Supplementary Information

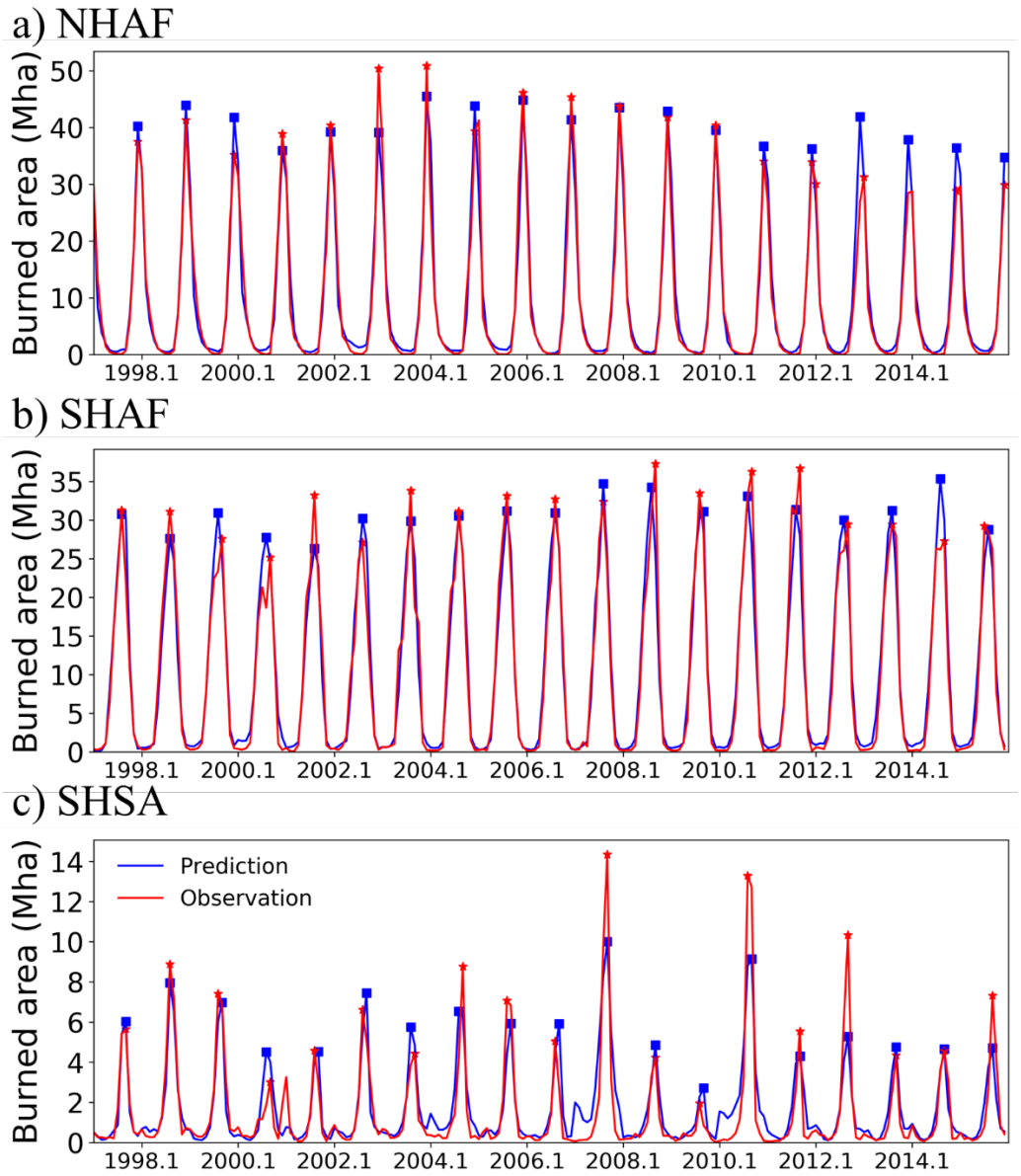

Fig. S1: Observed and AttentionFire modeled monthly total burned areas in NHAF, SHAF, and SHSA regions from 1997-2015. Peak fire month in each year and its corresponding burned areas are marked with red star (observations) and blue square (AttentionFire) markers.

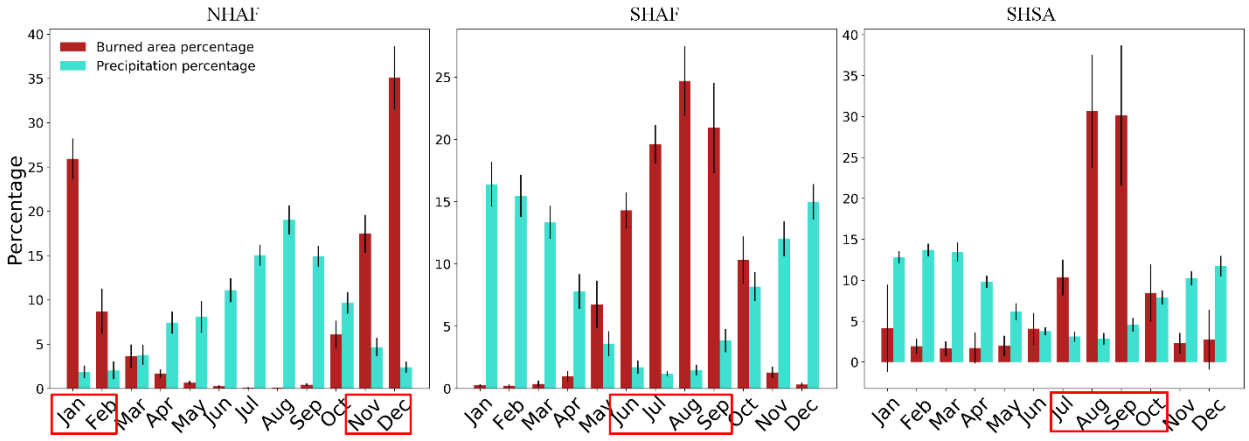

Fig. S2: Monthly mean precipitation and burned area percentage of yearly total amount. The months with top four largest burned areas are defined as the fire season (red box). Fire season accounts for 87.2%, 79.5%, and 79.6% of yearly total burned areas over NHAF, SHAF, and SHSA, respectively.

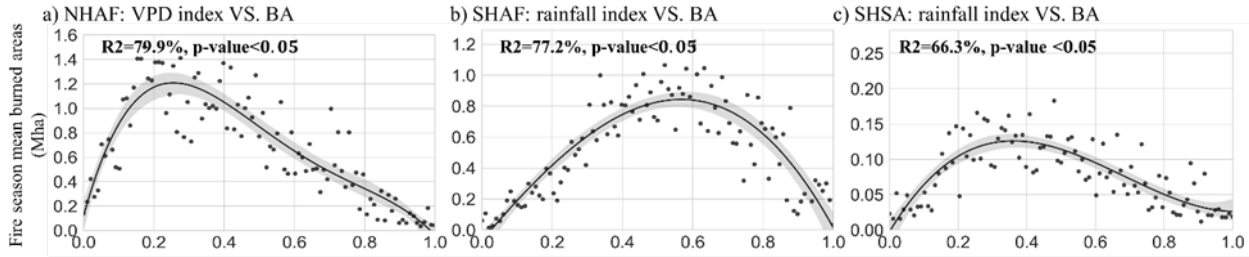

Fig. S3: Dependency between fire season mean burned area and rainfall or VPD scalars (standardized) in NHAF, SHAF, and SHSA regions. The x-axis is the weighted sum of each driving variable across time. The weight of each month is the calculated mean temporal attention weight of the driver at the corresponding 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.

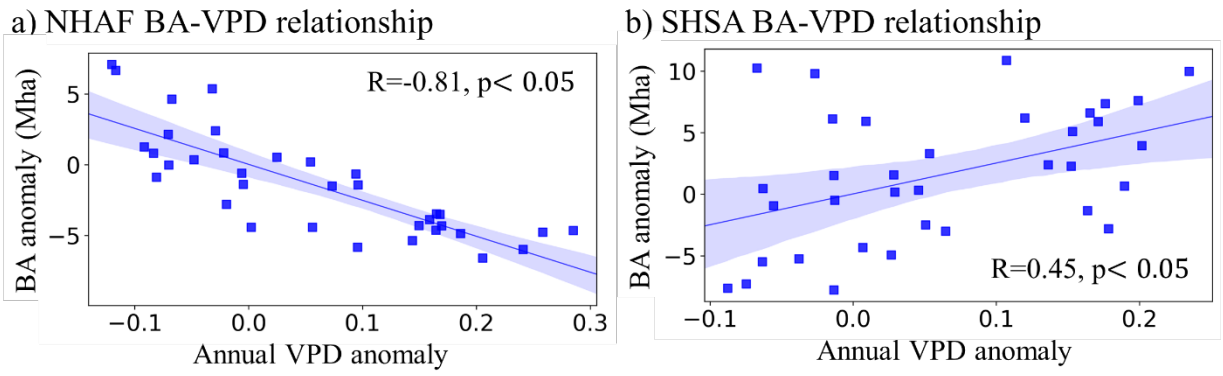

Fig. S4. Regression relationship between Burned area changes and vapor pressure deficit changes.

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. Datasets

Variables	dataset
Burned Area	Global Fire Emissions Database 4[3]: https://daac.ornl.gov/VEGETATION/guides/fire_emissions_v4.html
Climate	Precipitation (RAIN), temperature (TSA), surface air pressure (PA), specific humidity (SH), downward short-wave radiation (SW), wind speed (WIND) from NCEP-DOE Reanalysis 2[4]: https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html
Fuel condition and availability	Fuel moisture, coarse wood debris, total vegetation biomass, litter biomass are from ELM prognostic simulations [5]
Population	Population density [6]
Road density	Global maps of road density [7]
Livestock	Global maps of livestock density [8]
Land cover	Bare soil, Forest and, Grass percentage are from LUH2 [9]
VPD	Calculated based on air temperature, air pressure, and specific humidity[10]
Oceanic index	Niño Index (ONI), Atlantic multidecadal Oscillation (AMO) index, Tropical Northern Atlantic (TNA) Index, and Tropical Southern Atlantic (TSA) Index are from: https://psl.noaa.gov/data/climateindices/list/

Table S3. Ranked top-five important variables for future burned area changes in NHAF and SHSA.

NHAF		SHAF		SHSA	
Variable	Importance on trend	Variable	Importance on trend	Variable	Importance on trend
VPD	0.17	Population	0.35	VPD	0.35
Land cover	0.16	Temperature	0.16	Temperature	0.16
Temperature	0.11	VPD	0.15	Solar radiation	0.10
Population	0.10	Radiation	0.02	Precipitation	0.09
Precipitation	0.02	Land cover	0.01	Land cover	0.05

Note: A larger value in the table represents a more important variable for future burned area (BA) changes. To analyze interannual variations of a specific variable on future BA changes, firstly we iteratively surrogated the specific variable with its climatology while keeping the other variables the same; then we fed the new datasets into the model and reprojected the BA. We calculated the slope of BA trend in each grid of each region, and calculated the Pearson correlation coefficient (R) between the gridded slopes of predictions with raw datasets and surrogated datasets. If the coefficient R was larger, it means the surrogated variable was unimportant, therefore, we used 1-R (the value in the table) to show the importance of a specific variable on future BA changes (a larger value represents more important for future BA changes).

References

1 Chen, Y. *et al.* Forecasting fire season severity in South America using sea surface temperature anomalies. *Science* **334**, 787-791 (2011).

2 Yu, Y. *et al.* Quantifying the drivers and predictability of seasonal changes in African fire. *Nature communications* **11**, 1-8 (2020).

3 Giglio, L., Randerson, J. T. & Van Der Werf, G. R. Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *Journal of Geophysical Research: Biogeosciences* **118**, 317-328 (2013).

4 Kanamitsu, M. *et al.* Ncep–doe amip-ii reanalysis (r-2). *Bulletin of the American Meteorological Society* **83**, 1631-1644 (2002).

5 Zhu, Q. *et al.* Representing nitrogen, phosphorus, and carbon interactions in the E3SM Land Model: Development and global benchmarking. *Journal of Advances in Modeling Earth Systems*, doi: 10.1029/2018MS001571 (2019).

6 Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C. & Worley, B. A. LandScan: a global population database for estimating populations at risk. *Photogrammetric engineering and remote sensing* **66**, 849-857 (2000).

7 Meijer, J. R., Huijbregts, M. A., Schotten, K. C. & Schipper, A. M. Global patterns of current and future road infrastructure. *Environmental Research Letters* **13**, 064006 (2018).

8 Rothman-Ostrow, P., Gilbert, W. & Rushton, J. Tropical Livestock Units: Re-evaluating a Methodology. *Frontiers in Veterinary Science* **7**, 973 (2020).

9 Hurtt, G. C. *et al.* Harmonization of global land-use change and management for the period 850–2100 (LUH2) for CMIP6. *Geoscientific Model Development Discussions*, 1-65 (2020).

10 Bolton, D. The computation of equivalent potential temperature. *Monthly weather review* **108**, 1046-1053 (1980).