Supplement of Geosci. Model Dev., 16, 4867–4882, 2023
https://doi.org/10.5194/gmd-16-4867-2023-supplement

Geoscientific Model Development
Open Access
EGU

Supplement of

A gridded air quality forecast through fusing site-available machine learning predictions from RFSML v1.0 and chemical transport model results from GEOS-Chem v13.1.0 using the ensemble Kalman filter

Li Fang et al.

Correspondence to: Jianbing Jin (jianbing.jin@nuist.edu.cn) and Hong Liao (hongliao@nuist.edu.cn)

Table S1. Summary of selected features.

Region	NCP	PRD	SCB	YRD	FWP	REST
	$PM_{2.5}$	$PM_{2.5}$	$PM_{2.5}$	$PM_{2.5}$	$PM_{2.5}$	$PM_{2.5}$
Feature	v10	v10	d2m	v10	d2m	co
	co	pm2p5	tp	pm2p5	co	pm2p5

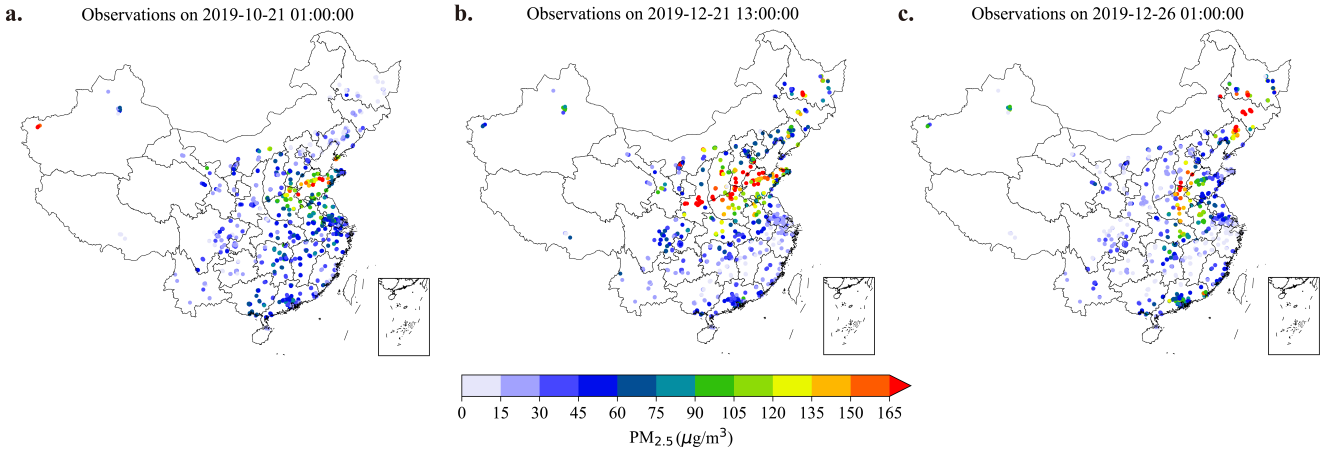

Figure S1. Snapshots of the $PM_{2.5}$ observations at 1074 air quality stations. Panels a-c are in parallel with the RFSML predictions shown in Fig.7 panels a-c.

$$\text{NMB} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{y}_t - y_t}{y_t}\cdot 100\% \tag{S1}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_t - y_t)^2} \tag{S2}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_t - y_t| \tag{S3}$$

$$\text{R} = \frac{E[(\hat{y}_t - E(\hat{y}_t))(y_t - E(y_t))]}{\sqrt{E[\hat{y}_t - E(\hat{y}_t)]^2}\cdot\sqrt{E[y_t - E(y_t)]^2}} \tag{S4}$$

where n is the size of the test dataset; $\hat{y}_t$ and y_t are the predicting values and ground observations of $PM_{2.5}$ concentration at time t.

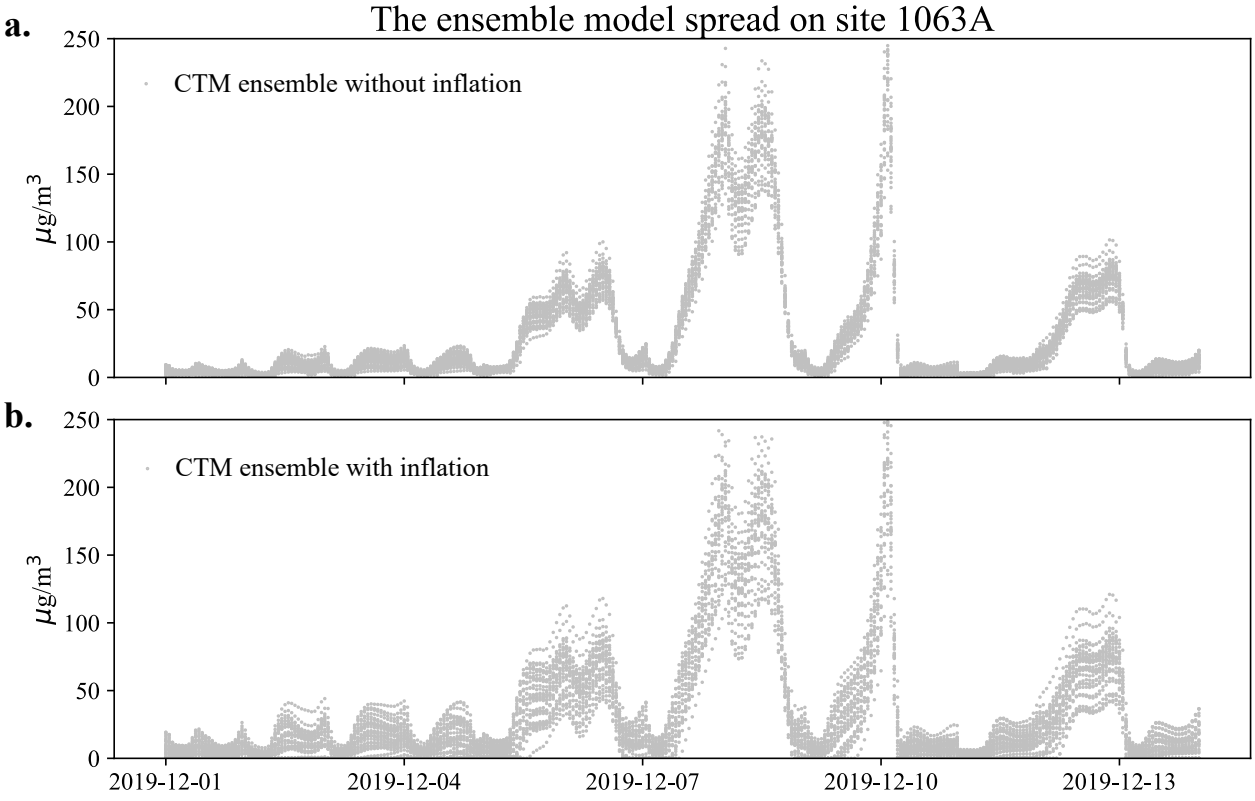

Figure S2. Panels a and b are the ensemble CTM model spread before and after inflation on an environmental monitoring station (Latitude:40.98°N, Longitude:117.95°E) in Chengde, Hebei Province.

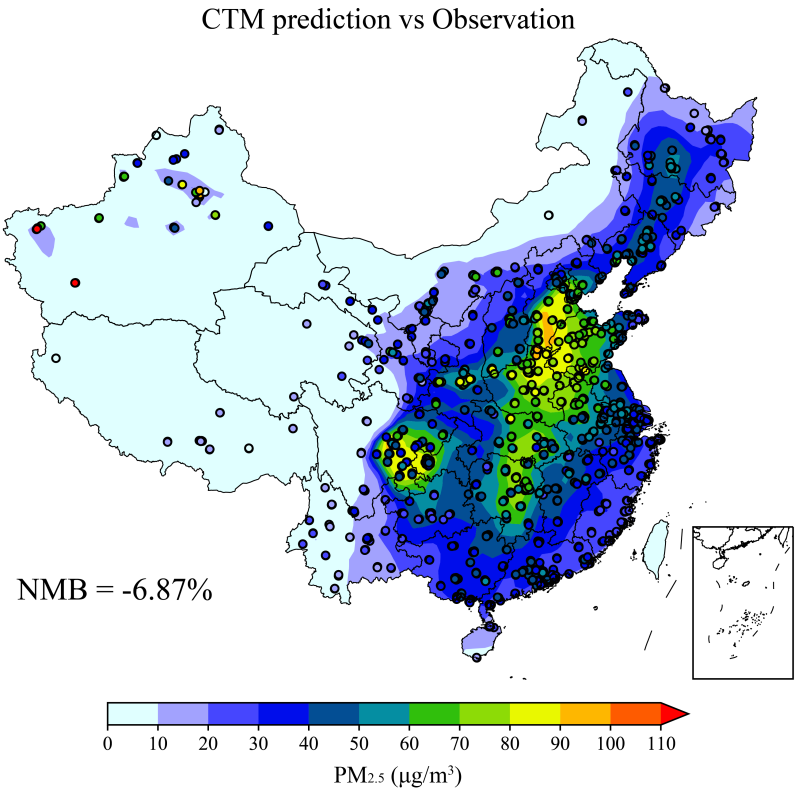

Figure S3. The average time results of CTM forecast versus Observations over the entire test period. Normalized mean bias (NMB) between the CTM forecast and observation is -6.87%.

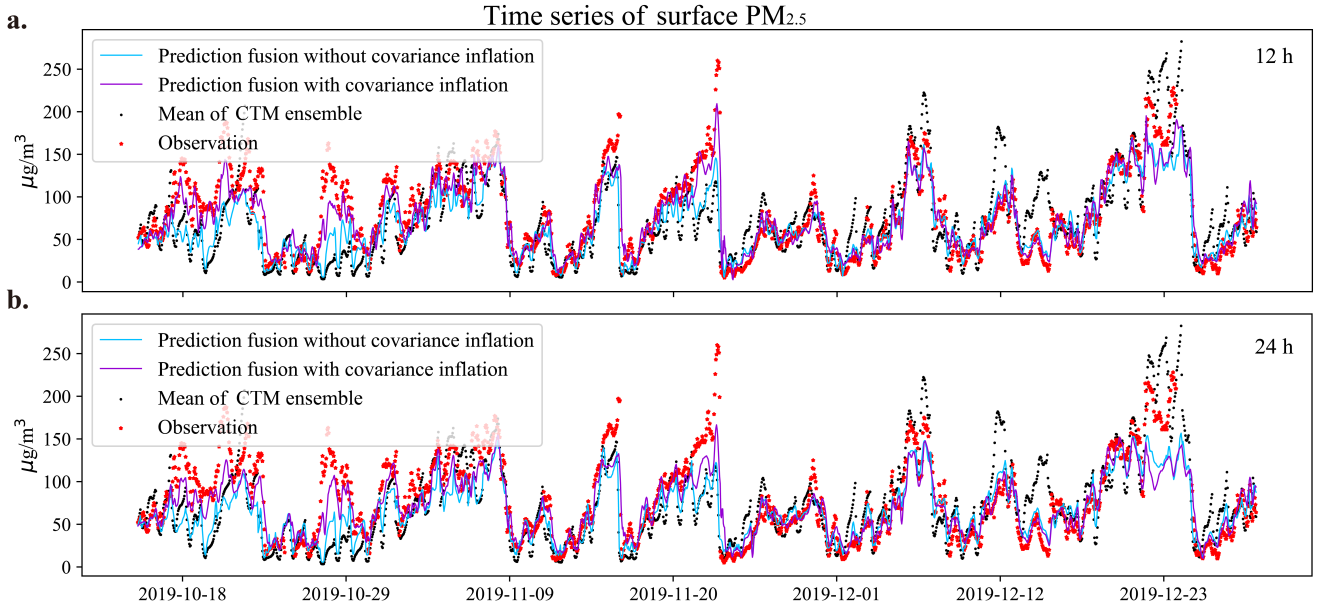

Figure S4. Time series of an environmental monitoring station (Latitude:34.65°N, Longitude:112.39°E) in Luoyang, Henan Province, which is one of the validation set. Markers of deep sky blue solid line, dark violet solid line, red star, and black dot represent prediction fusion without covariance inflation, prediction fusion with covariance inflation, ground observation, and mean of ensemble CTM predictions respectively. Subplots a and b represent forecasts 12 and 24 h ahead, respectively.

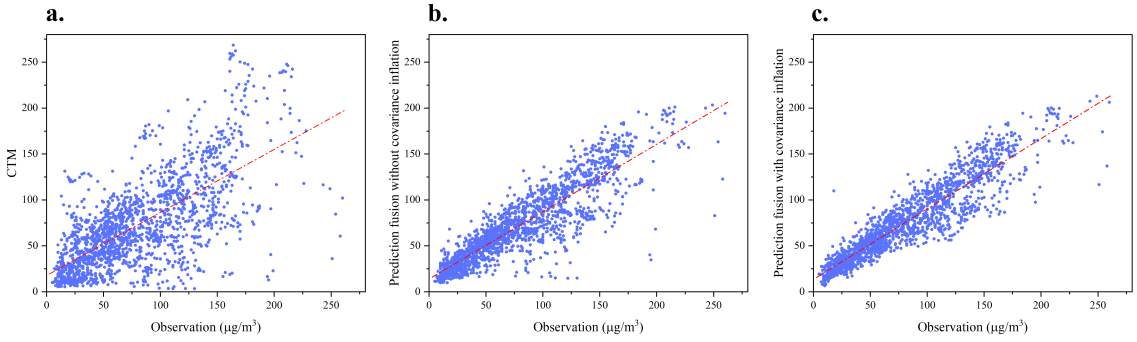

Figure S5. The distributions of model predictions and observations for a 6-hour predicting horizon. Panels (a), (b), and (c) illustrate the distributions between CTM, prediction fusion without covariance inflation, and prediction fusion with covariance inflation, and their corresponding ground observations, respectively.

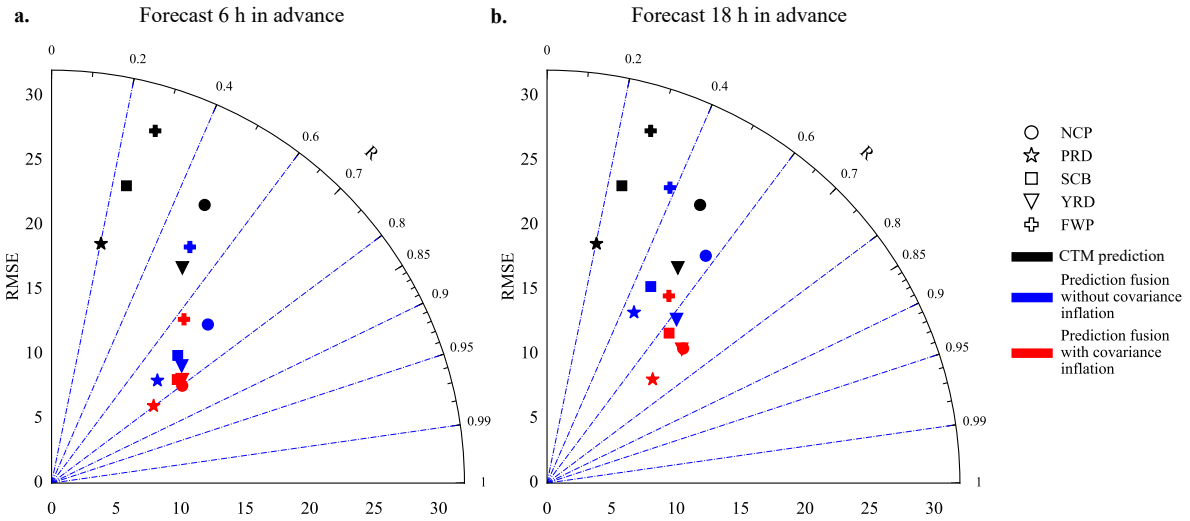

Figure S6. A modified Taylor Diagram that illustrates RMSE and R together. The regions of interest, including NCP, PRD, SCB, YRD, and FWP, are differentiated by unique markers of circle, star, square, triangle down, plus, and diamond, respectively. Additionally, the results from different approaches are visualized by distinct colors, with black, blue, and red indicating the results from CTM, prediction fusion without covariance inflation, and prediction fusion with covariance inflation, respectively. The diagram consists of two panels, representing forecasts for 6-hour and 18-hour time horizons, arranged in a left-to-right sequence.