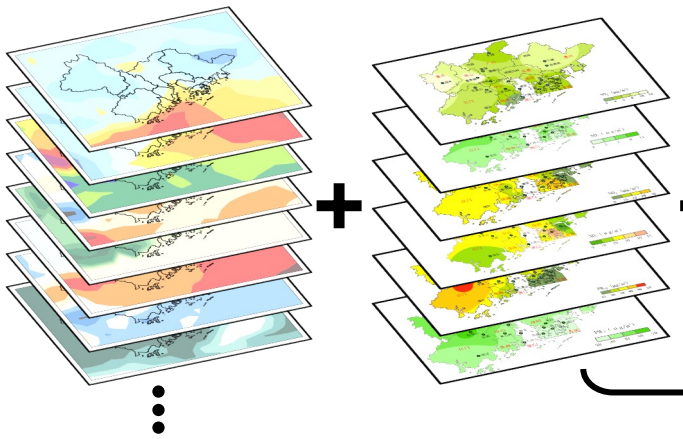
Step 1: Data acquisition & matching

Input variables

ERA5 meteorological data:
tp, tp5d, tp10d, tp30d, blh, d2m, t2m, u10, v10, msl, tcc, w850, lai

Air pollutant data:
SO_2_max, NO_2_max, PM10_max, SO_2_mean, NO_2_mean, PM10_mean

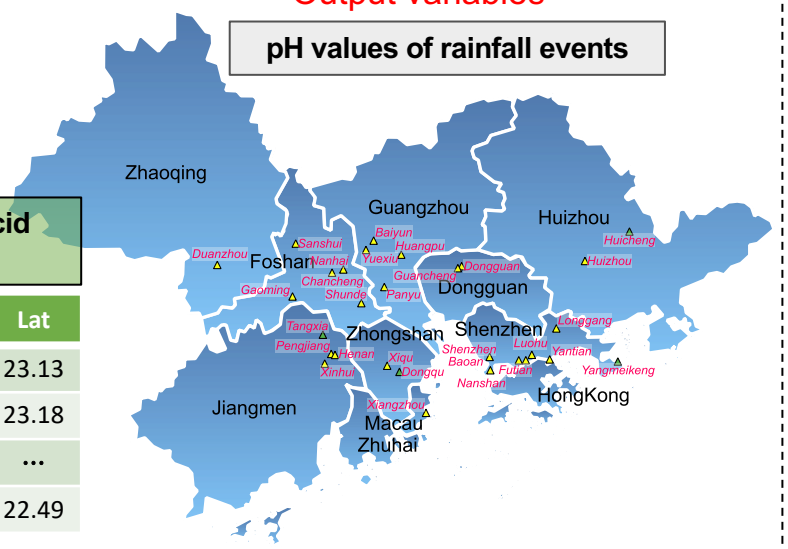


Location of acid rain stations

Station	Lon	Lat
Yuexiu	113.26	23.13
Baiyun	113.30	23.18
...	...	...
Dongqu	113.43	22.49

Output variables

pH values of rainfall events



Spatial and temporal matching

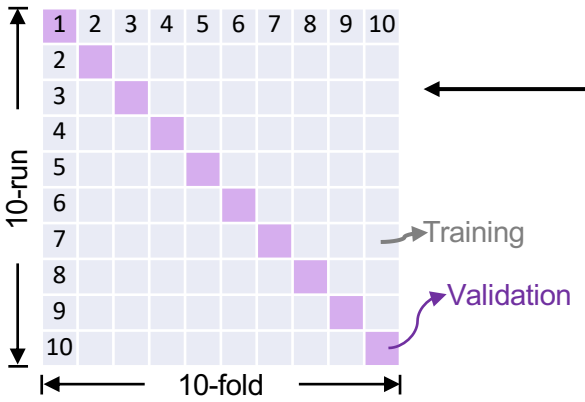
Step 2: XGBoost model establishment

Split original data into random training and testing sets

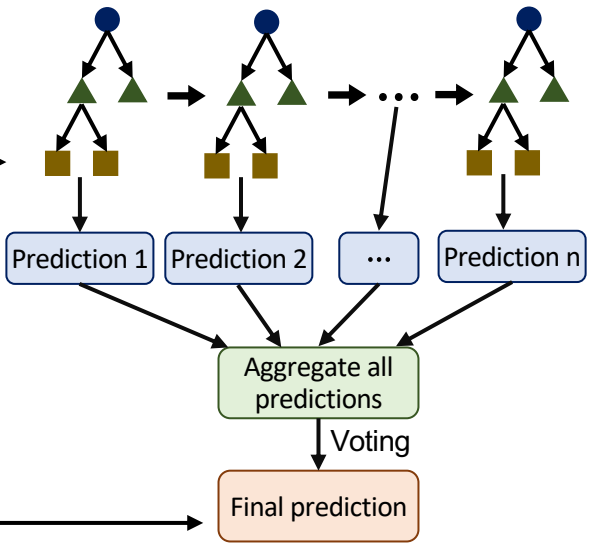
70% training set

30% testing set

10-fold cross-validation to determine the best parameters



Establishing XGBboost model



Testing the performance of the XGBoost model

Step 3: Model interpretation via SHAP

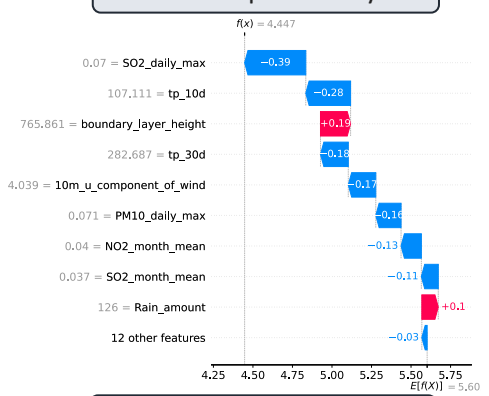
Model

Shap

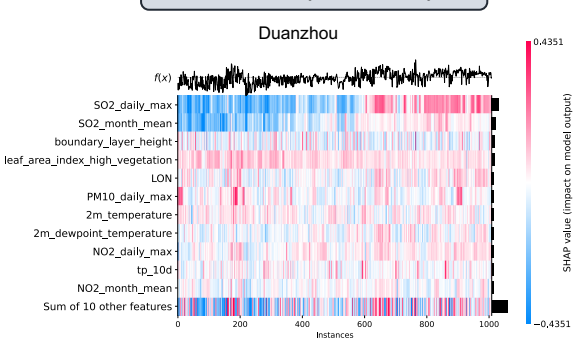
Shap.Explainer( )

Interpreting analyses

Local explainability



Cohort explainability



Global explainability