITFIRELPJmL5-7-10-fireORCHIDEE-MICTSSiB4-TR~~

~~RMSE 1.67 2.18 3.07 2.16 2.30 2.32 2.37 2.83 2.02 2.56 Spatial RMSE 0.50 0.87 1.02 0.82 0.85 0.96 0.89 1.24 0.75 1.32 Correlation 0.70 0.40 0.11 0.37 0.29 0.26 0.30 0.18 0.51 0.01 Spatial Correlation 0.89 0.61 0.32 0.62 0.58 0.43 0.53 0.40 0.69 -0.01 Monthly Correlation 0.81 0.55 0.11 0.39 0.49 0.18 0.35 -0.13 0.66 0.11 Yearly Correlation 0.72 0.80 0.07 0.29 0.57 0.38 -0.01 -0.34 0.69 -0.21~~

such as Temperate North America (TENA), Northern Hemishphere South America (NHSA) and ~~Equatorial Asia (EQAS)~~ Southern Hemisphere Africa (SHAF) are excellently modelled by BuRNN. In the majority of the regions, BuRNN captures the pattern of the interannual variability well~~, but consistently underestimates the amplitude and total burned area, for instance in Boreal North America (BONA), Central America (CEAM), Southern Hemishphere South America (SHSA), Europe (EURO), Northern Hemisphere Africa (NHAF), SHAF and Southeast Asia (SEAS)~~. In the Middle East (MIDE), BuRNN simulates the mean annual burned area well, while the interannual variability ~~is offespecially for the later years~~and long-term trend are off. In Boreal Asia (BOAS), our model simulates too little burned area, which is likely due to having a similar environmental setting as BONA where annual burned area is much lower. In Central Asia (CEAS) ~~, our simulations do not match the observed interannual variability or mean annual burned area. In Australia and New Zealand (AUST), total burned area is much lower than observed, while~~ and AUST, interannual variability is ~~reasonable in terms of general pattern (which years have high or low burned area), but again too low in amplitude~~reasonably modelled, but the highest burning years are underestimated e.g., 2001-2008 in CEAS and 1998-2002 in AUST. Global annual burned area is mostly dominated by the (African) savannah regions; ~~at this scale the discrepancy between observed and simulated global burned area is therefore mainly due to the underestimation of BuRNN in NHAF, SHAF and AUST. Nonetheless, in~~ therefore the ability of BuRNN to capture mean burned area, interannual variability and long-term trend is reflected in the good global performance Table A2. In most regions BuRNN ~~again~~ outperforms the process-based fire models over most metrics (Fig. 3). Lastly, we compare the distributions of observed and modelled burned area Fig. 9. We note that the rare high burned areas (>50% of land surface area) are generally not modelled by BuRNN.

Next, we repeat this evaluation procedure using the 2001-2019 FireCCI51 observational dataset as reference. We do this because our model is specifically trained to predict GFED5 burned area, while the process-based models are not. Although the absolute values between FireCCI51 and GFED5 differ, a similar pattern as Fig. 3 is observed when comparing BuRNN and the process-based models against FireCCI51 ~~, that is , BuRNN~~ (Fig. A8). BuRNN tends to outperform the process-based models, although less strongly than before. Especially the in the 3D RMSE BuRNN is often not the best performing model anymore. This makes sense as total burned area in FireCCI51 is about half of GFED5 so BuRNN is expected to make larger errors. In the correlation metrics however, BuRNN still clearly outperforms the process-based models ~~in most regions and for most metrics (Fig. A8)~~most of the time. There are two regions/metrics for which the FireCCI51 and GFED5 observational products show notable differences (Table A3). First the yearly correlation for SEAS between the two observational products is only 0.42. Second, the 3D correlation in CEAM is only 0.58, notably lower than all other regions and similar to the 3D correlation in EQAS. Therefore, any (dis)similarity of any model with any observational product should be taken with relatively large observational uncertainty in mind.

RMSE metrics vary in magnitude across different regions as they have different total burned areas. However, also the correlation metrics show large inter-regional differences. For example, in MIDE~~BuRNN outperforms all process-based models with a~~ , BuRNN has a very low yearly correlation of ~~0.42. In EQAS, six~~ 0.13. However, only a single process-based model

	AUST	BOAS	BONA	CEAM	CEAS	EQAS	EURO	MIDE	NHAF	NHSA	SEAS	SHAF	SHSA	TENA
RMSE	2.50	1.18	0.46	0.98	1.18	0.66	0.39	0.33	3.17	0.75	2.48	3.31	1.00	0.43
Spatial RMSE	0.64	0.27	0.10	0.39	0.27	0.33	0.11	0.13	0.99	0.31	0.90	1.07	0.43	0.13
Correlation	0.32	0.46	0.12	0.53	0.33	0.39	0.34	0.20	0.78	0.65	0.60	0.76	0.45	0.14
Spatial Correlation	0.71	0.70	0.30	0.55	0.63	0.41	0.64	0.35	0.90	0.77	0.60	0.88	0.55	0.30
Monthly Correlation	0.78	0.88	0.36	0.89	0.76	0.89	0.73	0.74	0.99	0.96	0.93	0.97	0.94	0.72
Yearly Correlation	0.87	0.66	0.49	0.85	0.67	0.95	0.69	0.13	0.80	0.81	0.73	0.67	0.76	0.87

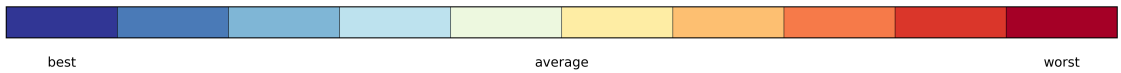

Figure 3. Regional evaluation scores of BuRNN. Colour scaling has been done based on the ranked values compared to the nine process-based fire models, with the minimum RMSE and maximum correlations coloured blue (best) and the highest RMSE and lowest correlation coloured red (worst).

~~AUSTBOASBONACEAMCEASEQASEUROMIDENHAFNHSASEASSHAFSHSATENA RMSE2.67 1.20 0.41 1.25 1.16 0.64 0.40 0.30 3.51 0.68 2.55 3.43 0.88 0.38 Spatial RMSE0.96 0.30 0.06 0.37 0.27 0.20 0.11 0.11 1.01 0.25 0.88 1.05 0.29 0.10 Correlation0.33 0.43 0.18 0.52 0.35 0.37 0.32 0.19 0.77 0.63 0.59 0.74 0.53 0.15 Spatial Correlation0.57 0.69 0.49 0.63 0.64 0.47 0.63 0.36 0.93 0.81 0.64 0.89 0.67 0.38 Monthly Correlation0.84 0.90 0.55 0.83 0.76 0.88 0.75 0.79 0.97 0.93 0.94 0.99 0.95 0.73 Yearly Correlation0.91 0.63 0.74 0.79 0.70 0.89 0.50 0.42 0.74 0.73 0.54 0.63 0.74 0.66~~

scores better in this metric. In SEAS, four out of the nine fire models outperform BuRNN in yearly correlation, but BuRNN still has a high ~~absolute score of 0.89~~ correlation of 0.73 in this region. ~~In general, the higher the average monthly burned area in a region, the better/easier predictions are for that region. This partially explains why regions with high burned areas, such as NHAF and SHAF have high monthly correlations as opposed to BONA and TENA where monthly correlations between modelled and observed burned area are lower.~~ Similar observations can be made over the spatial correlation, where NHAF and SHAF are ~~again~~ the regions with the best modelled spatial burned area and BONA and TENA the two regions where the spatial pattern is least well modelled of all regions~~, although the difference in correlation between best and worst region is smaller than with monthly correlations~~. The likely reason for the lower spatial correlation (both for BuRNN and the process-based models) in these regions is the stochastic nature of fires on these spatial and temporal scales. For example, large regions (many pixels) of Canadian forest are quasi-identical in terms of how their monthly input features look like. In these regions large fires are associated with periods of high fire weather danger, which usually occurs over many pixels on this scale. However, when a large fire event happens only a few pixels will see very high burned areas, where exactly these will occur is difficult to predict. Therefore, BuRNN and many process-based models do not predict these large fires in specific pixels but spread out the burned area over a larger area. This in turn leads to lower spatial predictive power in these regions. ~~Moreover, in a number of regions~~

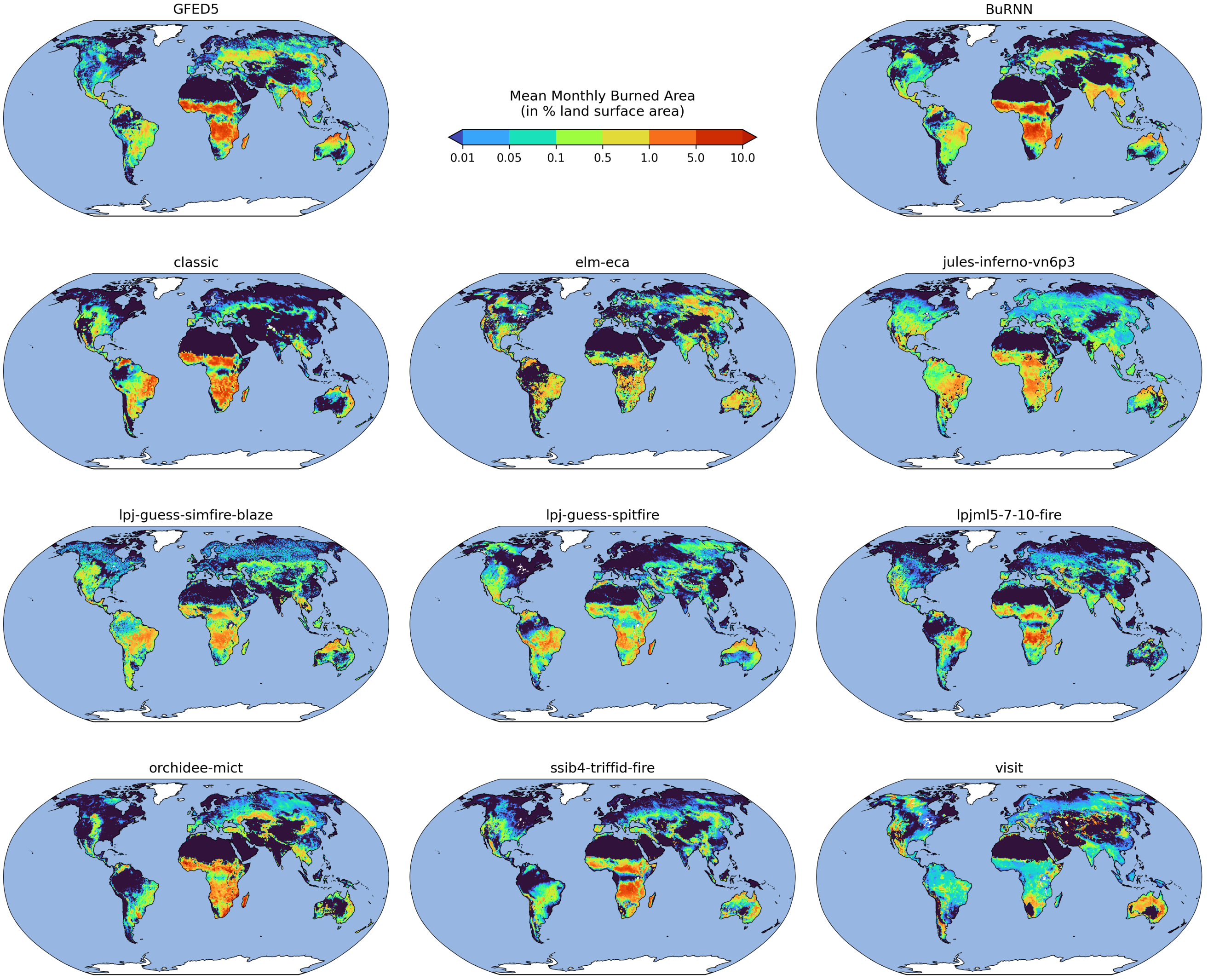

Figure 4. Mean monthly burned area (in % land surface area) over 1997-2019 for the GFED5 satellite product, BuRNN and the nine FireMIP models.

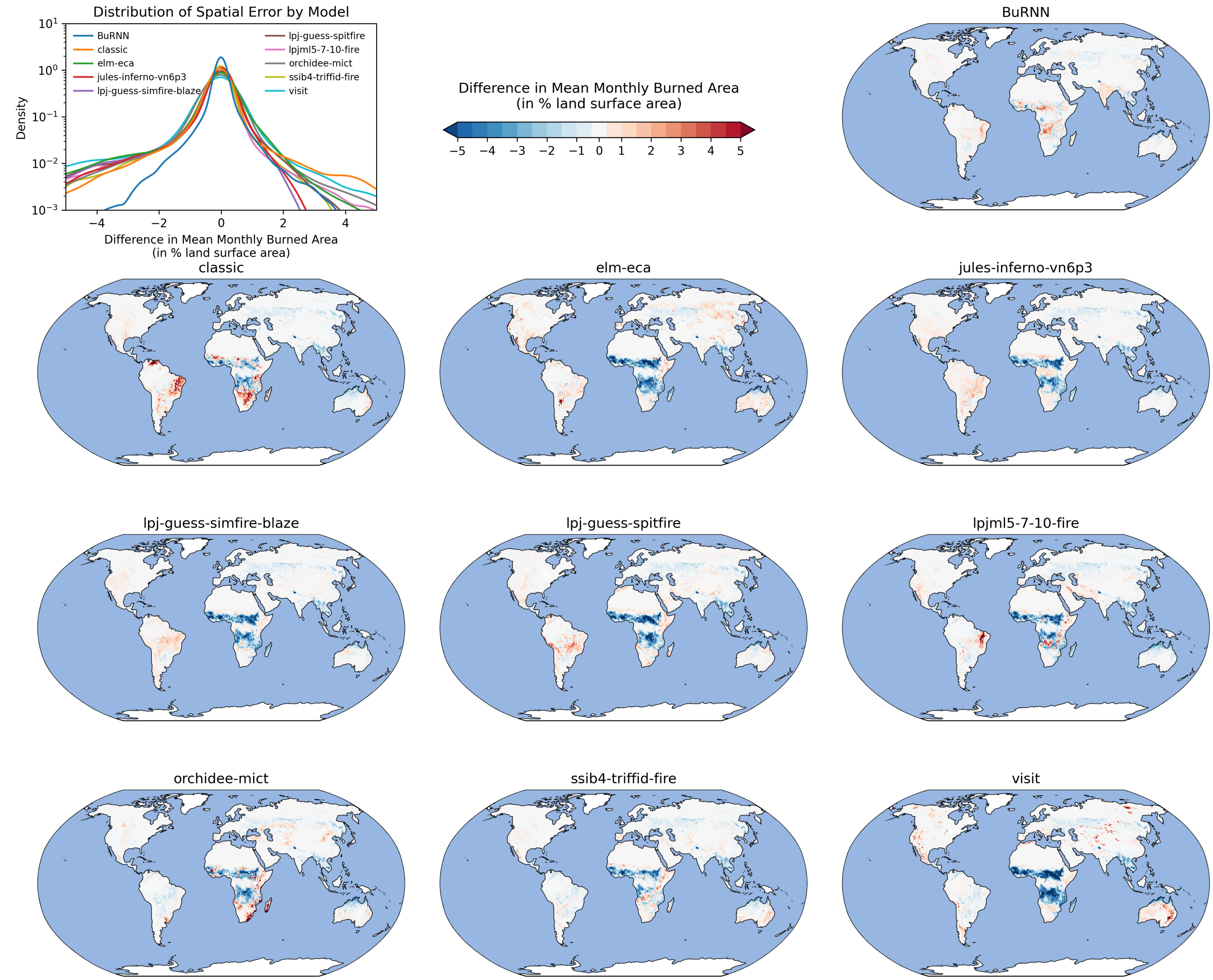

Figure 5. Spatial difference in mean monthly burned area (in % land surface area) over ~~the period~~ 1997-2019 ~~)~~ between GFED5 ~~observations~~ and the model simulations (including BuRNN). The left upper panel shows the distribution of pixel values per model, the more closely centered around 0, the better the modelled burned area pattern.

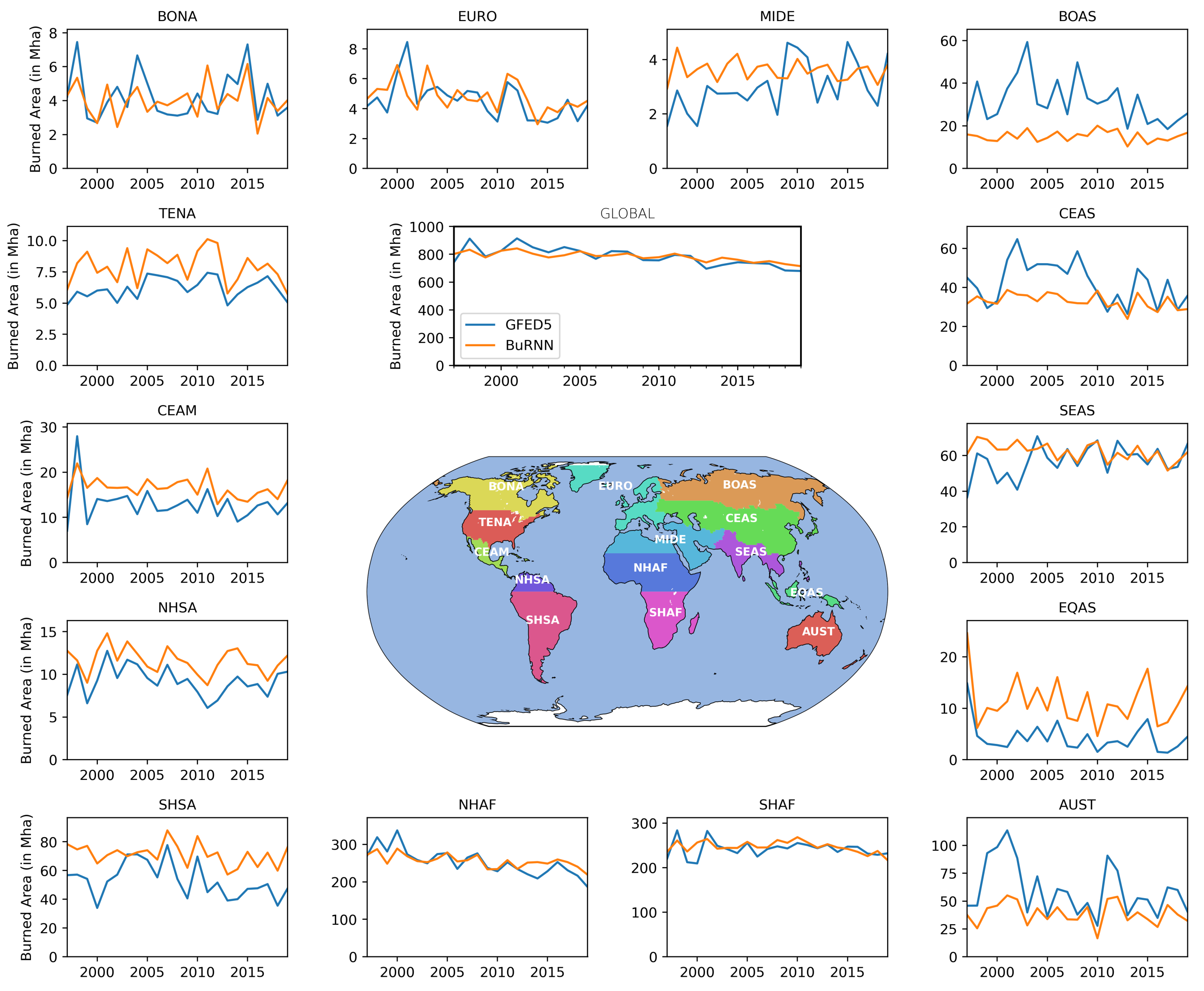

Figure 6. Annual sums (in mHa) of regional burned area by BuRNN (orange) and the GFED5 satellite observations (blue) for 1997-2019. The observed 2003-2019 annual trend is -8.45 ± 3.09 Mha per year ($\mu \pm$ 2SD), while BuRNN models -4.78 ± 2.21 Mha per year.

~~interannual variability is poorly modelled. Earlier research has shown that interannual variability in e.g., BONA, TENA, MIDE , BOAS and AUST is usually related to climate-related factors (Chuvieco et al., 2021). As BuRNN underestimates interannual variability in many of these regions, a possible improvement would involve including more climatic variables, such as vapour pressure deficit and evapotranspiration, during model training.~~

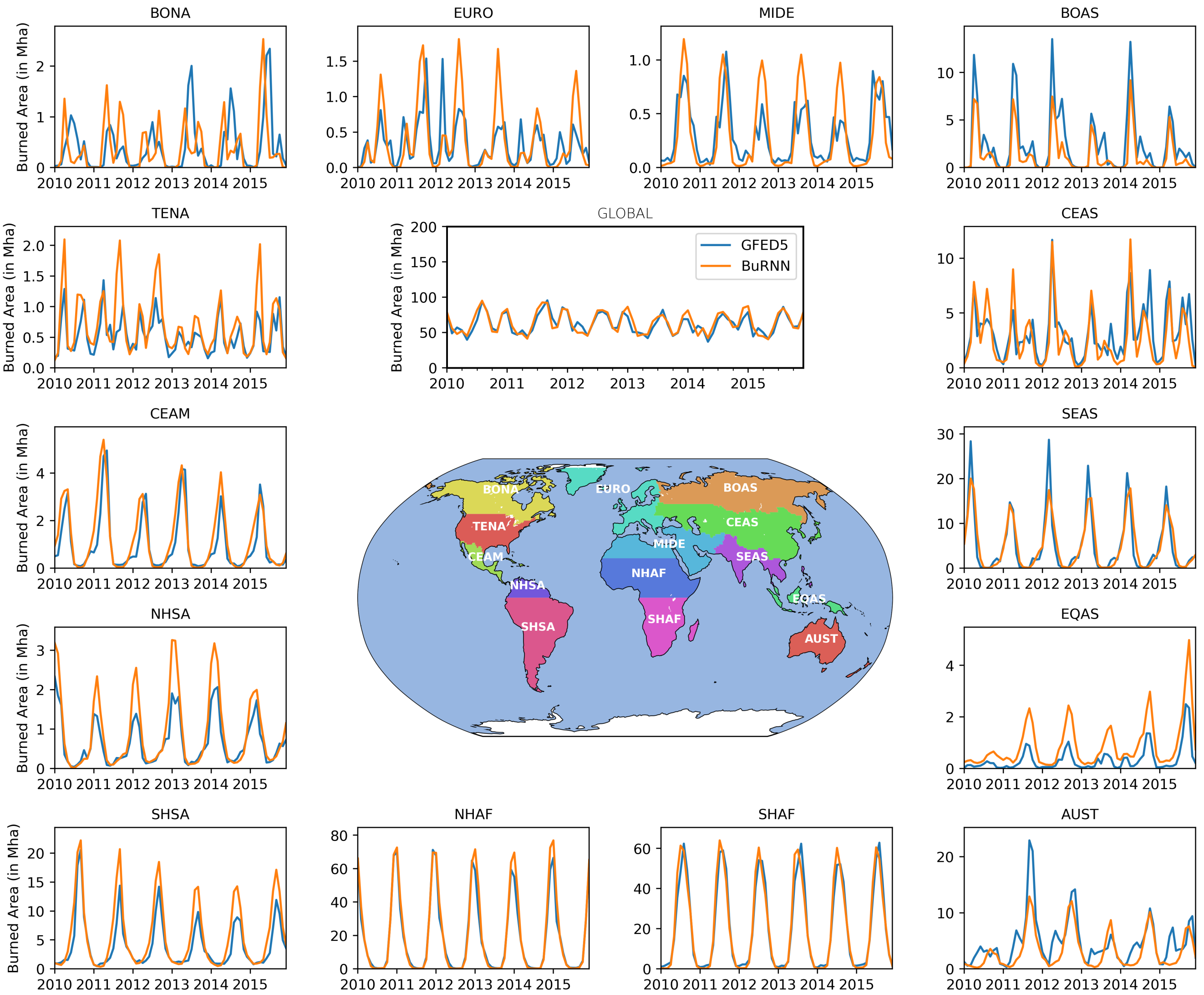

Figure 7. Monthly burned area (in mHa) per region by BuRNN (orange) and the GFED5 satellite observations (blue) for 2010-2015.

3.2 Drivers of BuRNN

We find that the ~~FWI and~~ climatic variables FWI and monthly mean of daily maximum temperature (~~tasmax) , the presence or absence of bare ground and C4 grasses, and GPP~~ temp) are the most impactful features ~~(??) across most regions~~ across all regions (Fig. 8). This suggests that although the Canadian FWI was originally designed to be used in Canadian forests, it can provide relevant information for many, if not all, regions in the world. However, important to consider here is that vegetation characteristics are spread across many more variables, giving each individual vegetation feature a lower importance. Moreover, we see that GDP, a variable often neglected in process-based fire models (Burton et al., 2024), often shows up high in the importance list. We also see bare ground as important indicator in all but one region (EQAS), which is to be expected as a high value of bare ground fraction should immediately render all other features for that grid cell irrelevant. Several regional differences in ~~SHAP values~~ feature importance can be observed. For example, ~~grassland indicators~~ C4 grasses show up in regions with considerable grassland~~fractions~~ /savannah coverage e.g., SHAF, SHAF, SHSA and AUST~~, but in EURO, TENA and BOASthe absence of African C4 grasses also shows up as a reducing factor i.e., because there are no African C4 grasses there, which it cannot have by default, burned area is lower than the global average. From the perspective of BuRNN this makes sense as the presence of African C4 grasses are often an indicator of high burned areas, so an absence is likely associated with lower than average burned area. We also note that NHSA, SHSA, NHAF, SHAF and AUST, which contain notable savannah regions, are the main regions where GPP is ranked higher than in most regions . This corroborates earlier assessments that state that savannah fires are predominantly limited by fuel and moisture availability (Lehmann et al., 2014; Alvarado et al., 2020; Takacs et al., 2021) . In MIDE and CEAS, both regions with large unvegetated deserts, bare ground is the most impactful feature. This can be explained by a high value for bare ground negating the relevance of all other features in a pixel. Additionally, bare ground is also among the top five predictors in~~ . Needleleaf trees (needletree) only show up in BONA and BOAS, and broadleaf evergreen trees are important in regions with important rainforests e.g., in South-America (NHSA and SHSA) and EQAS. Croplands show up in regions with noteable agricultural burning such as SEAS, EQAS and MIDE (Hall et al., 2024). Interestingly enough, in EURO croplands also are an important indicator, yet Europe does not have as extensive cropland fires as many other regions (Hall et al., 2024). Lastly, monthly average daily wind speed is often not present in the top indicators regionally. Even though wind speed is incredibly important for fire spread, it might (i) ~~BONA and BOAS, which have notable regions without or with little vegetation in the upper North, (ii) TENA, NHAF, SEAS and AUST, which have desert areas. In SEAS, the rural population is the most important predictor and it is the only region where this predictor even makes the top five.~~ be averaged out by the spatial (0.5 by 0.5) and temporal (monthly) scale we are working at, and (ii) it only affects burned area in the month it is actually burning. This is important as we take the average feature importance over many timesteps and hence is likely reduced by this aggregating operation. All of this indicates that BuRNN tends to prioritize specific features in their expected places.

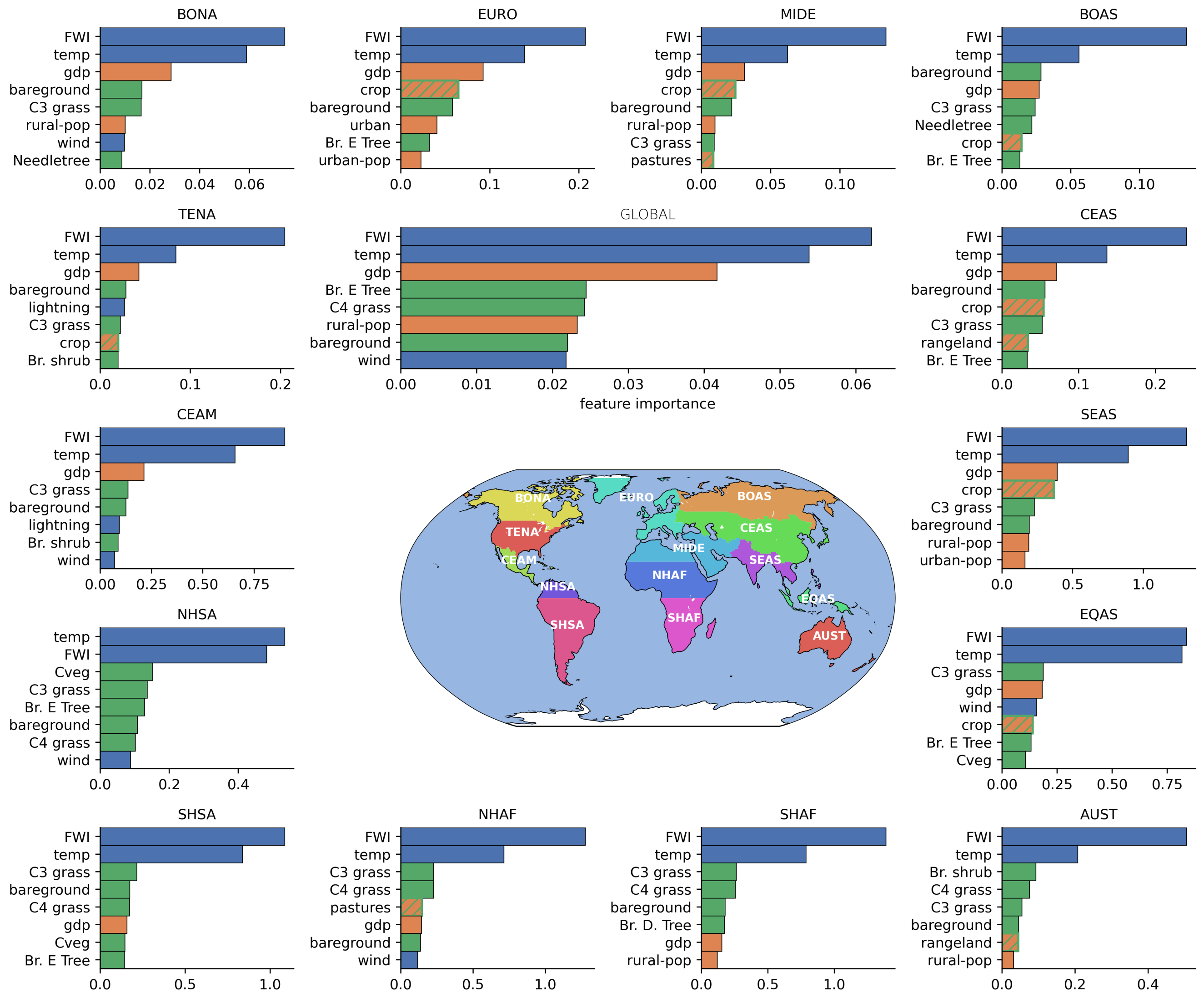

Figure 8. ~~SHAP values of~~ Feature importances for BuRNN per region, indicating which features ~~are~~ affect the predictions of BuRNN most ~~important~~ in each region. ~~SHAP~~ IG attribution values indicate the ~~importance of~~ strength by which a feature ~~in affecting~~ affects the prediction of a model compared to that feature's global mean. ~~Blue values show the effect of low values of the predictor on the prediction, while pink values indicate the effect of high values of the predictor. Features are ranked from~~ The top ~~to bottom in terms of total importance, the top five~~ eight are shown for each region, with climate variables in blue, vegetation variables in green and socio-economic variables in orange.

3.3 Application: a burned area reconstruction for the 20th Century

Fig. 9 shows the global and regional annual burned area as modelled by BuRNN for the period 1901-2019. BuRNN simulates that globally, from ~~1901-2000~~ 1901-1960 there has been a slight increase in burned area, which is mainly attributed to an increase in burned area in SHAF in that same period. ~~Noteable changes are found in TENA, NHSA EURO, MIDE and SHAF.~~ In TENA BuRNN simulates an increasing trend in burned area from 1901 until ∼1955 after which a ~~sharp~~ decline is observed from ∼1960 until ∼1990. In ~~NHSA a small but consistent increase in burned area is modelled for the entire 1901-2019 period. In~~ EURO a first period of high burned area with large interannual variability is modelled from 1901 until ∼1950, after which a stark declining trend is modelled by BuRNN. ~~the~~ The latter, more recent declining trend is also observed in the EFFIS database. ~~In MIDE burned area starts low from 1901 until ∼1930, after which is quickly increases until ∼1950, after which a slight negative trend is simulated.~~ Lastly, for SHAF a positive burned area trend is modelled for the 1901-2010 period, after which burned area again decreases in the last ∼10 years. Next, we also want to compare the 1982-1997 part of the BuRNN reconstruction to FireCCiLT11. Fig. A2 shows the ~~1982-2018~~ 1982-2017 regional annual burned area from BuRNN and FireCCiLT11. The annual burned area correlations for 1982-1993 and ~~1997-2018~~ 1997-2017 between FireCCiLT11 and BuRNN are listed in Table A4 along with the ~~1997-2018~~ 1997-2017 annual burned area correlations between GFED5 and FireCCiLT11. The annual correlation between the two products is relatively low (0.29). However, the uncertainty in burnt area estimates for this period is relatively high, and on average the correlation between BuRNN and FireCCiLT11 for the early period is higher than between the two observational products themselves for ~~1997-2018~~ 1997-2017 Table A4.

Additionally, we compare our reconstruction to regionally available burned area databases. Fig. 10 shows the burned area from EFFIS reported by 21 countries in the EU. Both correlation and bias between this EFFIS database and BuRNN is generally high, with BuRNN simulating higher burned areas than EFFIS. We note however that the reported burned area by EFFIS does not include cropland fires, as opposed to BuRNN, explaining ~~the high absolute biasas cropland fires account for ∼two-thirds to three quarters of the total burned area in Europe (Chen et al., 2023b)~~part of the absolute bias. In Fig. A3, a further comparison is made for 5 more regions (Canada, US, Brazil, Chile and Australia) where the correlation between national databases and BuRNN is only high in Brazil. The likely explanation for this discrepancy lies in the data collection. Correlation between BuRNN and EFFIS is high for individual countries, but is close to 0 when assessed over the 21 European countries combined for the entire period. As each national dataset inside the EFFIS database has a different start and end date, it makes calculating interannual variability inconsistent (unless we restrict the database to only those years available in all countries, which is 2017-2019). Similarly, many of the other national databases, like those in Canada, US and Australia, are composed of regional data sources that come available in different time periods mixed in with satellite images (usually LandSat) when available. In contrast, MapBiomas in Brazil has a single data source (LandSat) and thus does not suffer from this, there correlation (1985-1996) with BuRNN is high (~~0.78~~0.74). Therefore, we believe BuRNN shows a good correlation with these independent data sources whenever the data sources have consistent reporting of burned area. Moreover, in Europe a decreasing trend in annual burned area has been reported, especially in the Mediterranean (Rodrigues et al., 2013; Turco et al., 2016; Chen et al., 2023b). This is in line with the reconstruction of BuRNN.

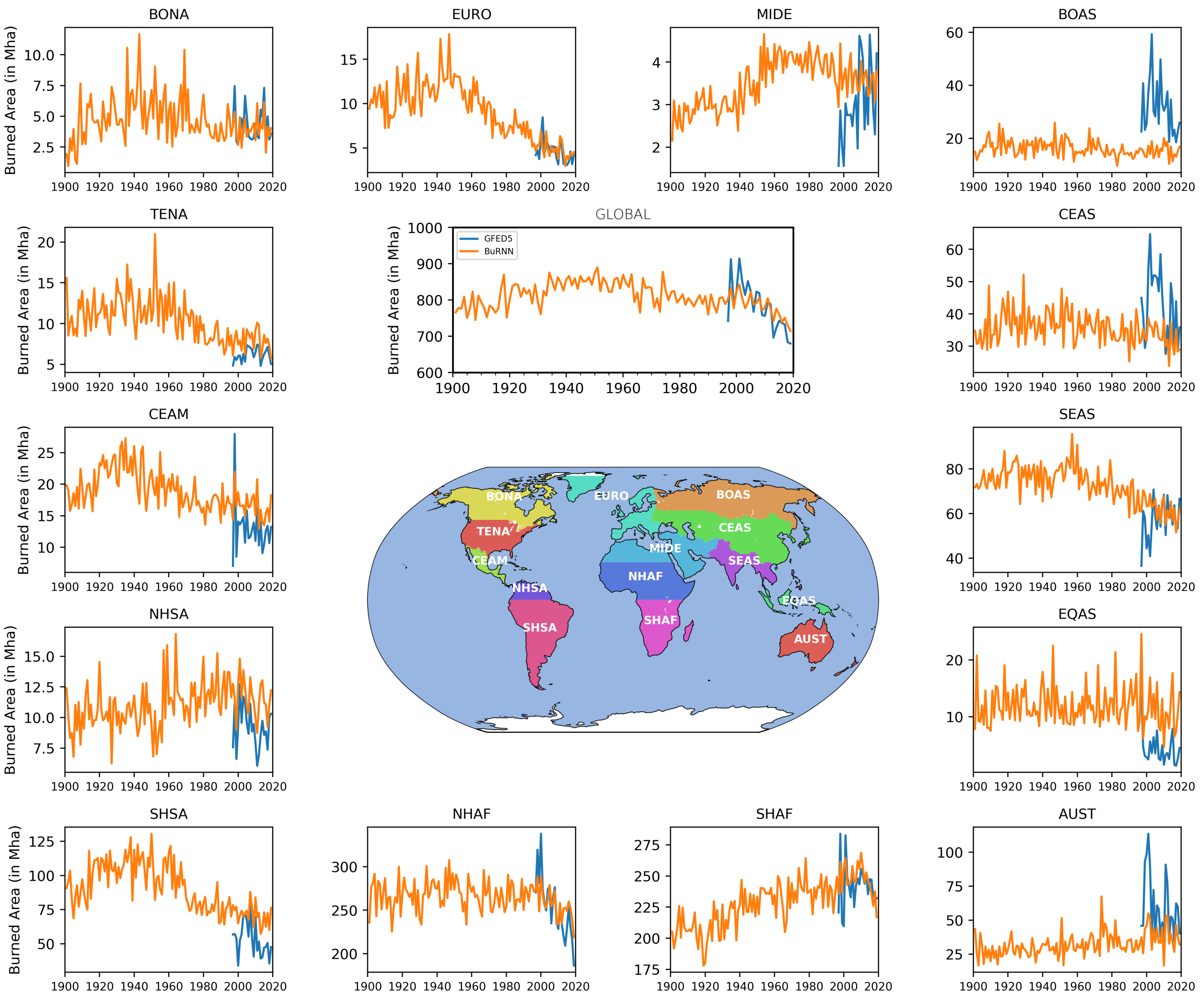

Figure 9. BuRNN's simulation of total annual burned burned area (in Mha) from 1901 to 2019 (orange) for each of the fourteen GFED5 regions and globally, along with the 1997-2019 GFED5 satellite-based burned area.

Figure 10. Annual sums (in ha) of national burned area by BuRNN (orange) and as reported by EFFIS. Methods for collecting and reporting burned area differ by country (and may differ throughout time)~~.~~ , the periods for which data have been plotted are between 1983 and 2019. Note that EFFIS does not include cropland fires.

4 Discussion

Scientific performance aside, BuRNN has a second benefit compared to process-based models i.e., speed and cost of running the model. Running the full 1901-2019 reconstruction (for all the 55 models) takes approximately an hour in total on a single CPU core on our HPC cluster. This is in stark contrast to the computational cost required to run fire-coupled DGVMs, which require hundreds up to tens of thousands of CPU hours. Of course, the major cost of running BuRNN is in the training phase, which typically takes around 10 hours on a single GPU (NVIDIA GeForce 1080Ti). Although the speed and performance of this first version of BuRNN are excellent, it does come at the expense of interpretability. As with most deep learning architectures, BuRNN does not physically relate drivers to responses. We have done effort to alleviate this through our ~~analysis of SHAP values~~XAI analysis, which approximates feature importance, but this understanding is not on par with our knowledge of the mechanisms in process-based models. Conversely, data-driven models can potentially contribute to improved process understanding: if we can unravel why and how BuRNN outperforms these process-based fire models, we can leverage that knowledge to improve the process-based models.

During training, we explicitly aimed to prevent overfitting and maximize generalisability in several ways. We employed a region-based cross validation to counteract the high spatial autocorrelation in our data, we used early stopping, applied normalization during preprocessing on the training data, batch normalization after the LSTM layer and dropout after the linear layer. We subsequently evaluated BuRNN in multiple ways over a number of metrics against multiple products. First, we evaluated the performance of BuRNN by assessing its error scores to GFED5, taking into account that for any region in the world, BuRNN has never seen data from that region before. Then we calculated spatiotemporal, spatial and temporal error scores and correlation of BuRNN to GFED5 and FireCCI51. We repeated this for the process-based fire models participating in ISIMIP3a and compared the relative performances, showing that in most regions over most metrics BuRNN outperforms state-of-the-art fire models. Our burned area reconstruction holds major promise for assessing spatial fire patterns in the pre-satellite era. To assess its quality, we compared our 1982-1993 reconstruction to the FireCCiLT11 remote sensing product and national census data. However, the low correlation between GFED5 and FireCCiLT11, highlights important observational uncertainty in the early satellite record, calling for caution when interpreting our AI-based reconstruction relative to FireCCiLT11 in this period. By comparing our reconstruction of BuRNN to national databases wherever available, we can potentially obtain a sense of regional product quality. We find particularly good correlation with national databases in the EU and Brazil. Databases from Canada, US, Chile and Australia showed poor correlation to the BuRNN reconstruction, likely caused by the heterogenous nature of these reference datasets. However, three main sources of uncertainty and drawbacks need to be raised. First, our model will learn relationships between population densities, GDP and fire occurrence. These might have changed over the last 120 years and nor BuRNN, nor the process-based models can account for this currently. Secondly, BuRNN also relies on three inputs from DGVMs, which are of course reliant on the performance of the model ensembles for these variables. Lastly, in BuRNN there are currently no fire-vegetation feedbacks, which are present in most process-based models.

5 Conclusions

Compared to process-based fire models, BuRNN pushes the state of the art in terms of simulation quality of burned area, demonstrating the potential for machine learning to improve the predictive capabilities in regional-to-global scale fire modelling. As fire behaviour is expected to have changed and continue to change due to climate change, understanding how they have evolved and will evolve is important for understanding our ecosystems, emissions and land use changes. BuRNN substantially improves our capabilities for simulating fire behaviour in all regions of the world compared to state-of-the-art process-based fire models. However, as a machine learning model its interpretability remains below that of conventional fire models. To address this limitation, we applied XAI to unravel some of the inner workings of BuRNN. From this, we conclude that in most regions, BuRNN prioritizes features that are relevant for that region. This includes, for example, ~~bare ground in regions with deserts and~~ FWI and temperature in all regions, C4 grasses in ~~all savannah regions~~regions with notable savannah areas and tree subtypes in regions with extensive forests. As an application, we apply BuRNN to reconstruct global monthly burned area at 0.5° x 0.5° spatial resolution over the period 1901-2019. While a valuable dataset for studying historical burned area patterns, it is a challenge to assess the quality of the product, given considerable discrepancy between different satellite-based burned area products and between the satellite products and national inventories. As the effects of climate change and socio-economic drivers on fire behaviour are largely unknown (quantitatively), BuRNN can aid in better unravelling past burned area patterns, which can improve carbon cycle modelling, help fire risk prevention and inform policy makers.

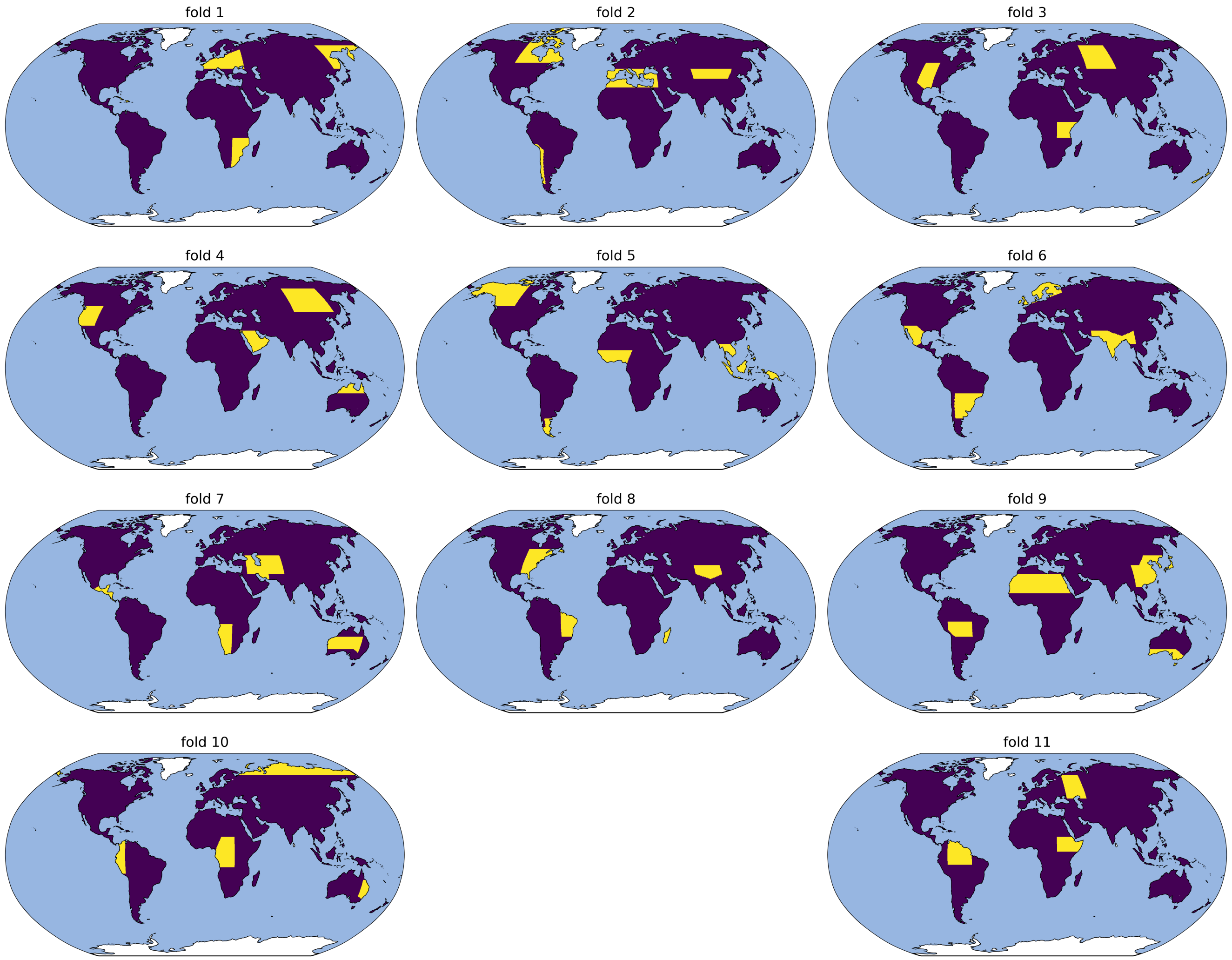

Figure A1. "Division of the 43 regions into 11 folds, used for training the models. The regions marked in yellow represent the 3-4 AR6 regions in that fold. During training we set each fold aside once, then train 5 models on the remaining 10 folds, each time with 8 folds as training and 2 folds as validation. E.g., Fold 1 is set aside as testing fold, then folds 2-3 are used as validation and folds 4-11 as training. Then, folds 4-5 are used as validation and folds 2-3 and 6-11 as training. This is followed by folds 6-7 as validation and folds 2-5 and 8-11 as training, etc."

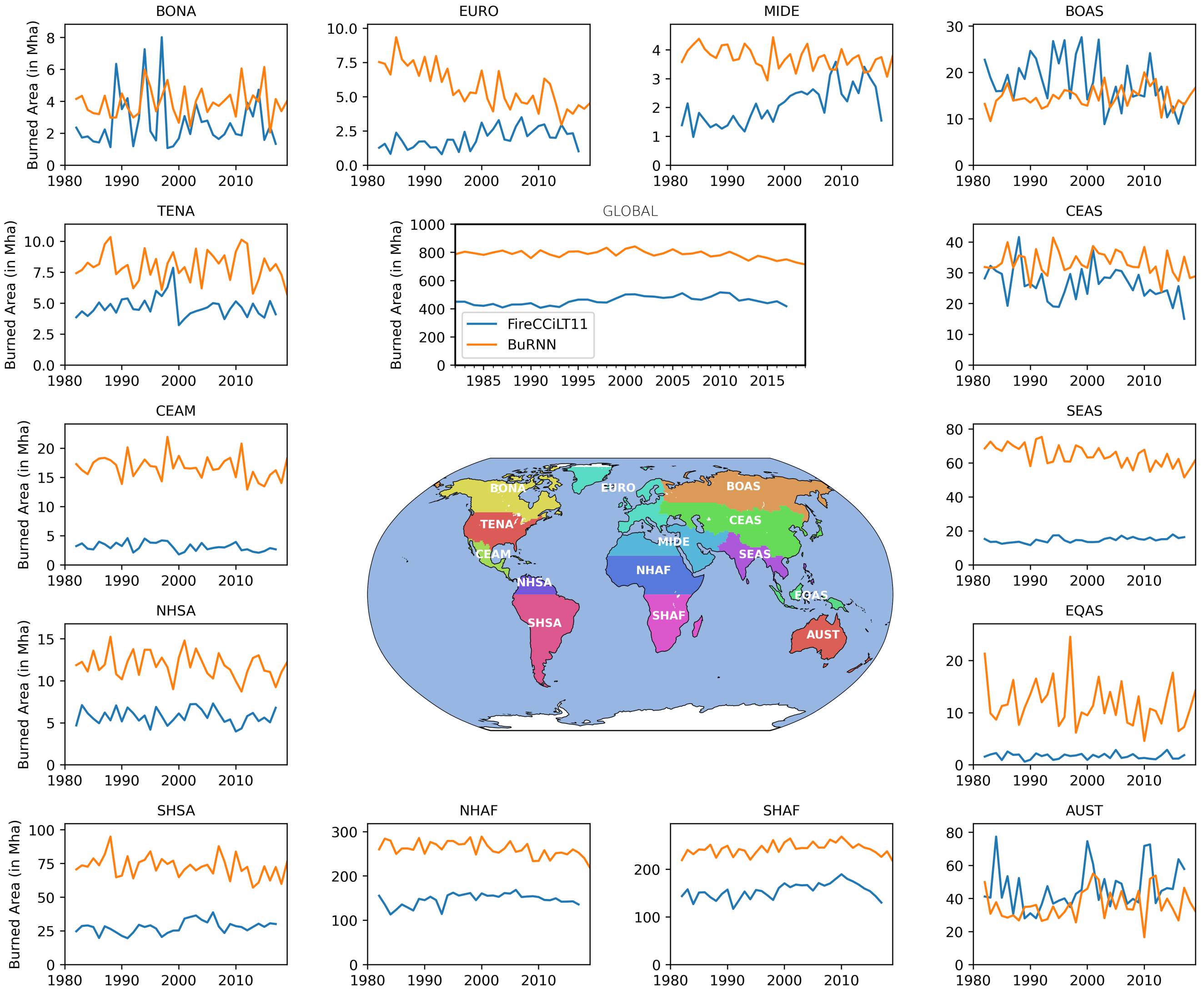

Figure A2. Annual sums of regional burned area by BuRNN (orange) and the FireCCiLT11 observations (blue) for 1982-2018.

Figure A3. Comparison of BuRNN to regional burned area databases. Note that in some regions managed and/or agricultural fires are not reported.

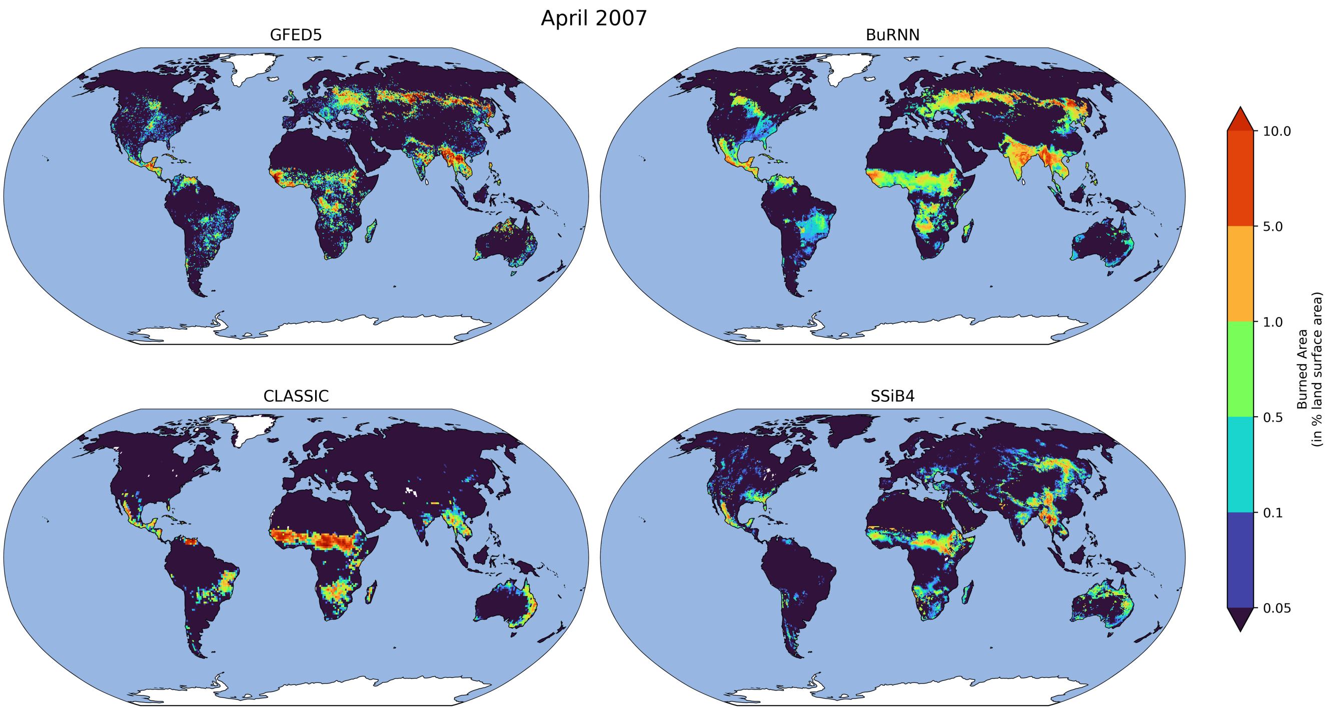

Figure A4. Comparison of BuRNN to GFED5 along with two process-based models (SSiB4 and CLASSIC) for April 2007.

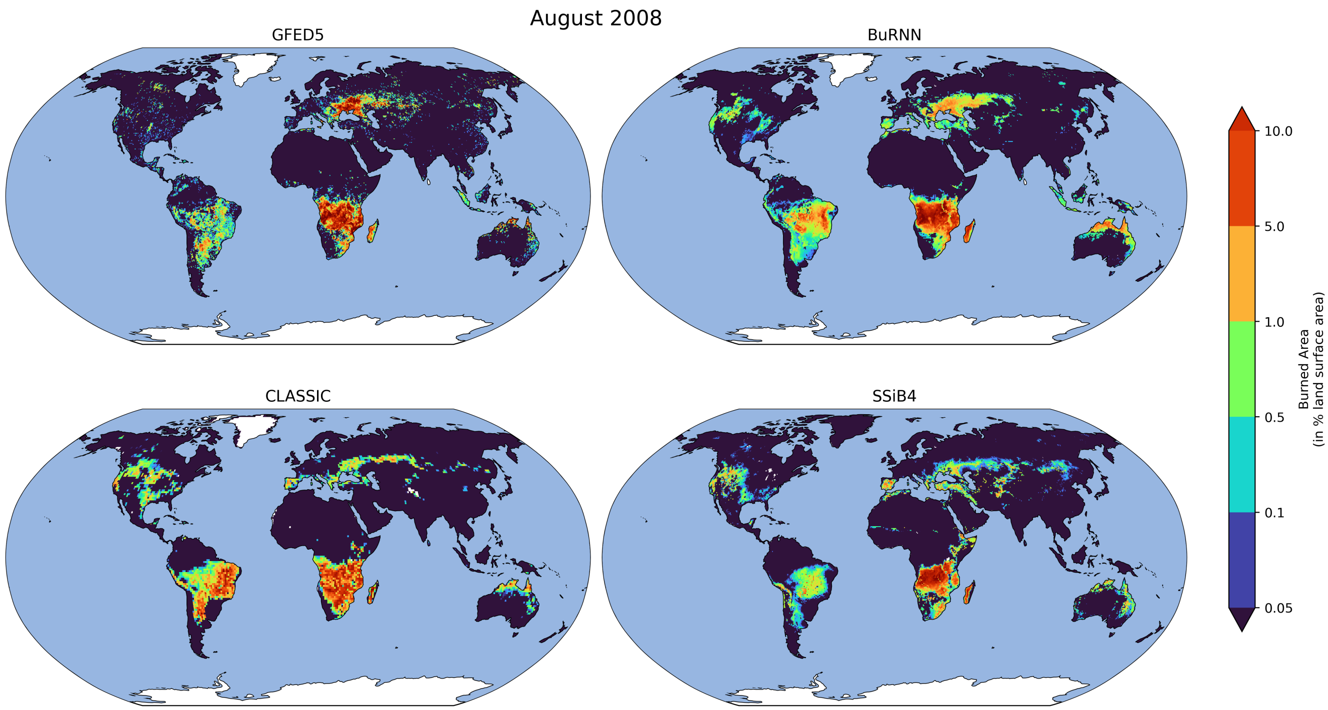

Figure A5. Comparison of BuRNN to GFED5 along with two process-based models (SSiB4 and CLASSIC) for August 2008.

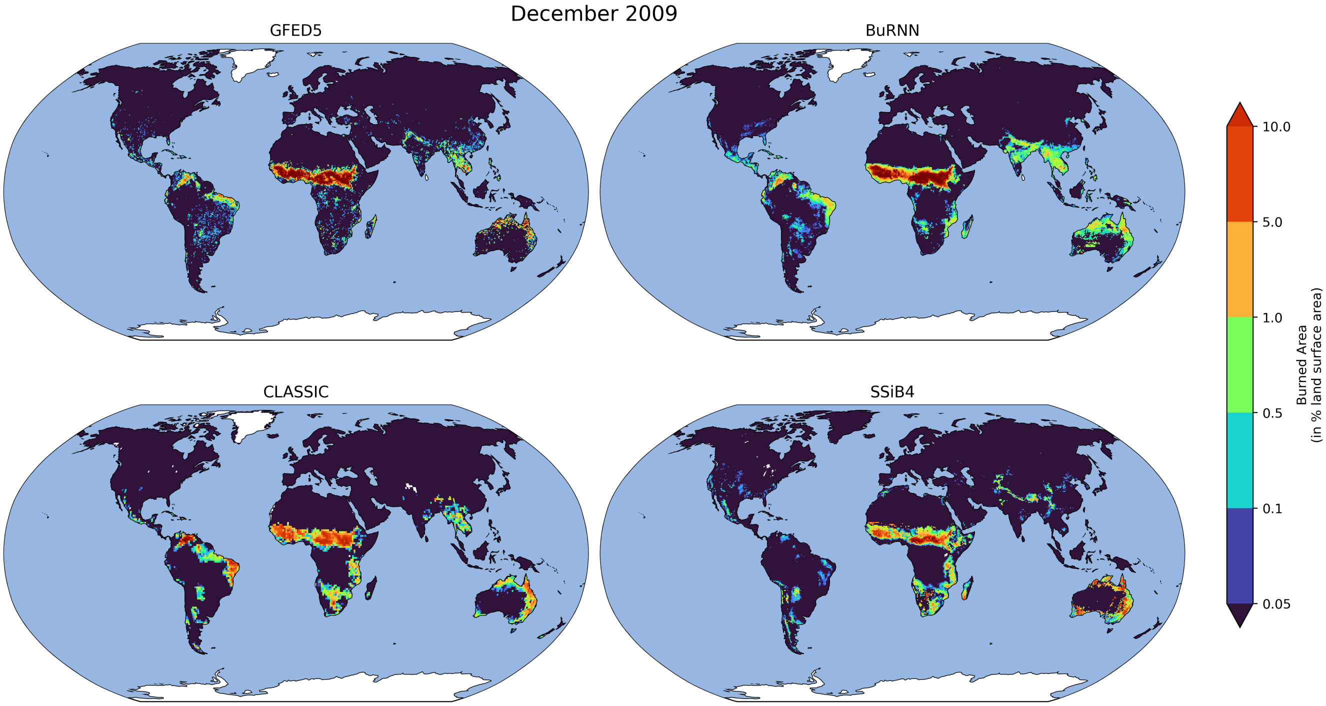

Figure A6. Comparison of BuRNN to GFED5 along with two process-based models (SSiB4 and CLASSIC) for December 2009.

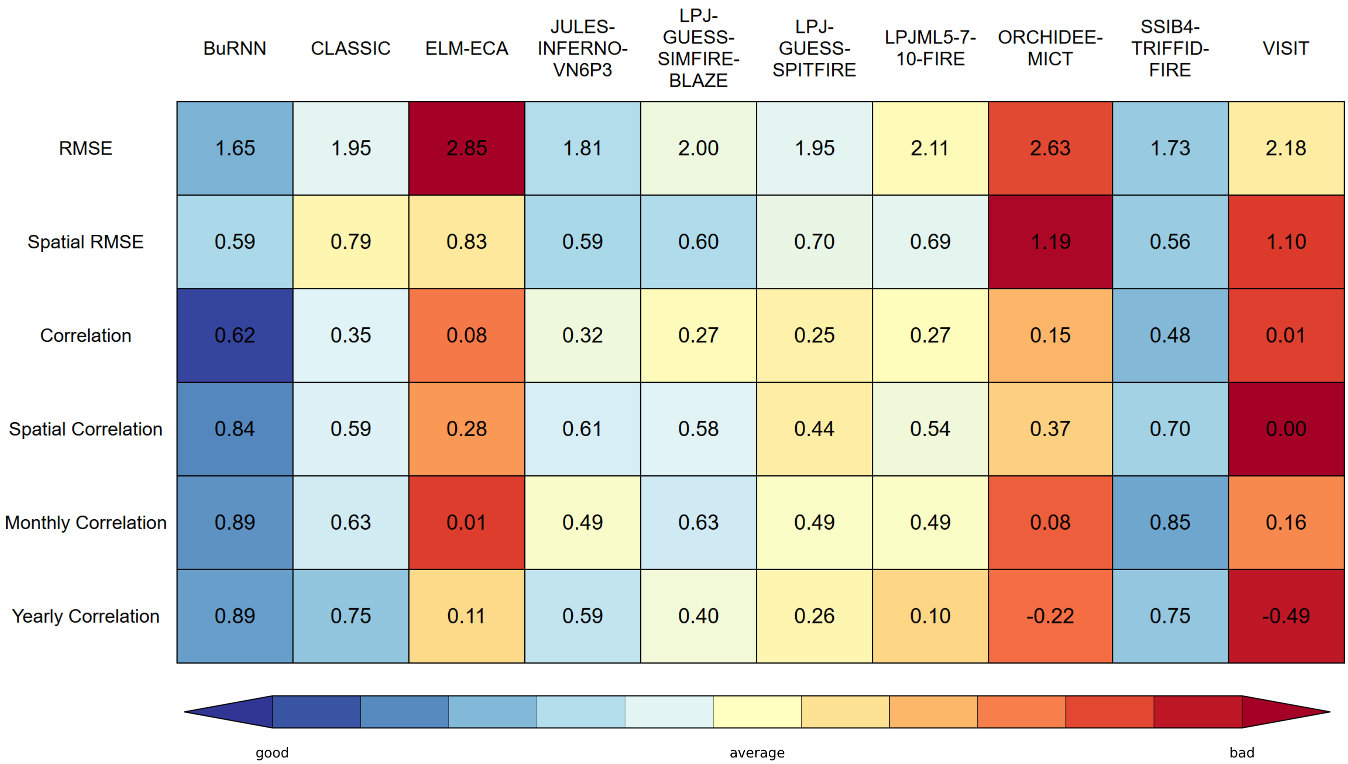

Figure A7. Global evaluation scores of BuRNN and the FireMIP models compared to FireCCI51 for 2003-2019. Colour scaling has been done based on the normalized values (value - row mean)/(row standard deviation) with the minimum and maximum values set to -2 and 2, respectively. Better scores (lower for RMSE and higher for Pearson correlation) are marked in blue, while worse performance is in red.

	AUST	BOAS	BONA	CEAM	CEAS	EQAS	EURO	Global	MIDE	NHAF	NHSA	SEAS	SHAF	SHSA	TENA
RMSE	2.58	0.95	0.48	0.92	1.12	0.68	0.33	1.65	0.26	3.44	0.84	1.78	3.85	1.11	0.50
Spatial RMSE	0.57	0.18	0.10	0.51	0.27	0.37	0.11	0.59	0.10	1.36	0.38	0.80	1.43	0.48	0.13
Correlation	0.31	0.36	0.06	0.30	0.21	0.27	0.29	0.62	0.16	0.71	0.59	0.35	0.67	0.38	0.09
Spatial Correlation	0.72	0.64	0.07	0.35	0.45	0.35	0.57	0.84	0.30	0.86	0.70	0.47	0.82	0.55	0.13
Monthly Correlation	0.81	0.82	0.11	0.83	0.53	0.90	0.81	0.89	0.77	0.95	0.92	0.77	0.98	0.88	0.64
Yearly Correlation	0.88	0.70	0.45	0.72	0.79	0.97	0.64	0.89	0.20	0.79	0.77	0.17	0.81	0.89	0.72

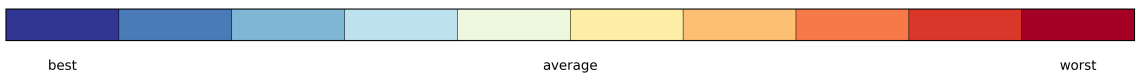

Figure A8. Regional evaluation scores of BuRNN compared to FireCCI51 for 2003-2019. Colour scaling has been done based on the ranked values compared to the nine process-based fire models, with the minimum RMSE and maximum correlations coloured blue (best) and the highest RMSE and correlation coloured red (worst).

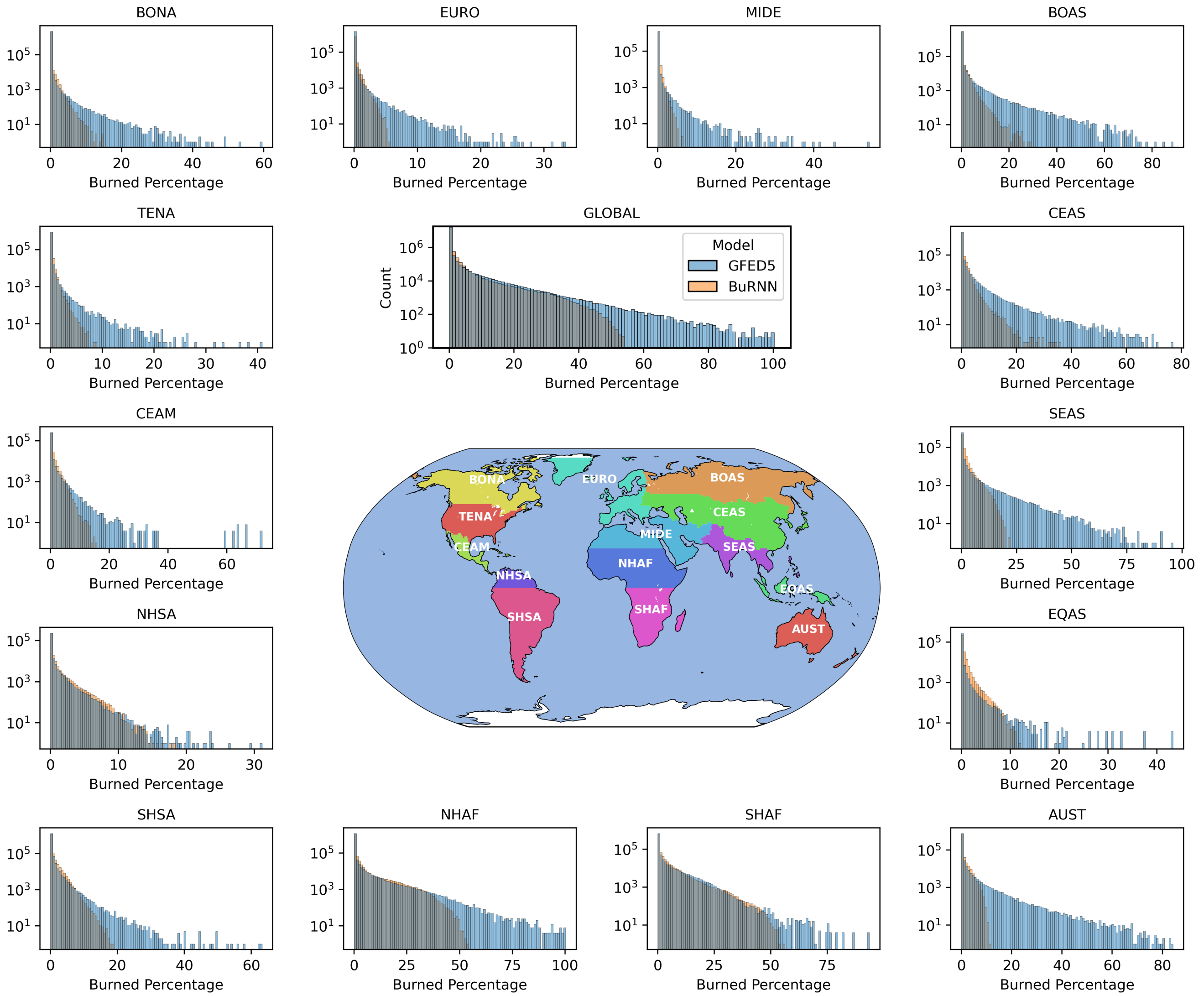

Figure 9. Histograms of the observed (blue) and modelled (orange) burned area (in % land surface area) for 1997-2019.

Table A1. Models used for the calculation of the ISIMIP Biome characteristics.

	CLASSIC	ELM-ECA	DLEM	JULES-ES-VN6P3	ORCHIDEE-MICT	SSiB4-TRIFFID-Fire	VISIT
cVeg		X				X	X
GPP	X	X	X	X	X	X	X
LAI	X	X			X	X	X

Table A2. ~~Evaluation scores~~ Theil-Sen slopes of ~~BuRNN~~ observed and ~~the FireMIP models in AUST~~modelled annual burned area for 2003-2019. ~~Colour scaling is based on the normalized values with the minimum and maximum values set to -2 and 2 (sigma). Better scores are marked~~ The cells depict mean $\pm$ 2SD of annual burned area trend in ~~blue, while worse performance is in red~~Mha per year.

Global
~~RMSE~~BONA
TENA
~~Monthly Correlation~~CEAM
NHSA
~~Yearly Correlation~~SHSA
~~Evaluation scores of BuRNN and the FireMIP models in BOAS. Colour scaling is based on the normalized values with the minimum and maximum val~~
~~RMSE~~MIDE
NHAF
SHAF
BOAS
~~Monthly Correlation~~CEAS
~~Yearly Correlation~~SEAS
~~Evaluation scores of BuRNN and the FireMIP models in BONA. Colour scaling is based on the normalized values with the minimum and maximum va~~
~~Yearly Correlation~~AUST

Table A3. ~~Evaluation scores~~ Regional comparison of ~~BuRNN and~~ the ~~FireMIP models in CEAM. Colour scaling is based on the normalized values with the minimum~~ observational products GFED5 and ~~maximum values set to -2 and 2 (sigma)~~FireCCI51.~~Better scores are marked in blue, while worse performance is in red.~~

RMSE

Table A4. Regional correlation of annual burned area of BuRNN and FireCCiLT11 (1982-1993 and 1997-2018) and between FireCCiLT11 and ~~GFED5~~ GFED5 (1997-2018).

	BuRNN-FireCCiLT11: 1982-1993	BuRNN-FireCCiLT11: 1997-2018	~~GFED5-FireCCiLT11~~GFED5-FireCCiLT11: 1997-2018
BONA	~~0.62~~ 0.52	~~-0.13~~ -0.01	-0.07
EURO	~~0.58~~ 0.67	~~0.21~~ 0.18	0.05
MIDE	~~-0.26~~ -0.08	~~-0.44~~ -0.41	0.29
BOAS	~~0.20~~ 0.15	~~-0.20~~ -0.19	-0.04
TENA	~~0.27~~ 0.32	~~0.22~~ 0.05	-0.23
CEAS	-0.28	~~0.30~~ 0.26	0.29
CEAM	~~0.60~~ 0.61	~~0.02~~ 0.07	0.12
SEAS	~~0.04~~ -0.01	~~-0.21~~ -0.59	0.32
NHSA	~~0.06~~ -0.01	~~0.08~~ 0.25	0.24
EQAS	~~-0.15~~ -0.17	~~-0.07~~ -0.10	-0.06
SHSA	~~0.21~~ 0.27	~~-0.05~~ -0.25	0.23
NHAF	~~0.21~~ 0.26	~~0.36~~ 0.47	0.52
SHAF	~~0.52~~ 0.55	~~0.50~~ 0.53	-0.06
AUST	~~0.32~~ -0.01	~~0.22~~ -0.06	0.19

~~Regional evaluation scores of BuRNN. Colour scaling has been done based on the ranked values compared to the with the minimum RMSE and maximum correlations coloured blue and the highest RMSE and correlation coloured red.~~

~~AUSTBOASBONACEAMCEASEQASEUROMIDENHAFNHSASEASSHAFSHSATENA~~

~~RMSE2.78 0.95 0.42 0.53 1.10 0.46 0.29 0.23 2.97 0.76 1.49 3.32 0.93 0.45 Spatial RMSE0.84 0.18 0.05 0.24 0.22 0.23 0.08 0.08 0.86 0.27 0.55 0.91 0.27 0.09 Correlation0.28 0.34 0.08 0.33 0.21 0.31 0.30 0.16 0.71 0.54 0.35 0.67 0.42 0.08 Spatial Correlation0.55 0.61 0.20 0.40 0.43 0.40 0.59 0.29 0.86 0.70 0.48 0.84 0.58 0.18 Monthly Correlation0.81 0.86 0.25 0.84 0.47 0.91 0.76 0.78 0.95 0.82 0.80 0.97 0.89 0.60 Yearly Correlation0.92 0.73 0.51 0.70 0.75 0.96 0.52 0.46 0.66 0.62 0.28 0.27 0.94 0.67~~

Code and data availability. All code for the pre-processing, training and post-processing of BuRNN is openly accessible on GitHub (https://github.com/VUB-HYDR/BuRNN)) and is archived on Zenodo under copyright license CC BY 4.0 (https://zenodo.org/records/17834206; Lampe, 2025a). The 1901-2019 burned area simulation of BuRNN is available on Zenodo as well along with all raw and pre-processed data to train BuRNN (https://zenodo.org/records/17778519; Lampe, 2025b). GFED5, HistLight and WGLC can be retrieved originally from Zenodo (https://zenodo.org/records/7668424, https://zenodo.org/records/6405396 and https://zenodo.org/records/15215319; Chen et al., 2023a; Kaplan and Lau, 2022b; Kaplan, 2025). The original ISIMIP data is also available through the ISIMIP data repository (https://data.isimip.org/). The CLM data is automatically generated during the pre-processing for a CLM model run.

Author contributions. SL, WT and EC conceptualized the study. SL, LG, BK, VH and BLS designed the model architecture. SL programmed and trained the model. SL, WT, LG, BK, SH and DK performed the analysis. WT supervised the project. All authors contributed to the final version of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. S.L. was supported by a PhD Fundamental Research Grant by Fonds Wetenschappelijk Onderzoek-Vlaanderen (FWO; 11M7725N). The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the FWO and the Flemish Government. W.T. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon Framework research and innovation programme (grant agreement No 101076909; ERC Consolidator Grant 'LACRIMA'). ECh has been funded by the European Space Agency's Climate Change Initiative (ESA CCI) programme (Contract No. 4000126706/19/I-NB). S.H. acknowledges support from the Max Planck Tandem group programme and from Universidad del Rosario within the programme of Fondos de arranque.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A Next-generation Hyperparameter Optimization Framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.

Alvarado, S. T., Andela, N., Silva, T. S. F., and Archibald, S.: Thresholds of fire response to moisture and fuel load differ between tropical savannas and grasslands across continents, Global Ecology and Biogeography, 29, 331–344, https://doi.org/https://doi.org/10.1111/geb.13034, 2020.

Archibald, S., Lehmann, C. E. R., Belcher, C. M., Bond, W. J., Bradstock, R. A., Daniau, A.-L., Dexter, K. G., Forrestel, E. J., Greve, M., He, T., Higgins, S. I., Hoffmann, W. A., Lamont, B. B., McGlinn, D. J., Moncrieff, G. R., Osborne, C. P., Pausas, J. G., Price, O., Ripley, B. S., Rogers, B. M., Schwilk, D. W., Simon, M. F., Turetsky, M. R., der Werf, G. R. V., and Zanne, A. E.: Biological and geophysical feedbacks with fire in the Earth system, Environmental Research Letters, 13, 033 003, https://doi.org/10.1088/1748-9326/aa9ead, 2018.

Bauters, M., Drake, T. W., Wagner, S., Baumgartner, S., Makelele, I. A., Bodé, S., Verheyen, K., Verbeeck, H., Ewango, C., Cizungu, L., et al.: Fire-derived phosphorus fertilization of African tropical forests, Nature Communications, 12, 5129, 2021.

Beck, P. S., Goetz, S. J., Mack, M. C., Alexander, H. D., Jin, Y., Randerson, J. T., and Loranty, M. M.: The impacts and implications of an intensifying fire regime on Alaskan boreal forest composition and albedo, Global Change Biology, 17, 2853–2866, 2011.

Beguería, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B.: Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring, International Journal of Climatology, 34, 3001–3023, https://doi.org/https://doi.org/10.1002/joc.3887, 2014.

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B.: Algorithms for hyper-parameter optimization, Advances in neural information processing systems, 24, 2011.

Bowman, D.: Wildfire science is at a loss for comprehensive data, Nature, 560, 7–8, 2018.

Bowman, D. M., Kolden, C. A., Abatzoglou, J. T., Johnston, F. H., van der Werf, G. R., and Flannigan, M.: Vegetation fires in the Anthropocene, Nature Reviews Earth & Environment, 1, 500–515, 2020.

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

Bracco, A., Brajard, J., Dijkstra, H. A., Hassanzadeh, P., Lessig, C., and Monteleoni, C.: Machine learning for the physics of climate, Nature Reviews Physics, 7, 6–20, 2025.

Brogan, D. J., Nelson, P. A., and MacDonald, L. H.: Reconstructing extreme post-wildfire floods: A comparison of convective and mesoscale events, Earth Surface Processes and Landforms, 42, 2505–2522, 2017.

Burton, C., Lampe, S., Kelley, D. I., Thiery, W., Hantson, S., Christidis, N., Gudmundsson, L., Forrest, M., Burke, E., Chang, J., et al.: Global burned area increasingly explained by climate change, Nature Climate Change, pp. 1–7, 2024.

Canadian Forest Service: National Burned Area Composite (NBAC), https://cwfis.cfs.nrcan.gc.ca, 2024.

Cao, Y., Yang, F., Tang, Q., and Lu, X.: An attention enhanced bidirectional LSTM for early forest fire smoke recognition, IEEE Access, 7, 154 732–154 742, 2019.

Carvalho, A., Monteiro, A., Flannigan, M., Solman, S., Miranda, A. I., and Borrego, C.: Forest fires in a changing climate and their impacts on air quality, Atmospheric Environment, 45, 5545–5553, 2011.

Chakrabarty, R. K., Shetty, N. J., Thind, A. S., Beeler, P., Sumlin, B. J., Zhang, C., Liu, P., Idrobo, J. C., Adachi, K., Wagner, N. L., et al.: Shortwave absorption by wildfire smoke dominated by dark brown carbon, Nature Geoscience, 16, 683–688, 2023.

Chen, J., Li, C., Ristovski, Z., Milic, A., Gu, Y., Islam, M. S., Wang, S., Hao, J., Zhang, H., He, C., et al.: A review of biomass burning: Emissions and impacts on air quality, health and climate in China, Science of the Total Environment, 579, 1000–1034, 2017.

Chen, Y., Hall, J., van Wees, D., Andela, N., Hantson, S., Giglio, L., van der Werf, G., Morton, D., and Randerson, J.: Global Fire Emissions Database (GFED5) Burned Area (0.1)[Data set], Zenodo, 2023a.

Chen, Y., Hall, J., van Wees, D., Andela, N., Hantson, S., Giglio, L., van der Werf, G. R., Morton, D. C., and Randerson, J. T.: Multi-decadal trends and variability in burned area from the fifth version of the Global Fire Emissions Database (GFED5), Earth System Science Data, 15, 5227–5259, https://doi.org/10.5194/essd-15-5227-2023, 2023b.

Chuvieco, E., Pettinari, M. L., Koutsias, N., Forkel, M., Hantson, S., and Turco, M.: Human and climate drivers of global biomass burning variability, Science of the Total Environment, 779, 146 361, 2021.

Chuvieco, E., Roteta, E., Sali, M., Stroppiana, D., Boettcher, M., Kirches, G., Storm, T., Khairoun, A., Pettinari, M. L., Franquesa, M., et al.: Building a small fire database for Sub-Saharan Africa from Sentinel-2 high-resolution images, Science of the total Environment, 845, 157 139, 2022.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., et al.: The twentieth century reanalysis project, Quarterly Journal of the Royal Meteorological Society, 137, 1–28, 2011.

Diniz-Filho, J. A. F., Rangel, T. F. L. V. B., and Bini, L. M.: Model selection and information theory in geographical ecology, Global Ecology and Biogeography, 17, 479–488, https://doi.org/https://doi.org/10.1111/j.1466-8238.2008.00395.x, 2008.

Falcon, W. and The PyTorch Lightning team: PyTorch Lightning, https://doi.org/10.5281/zenodo.3828935, 2019.

Franquesa, M., Stehman, S. V., and Chuvieco, E.: Assessment and characterization of sources of error impacting the accuracy of global burned area products, Remote Sensing of Environment, 280, 113 214, 2022.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., Le Quéré, C., Li, H., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Arneth, A., Arora, V., Bates, N. R., Becker, M., Bellouin, N., Berghoff, C. F., Bittig, H. C., Bopp, L., Cadule, P., Campbell, K., Chamberlain, M. A., Chandra, N., Chevallier, F., Chini, L. P., Colligan, T., Decayeux, J., Djeutchouang, L. M., Dou, X., Duran Rojas, C., Enyo, K., Evans, W., Fay, A. R., Feely, R. A., Ford, D. J., Foster, A., Gasser, T., Gehlen, M., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, O., Harris, I., Hefner, M., Heinke, J., Hurtt, G. C., Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A. K., Jarníková, T., Jersild, A., Jiang, F., Jin, Z., Kato, E., Keeling, R. F., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Lan, X., Lauvset, S. K., Lefèvre, N., Liu, Z., Liu, J., Ma, L., Maksyutov, S., Marland, G., Mayot, N., McGuire, P. C., Metzl, N., Monacci, N. M., Morgan, E. J., Nakaoka, S.-I., Neill, C., Niwa, Y., Nützel, T., Olivier, L., Ono, T., Palmer, P. I., Pierrot, D., Qin, Z., Resplandy, L., Roobaert, A., Rosan, T. M., Rödenbeck, C., Schwinger, J., Smallman, T. L., Smith, S. M., Sospedra-Alfonso, R., Steinhoff, T., Sun, Q., Sutton, A. J., Séférian, R., Takao, S., Tatebe, H., Tian, H., Tilbrook, B., Torres, O., Tourigny, E., Tsujino, H., Tubiello, F., van der Werf, G., Wanninkhof, R., Wang, X., Yang, D., Yang, X., Yu, Z., Yuan, W., Yue, X., Zaehle, S., Zeng, N., and Zeng, J.: Global Carbon Budget 2024, Earth System Science Data, 17, 965–1039, https://doi.org/10.5194/essd-17-965-2025, 2025.

Frieler, K., Volkholz, J., Lange, S., Schewe, J., Mengel, M., del Rocío Rivas López, M., Otto, C., Reyer, C. P., Karger, D. N., Malle, J. T., et al.: Scenario setup and forcing data for impact model evaluation and impact attribution within the third round of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP3a), Geoscientific Model Development, 17, 1–51, 2024.

Fritze, H., Smolander, A., Levula, T., Kitunen, V., and Mälkönen, E.: Wood-ash fertilization and fire treatments in a Scots pine forest stand: Effects on the organic layer, microbial biomass, and microbial activity, Biology and fertility of soils, 17, 57–63, 1994.

Gers, F. A., Schmidhuber, J., and Cummins, F.: Learning to forget: Continual prediction with LSTM, Neural computation, 12, 2451–2471, 2000.

Giglio, L., Randerson, J., Van der Werf, G., Kasibhatla, P., Collatz, G., Morton, D., and DeFries, R.: Assessing variability and long-term trends in burned area by merging multiple satellite fire products, Biogeosciences, 7, 1171–1186, 2010.

Giglio, L., Schroeder, W., and Justice, C. O.: The collection 6 MODIS active fire detection algorithm and fire products, Remote sensing of environment, 178, 31–41, 2016.

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., and Justice, C. O.: The Collection 6 MODIS burned area mapping algorithm and product, Remote sensing of environment, 217, 72–85, 2018.

Gincheva, A., Pausas, J. G., Edwards, A., Provenzale, A., Cerdà, A., Hanes, C., Royé, D., Chuvieco, E., Mouillot, F., Vissio, G., et al.: A monthly gridded burned area database of national wildland fire data, Scientific data, 11, 352, 2024.

Girona-García, A., Vieira, D. C., Silva, J., Fernández, C., Robichaud, P. R., and Keizer, J. J.: Effectiveness of post-fire soil erosion mitigation treatments: A systematic review and meta-analysis, Earth-Science Reviews, 217, 103 611, 2021.

Grant, L., Vanderkelen, I., Gudmundsson, L., Fischer, E., Seneviratne, S. I., and Thiery, W.: Global emergence of unprecedented lifetime exposure to climate extremes, Nature, 641, 374–379, 2025.

Hall, J. V., Argueta, F., Zubkova, M., Chen, Y., Randerson, J. T., and Giglio, L.: GloCAB: global cropland burned area from mid-2002 to 2020, Earth System Science Data, 16, 867–885, 2024.

Hanes, C. C., Wang, X., Jain, P., Parisien, M.-A., Little, J. M., and Flannigan, M. D.: Fire-regime changes in Canada over the last half century, Canadian Journal of Forest Research, 49, 256–269, 2019.

Hantson, S., Arneth, A., Harrison, S. P., Kelley, D. I., Prentice, I. C., Rabin, S. S., Archibald, S., Mouillot, F., Arnold, S. R., Artaxo, P., et al.: The status and challenge of global fire modelling, Biogeosciences, 13, 3359–3375, 2016.

Hantson, S., Kelley, D. I., Arneth, A., Harrison, S. P., Archibald, S., Bachelet, D., Forrest, M., Hickler, T., Lasslop, G., Li, F., et al.: Quantitative assessment of fire and vegetation properties in historical simulations with fire-enabled vegetation models from the Fire Model Intercomparison Project, Geoscientific Model Development Discussions, 2020, 1–25, 2020.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural computation, 9, 1735–1780, 1997.

Hodzic, A., Madronich, S., Bohn, B., Massie, S., Menut, L., and Wiedinmyer, C.: Wildfire particulate matter in Europe during summer 2003: meso-scale modeling of smoke emissions, transport and radiative effects, Atmospheric Chemistry and Physics, 7, 4043–4064, 2007.

Hunt, K. M., Matthews, G. R., Pappenberger, F., and Prudhomme, C.: Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States, Hydrology and Earth System Sciences, 26, 5449–5472, 2022.

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Goldewijk, K. K., et al.: Harmonization of global land-use change and management for the period 850–2100 (LUH2) for CMIP6, Geoscientific Model Development Discussions, 2020, 1–65, 2020.

Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cerezo-Mota, R., Cimadevilla, E., Cofiño, A. S., Di Luca, A., Faria, S. H., Gorodetskaya, I. V., Hauser, M., Herrera, S., Hennessy, K., Hewitt, H. T., Jones, R. G., Krakovska, S., Manzanas, R., Martínez-Castro, D., Narisma, G. T., Nurhati, I. S., Pinto, I., Seneviratne, S. I., van den Hurk, B., and Vera, C. S.: An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets, Earth System Science Data, 12, 2959–2970, https://doi.org/10.5194/essd-12-2959-2020, 2020.

Jacobs, L., Maes, J., Mertens, K., Sekajugo, J., Thiery, W., Van Lipzig, N., Poesen, J., Kervyn, M., and Dewitte, O.: Reconstruction of a flash flood event through a multi-hazard approach: focus on the Rwenzori Mountains, Uganda, Natural Hazards, 84, 851–876, 2016.

Kaplan, J. O.: The World Wide Lightning Location Network (WWLLN) Global Lightning Climatology (WGLC) and time series, https://doi.org/10.5281/zenodo.15215319, 2025.

Kaplan, J. O. and Lau, K. H.-K.: World Wide Lightning Location Network (WWLLN) Global Lightning Climatology (WGLC) and time series, 2022 update, Earth System Science Data, 14, 5665–5670, https://doi.org/10.5194/essd-14-5665-2022, 2022a.

Kaplan, J. O. and Lau, K. H.-K.: The HistLight global lightning stroke density reconstruction (1836–2015), https://doi.org/https://doi.org/10.5281/zenodo.6405395, 2022b.

Karevan, Z. and Suykens, J. A.: Transductive LSTM for time-series prediction: An application to weather forecasting, Neural Networks, 125, 1–9, 2020.

Khairoun, A., Mouillot, F., Chen, W., Ciais, P., and Chuvieco, E.: Coarse-resolution burned area datasets severely underestimate fire-related forest loss, Science of the Total Environment, 920, 170 599, 2024.

Lampe, S.: VUB-HYDR/BuRNN: Version 1.1, https://doi.org/10.5281/zenodo.17834206, 2025a.

Lampe, S.: BuRNN: A Data-Driven Fire Model, https://doi.org/10.5281/zenodo.17778519, 2025b.

Lange, S.: Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0), Geoscientific Model Development, 12, 3055–3070, https://doi.org/10.5194/gmd-12-3055-2019, 2019.

Lange, S., Volkholz, J., Geiger, T., Zhao, F., Vega, I., Veldkamp, T., Reyer, C. P., Warszawski, L., Huber, V., Jägermeyr, J., et al.: Projecting exposure to extreme climate impact events across six event categories and three spatial scales, Earth's Future, 8, e2020EF001 616, 2020.

Lange, S., Menz, C., Gleixner, S., Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Schmied, H. M., Hersbach, H., Buontempo, C., and Cagnazzo, C.: WFDE5 over land merged with ERA5 over the ocean (W5E5 v2.0), https://doi.org/10.48364/ISIMIP.342217, 2021.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., Van Kampenhout, L., Kennedy, D., et al.: The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty, Journal of Advances in Modeling Earth Systems, 11, 4245–4287, 2019.

Lawrence, P. J. and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), Journal of Geophysical Research: Biogeosciences, 112, 2007.

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., and Bretagnolle, V.: Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation, Global Ecology and Biogeography, 23, 811–820, https://doi.org/https://doi.org/10.1111/geb.12161, 2014.

Lehmann, C. E., Anderson, T. M., Sankaran, M., Higgins, S. I., Archibald, S., Hoffmann, W. A., Hanan, N. P., Williams, R. J., Fensham, R. J., Felfili, J., et al.: Savanna vegetation-fire-climate relationships differ among continents, Science, 343, 548–552, 2014.

Li, F., Val Martin, M., Andreae, M. O., Arneth, A., Hantson, S., Kaiser, J. W., Lasslop, G., Yue, C., Bachelet, D., Forrest, M., Kluzek, E., Liu, X., Mangeon, S., Melton, J. R., Ward, D. S., Darmenov, A., Hickler, T., Ichoku, C., Magi, B. I., Sitch, S., van der Werf, G. R., Wiedinmyer, C., and Rabin, S. S.: Historical (1700–2012) global multi-model estimates of the fire emissions from the Fire Modeling Intercomparison Project (FireMIP), Atmospheric Chemistry and Physics, 19, 12 545–12 567, https://doi.org/10.5194/acp-19-12545-2019, 2019.

Lizundia-Loiola, J., Otón, G., Ramo, R., and Chuvieco, E.: A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data, Remote Sensing of Environment, 236, 111 493, 2020.

Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems, edited by Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., vol. 30, Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf, 2017.

McLauchlan, K. K., Higuera, P. E., Miesel, J., Rogers, B. M., Schweitzer, J., Shuman, J. K., Tepley, A. J., Varner, J. M., Veblen, T. T., Adalsteinsson, S. A., Balch, J. K., Baker, P., Batllori, E., Bigio, E., Brando, P., Cattau, M., Chipman, M. L., Coen, J., Crandall, R., Daniels, L., Enright, N., Gross, W. S., Harvey, B. J., Hatten, J. A., Hermann, S., Hewitt, R. E., Kobziar, L. N., Landesmann, J. B., Loranty, M. M., Maezumi, S. Y., Mearns, L., Moritz, M., Myers, J. A., Pausas, J. G., Pellegrini, A. F. A., Platt, W. J., Roozeboom, J., Safford, H., Santos, F., Scheller, R. M., Sherriff, R. L., Smith, K. G., Smith, M. D., and Watts, A. C.: Fire as a fundamental ecological process: Research advances and frontiers, Journal of Ecology, 108, 2047–2069, https://doi.org/https://doi.org/10.1111/1365-2745.13403, 2020.

Meyer, H., Reudenbach, C., Wöllauer, S., and Nauss, T.: Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction, Ecological Modelling, 411, 108 815, https://doi.org/https://doi.org/10.1016/j.ecolmodel.2019.108815, 2019.

Otón, G., Lizundia-Loiola, J., Pettinari, M. L., and Chuvieco, E.: Development of a consistent global long-term burned area product (1982–2018) based on AVHRR-LTDR data, International Journal of Applied Earth Observation and Geoinformation, 103, 102 473, 2021.

Park, C. Y., Takahashi, K., Fujimori, S., Jansakoo, T., Burton, C., Huang, H., Kou-Giesbrecht, S., Reyer, C. P., Mengel, M., Burke, E., et al.: Attributing human mortality from fire PM2. 5 to climate change, Nature Climate Change, pp. 1–8, 2024.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, https://arxiv.org/abs/1912.01703, 2019.

Picotte, J. J., Bhattarai, K., Howard, D., Lecker, J., Epting, J., Quayle, B., Benson, N., and Nelson, K.: Changes to the Monitoring Trends in Burn Severity program mapping production procedures and data products, Fire Ecology, 16, 1–12, 2020.

Qi, D. and Majda, A. J.: Using machine learning to predict extreme events in complex systems, Proceedings of the National Academy of Sciences, 117, 52–59, 2020.

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

Reddy, D. S. and Prasad, P. R. C.: Prediction of vegetation dynamics using NDVI time series data and LSTM, Modeling Earth Systems and Environment, 4, 409–419, 2018.

Rodrigues, M., San Miguel, J., Oliveira, S., Moreira, F., and Camia, A.: An insight into spatial-temporal trends of fire ignitions and burned areas in the European Mediterranean countries, Journal of Earth Science and Engineering, 3, 497, 2013.

Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature machine intelligence, 1, 206–215, 2019.

Santurkar, S., Tsipras, D., Ilyas, A., and Madry, A.: How does batch normalization help optimization?, Advances in neural information processing systems, 31, 2018.

Sauer, I., Koch, J., Otto, C., and Volkholz, J.: ISIMIP3a GDP input data, https://doi.org/10.48364/ISIMIP.824555.1, 2024.

Shakesby, R. A.: Post-wildfire soil erosion in the Mediterranean: Review and future research directions, Earth-Science Reviews, 105, 71–100, 2011.

Short, K. C.: Spatial wildfire occurrence data for the United States, 1992-2020 [FPA_FOD_20221014], 2022.

Simon, M. F., Grether, R., de Queiroz, L. P., Skema, C., Pennington, R. T., and Hughes, C. E.: Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire, Proceedings of the National Academy of Sciences, 106, 20 359–20 364, 2009.

Simpson, K. J., Archibald, S., and Osborne, C. P.: Savanna fire regimes depend on grass trait diversity, Trends in Ecology & Evolution, 37, 749–758, 2022.

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J., McColl, C., Allan, R., Brohan, P., Yin, X., Smith, C., Spencer, L., et al.: An evaluation of the performance of the twentieth century reanalysis version 3, Journal of Climate, 34, 1417–1438, 2021.

Souza Jr, C. M., Z. Shimbo, J., Rosa, M. R., Parente, L. L., A. Alencar, A., Rudorff, B. F., Hasenack, H., Matsumoto, M., G. Ferreira, L., Souza-Filho, P. W., et al.: Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine, Remote Sensing, 12, 2735, 2020.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research, 15, 1929–1958, 2014.

Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic attribution for deep networks, in: International conference on machine learning, pp. 3319–3328, PMLR, 2017a.

Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic attribution for deep networks, in: International conference on machine learning, pp. 3319–3328, PMLR, 2017b.

Takacs, S., Schulte to Bühne, H., and Pettorelli, N.: What shapes fire size and spread in African savannahs?, Remote Sensing in Ecology and Conservation, 7, 610–620, 2021.

Teckentrup, L., Harrison, S. P., Hantson, S., Heil, A., Melton, J. R., Forrest, M., Li, F., Yue, C., Arneth, A., Hickler, T., et al.: Response of simulated burned area to historical changes in environmental and anthropogenic factors: a comparison of seven fire models, Biogeosciences, 16, 3883–3910, 2019.

Thiery, W., Lange, S., Rogelj, J., Schleussner, C.-F., Gudmundsson, L., Seneviratne, S. I., Andrijevic, M., Frieler, K., Emanuel, K., Geiger, T., et al.: Intergenerational inequities in exposure to climate extremes, Science, 374, 158–160, 2021.

Turco, M., Bedia, J., Di Liberto, F., Fiorucci, P., von Hardenberg, J., Koutsias, N., Llasat, M.-C., Xystrakis, F., and Provenzale, A.: Decreasing fires in mediterranean Europe, PLoS one, 11, e0150 663, 2016.

van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997–2016, Earth System Science Data, 9, 697–720, https://doi.org/10.5194/essd-9-697-2017, 2017.

van Wagner, C.: Development and structure of the Canadian forest fire weather index system, 35, Canadian Forestry Service, 1987.

Veraverbeke, S., Verstraeten, W. W., Lhermitte, S., Van De Kerchove, R., and Goossens, R.: Assessment of post-fire changes in land surface temperature and surface albedo, and their relation with fire–burn severity using multitemporal MODIS imagery, International Journal of Wildland Fire, 21, 243–256, 2012.

Volkholz, J. and Ostberg, S.: ISIMIP3a landuse input data, https://doi.org/10.48364/ISIMIP.571261.1, 2022.

Volkholz, J., Lange, S., Sauer, I., and Otto, C.: ISIMIP3a population input data, https://doi.org/10.48364/ISIMIP.822480.3, 2024.

Ward, D., Kloster, S., Mahowald, N., Rogers, B., Randerson, J., and Hess, P.: The changing radiative forcing of fires: global model estimates for past, present and future, Atmospheric Chemistry and Physics, 12, 10 857–10 886, 2012.

Yang, R., Hu, J., Li, Z., Mu, J., Yu, T., Xia, J., Li, X., Dasgupta, A., and Xiong, H.: Interpretable machine learning for weather and climate prediction: A review, Atmospheric Environment, p. 120797, 2024.

Zhu, C., Kobayashi, H., Kanaya, Y., and Saito, M.: Size-dependent validation of MODIS MCD64A1 burned area over six vegetation types in boreal Eurasia: Large underestimation in croplands, Scientific Reports, 7, 4181, 2017.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., et al.: A typology of compound weather and climate events, Nature reviews earth & environment, 1, 333–347, 2020.