Supplement of Atmos. Chem. Phys., 22, 3445–3468, 2022
https://doi.org/10.5194/acp-22-3445-2022-supplement

Atmospheric Chemistry and Physics
Open Access

Supplement of

Interpreting machine learning prediction of fire emissions and comparison with FireMIP process-based models

Sally S.-C. Wang et al.

Correspondence to: Sally S.-C. Wang (sing-chun.wang@pnnl.gov) and L. Ruby Leung (ruby.leung@pnnl.gov)

Supplement

Table S1. Hyperparameters used and the Grid search ranges tested in the ML model

	Tested ranges	**Final value**
Learning rate	0.1, 0.25, 0.35, 0.5, 0.8	0.35
Max_depth	3, 6, 9, 12	12
subsample	0.4, 0.6, 0.8, 1	1
nrounds	100, 150, 200, 250	150

Table S2. Standard deviation of meteorology and slopes of for the dependence of annual mean fire $PM_{2.5}$ emissions on meteorology for the southeastern US (SEUS) and western US (WUS)

	RH_SEUS	**Temp_SEUS**	**RH_WUS**	**Temp_WUS**
Standard deviation	1.594	1.385	2.038	1.982
GFED_slope	-0.014	0.017	-0.017	0.049
CLM_slope	-0.042	0.119	-0.036	0.064
JULES_slope	-0.038	0.160	-0.11	0.162

Table S3. Mean wind speed of Oct 2003 and 2007 over southwestern California for the two datasets.

Wind speed (m/s)	**Oct 2003**	**Oct 2007**	**Oct mean**
CRUNCEP	3.75	4.98	4.41
NARR	5.07	5.67	5.40

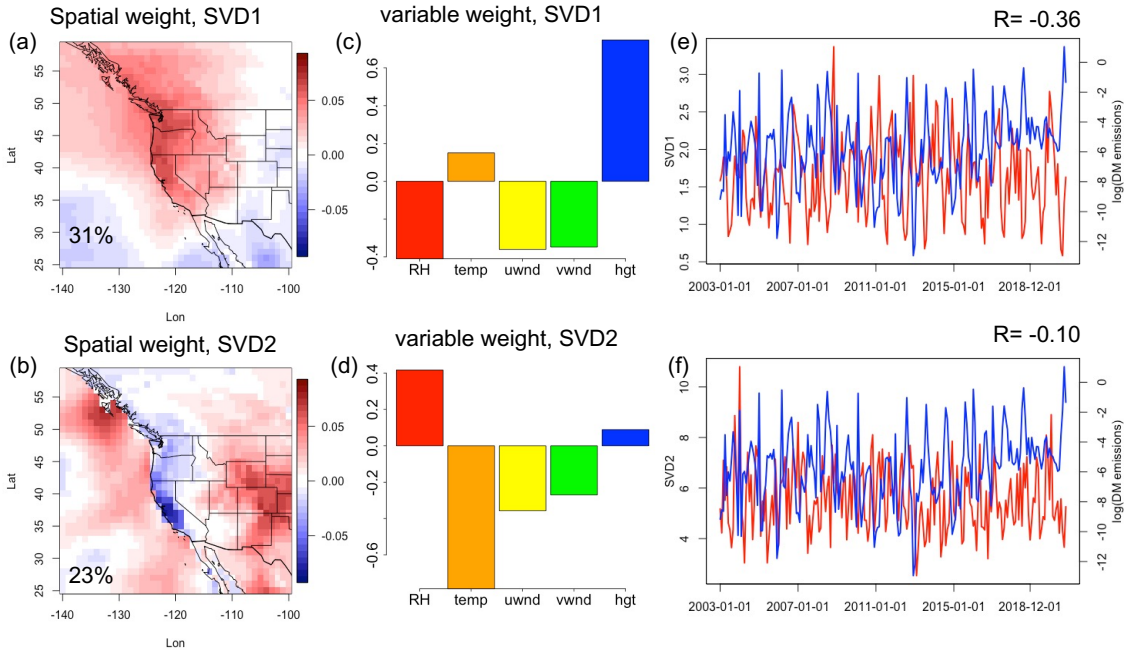

Figure S1. The spatial and variable weights of the first (a, c) and second (b, d) singular value decomposition (SVD) modes describing the spatial correlations between daily mean fire emissions over NCA and meteorological variables in the grid boxes in the western US from 2000 to 2017. The meteorological variables include temperature, relative humidity, precipitation, u and v wind speed. (e) Time series of monthly standard deviation of daily SVD1 (red) and logarithm of monthly total fire dry matter emissions (blue) for NCA. (f) Time series of monthly standard deviation of daily SVD2 (red) and logarithm of monthly total fire dry matter emissions (blue) for NCA.

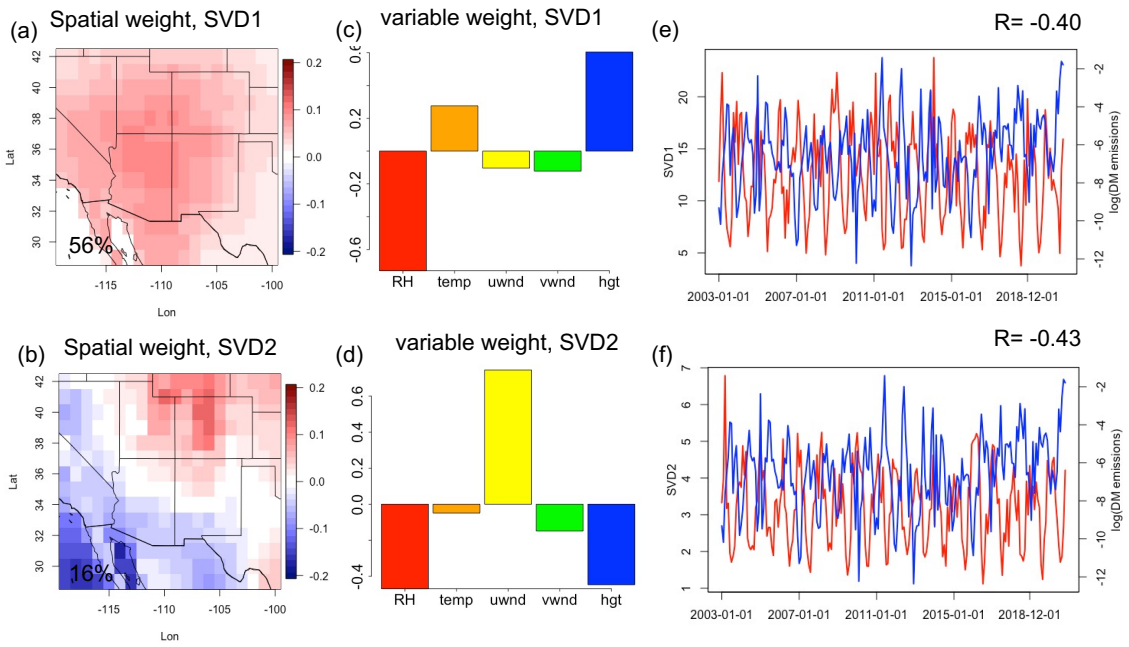

Figure S2. The spatial and variable weights of the first (a, c) and second (b, d) singular value decomposition (SVD) modes describing the spatial correlations between daily mean fire emissions over southern Rocky Mountains (SRM) and meteorological variables in the grid boxes in the southwestern US from 2000 to 2017. The meteorological variables include temperature, relative humidity, precipitation, u and v wind speed. (e) Time series of monthly standard deviation of daily SVD1 (red) and logarithm of monthly total fire dry matter emissions (blue) for SRM. (f) Time series of monthly standard deviation of daily SVD2 (red) and logarithm of monthly total fire dry matter emissions (blue) for SRM.

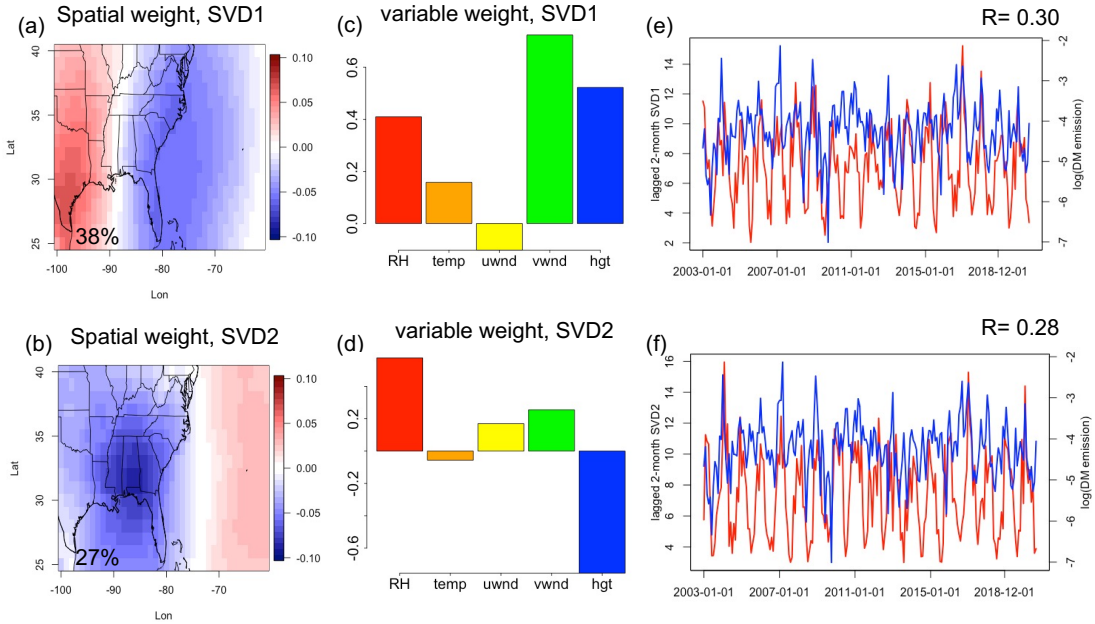

Figure S3. The spatial and variable weights of the first (a, c) and second (b, d) singular value decomposition (SVD) modes describing the spatial correlations between daily mean fire emissions over southeastern US (SEUS) and meteorological variables in the grid boxes in the SEUS from 2000 to 2017. The meteorological variables include temperature, relative humidity, precipitation, u and v wind speed. (e) Time series of monthly standard deviation of daily SVD1 (red) and logarithm of monthly total fire dry matter emissions (blue) for SEUS. (f) Time series of monthly standard deviation of daily SVD2 (red) and logarithm of monthly total fire dry matter emissions (blue) for SEUS.

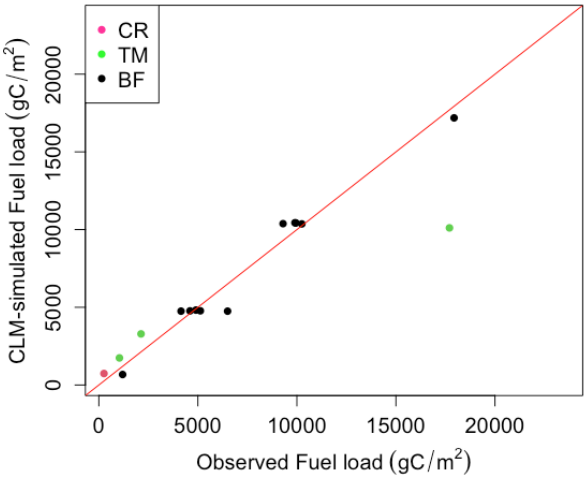

Figure S4. Scatter plot between the observed and CLM-simulated fuel load over CONUS. Three fire types are marked in pink (cropland), green (temperate forest), and black (boreal forest). The data is obtained from fuel consumption database (van der Werf et al., 2017; Van Leeuwen et al., 2014).

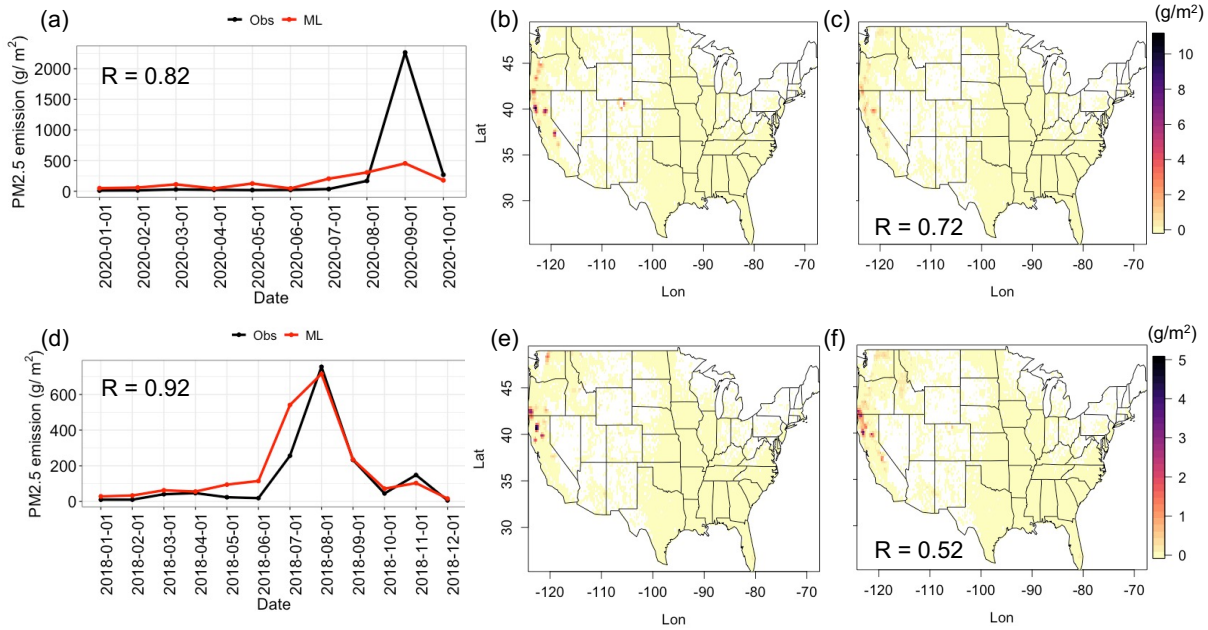

Figure S5. (a, d) Time series of the total fire $PM_{2.5}$ emissions over CONUS and spatial distributions of the annual mean fire $PM_{2.5}$ emission for (b, e) GFED, and (c, f) ML model (top panel: 2020 and bottom panel: 2018).

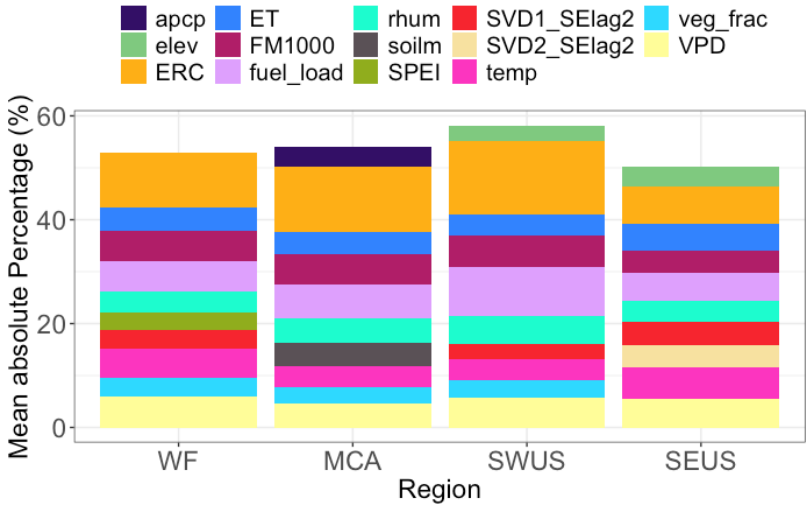

Figure S6. The top 10 variables based on the mean absolute percentage for the selected regions. The mean absolute percentage is calculated by the absolute SHAP value of the variable divided by the sum of the absolute SHAP values of all the variables.

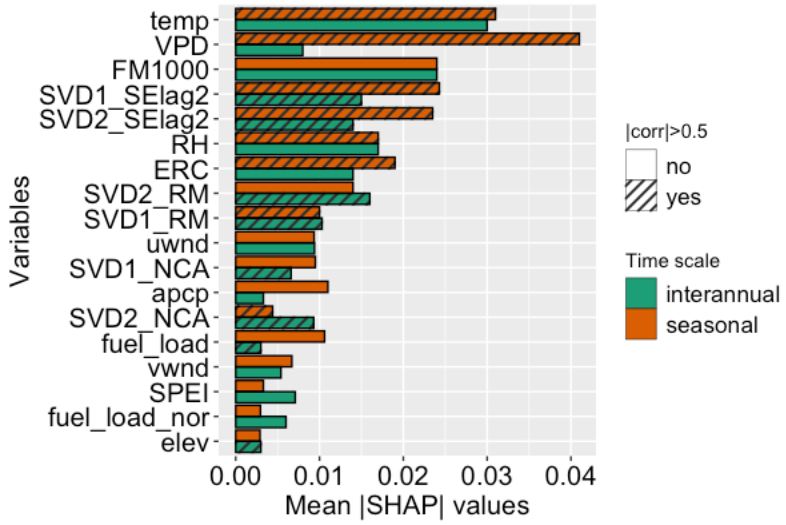

Figure S7. Variable importance in mean |SHAP| values at seasonal and interannual time scale. The variables and time scales which have correlation between time series of fire emissions and SHAP values larger than 0.5 are in stripes.

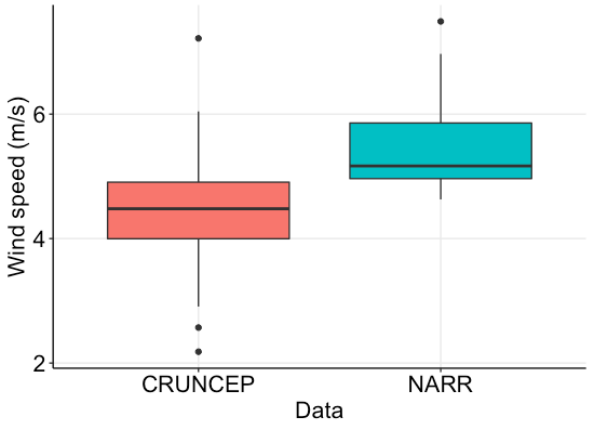

Figure S8. Distribution of wind speed for the days with large wind speeds (>4.5 m/s) in October during 2000-2012 for the CRUNCEP and NARR data. The NARR data is regridded to 0.5º × 0.5º, matching the spatial resolution of the CRUNCEP data.

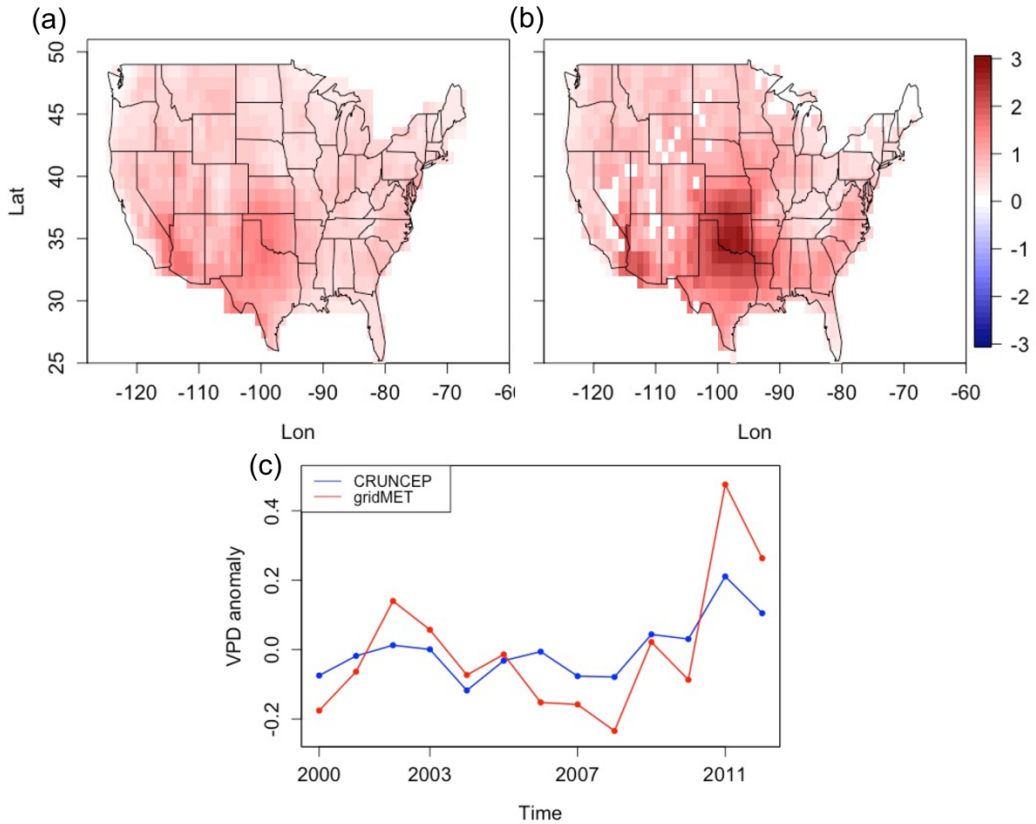

Figure S9. Top panel: Spatial map of VPD anomalies in JJA for (a) CRUNCEP and (b) gridMET. Bottom panel: (c) Time series of VPD anomalies in JJA from 2000 to 2012.

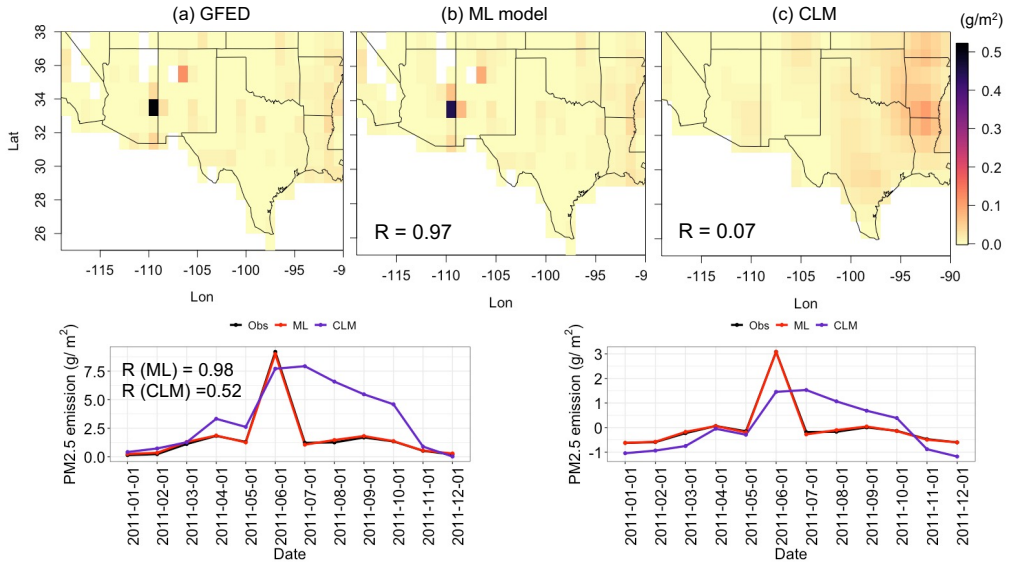

Figure S10. Top panel: Spatial distributions of the annual mean fire $PM_{2.5}$ emission in 2011 for (a) GFED, (b) ML model, and (c) CLM. Bottom panel: Time series of the (d) total fire $PM_{2.5}$ emissions and (e) normalized fire $PM_{2.5}$ emission over southern US domain during 2011.

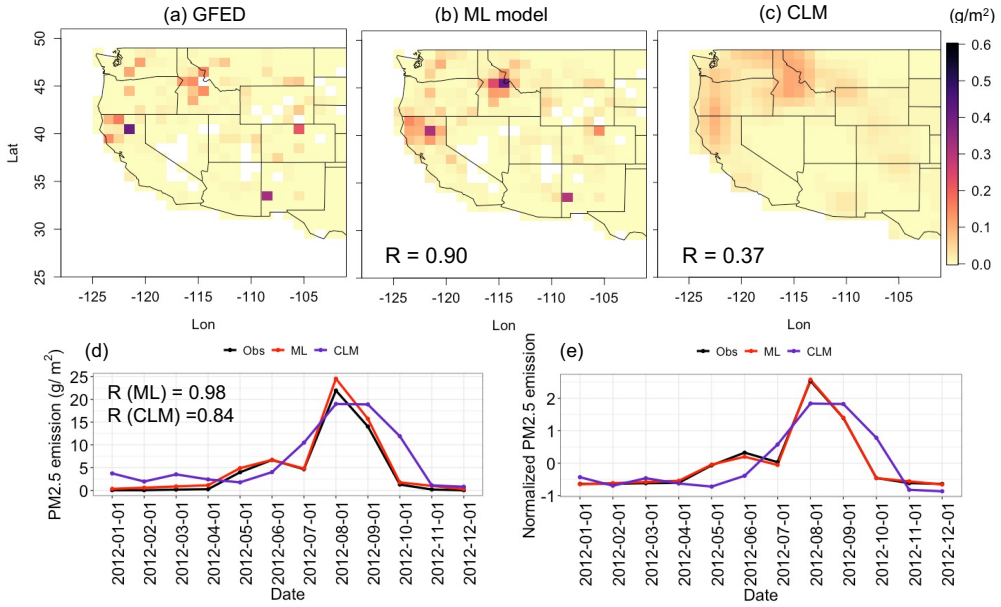

Figure S11. Top panel: Spatial distributions of the annual mean fire $PM_{2.5}$ emission in 2012 for (a) GFED, (b) ML model, and (c) CLM. Bottom panel: Time series of the (d) total fire $PM_{2.5}$ emissions and (e) normalized fire $PM_{2.5}$ emission over western US domain during 2012.