Supplementary Material

Improving Ground-Level NO₂ Estimation in China Using GEMS Measurements and a Nested Machine Learning Model

Naveed Ahmad[1], Changqing Lin[1,*], Alexis K.H. Lau,[1,2] Jhoon Kim[3], Fangqun Yu[4], Chengcai Li[5], Ying Li[6], Jimmy C.H. Fung[1,7], Xiang Qian Lao[8]

[1]Division of Environment and Sustainability, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China

[2]Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China

[3]Department of Atmospheric Sciences, Yonsei University, Seoul, 03722, Korea

[4]Atmospheric Sciences Research Center, State University of New York at Albany, Albany, NY 12226, US

[5]Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing 100871, China

[6]Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

[7]Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China

[8]Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China

Correspondence to: Changqing Lin (cqlin@ust.hk)

It includes Text S1 and Figs. S1-S6

Text S1: Hourly, seasonal, and annual correction factors

There was some missing data for satellite NO_2 VCDs measurements due to cloudy conditions between 08:00 AM and 03:00 PM for 2021. Therefore, we applied the correction factors, which represent the ratio between the average NO_2 from all ground measurements and the average ground-level NO_2 when satellite data was available (Eq. S1). These correction factors were used to obtain a bias-corrected estimation of satellite-derived ground-level NO_2 concentrations for each hour from 08:00 AM to 03:00 PM.

$$C_{f_k} = \frac{\frac{1}{m}\sum_{i=1}^{m} C_{g_{NO_{2_i}}}}{\frac{1}{n}\sum_{i=1}^{n} C_{g_{NO_{2_i}}}} \tag{S1}$$

Here, C_f represents the correction factor for hour k (each hour from 08:00 AM to 03:00 PM), C_g presents ground-measured NO_2 concentrations, m shows all ground observations of NO_2, and n corresponds to ground observations of NO_2 at hours when the satellite data were available. The station-based spatial distributions of correction factors for each hour from 08:00 AM to 03:00 PM are revealed in Fig. S4. As the predicted NO_2 concentrations in the study region were on a regular grid of $0.2° \times 0.4°$, the bilinear interpolation was applied to map the correction factors on the same regular grid of $0.2° \times 0.4°$ (Fig. S5). The bias-corrected ground-level NO_2 concentrations for each hour from 08:00 AM to 03:00 PM were then estimated using Eq. S2.

$$C_{NO_2} = C_{p_{NO_2}} \times C_f \tag{S2}$$

where, C_{NO2} represents the corrected satellite-estimated ground-level NO_2 concentrations, C_{PNO2} are initial predicted NO_2 concentrations, and C_f is correction factor for each hour from 08:00 AM to 03:00 PM.

Further, as the satellite data was available only during the daytime from 08:00 AM to 03:00 PM, and besides missing data because of clouds, there was also missing satellite data for nighttime and other hours of the day beyond 08:00 AM and 03:00 PM. Therefore, for seasonal correction factors, we used the seasonal average of all available ground-measured NO_2 concentrations for 24 hours and the seasonal average of ground-measured NO_2 when the satellite data was available. Subsequently, Eq. S1 was used to calculate the correction factors for each season separately (i.e., spring, summer, fall, and winter). Similarly, to obtain the annual correction factor, we estimated the annual average of all available ground-measured NO_2 concentrations for 24 hours and the annual average of ground-measured NO_2 when the satellite data was available. The spatial distributions of station-based and interpolated seasonal and annual correction factors are shown in Figs. S2 and S3. Then, Eq. S2 was used for the bias correction of seasonal and annual ground-level NO_2 concentrations.
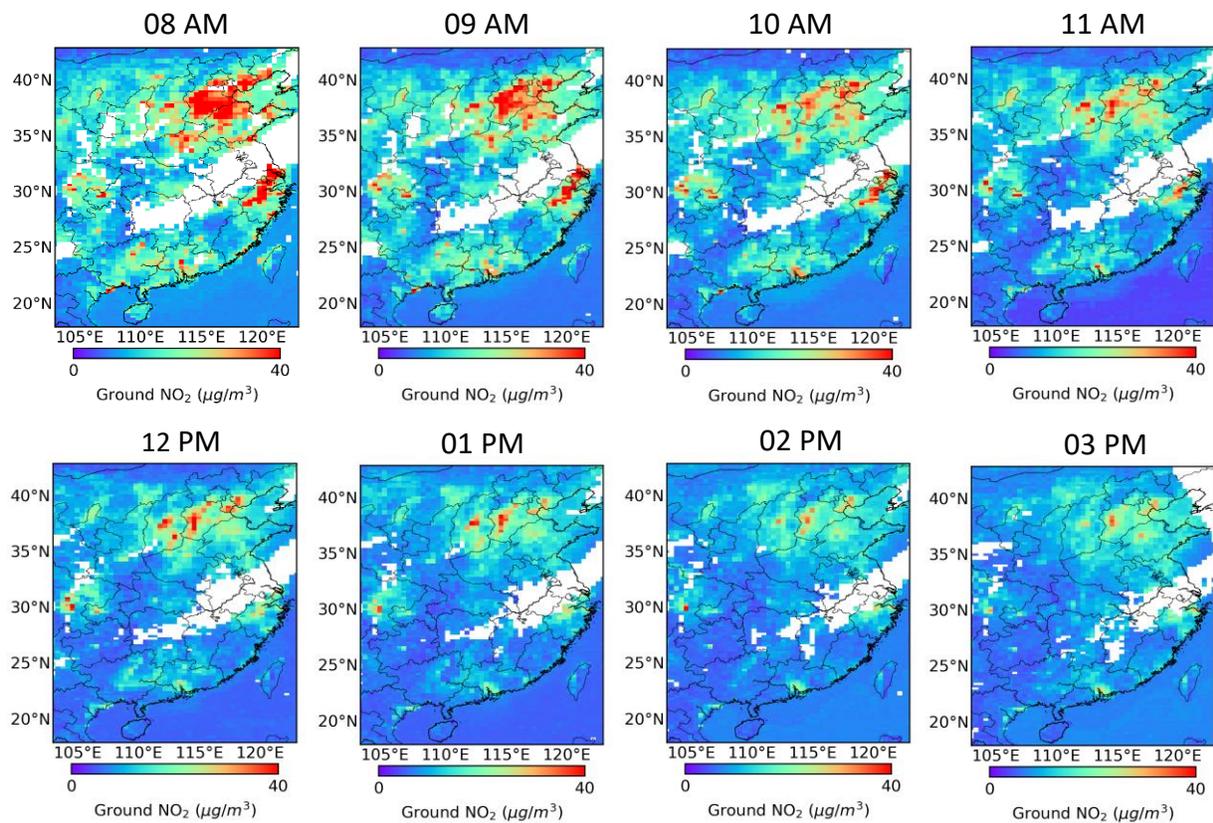
Figure S1: Spatial distributions of satellite-derived hourly ground-level NO₂ concentrations on 29th September 2021 for each hour from 08:00 AM to 03:00 PM (UTC+8).
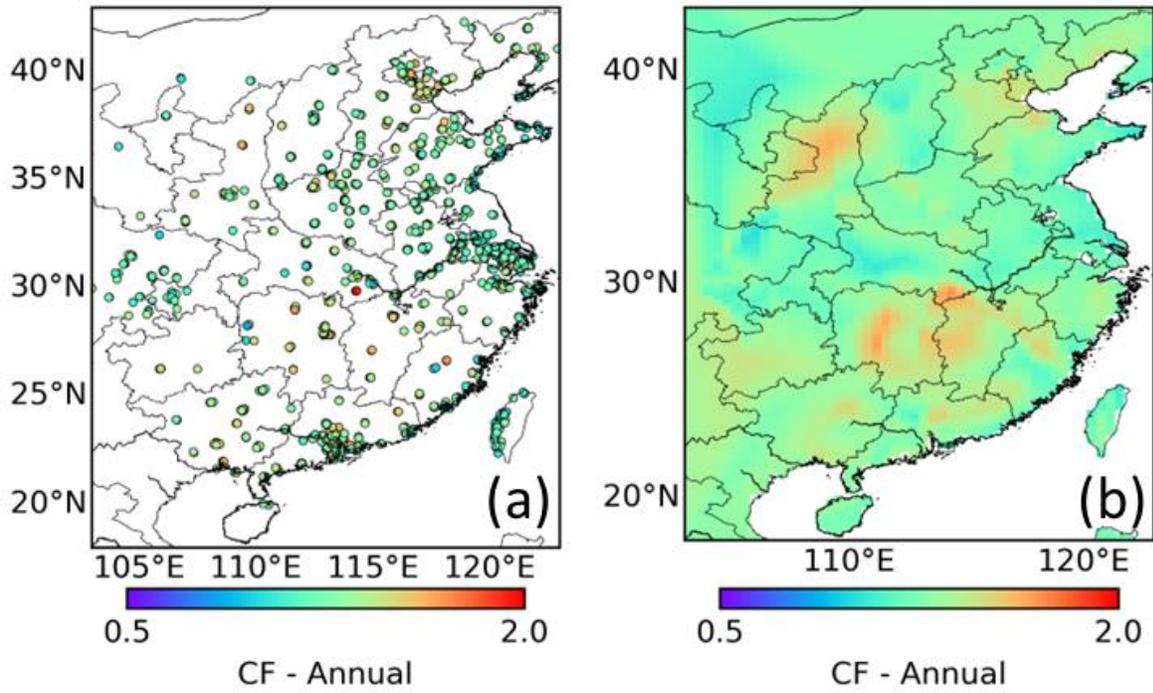
Figure S2: Spatial distribution of (a) station-based and (b) interpolated correction factors for estimating annual mean NO₂ concentration for 2021.
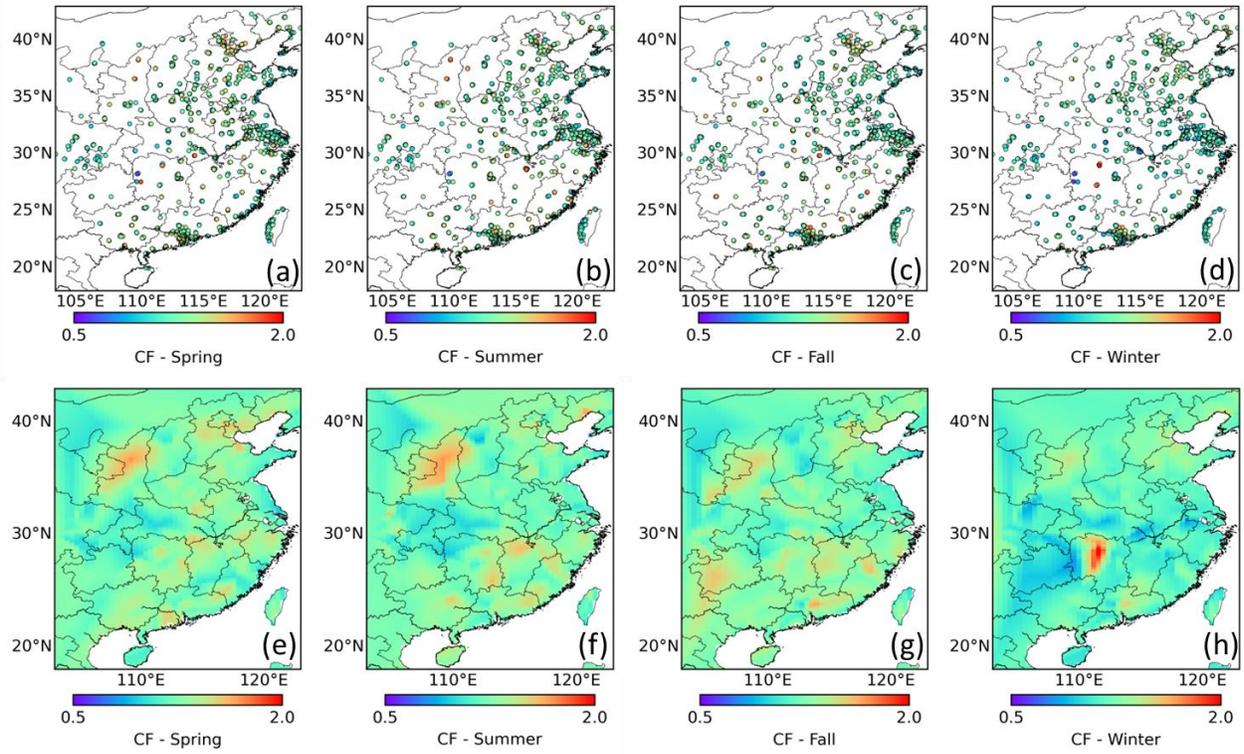
Figure S3: Spatial distributions of (a-d) station-based and (e-h) interpolated correction factors for estimating seasonal mean of NO₂ concentration for 2021.
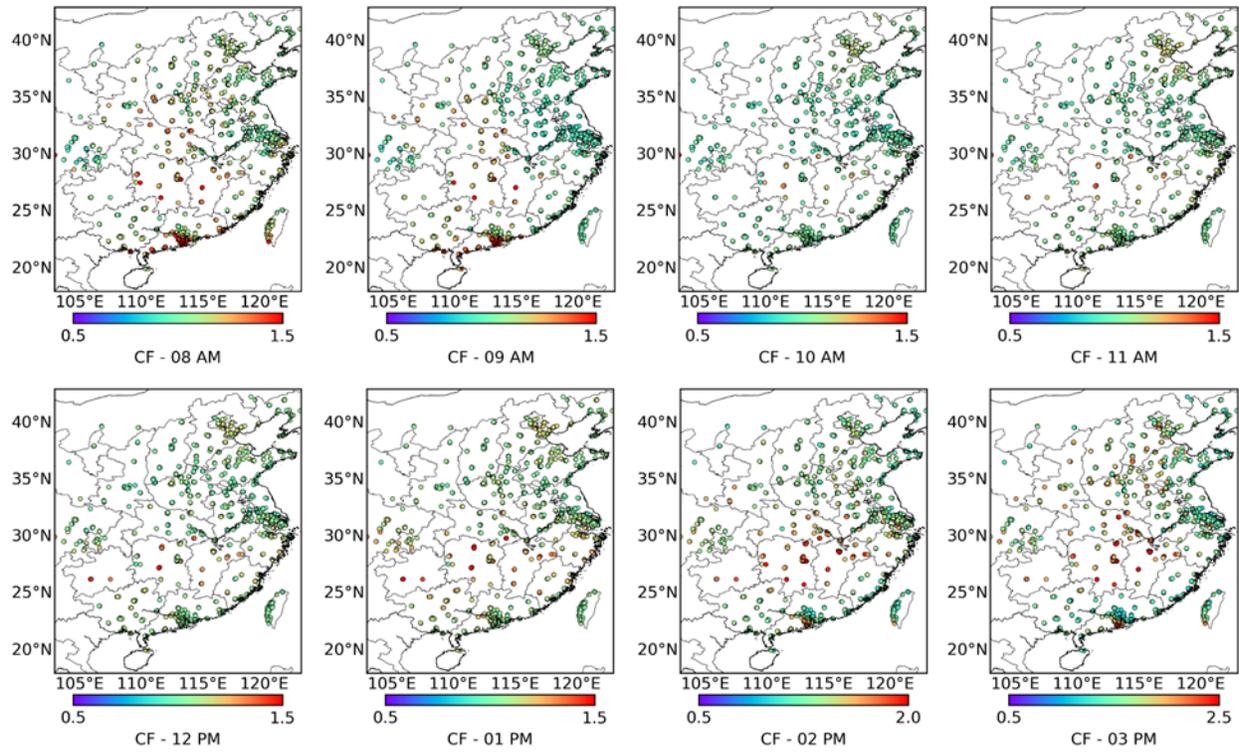
Figure S4: Spatial distributions of station-based correction factor between 08:00 AM and 03:00 PM (UTC+8) for 2021.
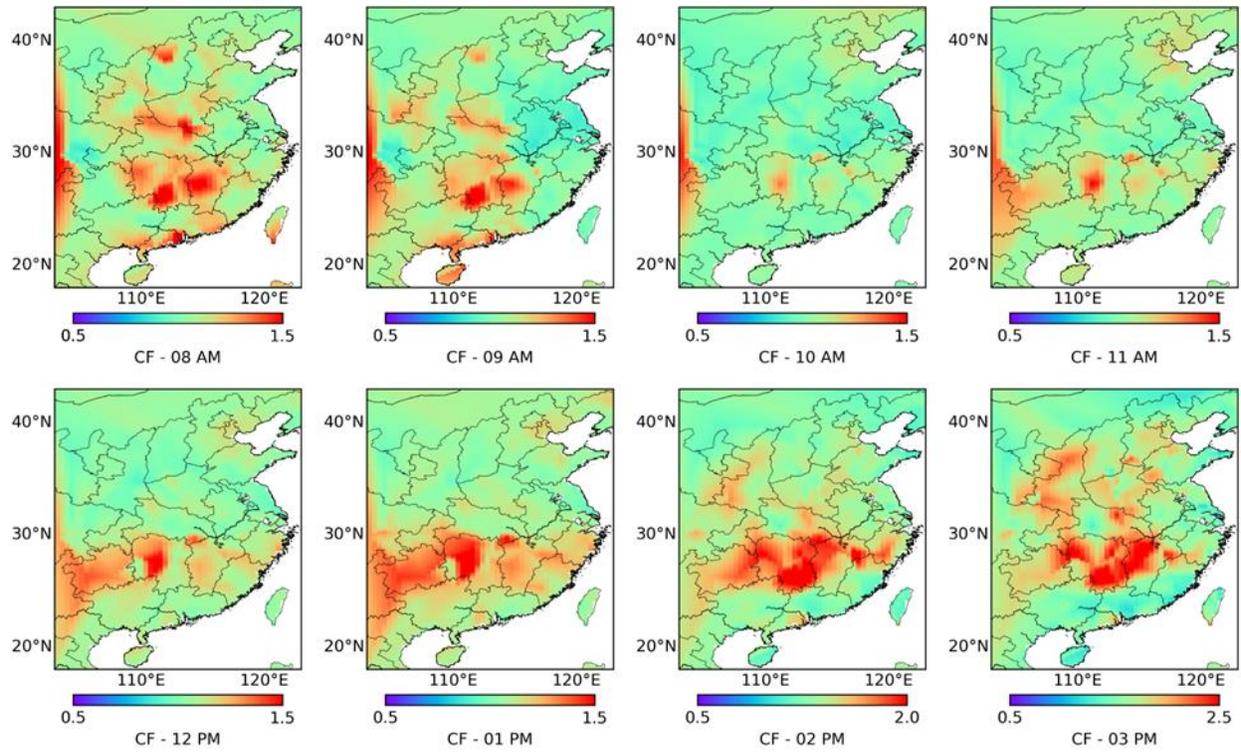
Figure S5: Spatial distributions of interpolated correction factor from individual hours between 08:00 AM and 03:00 PM (UTC+8) for 2021.
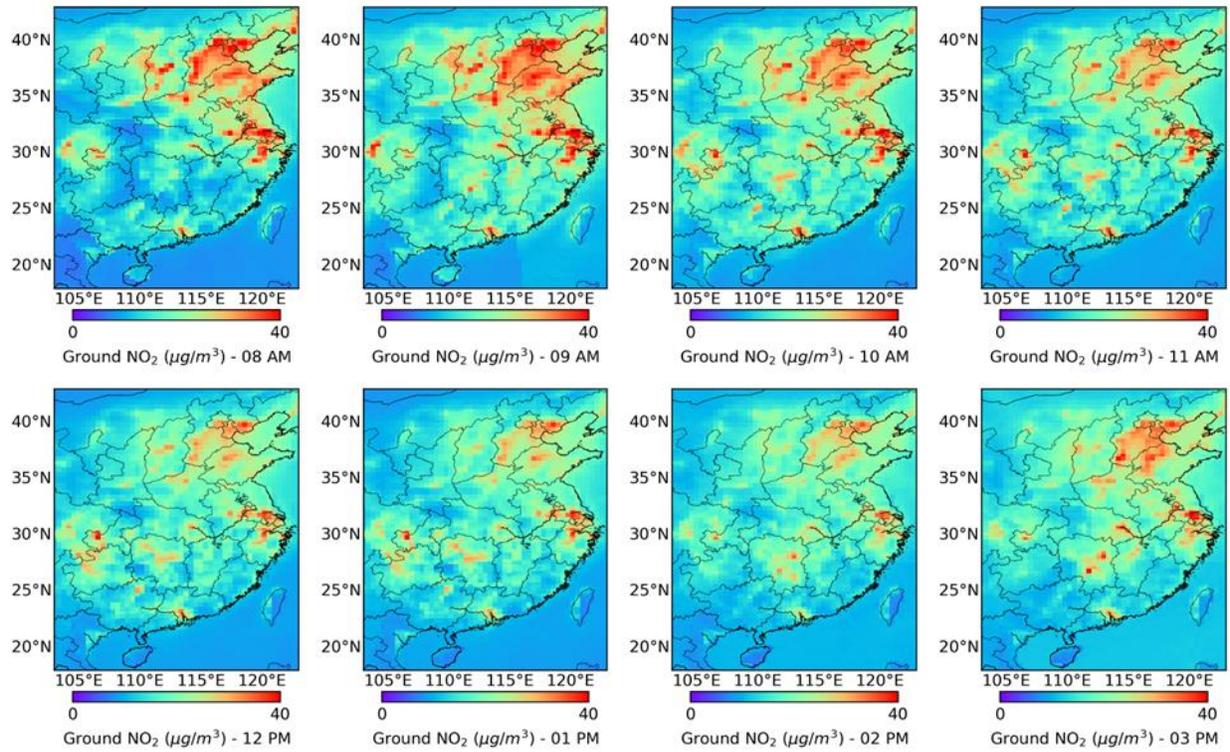
Figure S6: Spatial distributions of average of the ground-level NO₂ concentrations for each hour between 08:00 AM and 03:00 PM (UTC+8) for 2021.