Supplement of Geosci. Model Dev., 16, 869–884, 2023
https://doi.org/10.5194/gmd-16-869-2023-supplement

Geoscientific Model Development
Open Access

Supplement of

AttentionFire_v1.0: interpretable machine learning fire model for burned-area predictions over tropics

Fa Li et al.

Correspondence to: Qing Zhu (qzhu@lbl.gov)

Supplementary Information

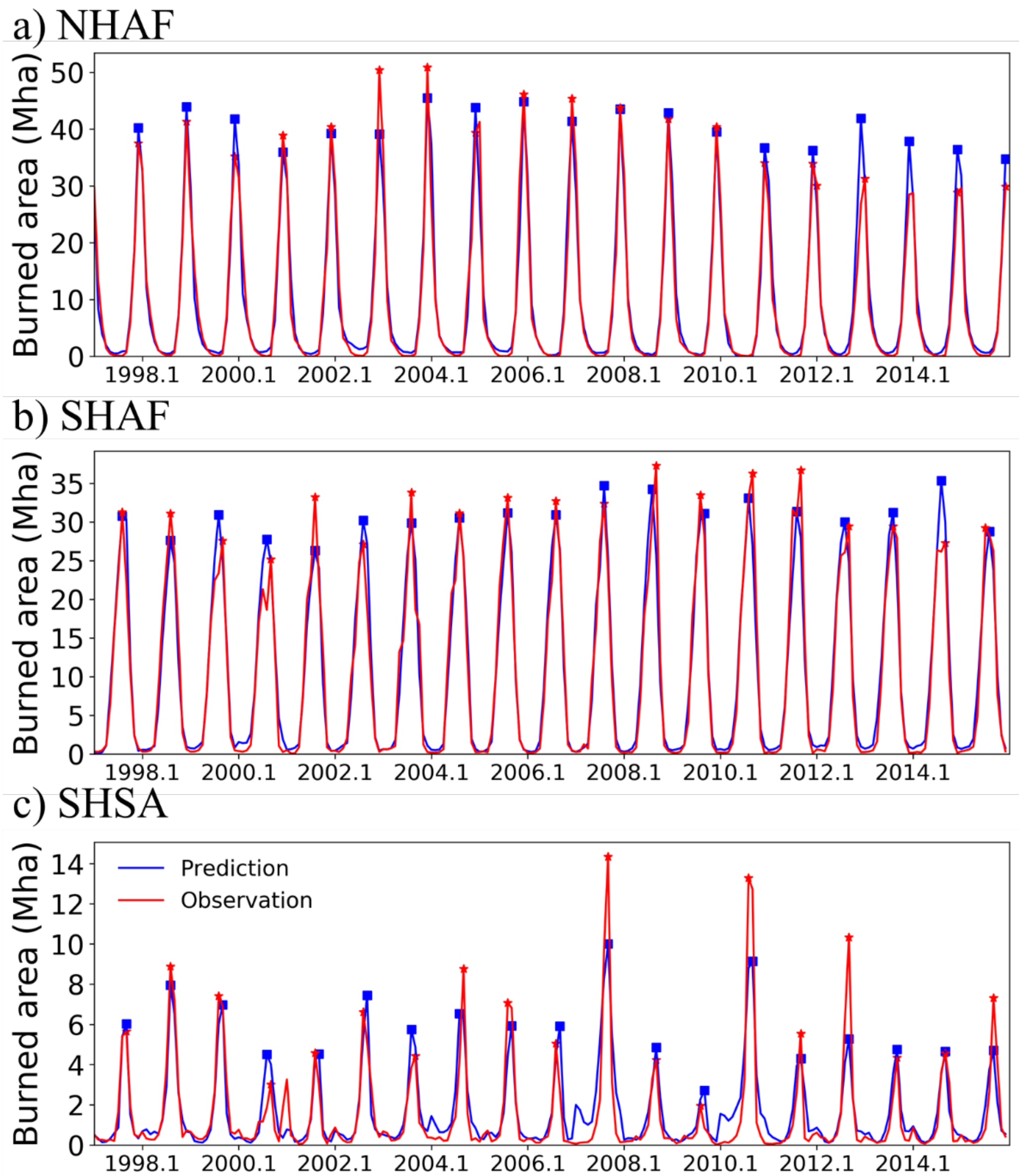

Fig. S1: Observed and AttentionFire modeled monthly total burned areas in NHAF, SHAF, and SHSA regions from 1997-2015. Peak fire month in each year and its corresponding burned areas are marked with red star (observations) and blue square (AttentionFire) markers.

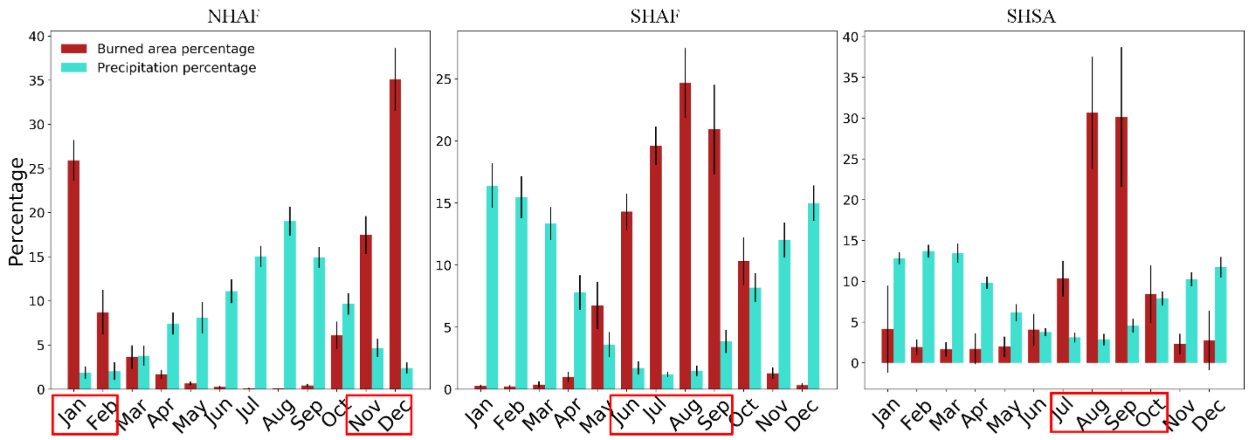

Fig. S2: Monthly mean precipitation and burned area percentage of yearly total amount. The months with top four largest burned areas are defined as the fire season (red box). Fire season accounts for 87.2%, 79.5%, and 79.6% of yearly total burned areas over NHAF, SHAF, and SHSA, respectively.

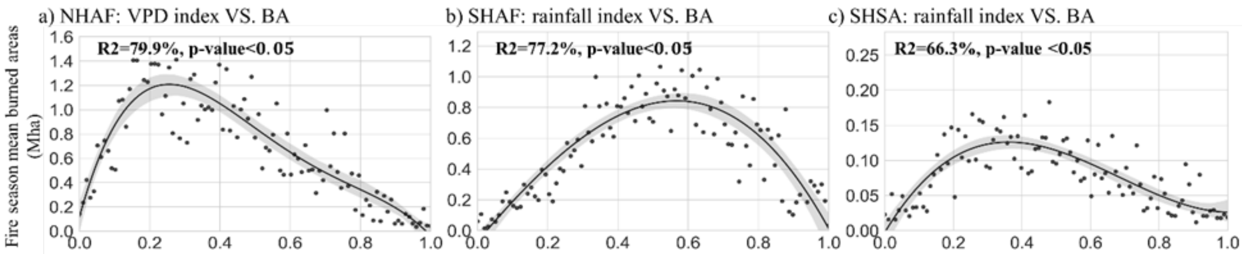

Fig. S3: Dependency between fire season mean burned area and rainfall or VPD scalars (standardized) in NHAF, SHAF, and SHSA regions. The x-axis is the weighted sum of each driving variable across time. The weight of each month is the calculated mean temporal attention weight of the driver at the corresponding month. The x-axis is evenly divided into 100 bins according to the range of weighted sum rainfall or VPD of different grids in each studied region and normalized to the range from 0 to 1. The fire season mean burned area in each bin is calculated. Each point in the figure represents the fire season mean burned area in the corresponding rainfall or VPD bin. The coefficient of determination (R^2) is the explained variance of polynomial fitted fire season mean burned areas and observations.

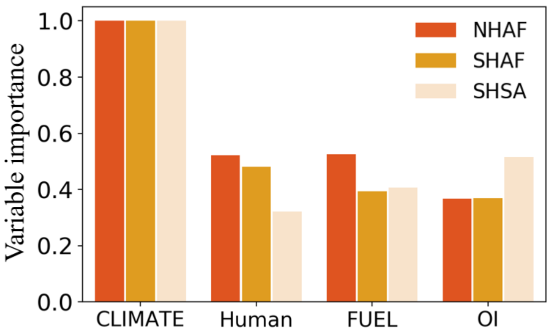

Fig. S4. Variable importance for fire-season burned area in Northern Hemisphere Africa

(NHAF), Southern Hemisphere Africa (SHAF), and Southern Hemisphere South America (SHSA). All drivers are divided into four categories: climate, human activities (e.g., population and road density, livestock), fuel, and ocean indices (OI). The importance score is normalized by the dominant variable importance score, and a larger score represents a more important variable.

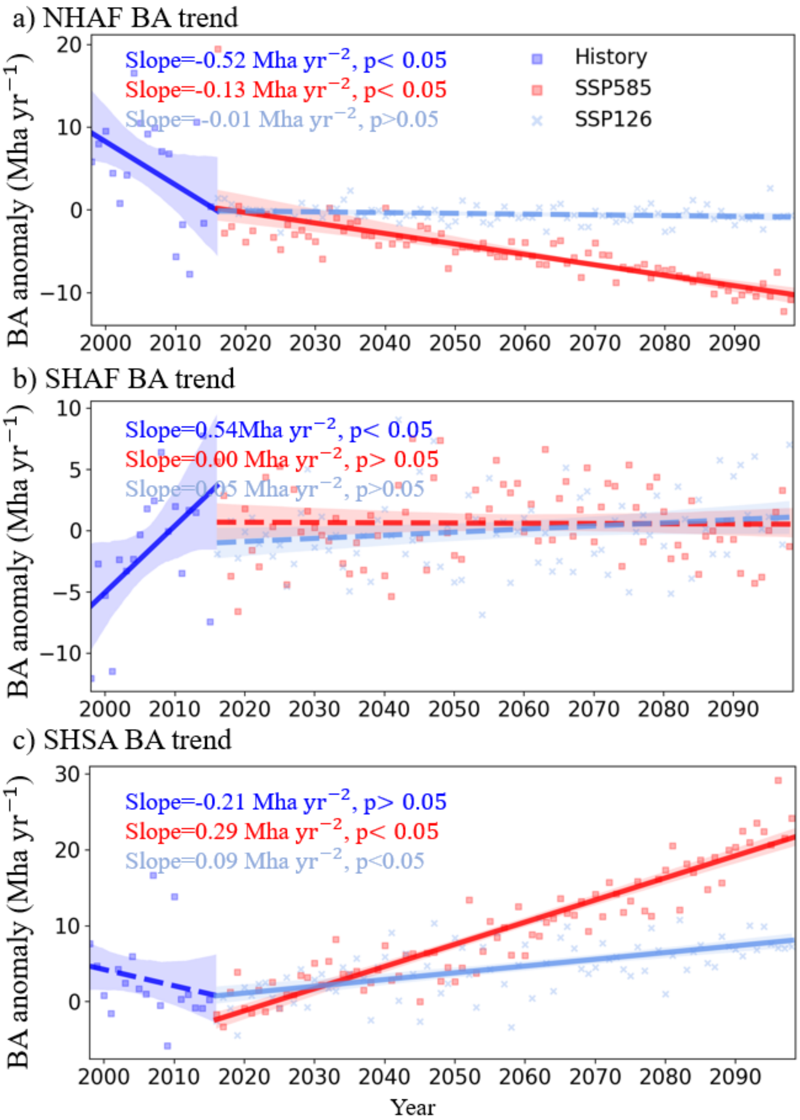

Fig. S5. Burned area changes in history and future. Deep blue and light blue lines represent burned area changes under SSP585 and SSP126, respectively, and red lines represent burned area changes in history. Solid lines represented significant (p <0.05) burned area trends while dashed lines represented non-significant trends.

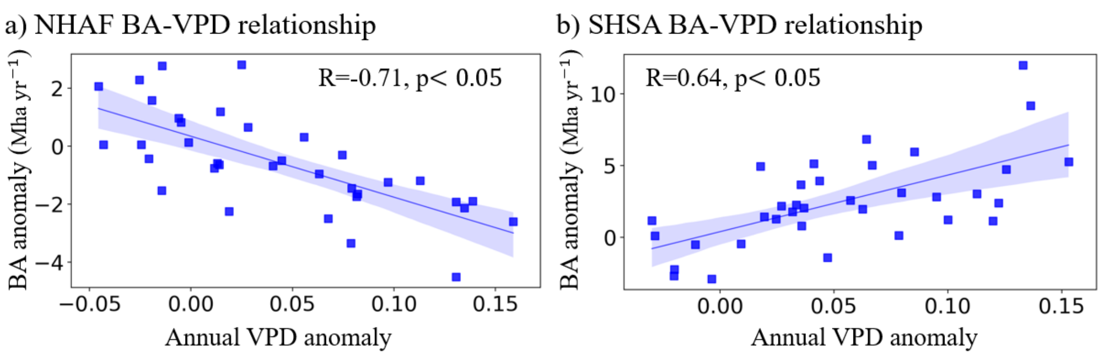

Fig. S6. Regression relationship between burned area changes and vapor pressure deficit changes in Northern Hemisphere Africa (NHAF), and Southern Hemisphere South America (SHSA) region.

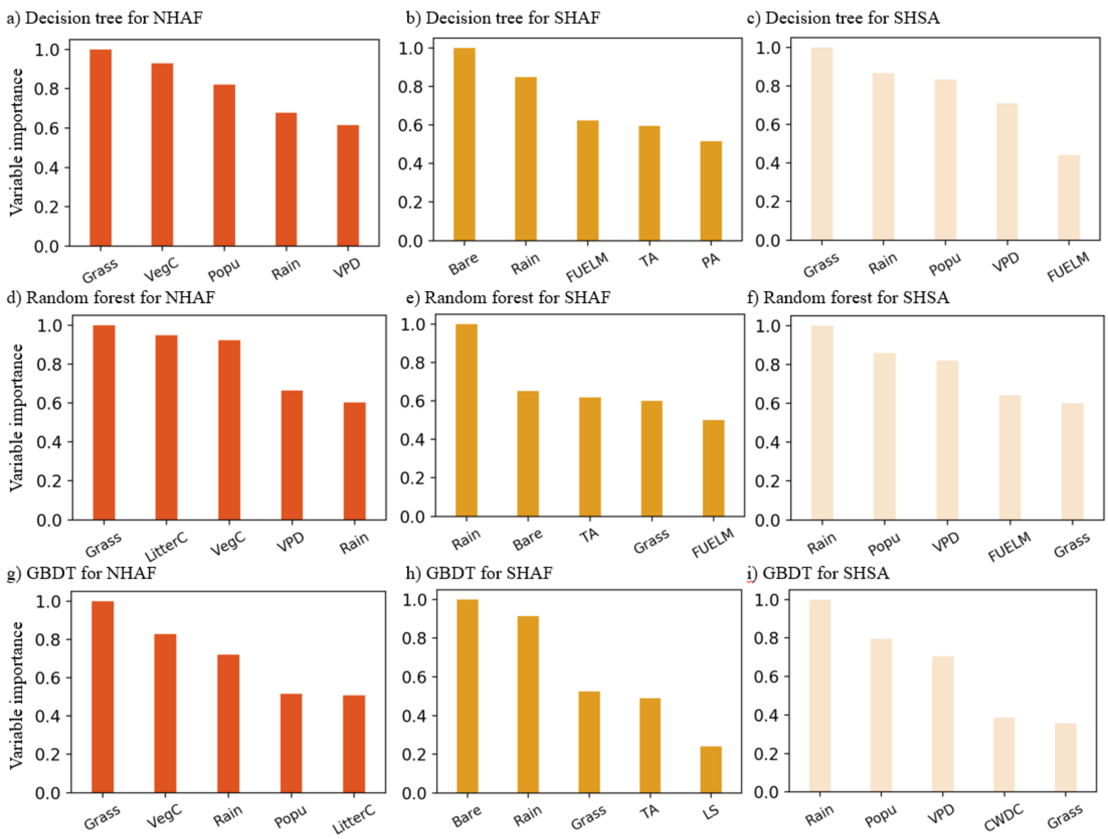

Fig. S7. Ranked top-five important variables for burned area in Northern Hemisphere Africa (NHAF) (a), Southern Hemisphere Africa (SHAF) (b), and Southern Hemisphere South America (SHSA) (c). For each region, the variable importance was normalized by the dominant variable importance. Three machine learning models that provide variable importance score are considered, including decision tree, random forest, and Gradient Boosting Decision Tree (GBDT). The full name of each abbreviated variable is listed in Table 2 in the main text.

Table S1. Model Hyperparameter settings

Model	Hyperparameter settings
Random Forest	Minimum leaf sample: [3,6,9] Number of trees: [20,30,40]
Decision Tree	Minimum leaf sample: [3,6,9] Maximum depth: 150
Gradient Boosting Decision Tree (GBDT)	Learning rate: 0.01 Maximum depth: [3,4,5] Number of trees: 100
ANN	Maximum iteration: 2,000 Number of neuros in hidden layers: [30,10] Batch size: 32 Activation: RELU Optimizer: SGD
LSTM	Dimension of hidden state vector: [8, 12, 16] Learning rate: initial value 0.01, and update by multiplying 0.8 each step. Batch size: 32 Optimizer: Adam with weight decay rate [10^{-3}, 10^{-4}, 5×10^{-4}, 10^{-5}] Sequence length: 12 Dropout rate: 0.1
AttentionFire	Dimension of hidden state vector: [8, 12, 16] Learning rate: initial value 0.01, and update by multiplying 0.8 each step. Batch size: 32 Optimizer: Adam with weight decay rate [10^{-3}, 10^{-4}, 5×10^{-4}, 10^{-5}] Sequence length: 12 Dropout rate: 0.1

Table S2. Model maximum computational cost across the three studied regions

Model	Memory usage (GB)	Time consumption (seconds)
Random Forest	3.6	445
Decision Tree		25
Gradient Boosting Decision Tree (GBDT)	4.5	552
ANN	2.2	137
LSTM	4.7	96
AttentionFire	5.3	568

Note: the computational cost is the maximum cost for wildfire model training in the three studied regions, including Northern Hemisphere Africa (NHAF), Southern Hemisphere Africa (SHAF), and Southern Hemisphere South America (SHSA). We acknowledge that the memory usage and time consumption could be different with different computational settings (e.g., GPU versus CPU, number of computational nodes, CPU frequency, and Python package versions) and different data size in the three regions. Since some baseline models (e.g., random forest and decision tree) were commonly trained on the CPU, the results here were derived with single core of Intel® Core™ i9-9900K CPU (3.60 GHz) instead of GPU, Python 3.7, and the maximum computational cost over the three regions.