Supplementary Information for Tackling Key Drivers and Predicting Fires in Boreal Peatland with a Two-step Machine Learning Framework

1. Data and validation metrics

1.1 Data List

Table S1 Data information for the response variables and explanatory variables

Groups	Variables	Datasets	Time Span and Resolution	Spatial Resolution	Citations
Response variables					
Fire Impacts	Burned area	GFED4.1s, FireCCI5.1	1997-2016, monthly	0.25x0.25	(Chuvieco et al., 2018; Randerson et al., 2015)
	C emission	GFED4.1s	1997-2016, monthly	0.25x0.25	(Randerson et al., 2015)
	Burn date	MCD45A1, MCD64A1	2001-2018	500m	(Giglio et al., 2018; Roy et al., 2008)
Explanatory variables					
Atmospheric	near-surface temperature (TMP)	CRU_ts4.04	1901-2018, monthly	0.5x0.5	(Harris et al., 2020)
	near-surface temperature minimum (TMN)	CRU_ts4.04	1901-2018, monthly	0.5x0.5	
	near-surface temperature maximum (TMX)	CRU_ts4.05	1901-2018, monthly	0.5x0.5	
	diurnal temperature range (DTR)	CRU_ts4.06	1901-2018, monthly	0.5x0.5	
	Precipitation (PRE)	CRU_ts4.04	1901-2019, monthly	0.5x0.5	
	Evapotranspiration (ET)	CRU_ts4.04	1901-2018, monthly	0.5x0.5	
	wet day frequency (WET)	CRU_ts4.04	1901-2018, monthly	0.5x0.5	
	vapor pressure (VAP)	CRU_ts4.04	1901-2018, monthly	0.5x0.5	
	cloud cover percentage (CLD)	CRU_ts4.04	1901-2018, monthly	0.5x0.5	
	ground frost frequency (FRT)	CRU_ts4.04	1901-2018, monthly	0.5x0.5	
	Palmer Drought Severity Index (PDSI)	CRU_ts4.04	1901-2018, monthly	0.5x0.5	
	Saturated vapor pressure	$6.112 \times e^{\frac{(22.46\ + TMP)}{(272.62 + TMP)}}$	1901-2018, monthly	0.5x0.5	(World Meteorological Organization, 2008)
	relative humidity (RH)	(VAP / SVP) x 100	1901-2018, monthly	0.5x0.5	
	Vapor pressure deficit (VPD)	SVP - VAP	1901-2018, monthly	0.5x0.5	
	2-m windspeed (WIN)	MERRA2	1980-2020, 1h	0.5 x 0.625	(World Meteorological Organization, 2008)
vegetation	GPP	Madani et al.-2020	1982-2016, monthly	0.083x0.083	(Madani and Parazoo, 2020)
	NDVI	GIMMS3g	1982-2015, monthly	0.083x0.083	(Pinzon and Tucker, 2014)

Table S1 continued

Groups	Variables	Datasets	Time Span and Resolution	Spatial Resolution	Citations
Soil	Soil moisture (SMroot and SMsurf)	GLEAM v3.3a, v3.3b, ECMWF	1980-2018, monthly	0.5x0.5	(Martens et al., 2017)
	Northern Peatland	Hugelius-2020	one period	10km	(Hugelius et al., 2020)
Socioeconomic	Population density (POPD)	HYDE v3.2	10000BCE - 2015CE	0.083x0.083	(Klein Goldewijk et al., 2017)

1.2 constructed climate variables

$$Saturated\ vapor\ pressure(SVP) = 6.112 \times e^{\frac{(22.46\ +TMP)}{(272.62+TMP)}}, \quad \text{(S1)}$$

$$relative\ humidity\ (HR) = \frac{VAP}{SAP} \times 100\%, \quad \text{(S2)}$$

$$vapor\ pressure\ deficit(VPD) = SVP - VAP, \quad \text{(S3)}$$

The MERRA-2 2-meter wind-speed product includes the eastward wind (U2M) and northward wind (V2M), whose synthetic wind-speed is calculated as:

$$Windspeed(WSP) = \sqrt{(U2M^2 + V2M^2)}, \quad \text{(S4)}$$

1.3 validation metrics

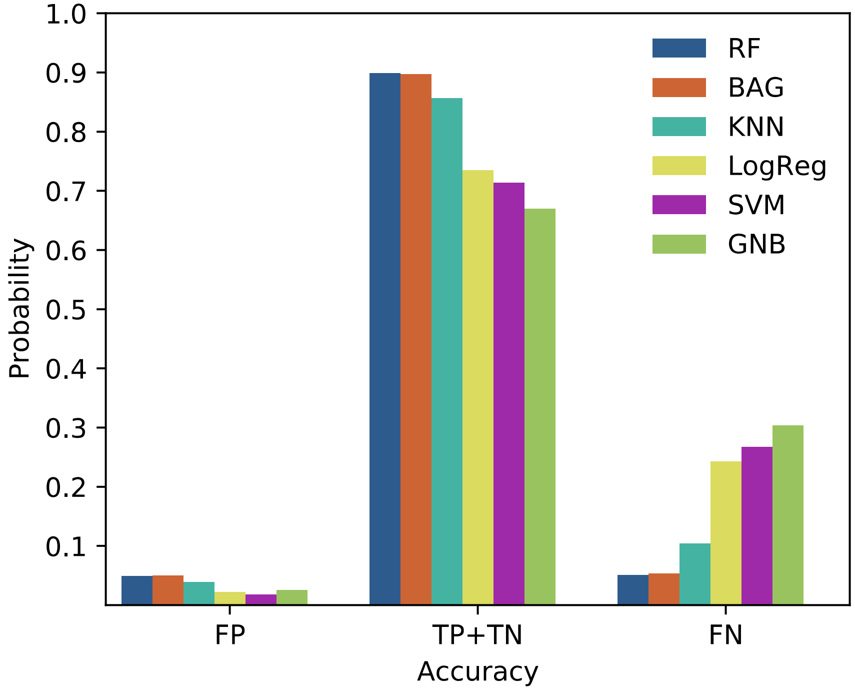

Figure S1 The histogram plots of accuracy metrices between ML predicted and observed fire/no-fire classes based on FireCCI burned area dataset. The FN stands for False Negative prediction, whose value is -1, which means that observed fires are wrongly predicted as no-fires; TP and FN stand for Ture Positive and False Negative predictions respectively, whose value is 0, meaning fires or no-fires are both correctly predicted; and FP stands for False Positive prediction, whose values is 1, meaning observed no-fire months are wrongly predicted as fire months.

Table S 2 The testing evaluation metrices of simulations with different datasets; the mean value and standardized error are calculated from multiple machine learning techniques

Simulations	Data	Accuracy	Recall	Precision	AUC	PPV	FDR	FOR	NPV
all	FireCCI_BA	0.81 ± 0.08	0.71 ± 0.12	0.43 ± 0.13	0.77 ± 0.03	0.43 ± 0.13	0.57 ± 0.13	0.05 ± 0.01	0.95 ± 0.01
no-humi	FireCCI_BA	0.78 ± 0.09	0.68 ± 0.11	0.37 ± 0.11	0.74 ± 0.02	0.37 ± 0.11	0.63 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-pre	FireCCI_BA	0.79 ± 0.09	0.70 ± 0.11	0.40 ± 0.11	0.75 ± 0.02	0.40 ± 0.11	0.60 ± 0.11	0.05 ± 0.01	0.95 ± 0.01
no-soimoi	FireCCI_BA	0.79 ± 0.09	0.68 ± 0.13	0.40 ± 0.11	0.75 ± 0.03	0.40 ± 0.11	0.60 ± 0.11	0.05 ± 0.02	0.95 ± 0.02
no-tmp	FireCCI_BA	0.79 ± 0.09	0.71 ± 0.11	0.40 ± 0.11	0.76 ± 0.02	0.40 ± 0.11	0.60 ± 0.11	0.05 ± 0.01	0.95 ± 0.01
no-tmp-hmi	FireCCI_BA	0.78 ± 0.08	0.67 ± 0.08	0.36 ± 0.10	0.73 ± 0.02	0.36 ± 0.10	0.64 ± 0.10	0.06 ± 0.01	0.94 ± 0.01
no-tmp-pre	FireCCI_BA	0.79 ± 0.08	0.70 ± 0.11	0.39 ± 0.11	0.75 ± 0.02	0.39 ± 0.11	0.61 ± 0.11	0.05 ± 0.01	0.95 ± 0.01
no-tmp-pre-hmi	FireCCI_BA	0.78 ± 0.08	0.66 ± 0.07	0.35 ± 0.10	0.73 ± 0.02	0.35 ± 0.10	0.65 ± 0.10	0.06 ± 0.01	0.94 ± 0.01
no-tmp-smo	FireCCI_BA	0.79 ± 0.09	0.70 ± 0.12	0.39 ± 0.11	0.75 ± 0.02	0.39 ± 0.11	0.61 ± 0.11	0.05 ± 0.01	0.95 ± 0.01
all	GFED_BA	0.83 ± 0.07	0.78 ± 0.03	0.53 ± 0.13	0.81 ± 0.03	0.53 ± 0.13	0.47 ± 0.13	0.05 ± 0.00	0.95 ± 0.00
no-humi	GFED_BA	0.78 ± 0.07	0.73 ± 0.06	0.45 ± 0.11	0.76 ± 0.02	0.45 ± 0.11	0.55 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-pre	GFED_BA	0.80 ± 0.07	0.73 ± 0.08	0.48 ± 0.11	0.77 ± 0.03	0.48 ± 0.11	0.52 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-soimoi	GFED_BA	0.80 ± 0.07	0.73 ± 0.08	0.48 ± 0.11	0.77 ± 0.02	0.48 ± 0.11	0.52 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-tmp	GFED_BA	0.80 ± 0.07	0.74 ± 0.06	0.48 ± 0.11	0.78 ± 0.02	0.48 ± 0.11	0.52 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-tmp-hmi	GFED_BA	0.79 ± 0.06	0.71 ± 0.05	0.46 ± 0.10	0.76 ± 0.02	0.46 ± 0.10	0.54 ± 0.10	0.07 ± 0.01	0.93 ± 0.01
no-tmp-pre	GFED_BA	0.80 ± 0.07	0.74 ± 0.07	0.48 ± 0.11	0.78 ± 0.02	0.48 ± 0.11	0.52 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-tmp-pre-hmi	GFED_BA	0.80 ± 0.06	0.72 ± 0.03	0.47 ± 0.10	0.77 ± 0.03	0.47 ± 0.10	0.53 ± 0.10	0.06 ± 0.00	0.94 ± 0.00
no-tmp-smo	GFED_BA	0.80 ± 0.07	0.74 ± 0.06	0.47 ± 0.11	0.77 ± 0.02	0.47 ± 0.11	0.53 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
all	GFED_C	0.83 ± 0.07	0.78 ± 0.03	0.53 ± 0.13	0.81 ± 0.03	0.53 ± 0.13	0.47 ± 0.13	0.05 ± 0.00	0.95 ± 0.00
no-humi	GFED_C	0.78 ± 0.07	0.73 ± 0.06	0.45 ± 0.11	0.76 ± 0.02	0.45 ± 0.11	0.55 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-pre	GFED_C	0.80 ± 0.07	0.73 ± 0.08	0.48 ± 0.11	0.77 ± 0.03	0.48 ± 0.11	0.52 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-soimoi	GFED_C	0.80 ± 0.07	0.73 ± 0.08	0.48 ± 0.11	0.77 ± 0.02	0.48 ± 0.11	0.52 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-tmp	GFED_C	0.80 ± 0.07	0.74 ± 0.06	0.48 ± 0.11	0.78 ± 0.02	0.48 ± 0.11	0.52 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-tmp-hmi	GFED_C	0.79 ± 0.06	0.71 ± 0.05	0.46 ± 0.10	0.76 ± 0.02	0.46 ± 0.10	0.54 ± 0.10	0.07 ± 0.01	0.93 ± 0.01
no-tmp-pre	GFED_C	0.80 ± 0.07	0.74 ± 0.07	0.48 ± 0.11	0.78 ± 0.02	0.48 ± 0.11	0.52 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-tmp-pre-hmi	GFED_C	0.79 ± 0.06	0.71 ± 0.04	0.46 ± 0.10	0.76 ± 0.02	0.46 ± 0.10	0.54 ± 0.10	0.07 ± 0.00	0.93 ± 0.00
no-tmp-smo	GFED_C	0.80 ± 0.07	0.74 ± 0.06	0.47 ± 0.11	0.77 ± 0.02	0.47 ± 0.11	0.53 ± 0.11	0.06 ± 0.01	0.94 ± 0.01

Simulations	Data	Accuracy	Recall	Precision	AUC	PPV	FDR	FOR	NPV
all	MCD45A1	0.89 ± 0.06	0.89 ± 0.07	0.94 ± 0.02	0.88 ± 0.05	0.94 ± 0.02	0.06 ± 0.02	0.20 ± 0.09	0.80 ± 0.09
no-humi	MCD45A1	0.88 ± 0.04	0.89 ± 0.05	0.92 ± 0.02	0.87 ± 0.04	0.92 ± 0.02	0.08 ± 0.02	0.21 ± 0.07	0.79 ± 0.07
no-pre	MCD45A1	0.87 ± 0.05	0.88 ± 0.07	0.92 ± 0.02	0.87 ± 0.05	0.92 ± 0.02	0.08 ± 0.02	0.21 ± 0.09	0.79 ± 0.09
no-soimoi	MCD45A1	0.87 ± 0.05	0.88 ± 0.07	0.93 ± 0.02	0.87 ± 0.04	0.93 ± 0.02	0.07 ± 0.02	0.21 ± 0.08	0.79 ± 0.08
no-tmp	MCD45A1	0.87 ± 0.06	0.87 ± 0.09	0.93 ± 0.03	0.87 ± 0.05	0.93 ± 0.03	0.07 ± 0.03	0.22 ± 0.10	0.78 ± 0.10
no-tmp-hmi	MCD45A1	0.86 ± 0.07	0.86 ± 0.09	0.92 ± 0.03	0.86 ± 0.06	0.92 ± 0.03	0.08 ± 0.03	0.24 ± 0.10	0.76 ± 0.10
no-tmp-pre	MCD45A1	0.87 ± 0.06	0.87 ± 0.09	0.93 ± 0.03	0.87 ± 0.06	0.93 ± 0.03	0.07 ± 0.03	0.23 ± 0.10	0.77 ± 0.10
no-tmp-pre-hmi	MCD45A1	0.86 ± 0.07	0.86 ± 0.09	0.93 ± 0.03	0.86 ± 0.06	0.93 ± 0.03	0.07 ± 0.03	0.24 ± 0.10	0.76 ± 0.10
no-tmp-smo	MCD45A1	0.87 ± 0.06	0.87 ± 0.09	0.93 ± 0.03	0.86 ± 0.05	0.93 ± 0.03	0.07 ± 0.03	0.23 ± 0.10	0.77 ± 0.10
all	MCD64A1	0.79 ± 0.08	0.70 ± 0.09	0.41 ± 0.11	0.75 ± 0.03	0.41 ± 0.11	0.59 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-humi	MCD64A1	0.75 ± 0.09	0.63 ± 0.11	0.36 ± 0.10	0.70 ± 0.03	0.36 ± 0.10	0.64 ± 0.10	0.07 ± 0.01	0.93 ± 0.01
no-pre	MCD64A1	0.77 ± 0.09	0.66 ± 0.11	0.38 ± 0.10	0.72 ± 0.03	0.38 ± 0.10	0.62 ± 0.10	0.07 ± 0.01	0.93 ± 0.01
no-soimoi	MCD64A1	0.76 ± 0.10	0.63 ± 0.13	0.38 ± 0.10	0.71 ± 0.03	0.38 ± 0.10	0.62 ± 0.10	0.07 ± 0.01	0.93 ± 0.01
no-tmp	MCD64A1	0.77 ± 0.09	0.65 ± 0.12	0.38 ± 0.10	0.72 ± 0.03	0.38 ± 0.10	0.62 ± 0.10	0.07 ± 0.01	0.93 ± 0.01
no-tmp-hmi	MCD64A1	0.77 ± 0.08	0.62 ± 0.09	0.37 ± 0.10	0.71 ± 0.03	0.37 ± 0.10	0.63 ± 0.10	0.07 ± 0.01	0.93 ± 0.01
no-tmp-pre	MCD64A1	0.77 ± 0.09	0.67 ± 0.09	0.39 ± 0.11	0.73 ± 0.03	0.39 ± 0.11	0.61 ± 0.11	0.06 ± 0.01	0.94 ± 0.01
no-tmp-pre-hmi	MCD64A1	0.77 ± 0.08	0.62 ± 0.08	0.37 ± 0.10	0.70 ± 0.03	0.37 ± 0.10	0.63 ± 0.10	0.07 ± 0.01	0.93 ± 0.01
no-tmp-smo	MCD64A1	0.76 ± 0.09	0.65 ± 0.10	0.37 ± 0.11	0.71 ± 0.03	0.37 ± 0.11	0.63 ± 0.11	0.07 ± 0.01	0.93 ± 0.01

Table S 3 random forest performances in different simulations with different datasets

Dataset	Model	Simulation	Type	Accuracy	Recall	Precision	F1-score	AUC	PPV	FDR	FOR	NPV
FireCCI_BA	RF	all	testing	0.90	0.62	0.61	0.61	0.78	0.61	0.39	0.06	0.94
FireCCI_BA	RF	no-tmp	testing	0.89	0.60	0.56	0.58	0.77	0.56	0.44	0.06	0.94
FireCCI_BA	RF	no-pre	testing	0.89	0.60	0.55	0.57	0.76	0.55	0.45	0.06	0.94
FireCCI_BA	RF	no-humi	testing	0.88	0.57	0.53	0.55	0.75	0.53	0.47	0.06	0.94
FireCCI_BA	RF	no-soimoi	testing	0.89	0.60	0.55	0.57	0.76	0.55	0.45	0.06	0.94
FireCCI_BA	RF	no-tmp-pre	testing	0.89	0.60	0.55	0.57	0.76	0.55	0.45	0.06	0.94
FireCCI_BA	RF	no-tmp-hmi	testing	0.87	0.59	0.51	0.55	0.75	0.51	0.49	0.06	0.94
FireCCI_BA	RF	no-tmp-smo	testing	0.89	0.60	0.56	0.58	0.77	0.56	0.44	0.06	0.94
FireCCI_BA	RF	no-tmp-pre-hmi	testing	0.87	0.58	0.51	0.54	0.75	0.51	0.49	0.06	0.94
GFED_BA	RF	all	testing	0.90	0.74	0.70	0.72	0.84	0.70	0.30	0.05	0.95
GFED_BA	RF	no-tmp	testing	0.88	0.68	0.63	0.65	0.80	0.63	0.37	0.07	0.93
GFED_BA	RF	no-pre	testing	0.88	0.68	0.63	0.65	0.80	0.63	0.37	0.07	0.93
GFED_BA	RF	no-humi	testing	0.87	0.67	0.60	0.63	0.79	0.60	0.40	0.07	0.93
GFED_BA	RF	no-soimoi	testing	0.88	0.68	0.62	0.65	0.80	0.62	0.38	0.07	0.93
GFED_BA	RF	no-tmp-pre	testing	0.88	0.68	0.63	0.66	0.80	0.63	0.37	0.07	0.93
GFED_BA	RF	no-tmp-hmi	testing	0.87	0.67	0.60	0.64	0.79	0.60	0.40	0.07	0.93
GFED_BA	RF	no-tmp-smo	testing	0.88	0.68	0.63	0.65	0.80	0.63	0.37	0.07	0.93
GFED_BA	RF	no-tmp-pre-hmi	testing	0.88	0.68	0.62	0.65	0.80	0.62	0.38	0.07	0.93
GFED_C	RF	all	testing	0.90	0.74	0.70	0.72	0.84	0.70	0.30	0.05	0.95
GFED_C	RF	no-tmp	testing	0.88	0.68	0.63	0.65	0.80	0.63	0.37	0.07	0.93
GFED_C	RF	no-pre	testing	0.88	0.68	0.63	0.65	0.80	0.63	0.37	0.07	0.93
GFED_C	RF	no-humi	testing	0.87	0.67	0.60	0.63	0.79	0.60	0.40	0.07	0.93
GFED_C	RF	no-soimoi	testing	0.88	0.68	0.62	0.65	0.80	0.62	0.38	0.07	0.93
GFED_C	RF	no-tmp-pre	testing	0.88	0.68	0.63	0.66	0.80	0.63	0.37	0.07	0.93
GFED_C	RF	no-tmp-hmi	testing	0.87	0.67	0.60	0.64	0.79	0.60	0.40	0.07	0.93
GFED_C	RF	no-tmp-smo	testing	0.88	0.68	0.63	0.65	0.80	0.63	0.37	0.07	0.93
GFED_C	RF	no-tmp-pre-hmi	testing	0.87	0.66	0.60	0.63	0.79	0.60	0.40	0.07	0.93
MCD45A1	RF	all	testing	0.94	0.94	0.96	0.95	0.93	0.96	0.04	0.11	0.89

Dataset	Model	Simulation	Type	Accuracy	Recall	Precision	F1-score	AUC	PPV	FDR	FOR	NPV
MCD45A1	RF	no-tmp	testing	0.93	0.93	0.95	0.94	0.92	0.95	0.05	0.13	0.87
MCD45A1	RF	no-pre	testing	0.92	0.94	0.95	0.94	0.92	0.95	0.05	0.13	0.87
MCD45A1	RF	no-humi	testing	0.92	0.94	0.95	0.94	0.92	0.95	0.05	0.12	0.88
MCD45A1	RF	no-soimoi	testing	0.92	0.94	0.95	0.94	0.92	0.95	0.05	0.13	0.87
MCD45A1	RF	no-tmp-pre	testing	0.93	0.94	0.96	0.95	0.92	0.96	0.04	0.12	0.88
MCD45A1	RF	no-tmp-hmi	testing	0.92	0.93	0.95	0.94	0.92	0.95	0.05	0.13	0.87
MCD45A1	RF	no-tmp-smo	testing	0.92	0.93	0.95	0.94	0.92	0.95	0.05	0.13	0.87
MCD45A1	RF	no-tmp-pre-hmi	testing	0.92	0.93	0.95	0.94	0.92	0.95	0.05	0.13	0.87
MCD64A1	RF	all	testing	0.88	0.60	0.56	0.58	0.76	0.56	0.44	0.07	0.93
MCD64A1	RF	no-tmp	testing	0.87	0.55	0.53	0.54	0.74	0.53	0.47	0.08	0.92
MCD64A1	RF	no-pre	testing	0.86	0.54	0.52	0.53	0.73	0.52	0.48	0.08	0.92
MCD64A1	RF	no-humi	testing	0.86	0.52	0.52	0.52	0.72	0.52	0.48	0.08	0.92
MCD64A1	RF	no-soimoi	testing	0.86	0.52	0.52	0.52	0.72	0.52	0.48	0.08	0.92
MCD64A1	RF	no-tmp-pre	testing	0.87	0.56	0.55	0.56	0.74	0.55	0.45	0.07	0.93
MCD64A1	RF	no-tmp-hmi	testing	0.86	0.54	0.52	0.53	0.73	0.52	0.48	0.08	0.92
MCD64A1	RF	no-tmp-smo	testing	0.86	0.53	0.52	0.52	0.72	0.52	0.48	0.08	0.92
MCD64A1	RF	no-tmp-pre-hmi	testing	0.86	0.54	0.52	0.53	0.73	0.52	0.48	0.08	0.92

Table S4 validation accuracy for the ALL-simulation with SMOTE using RF

Data	Sim.	Step	Accuracy	Recall	Precision	F1-score	AUC	PPV	FDR	FOR	NPV
FireCCI_BA	ALL	testing	0.90	0.62	0.61	0.61	0.78	0.61	0.39	0.06	0.94
GFED_BA	ALL	testing	0.90	0.74	0.70	0.72	0.84	0.70	0.30	0.05	0.95
GFED_C	ALL	testing	0.90	0.74	0.70	0.72	0.84	0.70	0.30	0.05	0.95
MCD45A1	ALL	testing	0.94	0.94	0.96	0.95	0.93	0.96	0.04	0.11	0.89
MCD64A1	ALL	testing	0.88	0.60	0.56	0.58	0.76	0.56	0.44	0.07	0.93

Table S5 validation accuracy for the ALL-simulation without SMOTE using RF

Data	Sim.	Step	Accuracy	Recall	Precision	F1-score	AUC	PPV	FDR	FOR	NPV
FireCCI_BA	ALL	testing	0.91	0.42	0.77	0.54	0.70	0.77	0.23	0.08	0.92
GFED_BA	ALL	testing	0.91	0.61	0.81	0.69	0.79	0.81	0.19	0.08	0.92
GFED_C	ALL	testing	0.91	0.61	0.81	0.69	0.79	0.81	0.19	0.08	0.92
MCD45A1	ALL	testing	0.94	0.96	0.96	0.96	0.93	0.96	0.04	0.09	0.91
MCD64A1	ALL	testing	0.89	0.42	0.72	0.53	0.70	0.72	0.28	0.09	0.91

2. Research Area

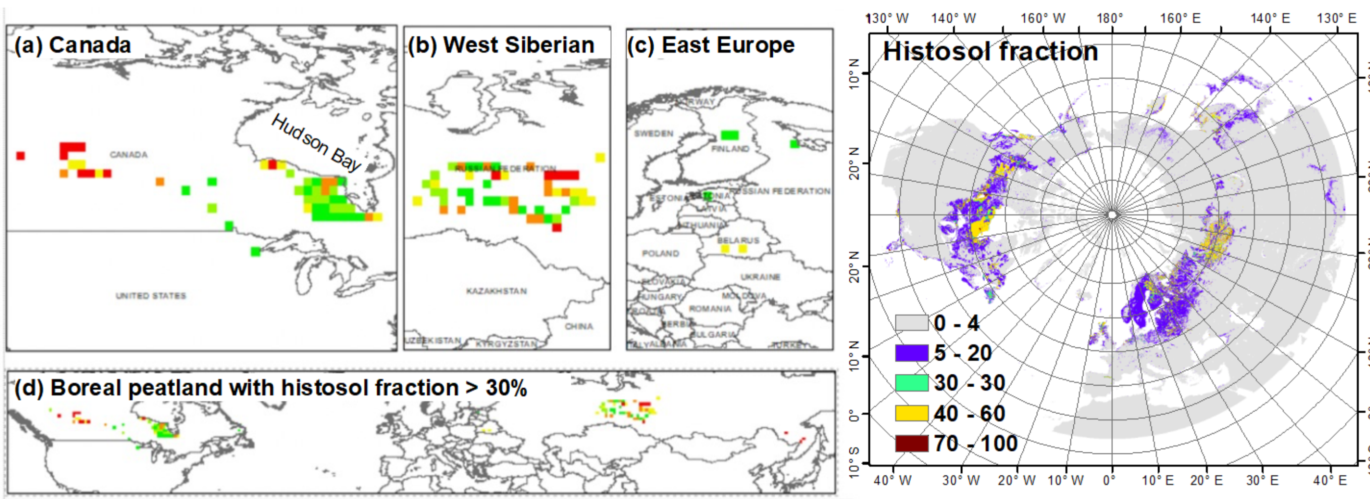

Figure S2 Research Area

3. Spatial fire counts (training and testing)

3.1 FireCCI BA (training and testing)

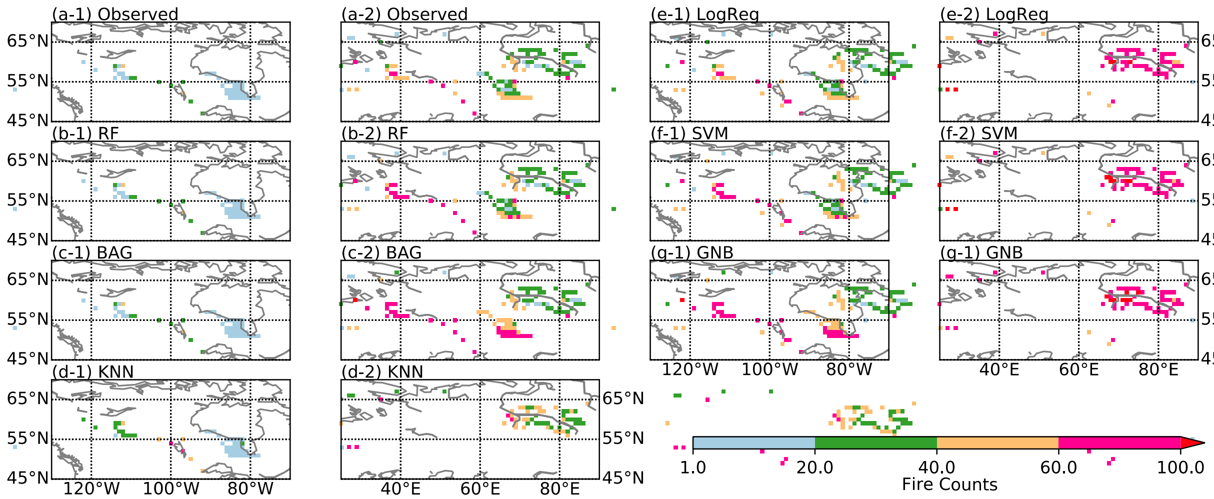

Figure S 3 The spatial map of fire counts from observations (a-1, a-2), and multiple MLTs' predictions (b-1 to g-2) with FireCCI as the target variable

3.2 GFED BA (training and testing)

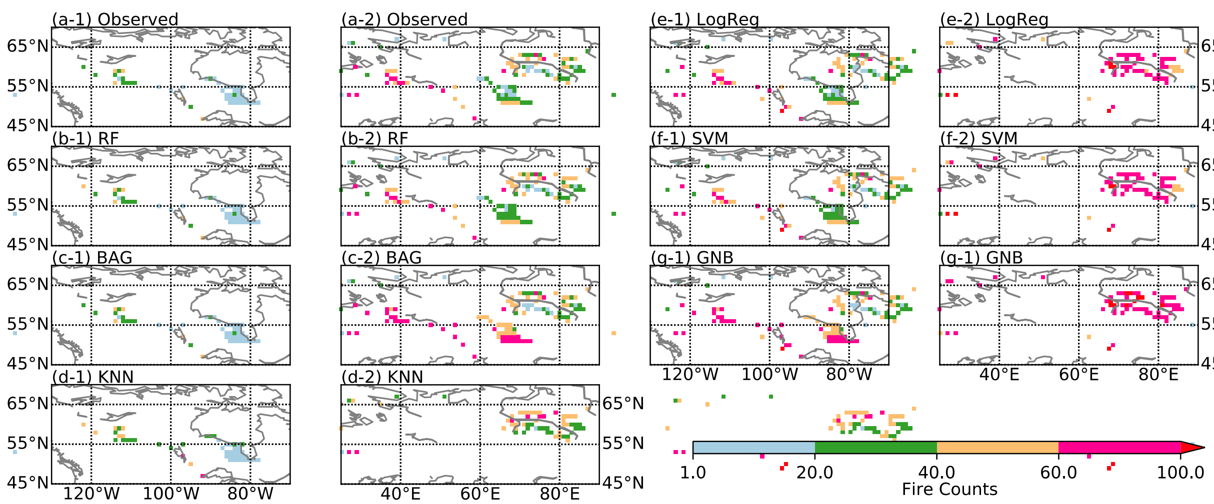

Figure S4 The spatial map of fire counts from observations (a-1, a-2), and multiple MLTs' predictions (b-1 to g-2) with GFED burned area as the target variable

3.3 GFED C emission (training and testing)

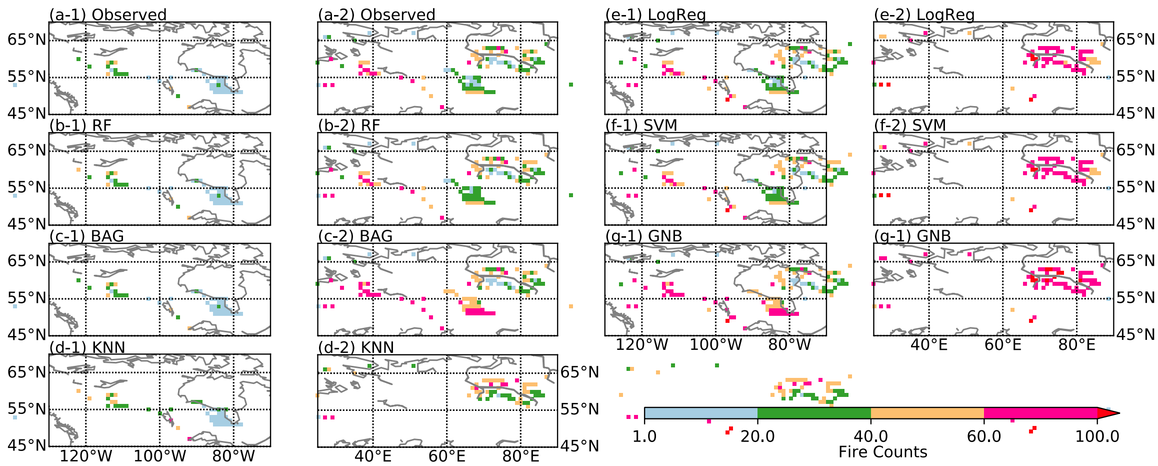

Figure S 5 The spatial map of fire counts from observations (a-1, a-2), and multiple MLTs' predictions (b-1 to g-2) with GFED C emissions as the target variable

3.4 MCD64A1 fire counts (training and testing)

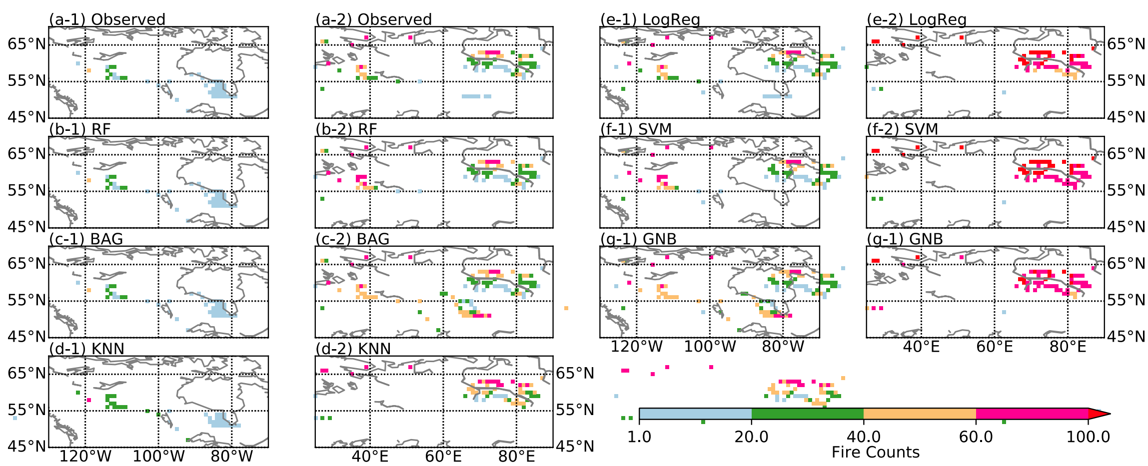

Figure S 6 The spatial map of fire counts from observations (a-1, a-2), and multiple MLTs' predictions (b-1 to g-2) with MCD64A1 fire detection as the target variable

3.5 MCD45A1 fire counts (training and testing)

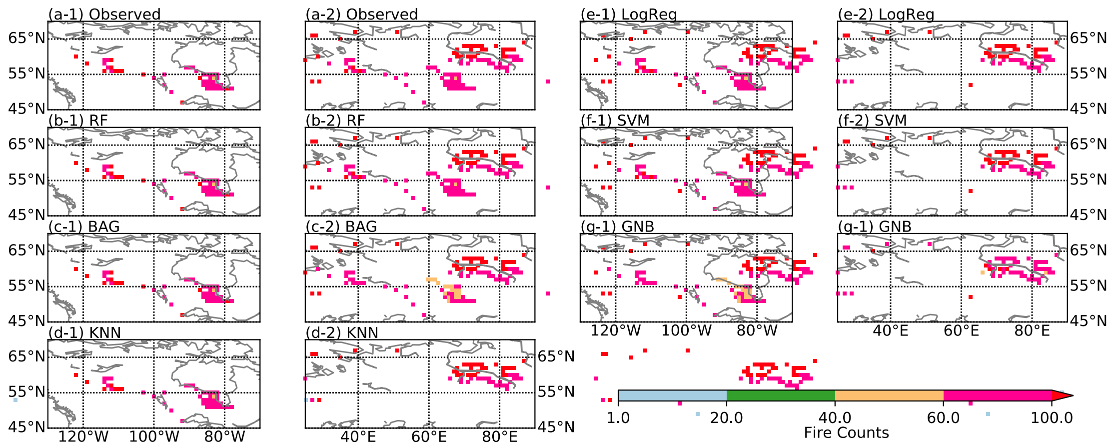

Figure S 7 The spatial map of fire counts from observations (a-1, a-2), and multiple MLTs' predictions (b-1 to g-2) with MCD45A1 fire detection as the target variable

4. Spatial fire counts (testing only)

4.1 FireCCI BA testing

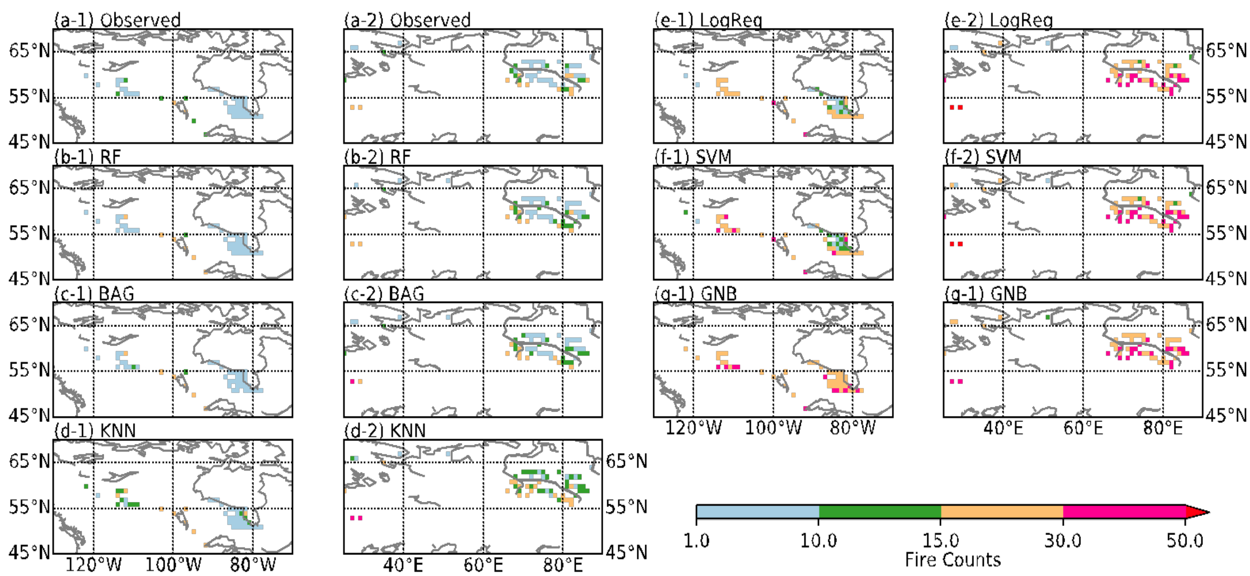

Figure S 8 Spatial distribution of observed and ML modeling predicted fire counts in validation with FireCCI burned area data.

4.2 GFED BA testing

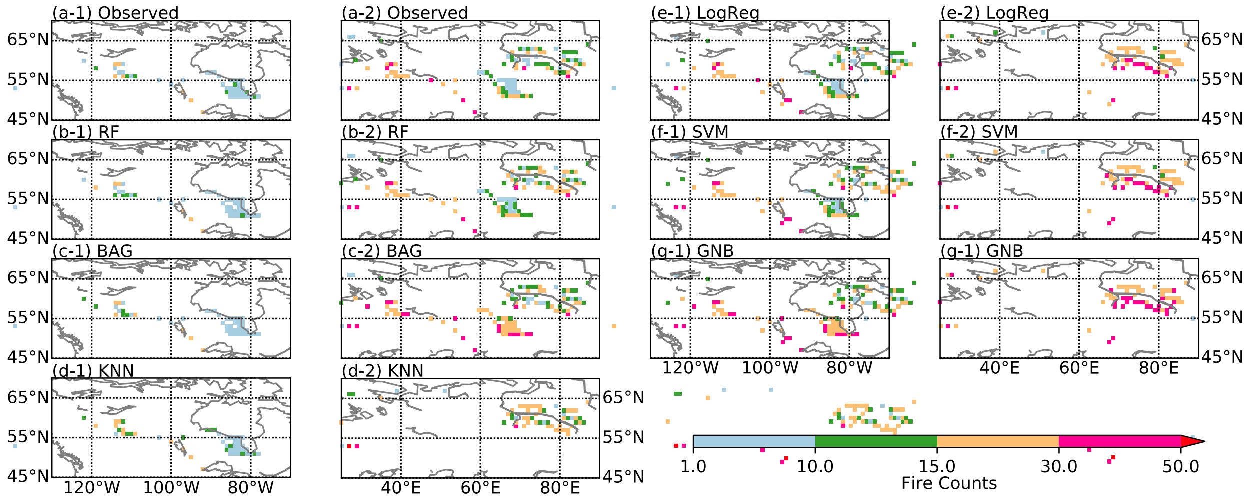

Figure S9 The spatial map of fire counts from observations (a-1, a-2), and multiple MLTs' predictions (b-1 to g-2) in the testing stage with GFED burned area as the target variable

4.3 GFED C emission testing

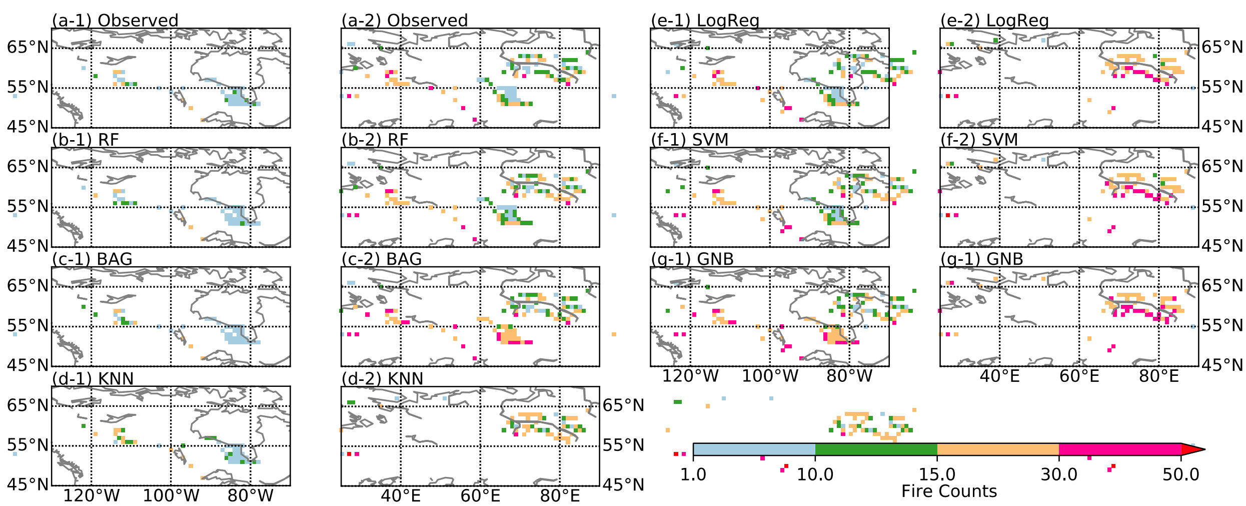

Figure S10 The spatial map of fire counts from observations (a-1, a-2), and multiple MLTs' predictions (b-1 to g-2) in the testing stage with GFED C emissoions as the target variable

4.4 MCD64A1 fire counts testing

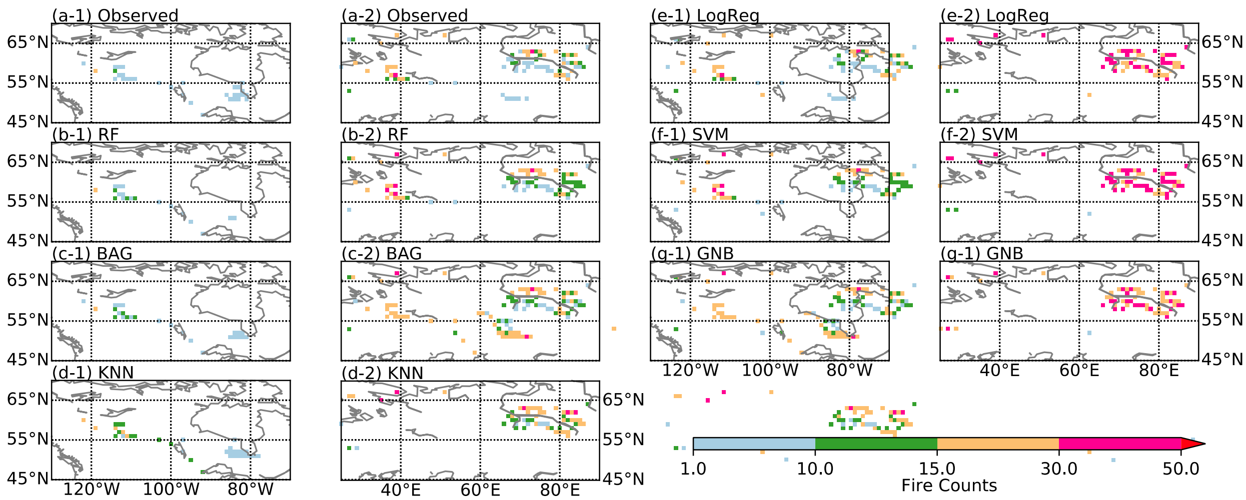

Figure S 11 The spatial map of fire counts from observations (a-1, a-2), and multiple MLTs' predictions (b-1 to g-2) in the testing stage with MCD64A1 as the target variable

4.5 MCD45A1 fire counts testing

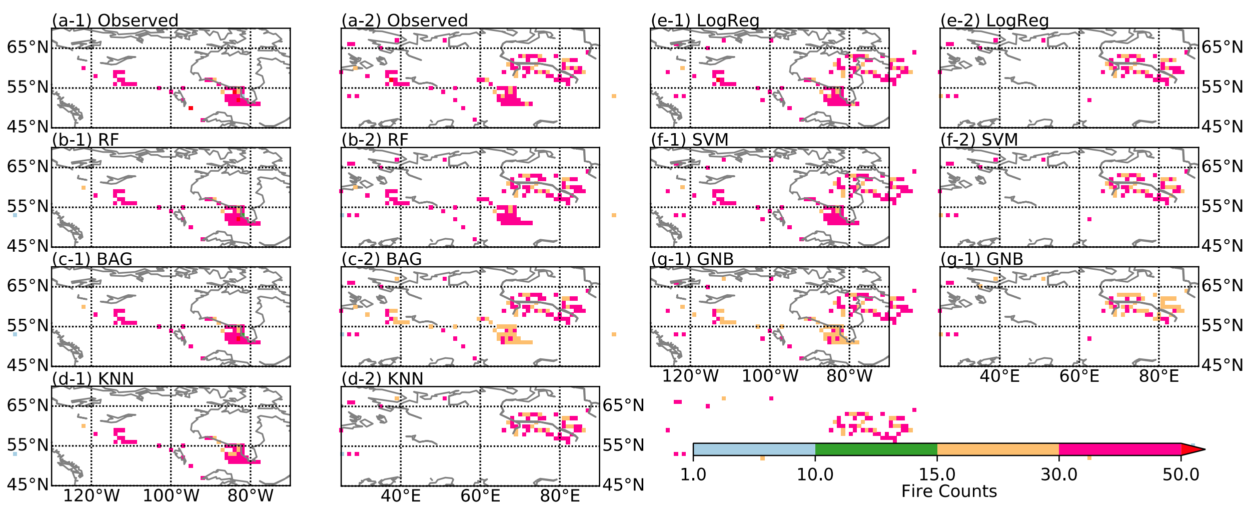

Figure S 12 The spatial map of fire counts from observations (a-1, a-2), and multiple MLTs' predictions (b-1 to g-2) in the testing stage with MCD45A1 as the target variable

5. fire counts (seasonality, testing only)

5.1 GFED BA seasonality (testing)

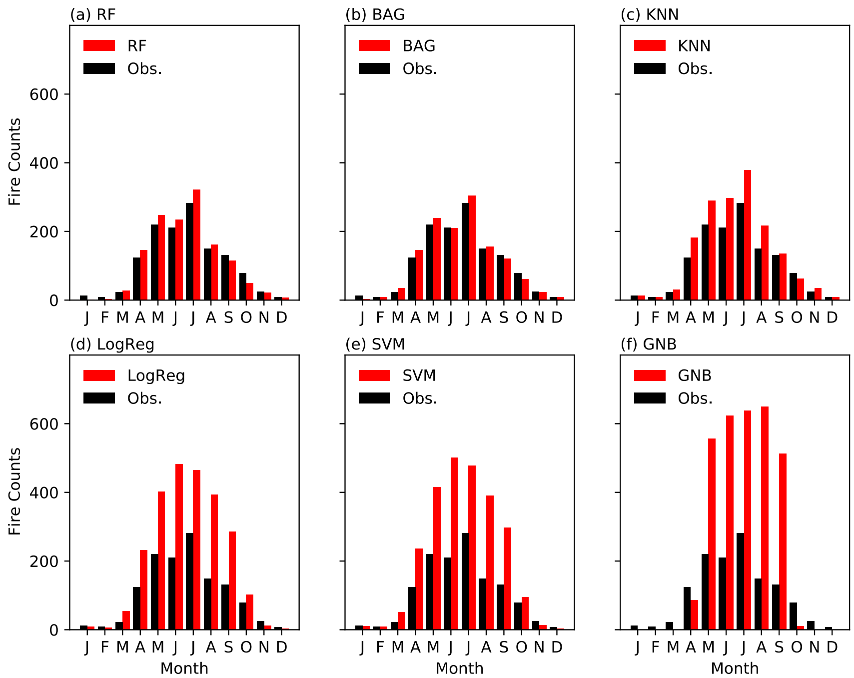

Figure S 13 Seasonality of fire counts prediction in testing stage with GFED burned area data

5.2 GFED C seasonality (testing)

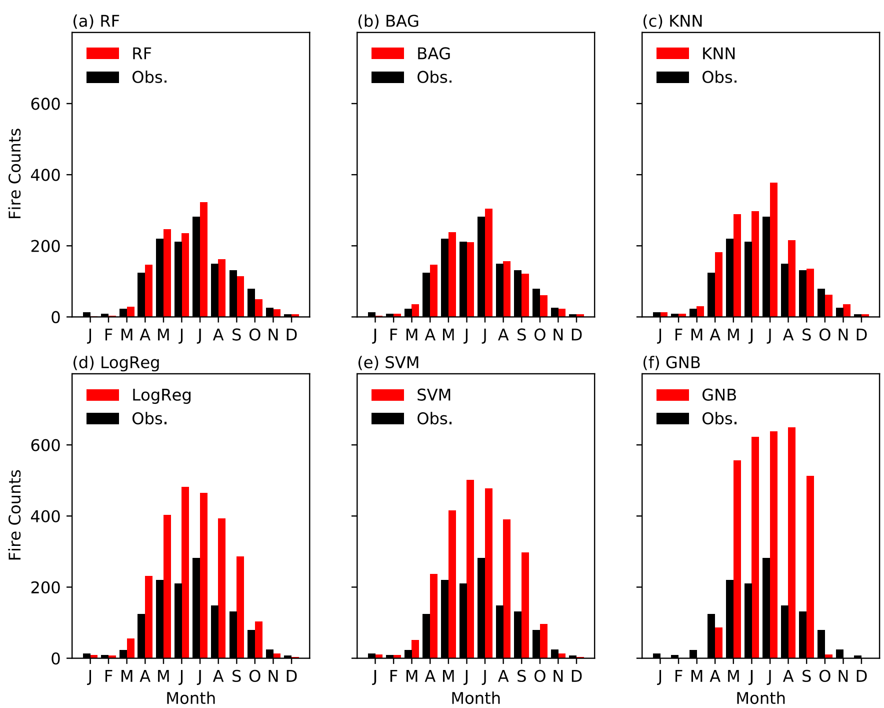

Figure S 14 Seasonality of fire counts prediction in testing stage with GFED burned area data

5.3 MCD64 seasonality (testing)

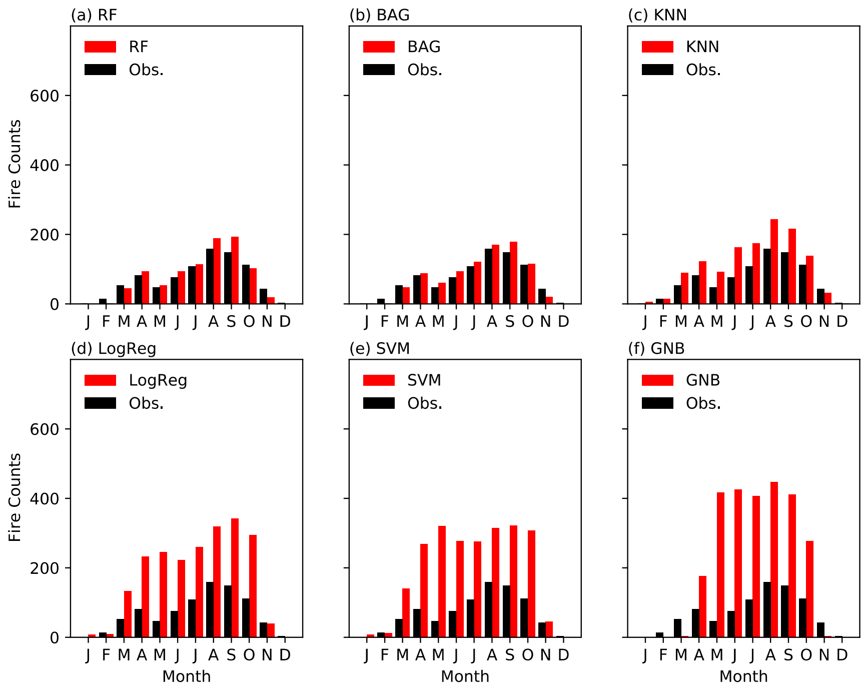

Figure S 15 Seasonality of fire counts prediction in testing stage with MCD64A1 data

5.4 MCD45 seasonality (testing)

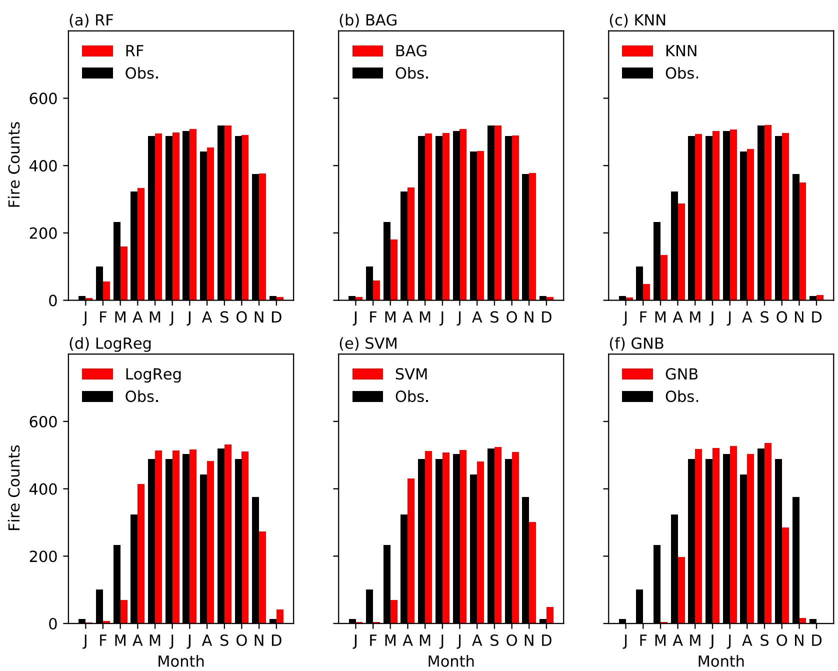

Figure S 16 Seasonality of fire counts prediction in testing stage with MCD45A1 data

6. fire counts (Seasonality, testing and training)

6.1 FireCCI seasonality (training and testing)

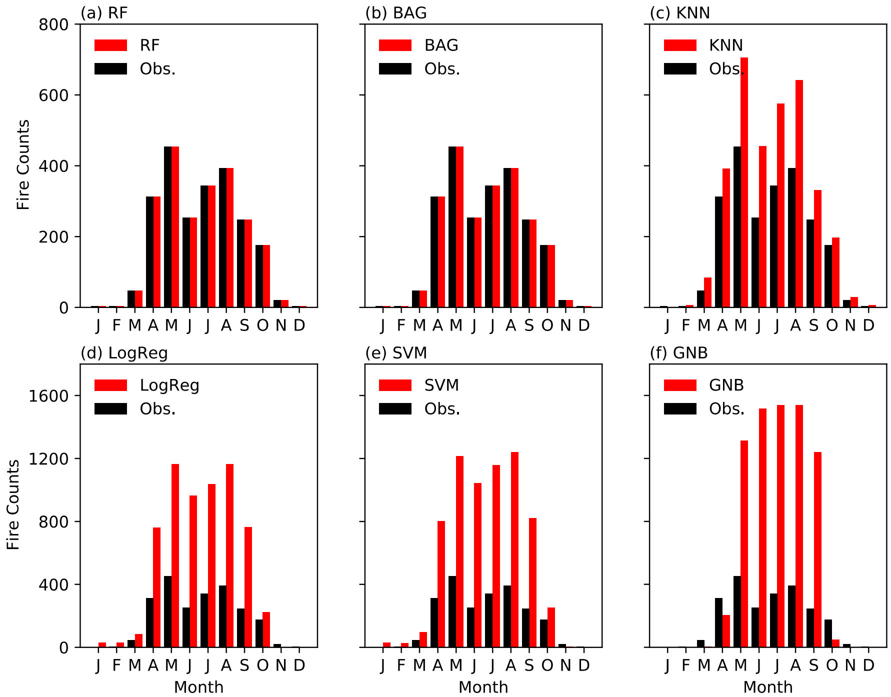

Figure S 17 Seasonality of fire counts prediction in both testing and training stage with FireCCI data

6.2 GFED BA seasonality (training and testing)

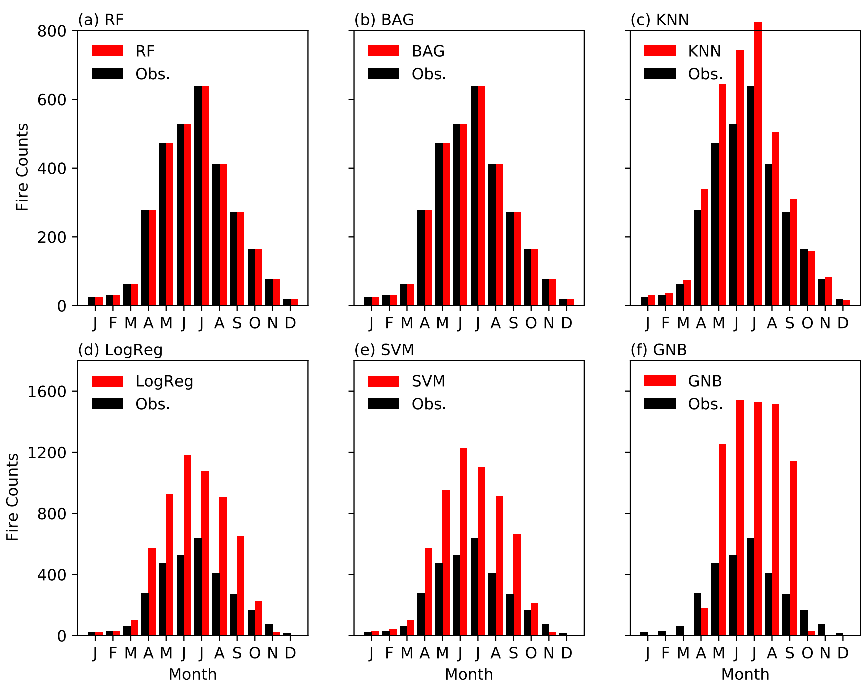

Figure S18 Seasonality of fire counts prediction in both testing and training stage with GFED BA data

6.3 GFED C seasonality (training and testing)

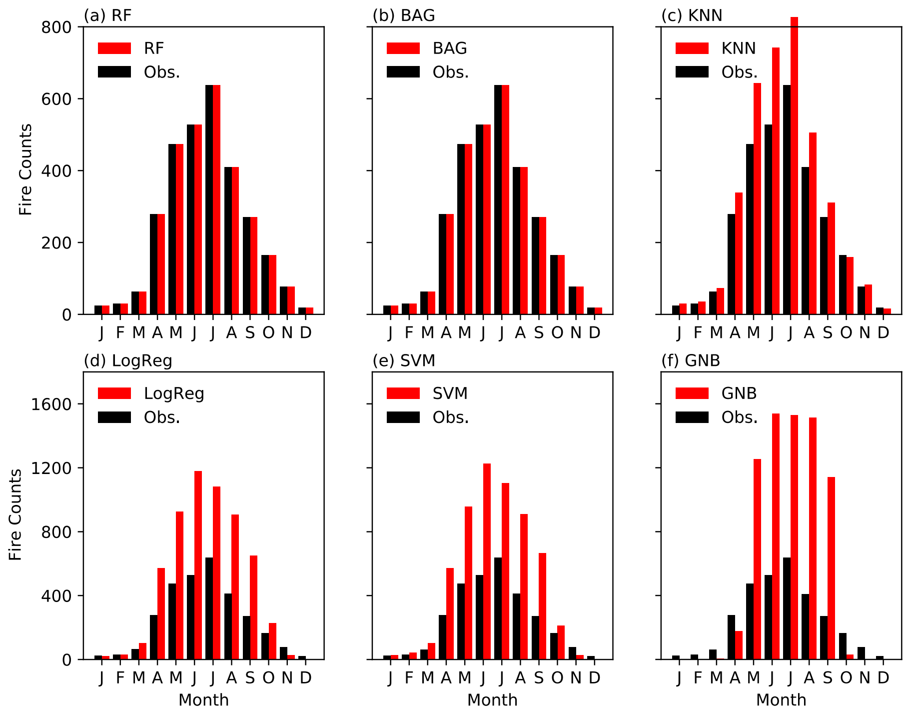

Figure S 19 Seasonality of fire counts prediction in both testing and training stage with GFED C data

6.4 MCD64 seasonality (training and testing)

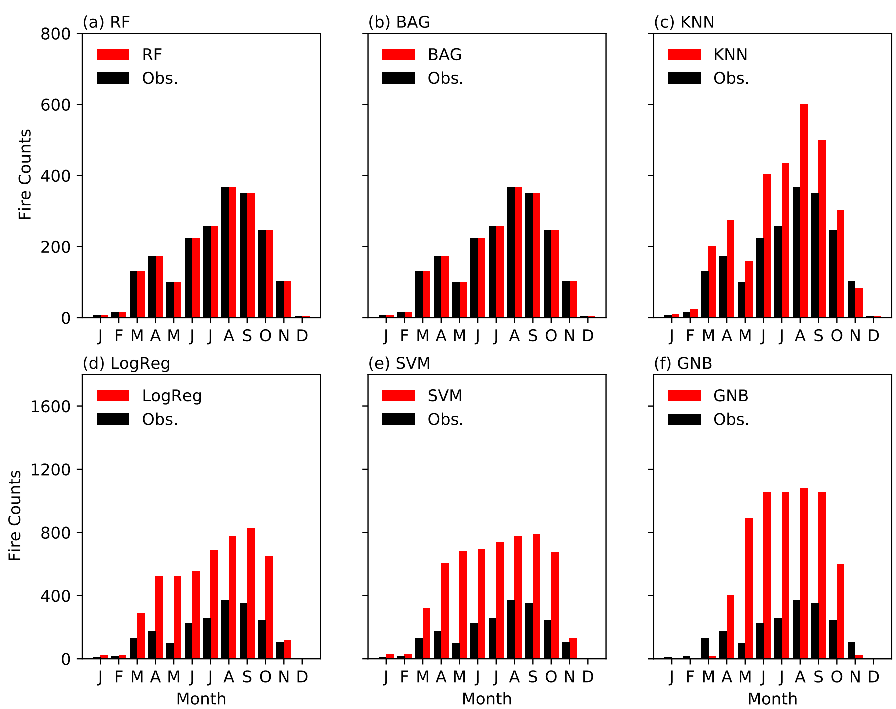

Figure S20 Seasonality of fire counts prediction in both testing and training stage with MCD64A1 data

6.5 MCD45 seasonality (training and testing)

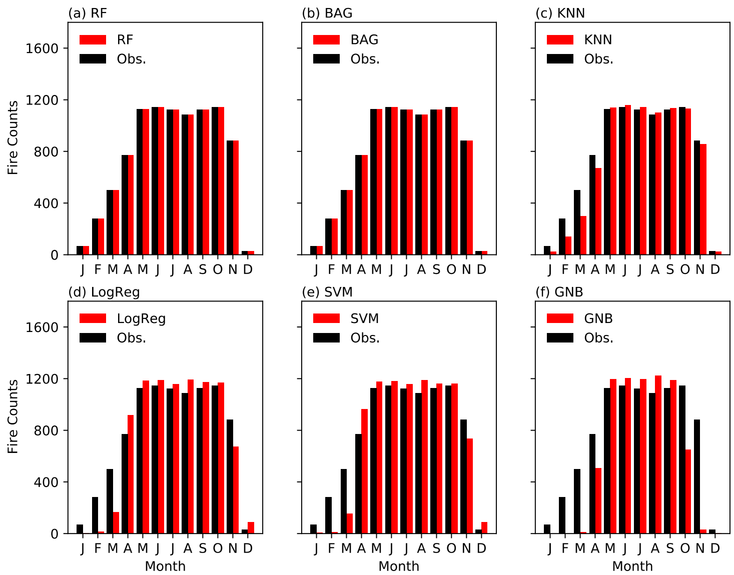

Figure S 21 Seasonality of fire counts prediction in both testing and training stage with MCD45A1 data

7. The importance rank of variables in multiple datasets

7.1 Factor importance rank from GFED BA

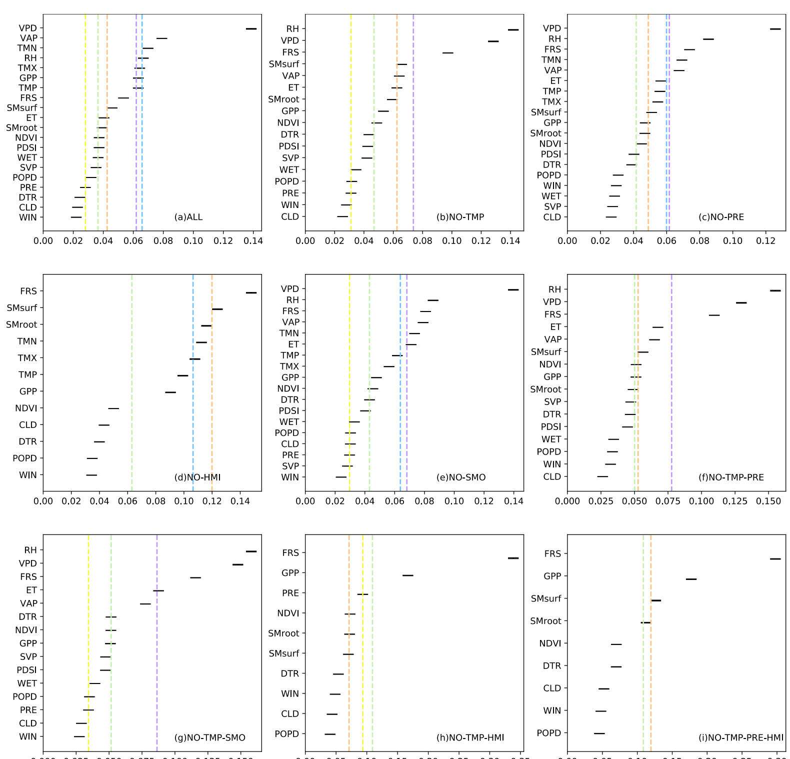

Figure S22 The importance rank of factors with GFED burned area data

7.2 Factor importance rank from GFED C

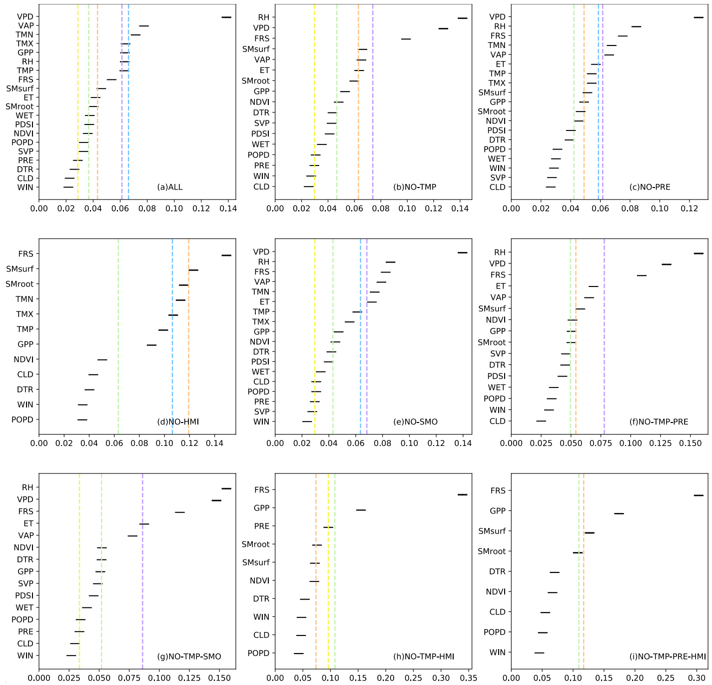

Figure S 23 The importance rank of factors with GFED C emission data

7.3 Factor importance rank from MCD64A1

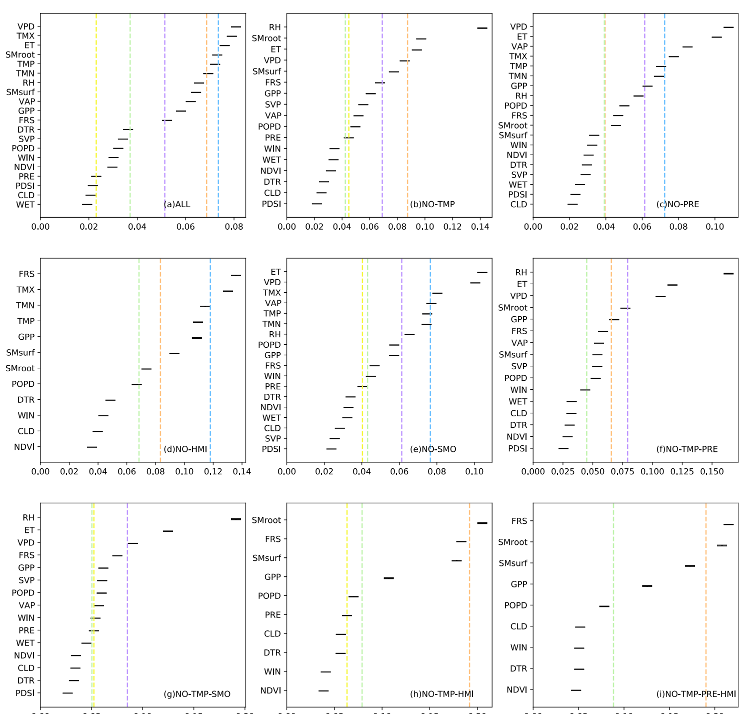

Figure S 24 The importance rank of factors with MCD64A1 data

7.4 Factor importance rank from MCD45A1

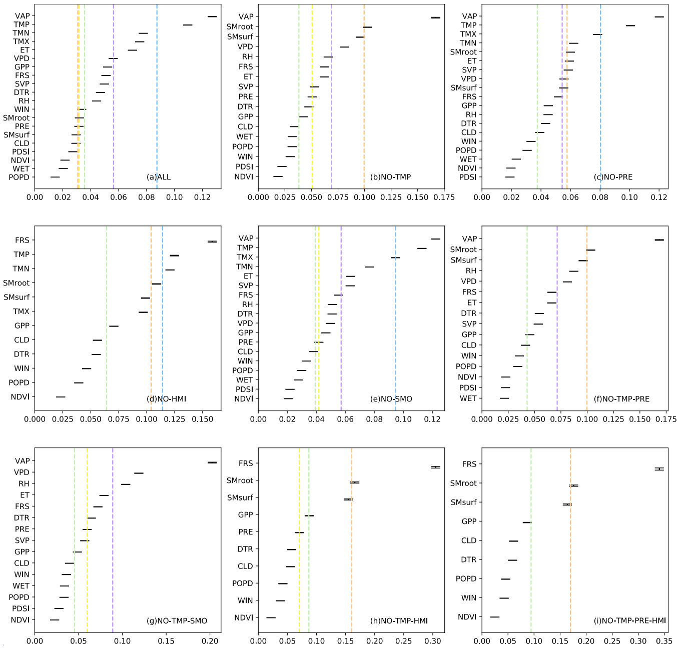

Figure S 25 The importance rank of factors with MCD45A1 data

8. fire impacts (testing only)

8.1 FireCCI burned area predictions in testing se

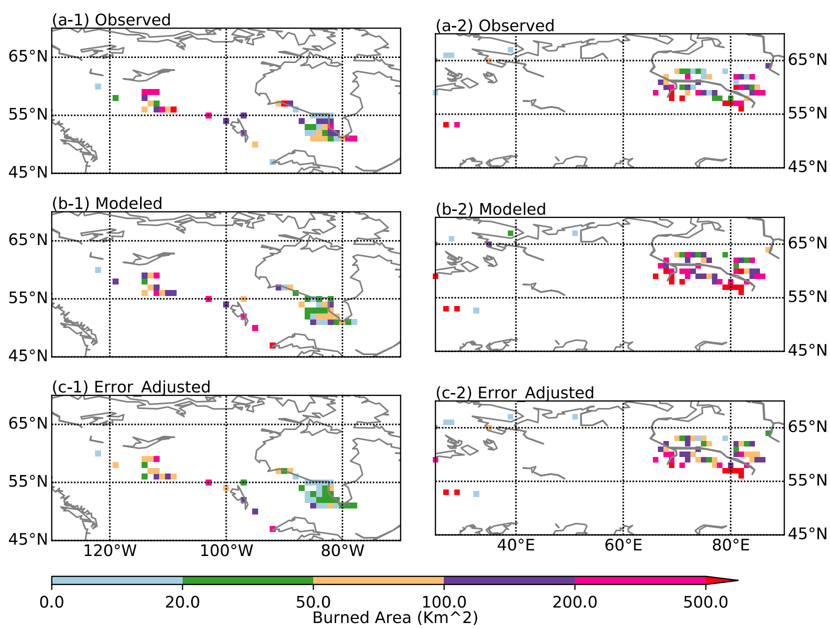

Figure S 26 Spatial distributions of the observed, stacked machine learning predictions and the error-adjusted FireCCI burned area with validation datasets. The left column (a-1, b-1, c-1) represents the peatland fire BA located in Canada, especially the Hudson Bay Lowland, and the right columns (a-2, b-2, c-2) represents fire impacts in the West Siberian area.

8.2 GFED BA predictions in testing set

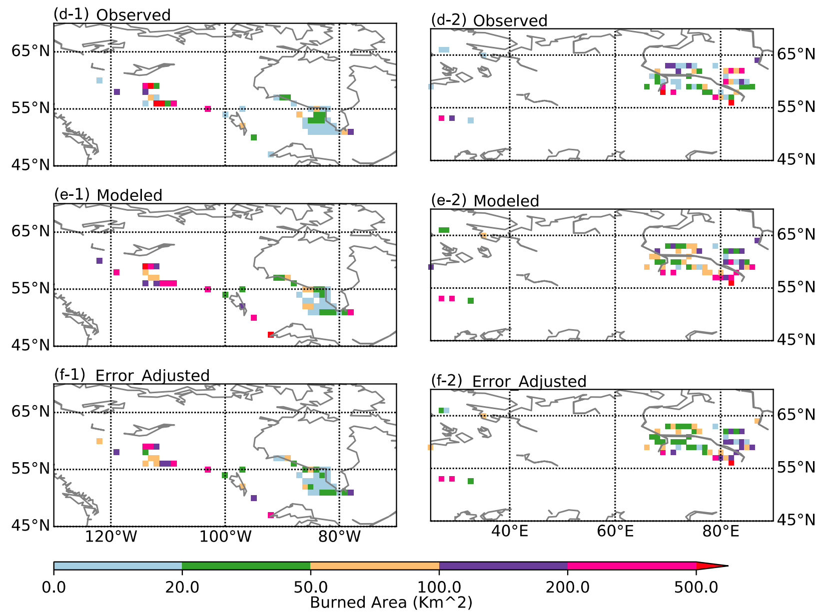

Figure S 27 Spatial distributions of the observed, stacked machine learning predictions and the error-adjusted GFED BA with validation datasets. The left column (a-1, b-1, c-1) represents the peatland fire BA located in Canada, especially the Hudson Bay Lowland, and the right columns (a-2, b-2, c-2) represents fire impacts in the West Siberian area.

8.3 GFED C emission prediction in testing set

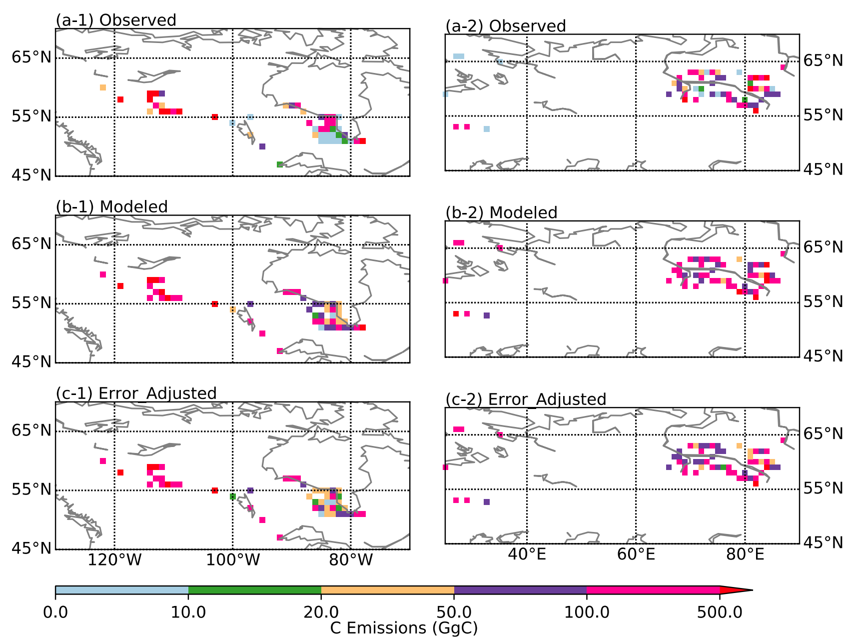

Figure S 28 Spatial distributions of the observed, stacked machine learning predictions and the error-adjusted GFED C emissions with validation datasets. The left column (a-1, b-1, c-1) represents the peatland fire C emissions located in Canada, especially the Hudson Bay Lowland, and the right columns (a-2, b-2, c-2) represents fire impacts in the West Siberian area.

9. fire impacts (testing only)

9.1 FireCCI burned area predictions in testing set

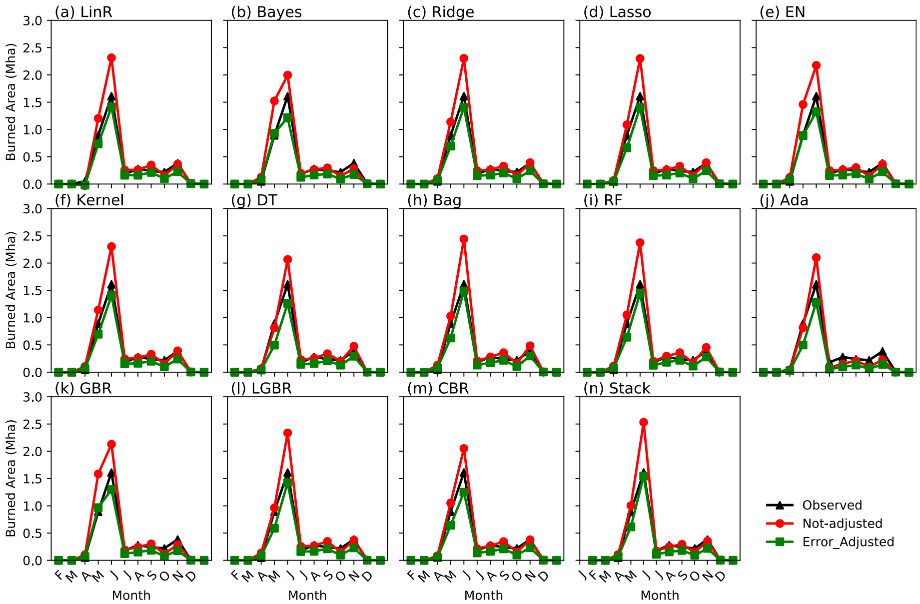

Figure S 29 Seasonality of the observed, modeled, and error adjusted FireCCI burned area from multiple machine learning leaners: (a) Linear Regression; (b) Bayesian linear Regression; (c) Ridge regression; (d)Lasso regression; (e)Elastic Net; (f)Kernel ridge regression; (g)Decision tree; (h)Bagging; (i)Random forests; (j) Adaptive boosting regression; (k)Gradient boosting regression; (l)Light gradient boosting regression; (m)Cat boosting regression; and (n) stacking.

9.2 GFED BA predictions in testing set

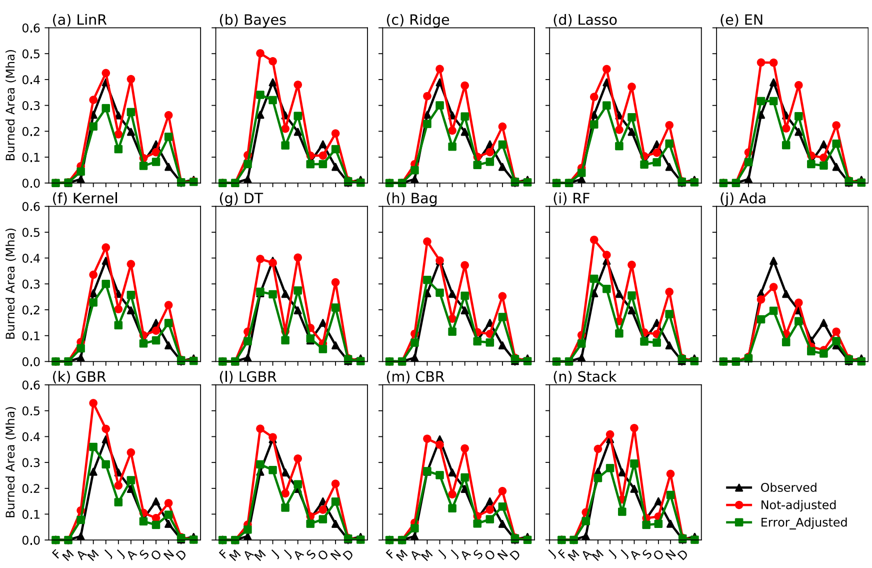

Figure S 30 Seasonality of the observed, modeled, and error adjusted GFED burned area from multiple machine learning leaners: (a) Linear Regression; (b) Bayesian linear Regression; (c) Ridge regression; (d)Lasso regression; (e)Elastic Net; (f)Kernel ridge regression; (g)Decision tree; (h) Bagging; (i) Random forests; (j) Adaptive boosting regression; (k)Gradient boosting regression; (l)Light gradient boosting regression; (m)Cat boosting regression; and (n) stacking.

9.3 GFED C emission prediction in testing set

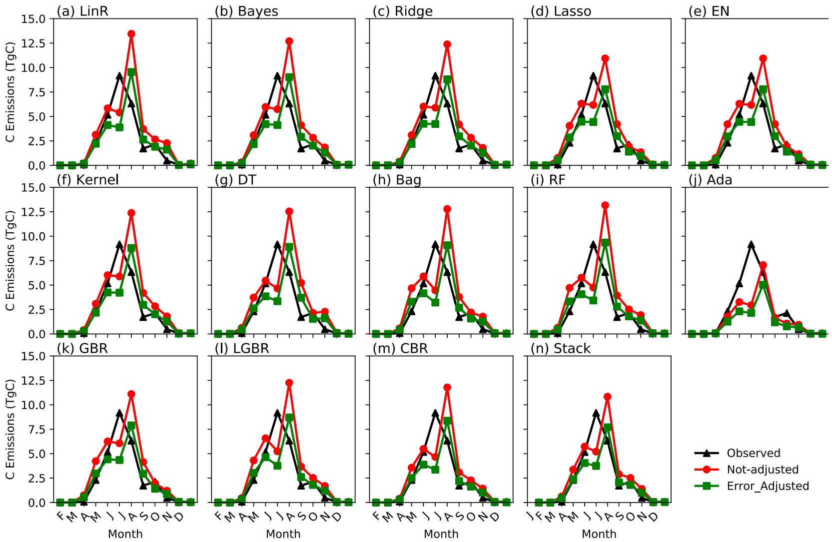

Figure S 31 Seasonality of the observed, modeled, and error adjusted GFED C emission from multiple machine learning leaners: (a) Linear Regression; (b) Bayesian linear Regression; (c) Ridge regression; (d)Lasso regression; (e)Elastic Net; (f)Kernel ridge regression; (g)Decision tree; (h)Bagging; (i)Random forests; (j) Adaptive boosting regression; (k)Gradient boosting regression; (l)Light gradient boosting regression; (m)Cat boosting regression; and (n) stacking.

10. fire impacts (training and testing)

10.1 FireCCI burned area

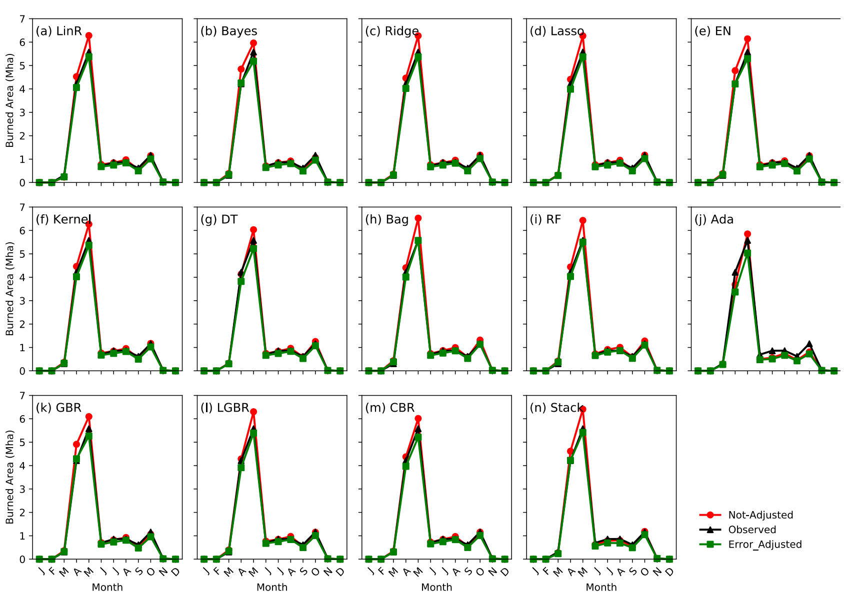

Figure S 32 Seasonality of the observed, not-adjusted, and error adjusted FireCCI BA from multiple machine learning leaners: (a) Linear Regression; (b) Bayesian linear Regression; (c) Ridge regression; (d)Lasso regression; (e)Elastic Net; (f)Kernel ridge regression; (g)Decision tree; (h)Bagging; (i)Random forests; (j) Adaptive boosting regression; (k)Gradient boosting regression; (l)Light gradient boosting regression; (m)Cat boosting regression; and (n) stacking.

10.2 GFED-BA

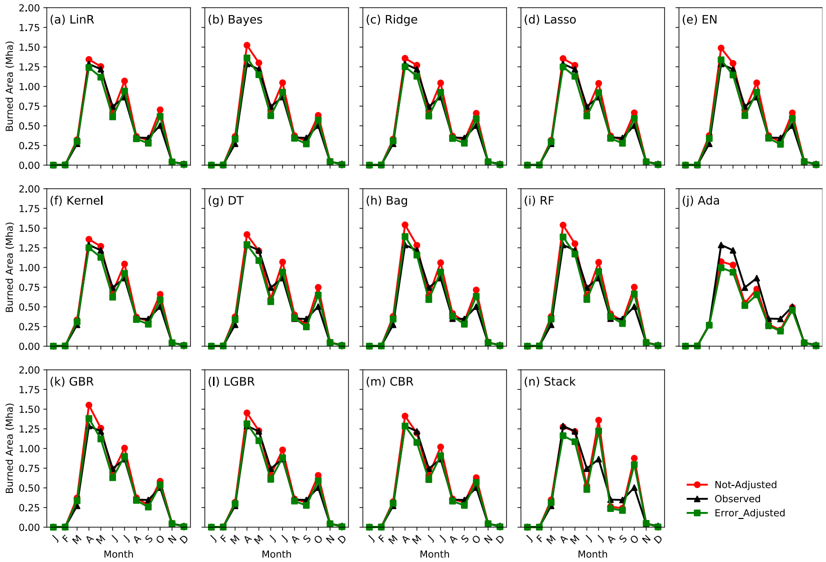

Figure S 33 Seasonality of the observed, not-adjusted, and error adjusted GFED BA from multiple machine learning leaners: (a) Linear Regression; (b) Bayesian linear Regression; (c) Ridge regression; (d)Lasso regression; (e)Elastic Net; (f)Kernel ridge regression; (g)Decision tree; (h)Bagging; (i)Random forests; (j) Adaptive boosting regression; (k)Gradient boosting regression; (l)Light gradient boosting regression; (m)Cat boosting regression; and (n) stacking.

10.3 GFED-C

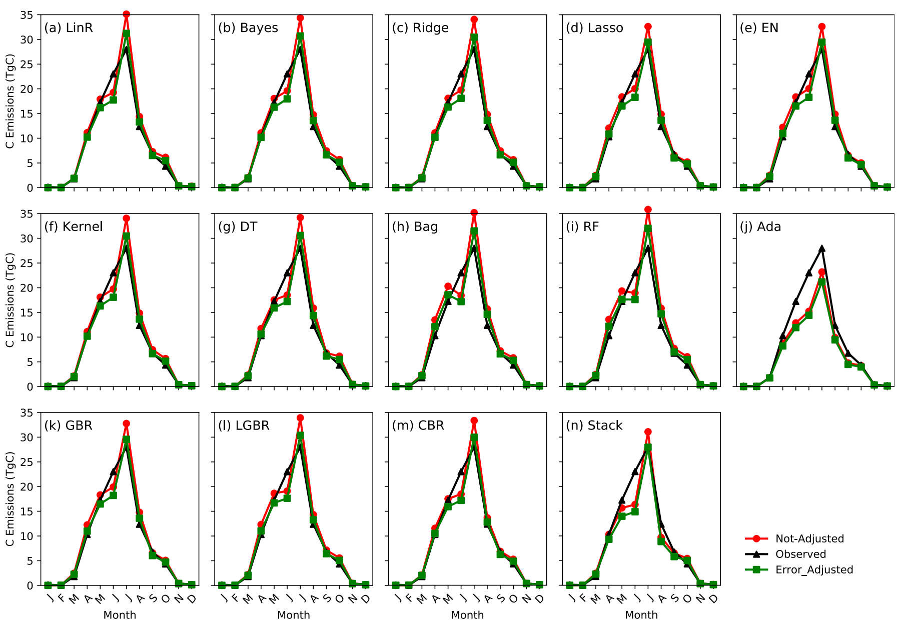

Figure S 34 Seasonality of the observed, not-adjusted, and error adjusted GFED C emissions from multiple machine learning leaners: (a) Linear Regression; (b) Bayesian linear Regression; (c) Ridge regression; (d)Lasso regression; (e)Elastic Net; (f)Kernel ridge regression; (g)Decision tree; (h)Bagging; (i)Random forests; (j) Adaptive boosting regression; (k)Gradient boosting regression; (l)Light gradient boosting regression; (m)Cat boosting regression; and (n) stacking.

11. Data correlations

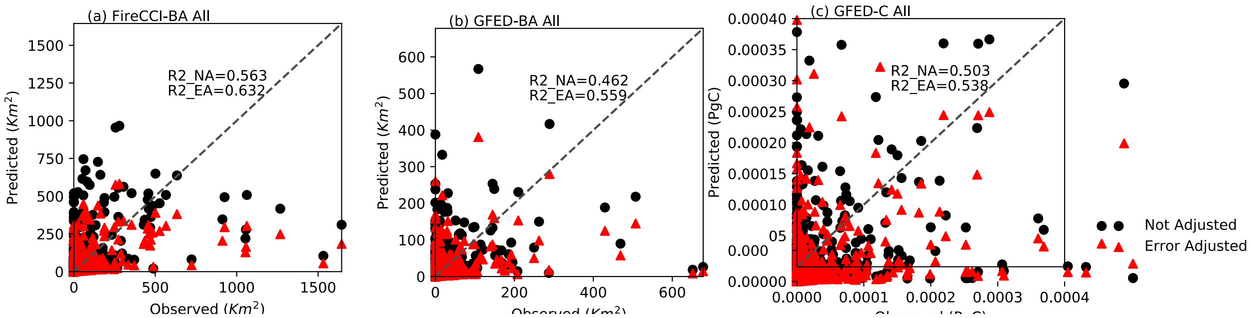

Figure S35 scatter plot based on the observational data (x-axis), the model predicted (not-adjusted) data, and error-adjusted predicted data (y-axis) from FireCCI (FireCCI-BA All) (a), GFED burned area (GFED-BA All) (b), and GFED C emissions (GFED-C All) (c); the black dot and red triangle means scatters from the original model predicted data (not adjusted) and error-adjusted data.