Supplementary Information (SI) for Environmental Science: Atmospheres.

Supplementary information

Fine particulate air pollution estimation in Ouagadougou using satellite aerosol optical depth and meteorological parameters

Joe Adabouk Amooli[a,b], Daniel M Westervelt*[b], Kwame Oppong Hackman[c], Bernard Nana[d]

Table S1: % values of data availability of $PM_{2.5}$ from the embassy dataset for the measurement periods

Month	% of Data Availability
Jan	65
Feb	100
March	100
April	100
May	100
June	20
July	14
August	86
September	100
October	17
November	53
December	65

Table S2: % values of data availability of satellite AOD over the data collection period.

Year	% of Data Availability	Year	% of Data Availability
2000	45	2012	60
2001	46	2013	61
2002	50	2014	56
2003	58	2015	63
2004	60	2016	58
2005	59	2017	62
2006	61	2018	56
2007	55	2019	55
2008	58	2020	58
2009	59	2021	63
2010	57	2022	57
2011	59		

Table S3*:* Pearson correlation coefficients between observed and satellite weather parameters at Ouagadougou International Airport

Parameters	Dry season r	Rainy Season r
Observed precipitation and CHIRPS precipitation (resampled to 1km resolution)	0.83	0.79
Observed relative humidity and Era5-Land relative humidity (resampled to 1km resolution)	0.87	0.89
Observed temperature and Era5-Land surface	0.97	0.84

temperature		
Observed wind speed and Era5-Land wind speed (resampled to 1km resolution)	0.90	0.88
Observed wind direction and Era5-Land wind direction (resampled to 1km resolution)	0.85	0.81

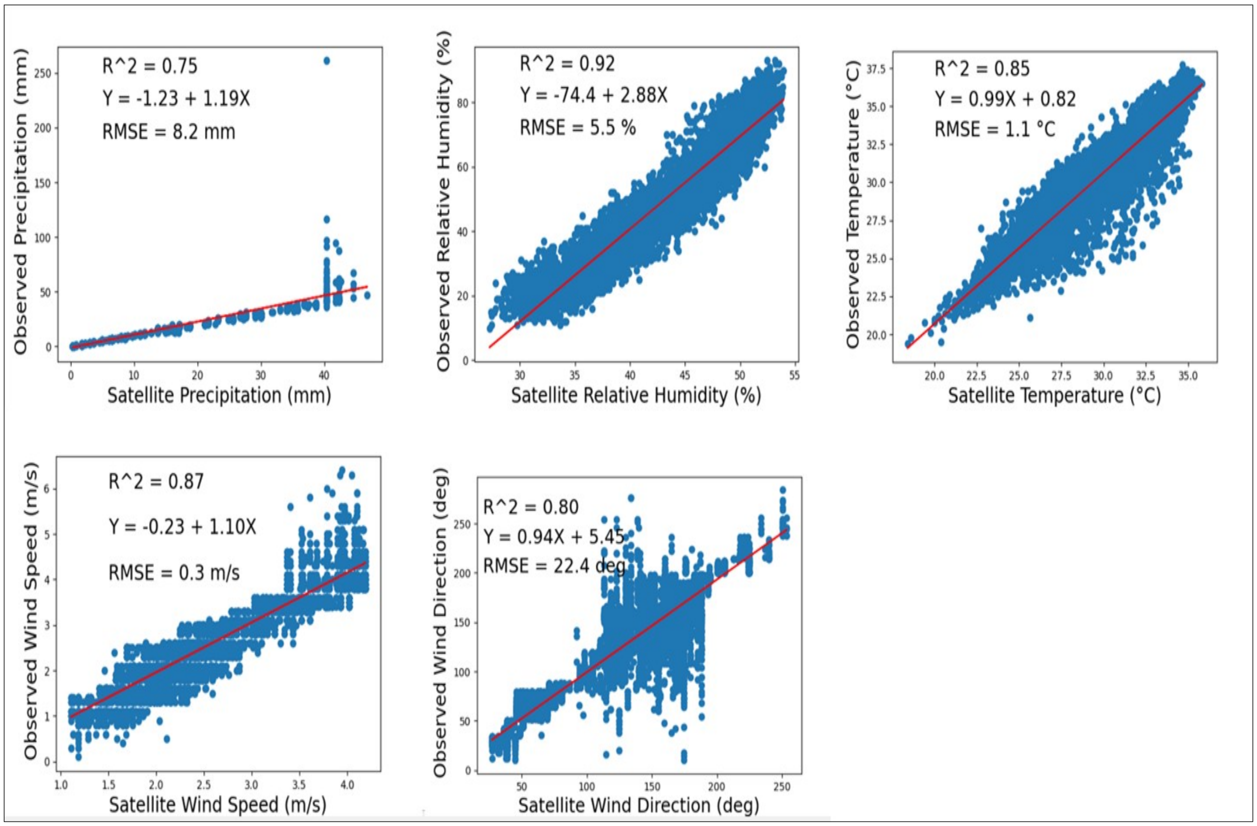

Figure S1: Simple linear models for correcting satellite data in Ouagadougou

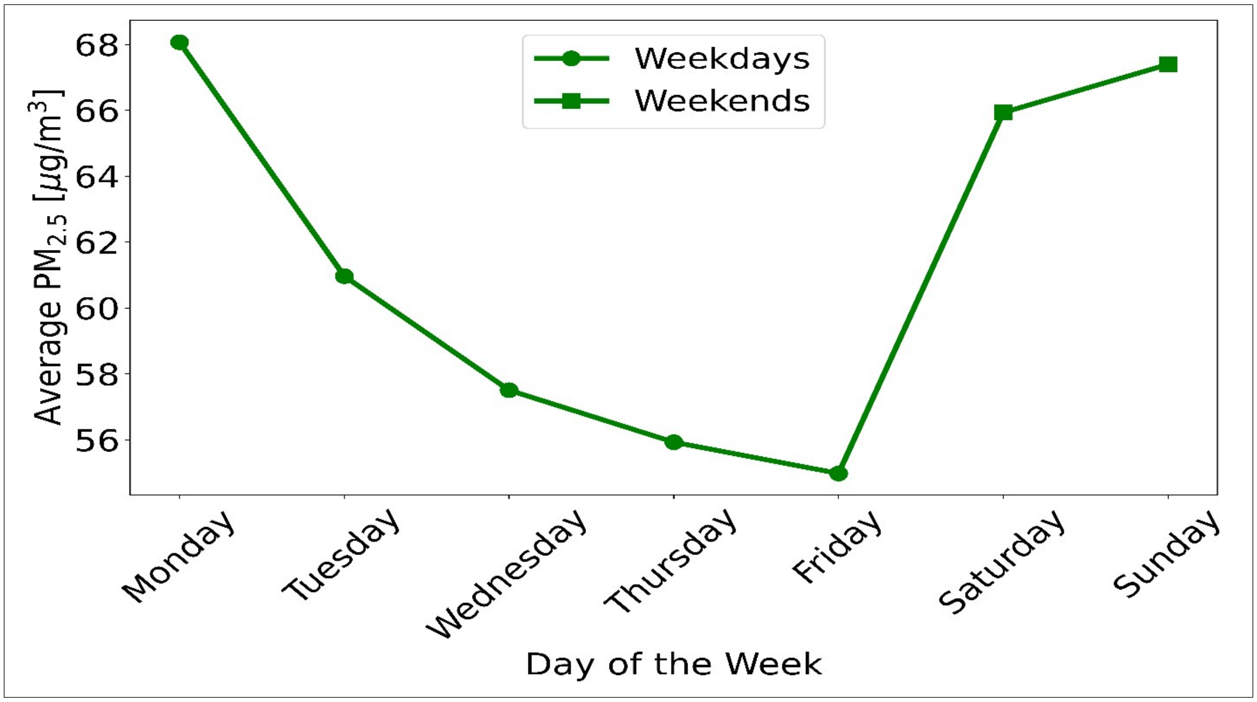

Figure S2: Temporal changes of $PM_{2.5}$ on weekdays and weekends

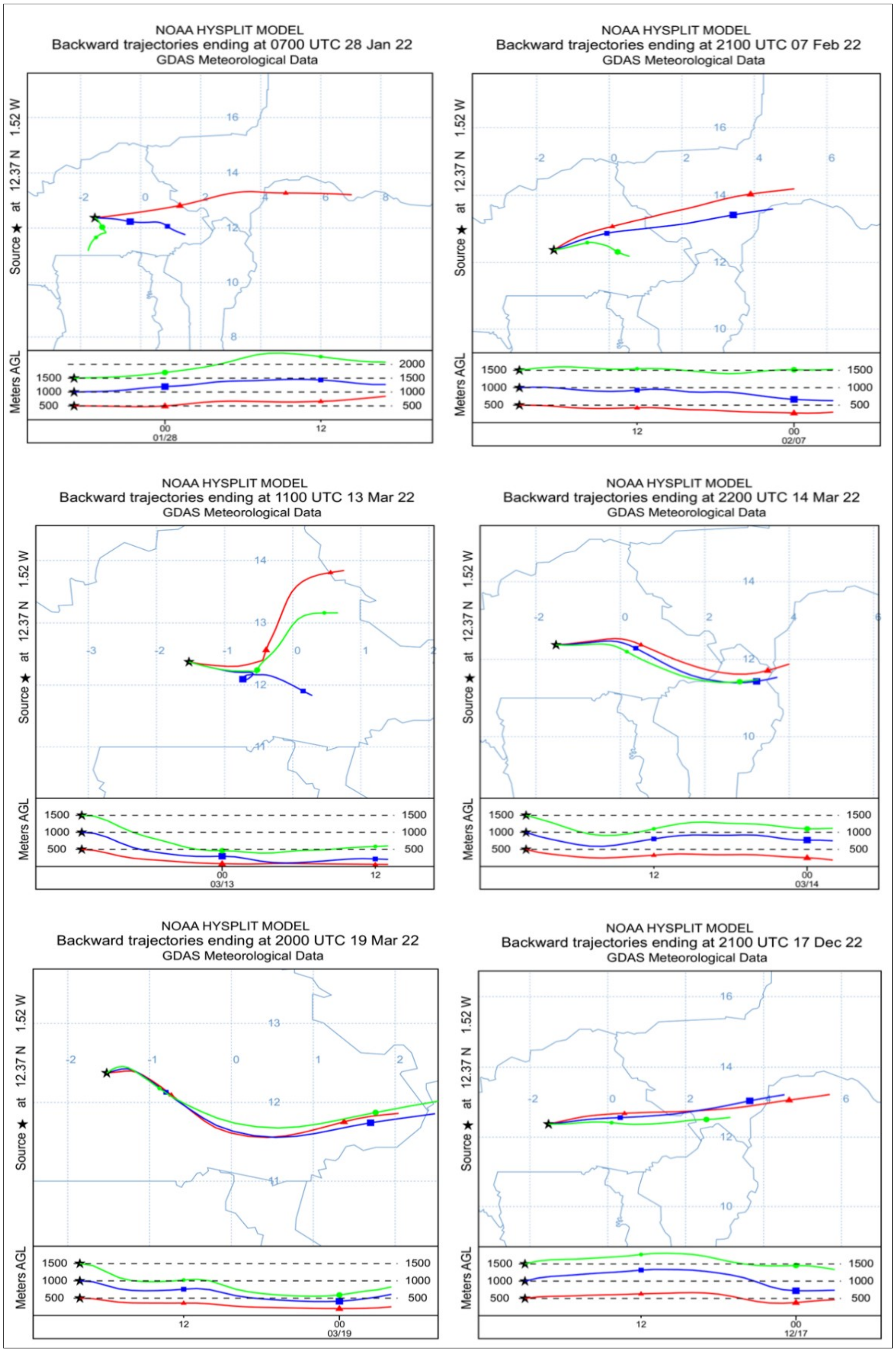

Figure S3: Potential source regions and contributions to $PM_{2.5}$ from Trajstat model (NOAA HYSPLIT model) for extremely polluted days in Ouagadougou at three receptor heights 500 m (red), 1000 m (blue), and 1500 m (green).

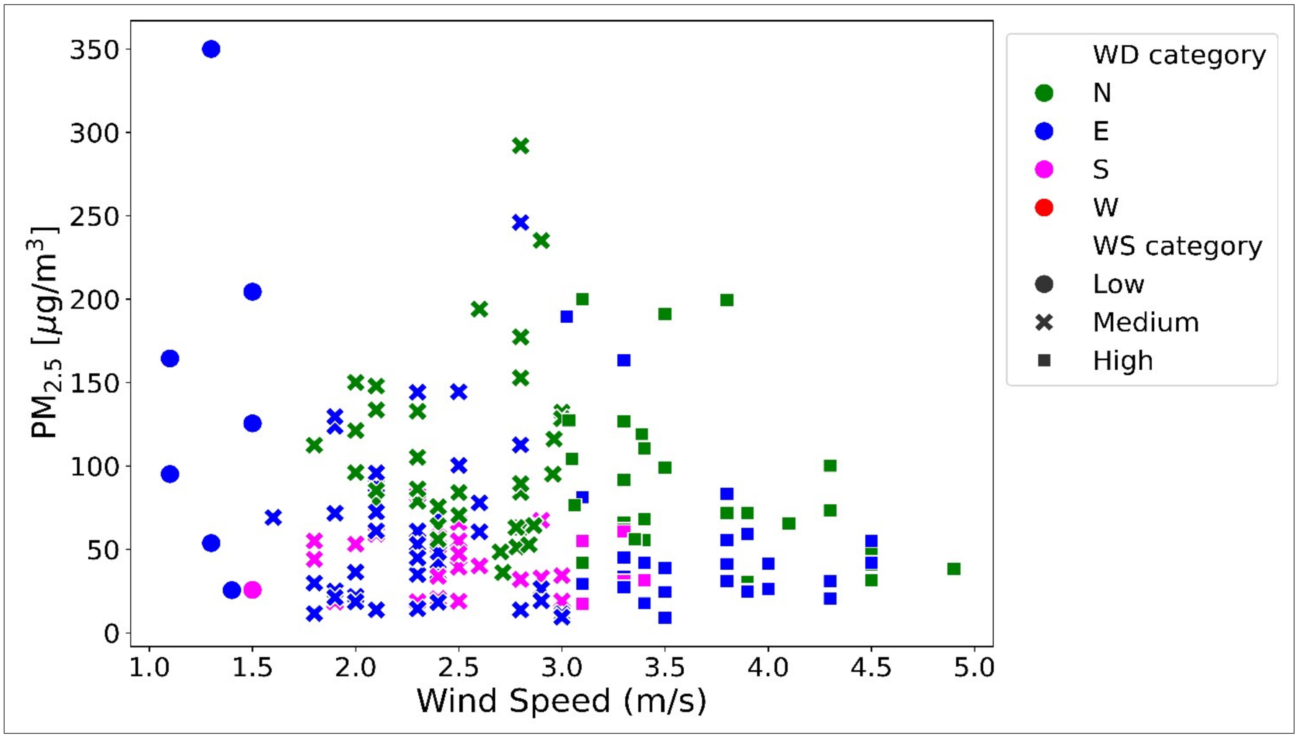

Figure S4: Conditional bivariate analysis of the relationship between wind speed, wind direction, and $PM_{2.5}$ concentrations

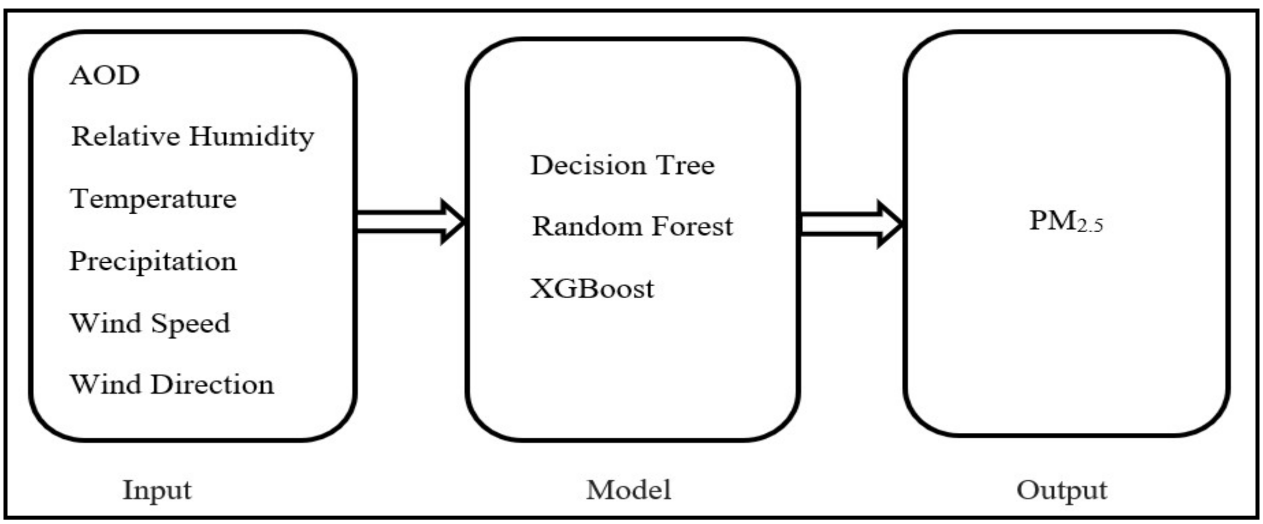

Figure S5*:* Schematic of AOD and weather parameters as model input

Table S4*:* Models and hyperparameters set to obtain optimal performance.

Model	**Hyperparameters set**
Decision tree	ccp_alpha = 0.01

	max_depth = 5 max_features = 'sqrt' min_samples_split = 2 random_state = 42 splitter = 'best'
Random forest	max_depth =7 n_estimators = 50 max_features = ‘sqrt’ ccp_alpha=0.01 min_samples_split = 4 min_samples_leaf=2
XGBoost	base_score = 0.5 booster = gbtree callbacks = None colsample_bylevel = None colsample_bynode = None colsample_bytree = 0.5 early_stopping_rounds = None enable_categorical = False eval_metric = None feature_types = None gamma = 0.4 gpu_id = None grow_policy = None importance_type = None interaction_constraints = None learning_rate =0.1 max_bin = None max_cat_threshold = None max_cat_to_onehot = None max_delta_step = None max_depth = 7 max_leaves = None min_child_weight = 7 missing = nan monotone_constraints = None n_estimators = 100 n_jobs = None num_parallel_tree = None predictor = None random_state = 42

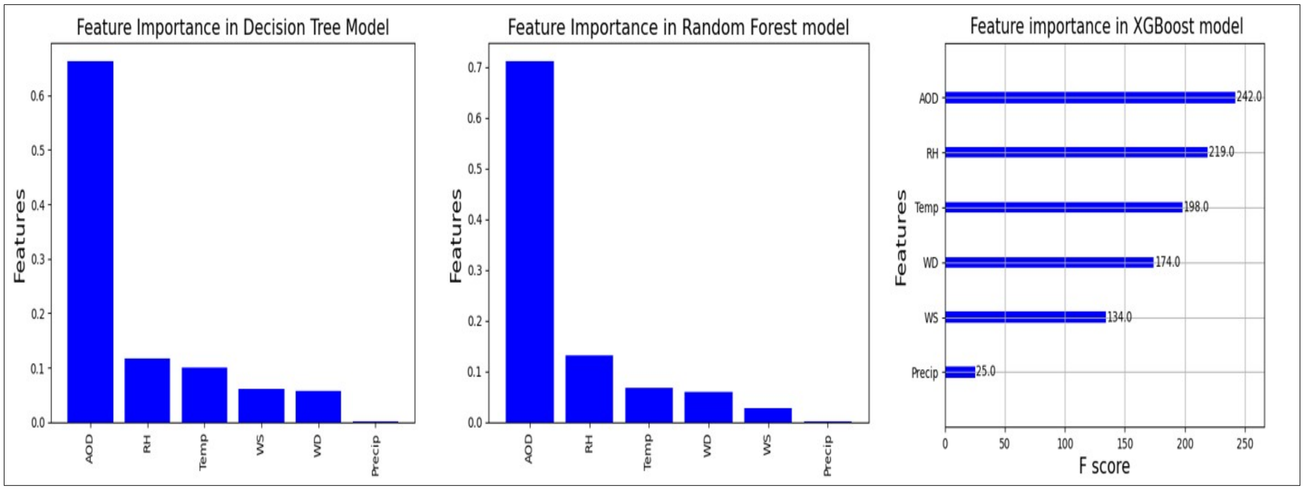

Figure S6: Feature importance in the DT, RF, and XGBoost models

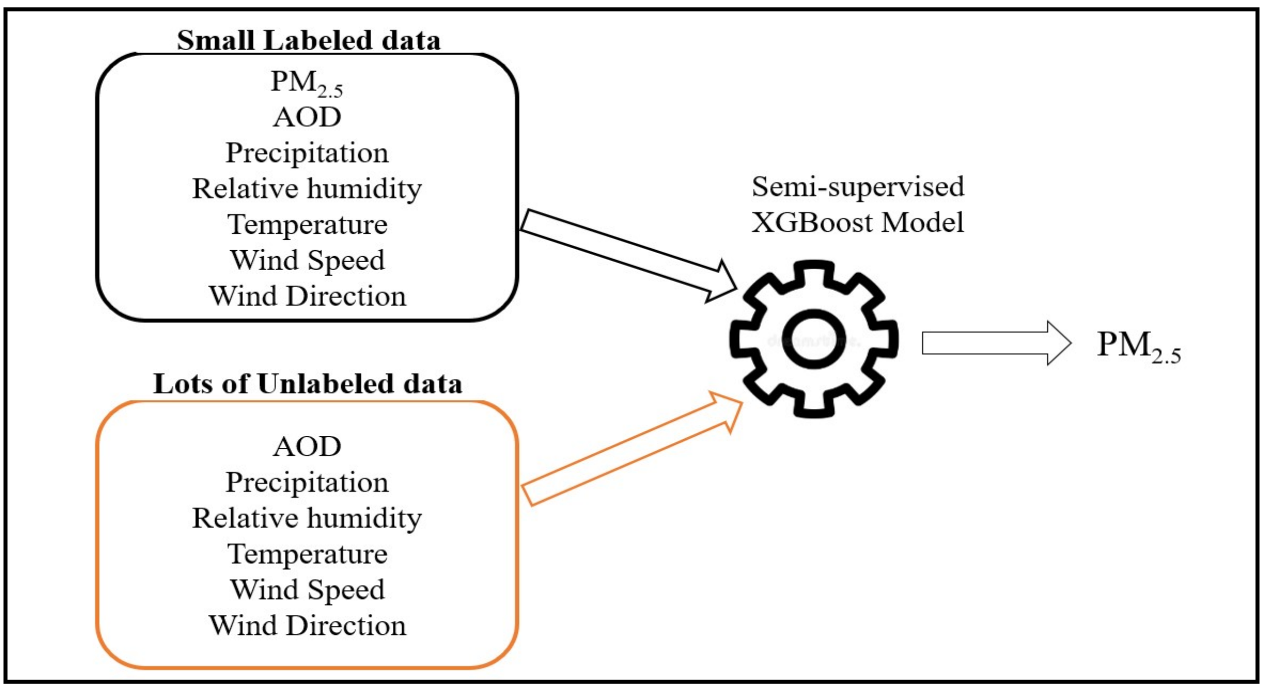

Figure S7: Schematic of semi-supervised XGBoost

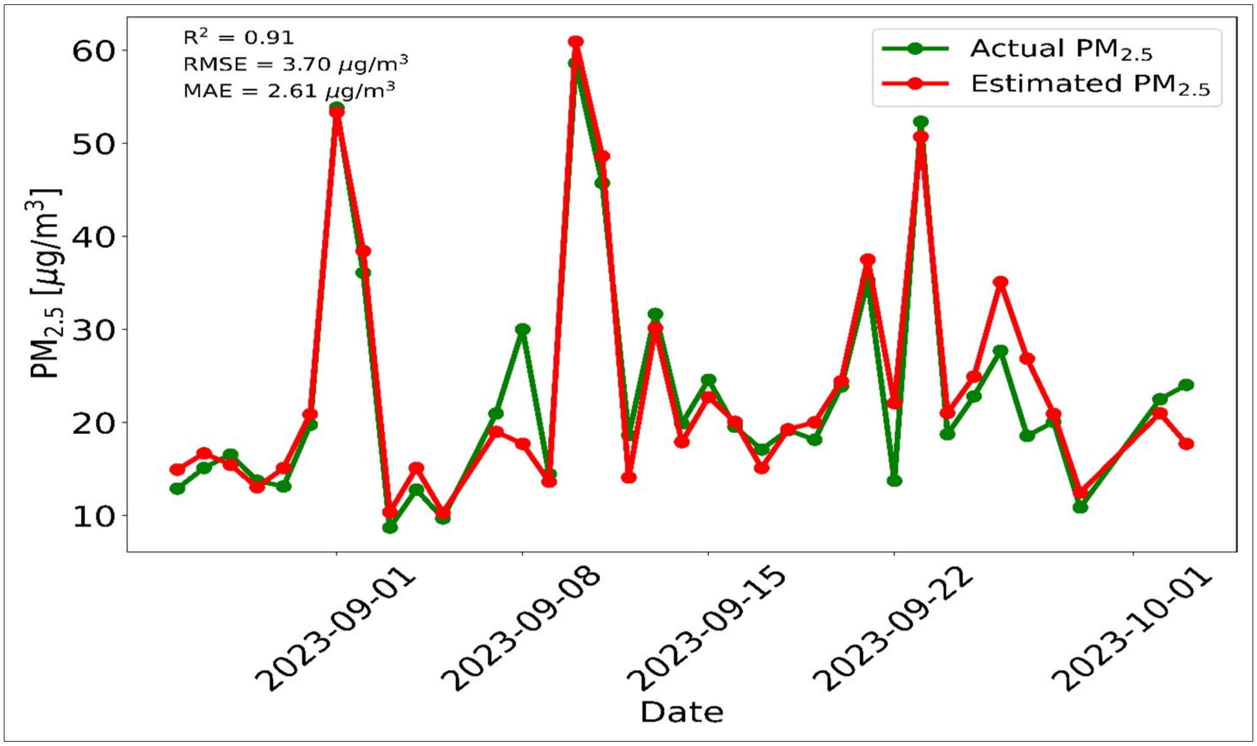

Figure S8: Validation of the semi-supervised XGBoost model using an independent dataset (TEOM 1400a, a federal equivalent method gravimetric $PM_{2.5}$ monitor located at Université Joseph Ki-Zerbo in Ouagadougou)