Supplementary Material for 'Investigating the impact of meteorology and emissions on $PM_{2.5}$ and PM_{10} in Delhi using machine learning'

SM 1 - Station list for the data on different pollutants – $PM_{2.5}$, PM_{10}, and CO

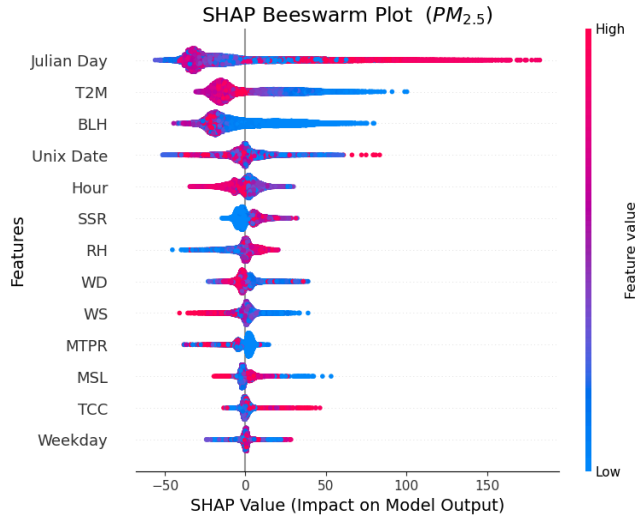
S.no.	Station Name	$PM_{2.5}$	PM_{10}	CO
1.	Anand Vihar, New Delhi - DPCC	✓	✓	✓
2.	Ashok Vihar, Delhi - DPCC	✓	✓	✓
3.	Aya Nagar, New Delhi - IMD	✓	✓	✓
4.	CRRI Mathura Road, New Delhi - IMD	✓	✓	✓
5.	Dr. Karni Singh Shooting Range, Delhi - DPCC	✓	✓	✓
6.	DTU, New Delhi - CPCB	✓	✓	✓
7.	Dwarka-Sector 8, Delhi - DPCC	✓	✓	✓
8.	IGI Airport (T3), Delhi - IMD	✓	✓	✓
9.	IHBAS, Dilshad Garden, New Delhi - CPCB	✓	×	✓
10.	ITO, New Delhi - CPCB	✓	✓	✓
11.	Jahangirpuri, Delhi - DPCC	✓	✓	✓
12.	Jawaharlal Nehru Stadium, Delhi - DPCC	✓	✓	✓
13.	Lodhi Road, New Delhi - IMD	✓	✓	✓
14.	Major Dhyan Chand National Stadium, Delhi - DPCC	✓	✓	✓
15.	Mandir Marg, New Delhi - DPCC	✓	✓	✓
16.	Najafgarh, Delhi - DPCC	✓	✓	✓
17.	Narela, Delhi - DPCC	✓	✓	✓
18.	Nehru Nagar, Delhi - DPCC	✓	✓	✓
19.	North Campus, DU, Delhi - IMD	✓	✓	✓
20.	NSIT Dwarka, Delhi - CPCB	✓	×	✓
21.	Okhla Phase-2, Delhi - DPCC	✓	✓	✓
22.	Patparganj, Delhi - DPCC	✓	✓	✓
23.	Punjabi Bagh, Delhi - DPCC	✓	✓	✓
24.	Pusa, Delhi - IMD	✓	✓	✓
25.	R K Puram, Delhi - DPCC	✓	✓	✓

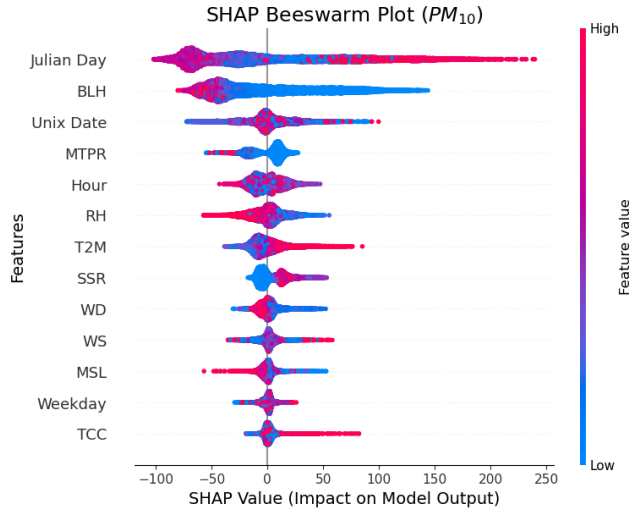
26.	Rohini, Delhi - DPCC	✓	✓	✓
27.	Shadipur, Delhi - CPCB	✓	×	✓
28.	Sirifort, Delhi - CPCB	✓	✓	✓
29.	Sonia Vihar, Delhi - DPCC	✓	✓	✓
30.	Vivek Vihar, Delhi - DPCC	✓	✓	✓
31.	Wazirpur, Delhi - DPCC	✓	✓	✓

Table S1. Continuous Ambient Air Quality Monitoring Station (CAAQMS) names for the pollutants, along with the availability for each pollutant.

SM 2 - SHAP (SHapley Additive exPlanations) Values

$PM_{2.5}$, PM_{10} and CO





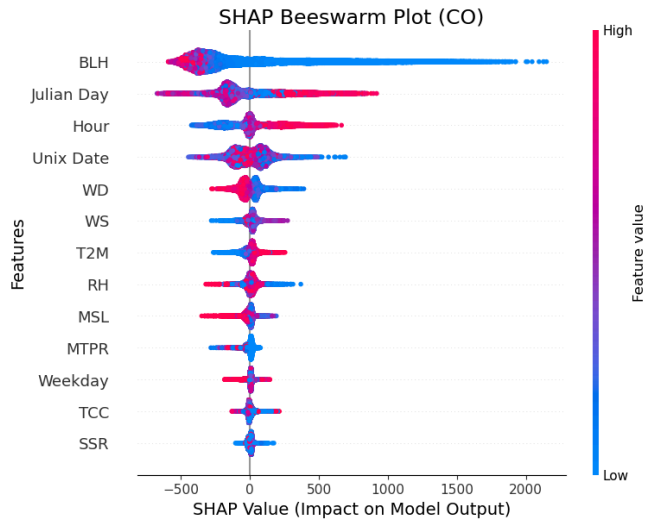


Figure S1. SHAP beeswarm plots showing feature importance for model predictions of $PM_{2.5}$, PM_{10}, and CO. SHAP values display the impact of increasing or decreasing the feature value on the model output. For instance, low BLH values (blue points) are linked with high pollutant levels (positive SHAP values), which is intuitive as a lower boundary layer height results in a weaker vertical pollutant dispersion.

For the three pollutants, $PM_{2.5}$, PM_{10}, and CO, temporal variables like Julian day, hour, and Unix date have a very strong impact, indicating strong seasonal and diurnal patterns in emissions and concentrations. Of meteorological variables, boundary layer height (BLH), temperature (T2M), relative humidity (RH), and solar radiation (SSR) have high impacts on predictions. It is important to note how RH behaves differently for $PM_{2.5}$ and PM_{10}. For $PM_{2.5}$, higher values of RH lead to higher concentrations. In contrast, for PM_{10}, we observe that the opposite holds, with higher RH values reducing the concentration.

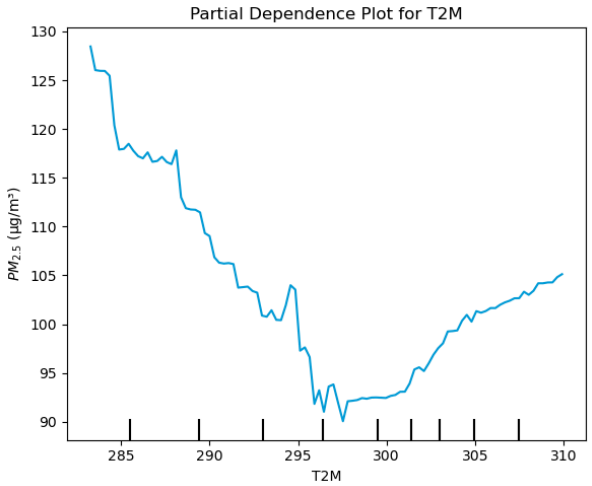
SM 3 - Hyperparameter Tuning

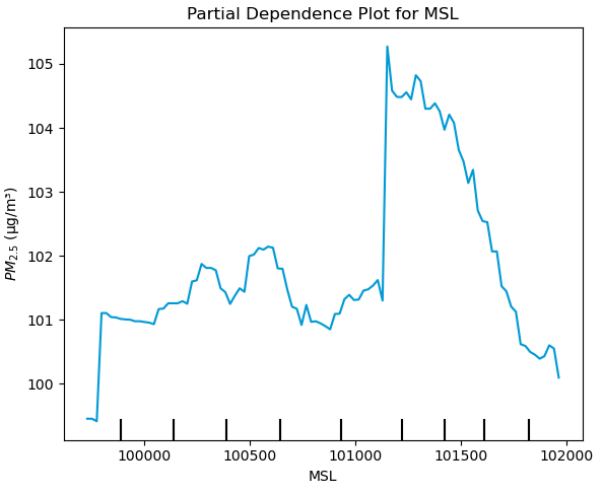
To optimise our models, we perform hyperparameter tuning in two steps. First, we use RandomizedSearchCV from scikit-learn to arrive at a near-optimal set of hyperparameters. We then narrow down our search around these hyperparameters using GridSearchCV. We optimise the number of trees, learning rate, the maximum depth to which the tree is allowed to grow, and the minimal number of samples to be present at a leaf node using cross-validation.

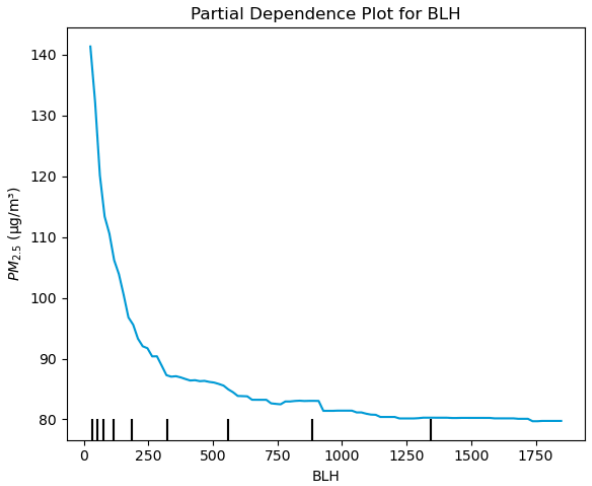
SM 4 - Partial Dependence Plots

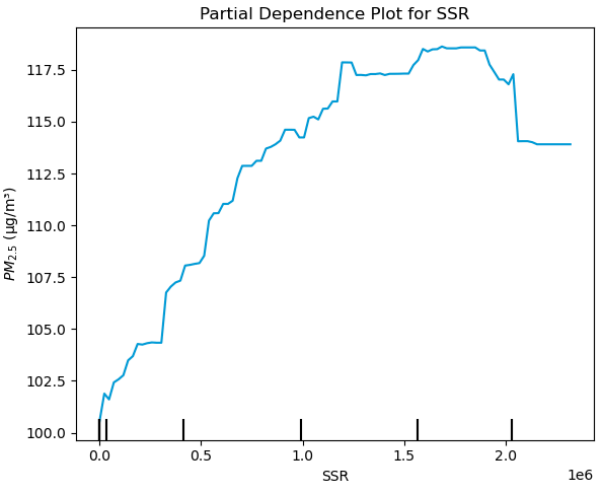
Partial Dependence (PD) plots provide a visualisation of the impact of the meteorological and temporal variables on the pollutant concentrations predicted by the model. PD plots show that temperature and RH affected the predictions of $PM_{2.5}$ and PM_{10} differently, with higher temperature and RH increasing PM10 and lowering $PM_{2.5}$.

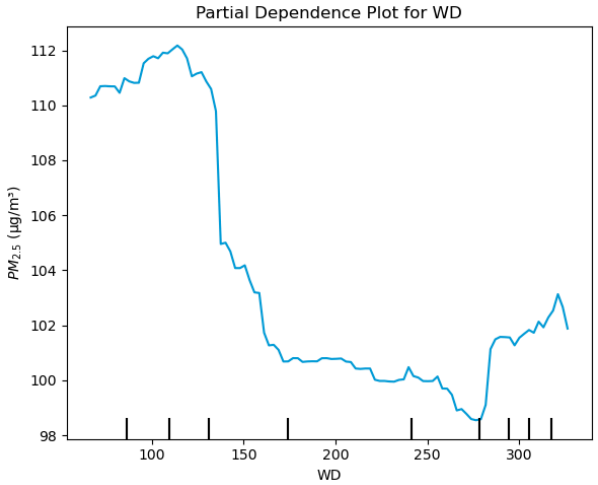
$PM_{2.5}$

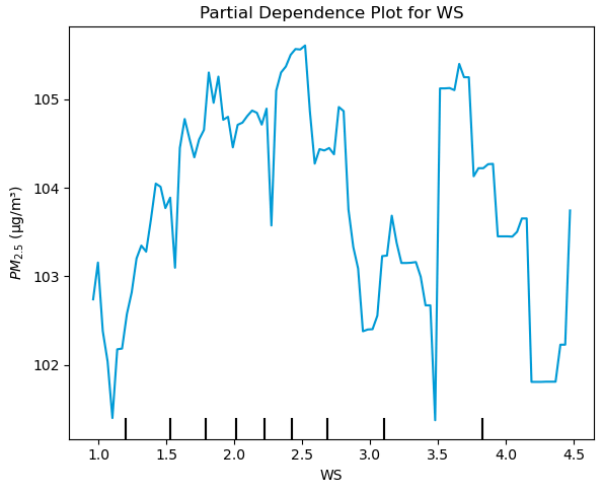

Partial Dependence Plot for T2M
$PM_{2.5}$ (μg/m³)
T2M

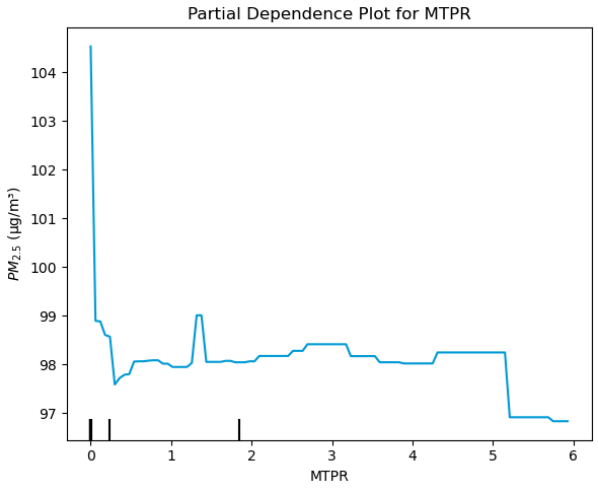

Partial Dependence Plot for MSL
$PM_{2.5}$ (μg/m³)
MSL

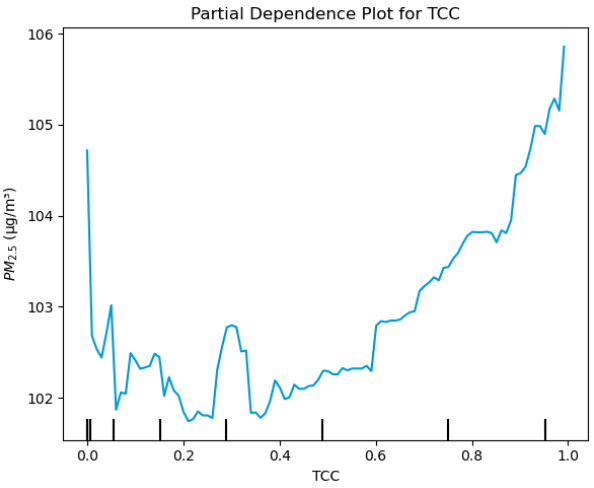

Partial Dependence Plot for BLH
$PM_{2.5}$ (μg/m³)
BLH

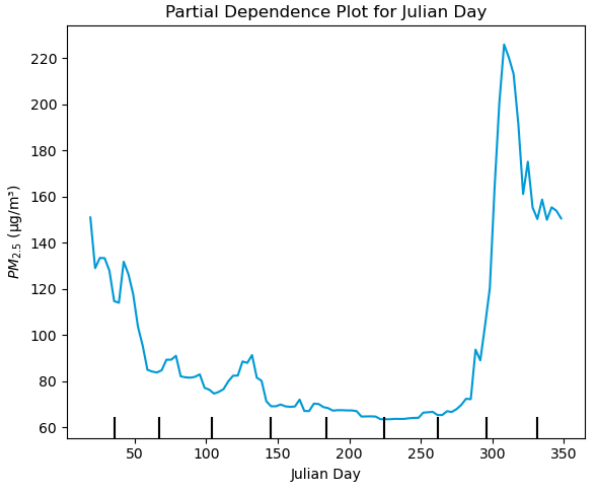

Partial Dependence Plot for SSR
$PM_{2.5}$ (μg/m³)
SSR
1e6

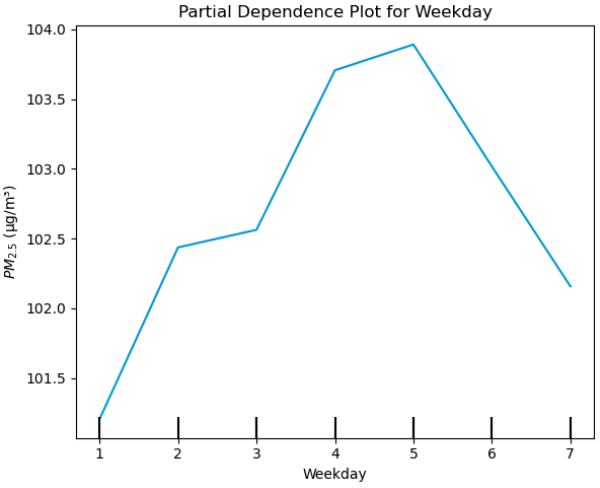

Partial Dependence Plot for WD
$PM_{2.5}$ (μg/m³)
WD


Partial Dependence Plot for WS
$PM_{2.5}$ (μg/m³)
WS


Partial Dependence Plot for MTPR
$PM_{2.5}$ (μg/m³)
MTPR


Partial Dependence Plot for TCC
$PM_{2.5}$ (μg/m³)
TCC


Partial Dependence Plot for Julian Day
$PM_{2.5}$ (μg/m³)
Julian Day


Partial Dependence Plot for Weekday
$PM_{2.5}$ (μg/m³)
Weekday

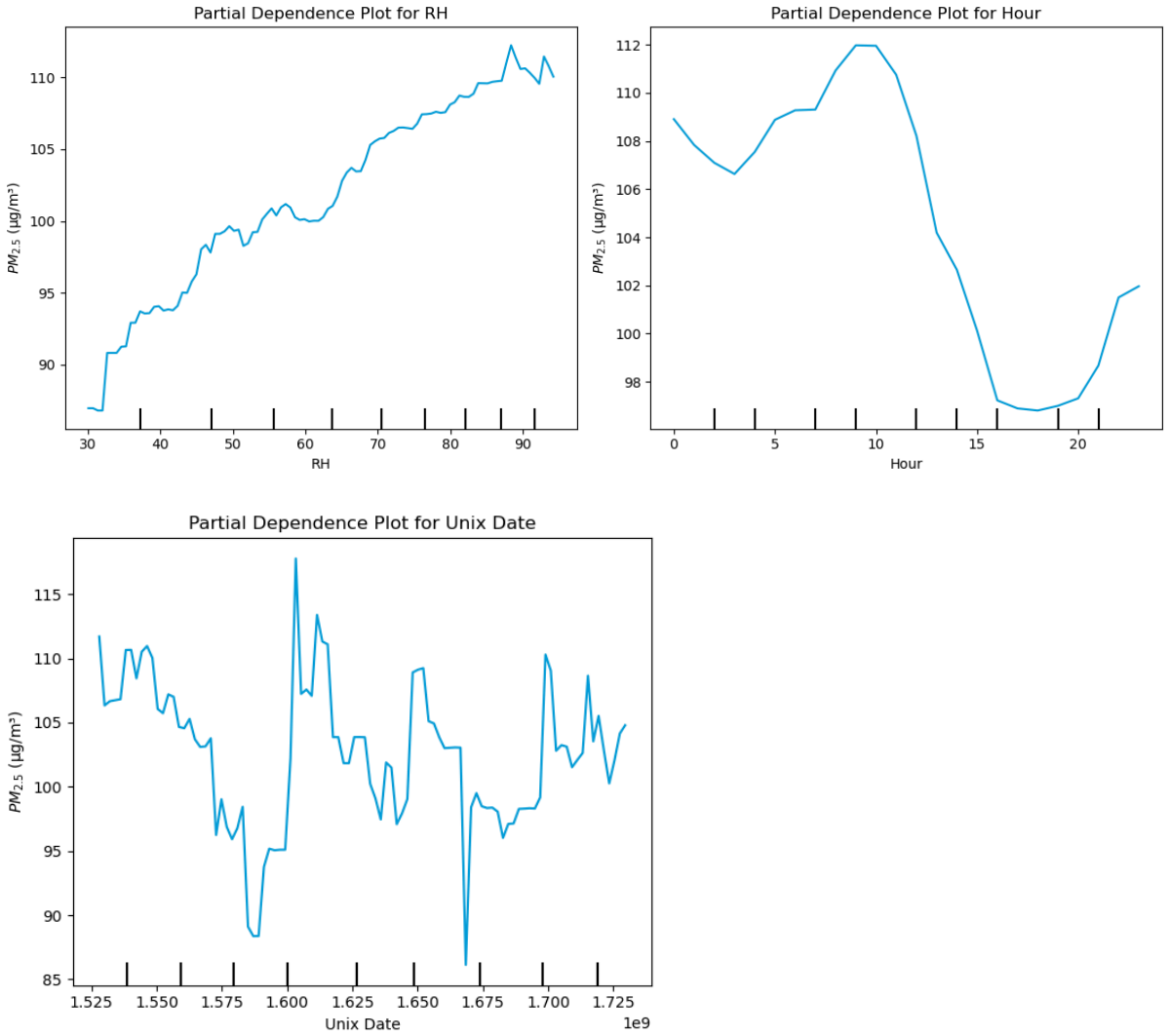
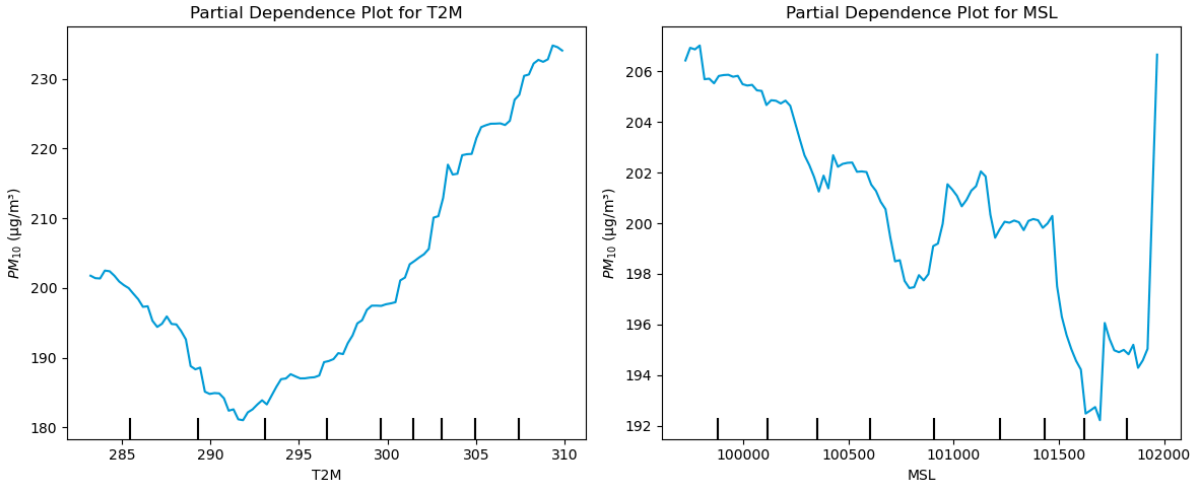
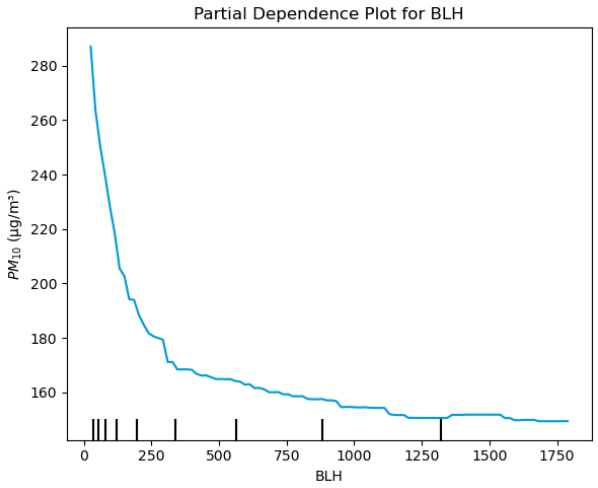
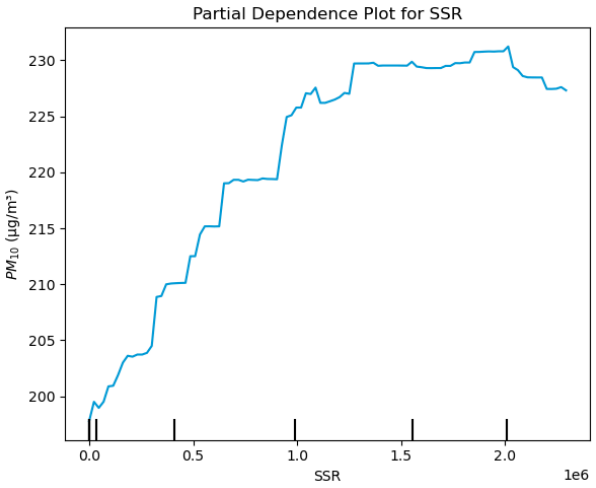


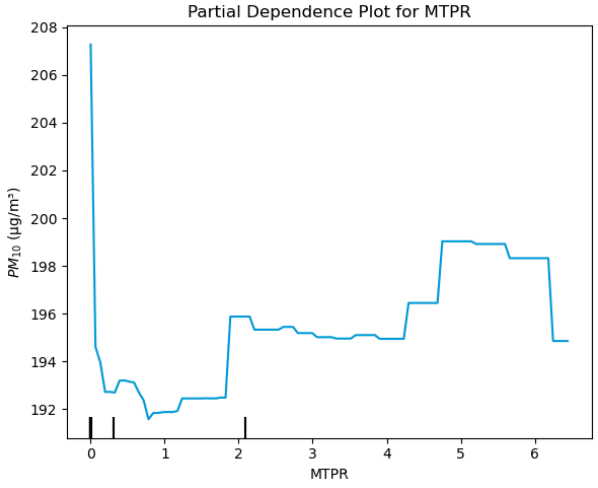
Figure S2. Partial Dependence plots for meteorological and temporal variables for $PM_{2.5}$, displaying the impact of the input variables on the PM_{10} concentrations predicted by the model.

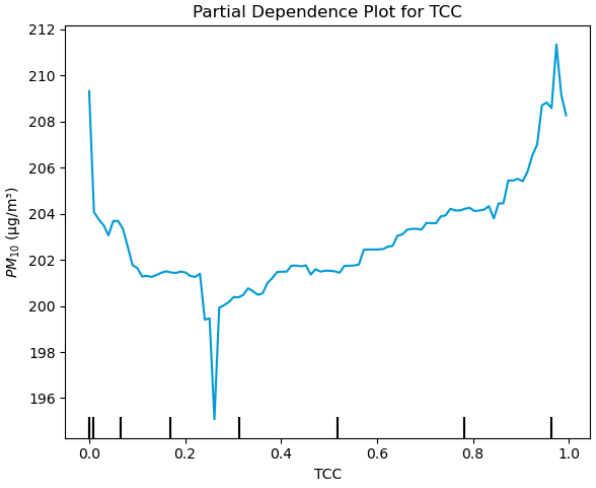
PM_{10}

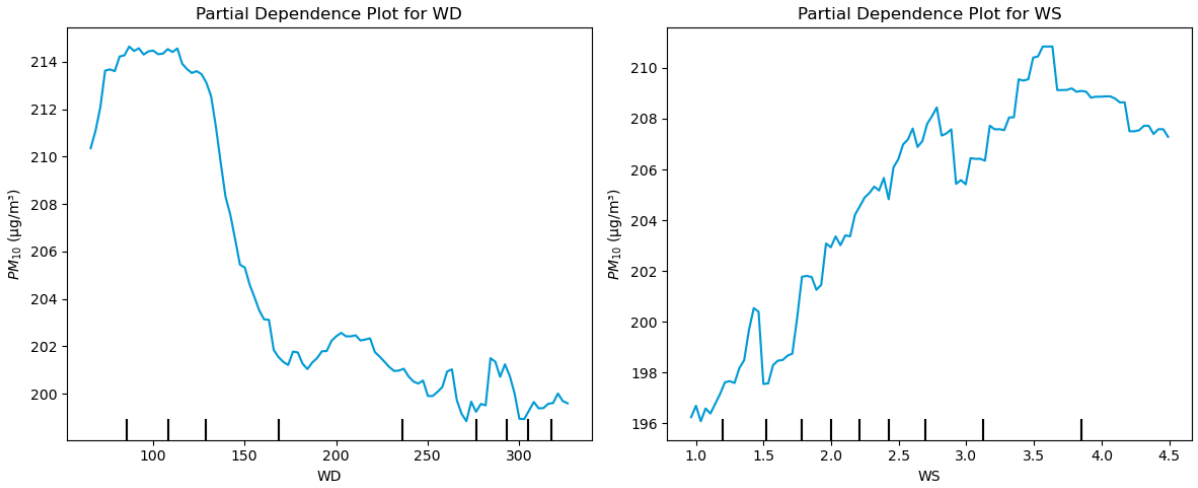



Partial Dependence Plot for BLH
PM_{10} (μg/m³)
BLH

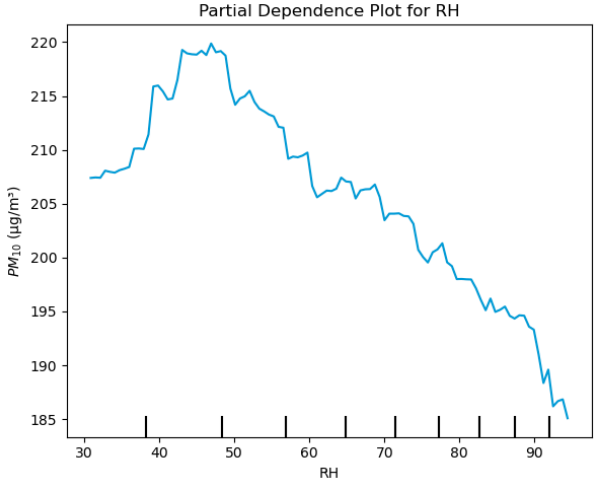

Partial Dependence Plot for SSR
PM_{10} (μg/m³)
SSR
1e6

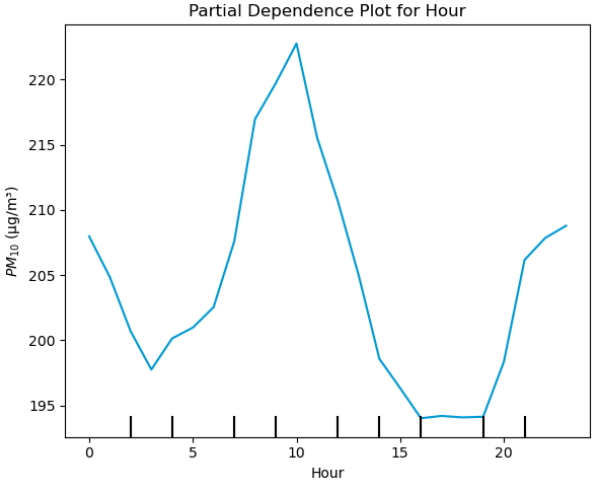

Partial Dependence Plot for MTPR
PM_{10} (μg/m³)
MTPR

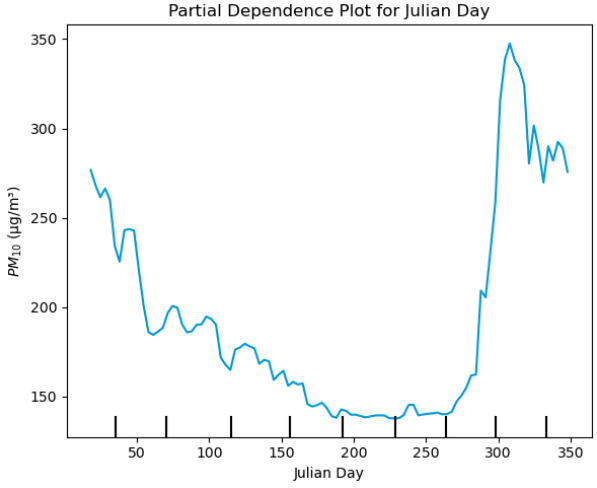

Partial Dependence Plot for TCC
PM_{10} (μg/m³)
TCC

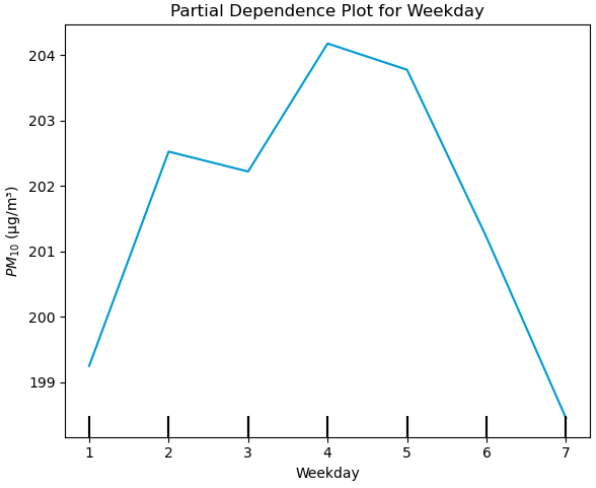

Partial Dependence Plot for WD
PM_{10} (μg/m³)
WD

Partial Dependence Plot for WS
PM_{10} (μg/m³)
WS


Partial Dependence Plot for RH
PM_{10} (μg/m³)
RH


Partial Dependence Plot for Hour
PM_{10} (μg/m³)
Hour


Partial Dependence Plot for Julian Day
PM_{10} (μg/m³)
Julian Day


Partial Dependence Plot for Weekday
PM_{10} (μg/m³)
Weekday

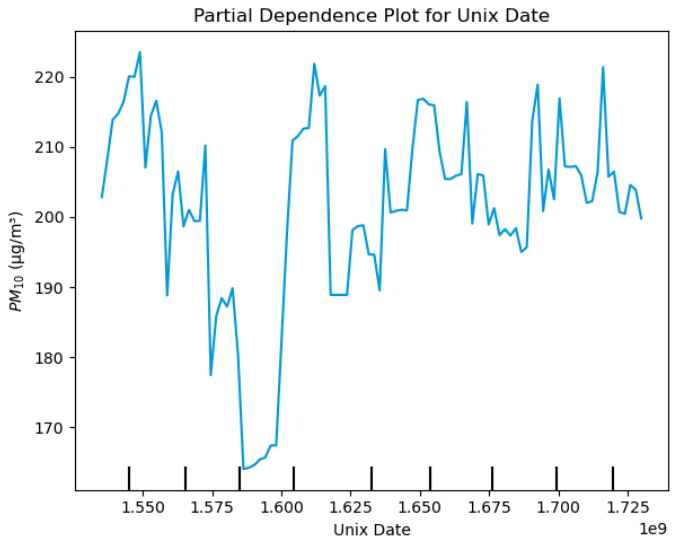
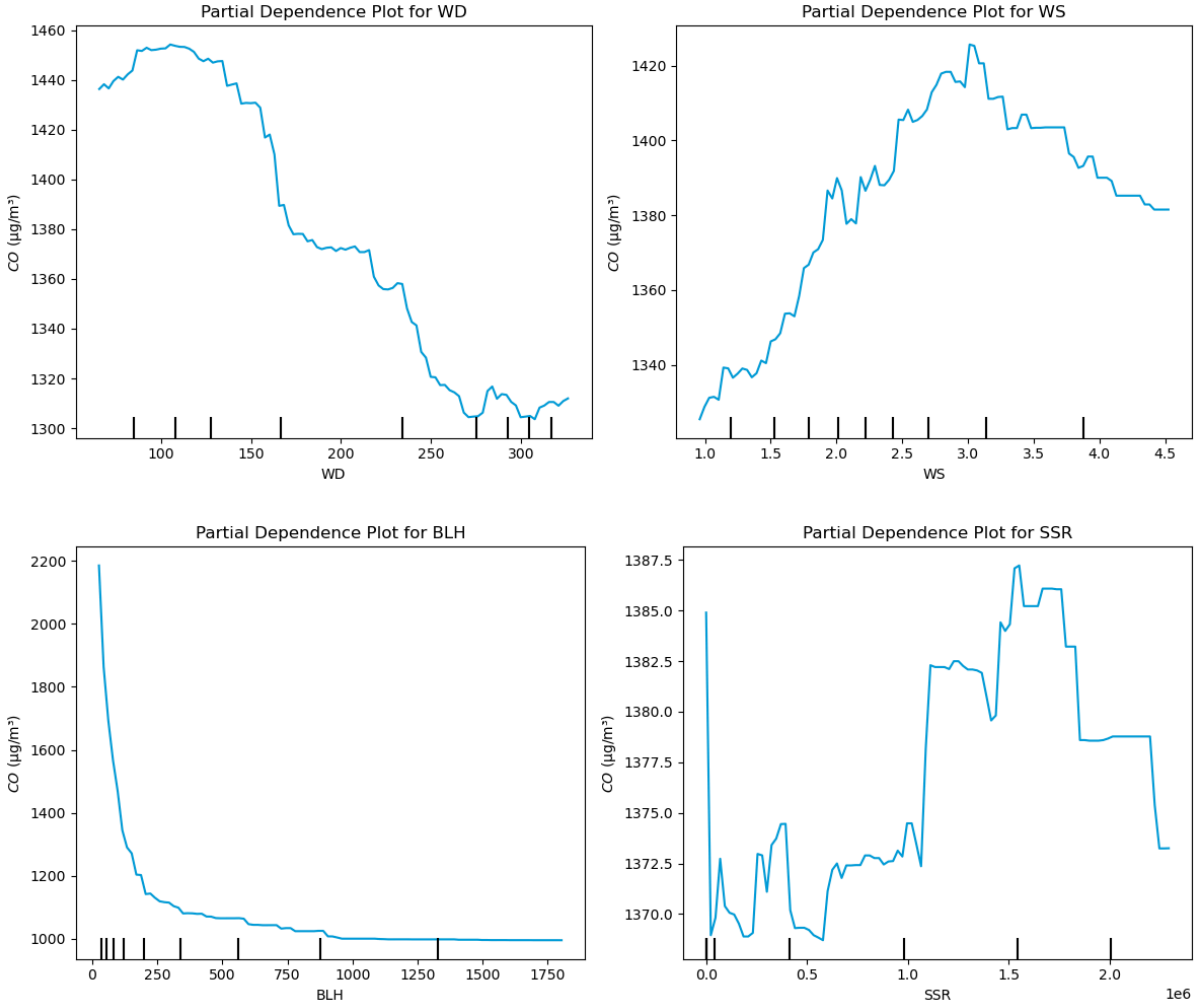
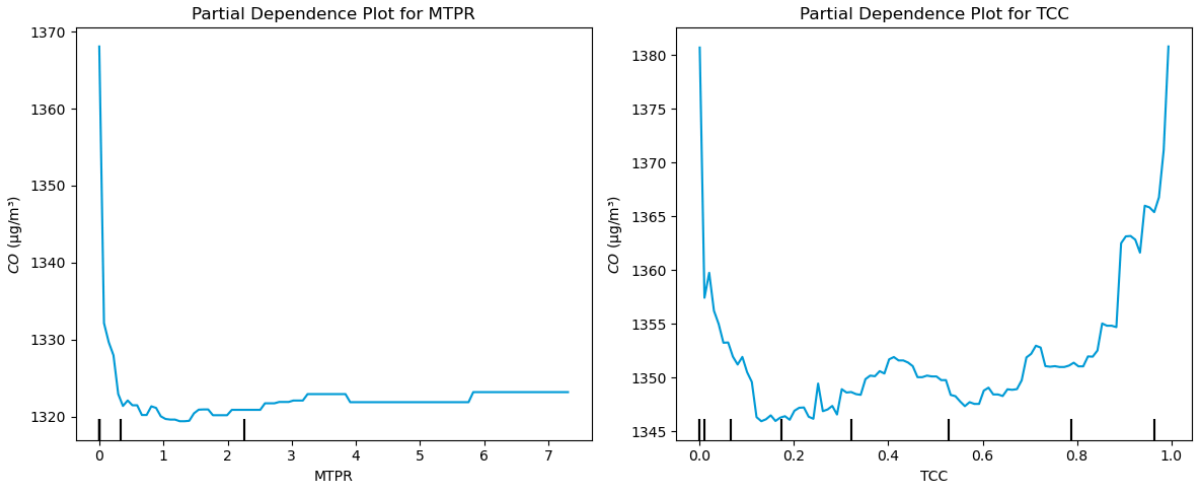


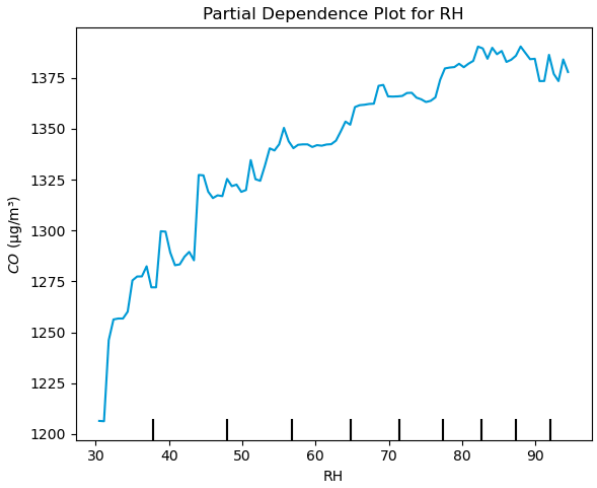
Figure S3. Partial Dependence plots for meteorological and temporal variables for PM_{10}, displaying the impact of the input variables on the PM_{10} concentrations predicted by the model.

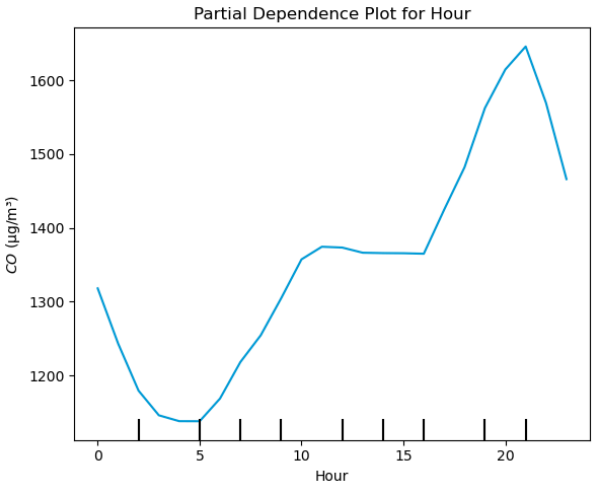
CO

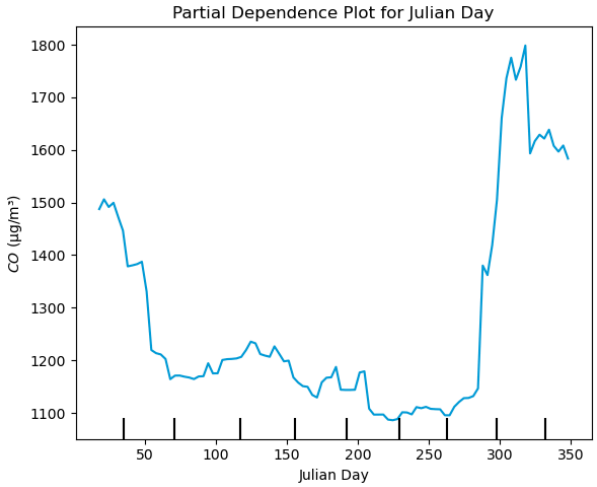

Partial Dependence Plot for WD
CO (µg/m³)
WD
Partial Dependence Plot for WS
CO (µg/m³)
WS
Partial Dependence Plot for BLH
CO (µg/m³)
BLH
Partial Dependence Plot for SSR
CO (µg/m³)
SSR
1e6

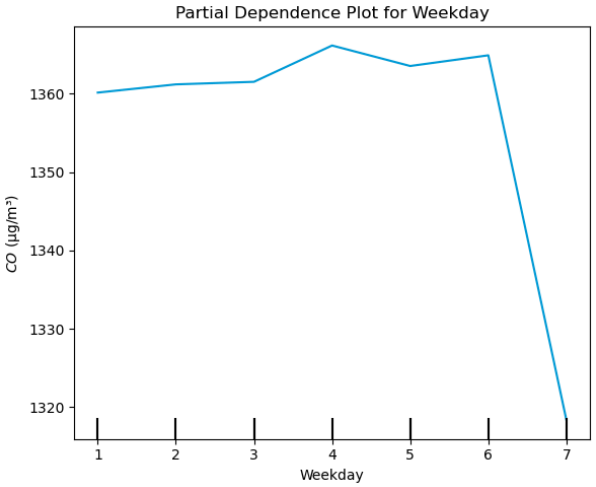

Partial Dependence Plot for MTPR
CO (µg/m³)
MTPR

Partial Dependence Plot for TCC
CO (µg/m³)
TCC


Partial Dependence Plot for RH
CO (µg/m³)
RH


Partial Dependence Plot for Hour
CO (µg/m³)
Hour


Partial Dependence Plot for Julian Day
CO (µg/m³)
Julian Day


Partial Dependence Plot for Weekday
CO (µg/m³)
Weekday

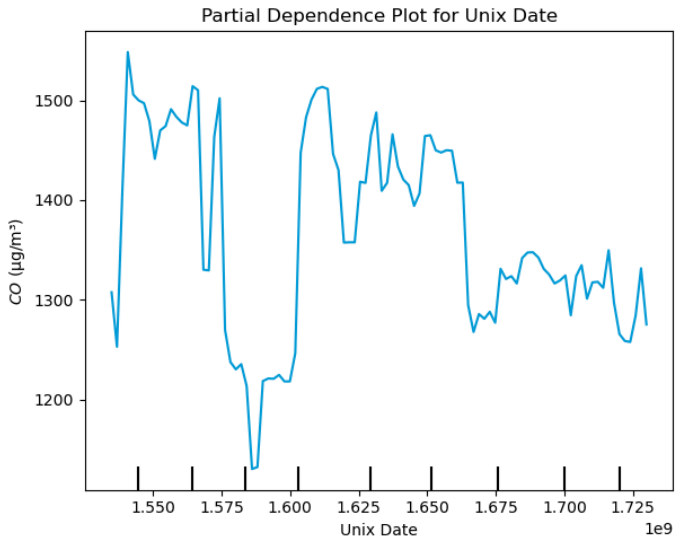


Figure S4. Partial Dependence plots for meteorological and temporal variables for CO, denoting how each variable affects the CO concentration.