ARTICLE

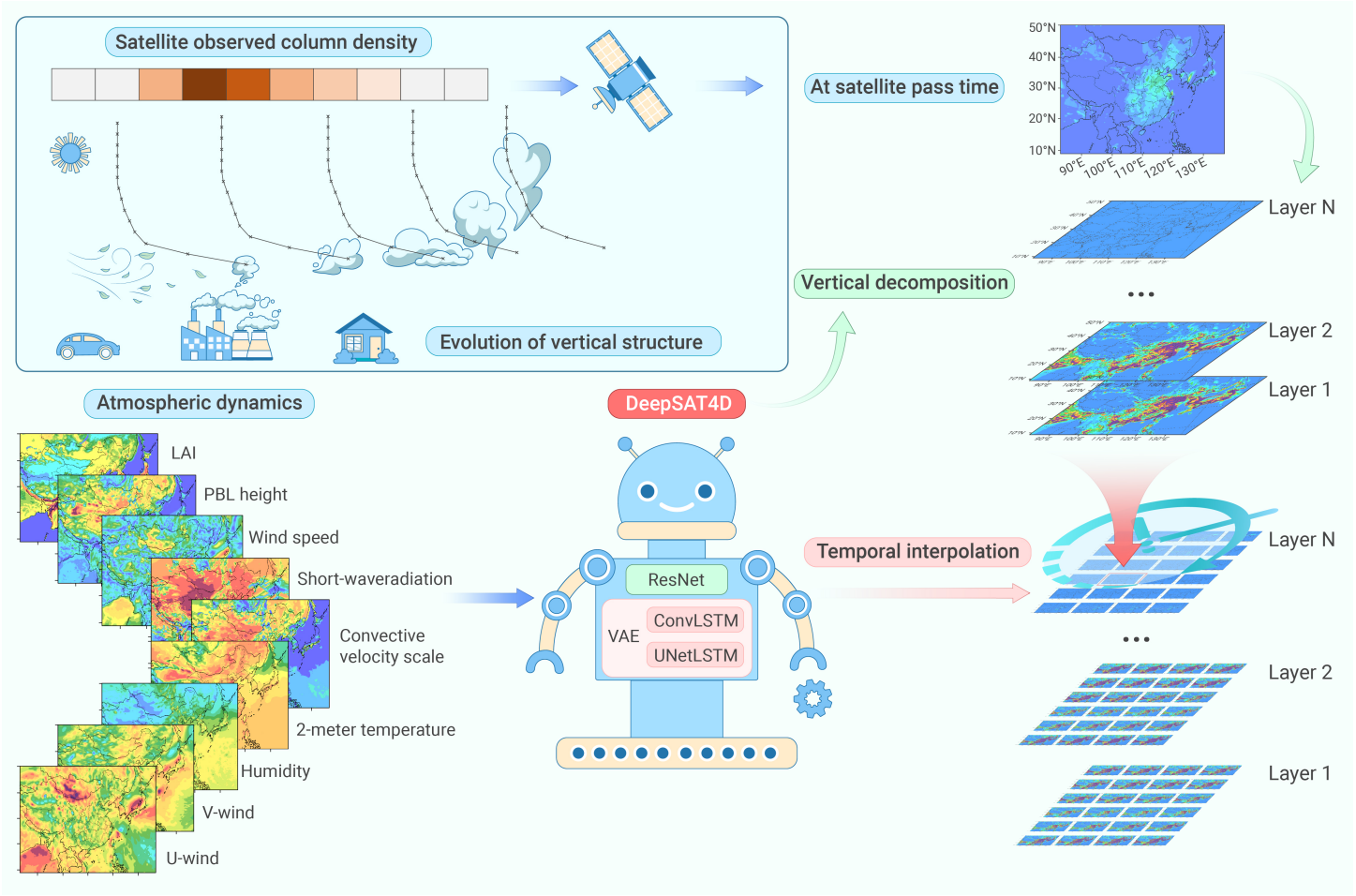
DeepSAT4D: Deep learning empowers four-dimensional atmospheric chemical concentration and emission retrieval from satellite

Siwei Li[1,3,4,*] **and Jia Xing**[2,*]

*Correspondence: siwei.li@whu.edu.cn (S.L.); jxing3@utk.edu (J.X.)

Received: November 29, 2023; Accepted: March 12, 2024; Published Online: March 18, 2024; https://doi.org/10.59717/j.xinn-geo.2024.100061

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- A deep learning model (DeepSAT4D) is developed to retrieve 4D chemical concentrations from satellite.
- The DeepSAT4D can regenerate dynamic evolution of vertical structure of atmospheric chemicals.
- The DeepSAT4D was applied to retrieve 2017-2021 4D NO_2 concentrations and NOx emissions in China.

ARTICLE

www.the-innovation.org/geoscience

DeepSAT4D: Deep learning empowers four-dimensional atmospheric chemical concentration and emission retrieval from satellite

Siwei Li[1,3,4,*] **and Jia Xing**[2,*]

[1]Hubei Key Laboratory of Quantitative Remote Sensing of Land and Atmosphere, School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430000, China

[2]Department of Civil and Environmental Engineering, the University of Tennessee, Knoxville 37996, USA

[3]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

[4]Hubei Luojia Laboratory, Wuhan University, Wuhan 430079, China

*Correspondence: siwei.li@whu.edu.cn (S.L.); jxing3@utk.edu (J.X.)

Received: November 29, 2023; Accepted: March 12, 2024; Published Online: March 18, 2024; https://doi.org/10.59717/j.xinn-geo.2024.100061

Citation: Li S. and Xing J. (2024). DeepSAT4D: Deep learning empowers four-dimensional atmospheric chemical concentration and emission retrieval from satellite. The Innovation Geoscience **2(1)**: 100061.

Accurate measurement of atmospheric chemicals is essential for understanding their impact on human health, climate, and ecosystems. Satellites provide a unique advantage by capturing data across the entire atmosphere, but their measurements often lack vertical details. Here, we introduce DeepSAT4D, an innovative method that efficiently reconstructs 4D chemical concentrations from satellite data. It achieves this by regenerating the dynamic evolution of vertical structure, intricately linked to complex atmospheric processes such as plume rise and transport, using advanced deep learning techniques. Its application with the Ozone Monitoring Instrument - Nitrogen Dioxide, a commonly used satellite product, demonstrates good agreement with ground-based monitoring sites in China from 2017 to 2021. Additionally, DeepSAT4D successfully captures emission reductions during 2020-pandemic shutdown. These findings emphasize DeepSAT4D's potential to enhance our understanding of the complete atmospheric chemical composition and to provide improved assessments of its impact on human health and Earth's ecosystem in the future.

INTRODUCTION

Accurate knowledge of atmospheric chemicals across both space and time is crucial for a deeper understanding of their effects on human health, the environment, climate, and ecosystems.[1-3] Given the limited scope of ground-based measurements, satellites emerge as the exclusive means to continuously collect direct measurements of atmospheric chemicals with global spatial coverage. Noteworthy satellites such as the Global Ozone Monitoring Experiment (GOME),[4] Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY),[5] and Ozone Monitoring Instrument (OMI)[6] are deployed for measuring atmospheric gaseous species. These instruments mainly provide the total column density aggregated across all vertical layers and are limited to a single measurement per day at certain satellite pass times. Although, some satellite instruments have limb-viewing as the Microwave Limb Sounder (MLS)[7] which could provide vertical information, their capabilities are mostly limited to measuring species in the upper troposphere and the lower stratosphere with relative lower spatial and temporal resolution. This limitation significantly constrains our ability to comprehensively assess atmospheric chemical composition, especially across various vertical and temporal dimensions.

Previous studies have endeavored to improve the accuracy of ground-level concentration estimations, recognizing the significance of ground-level concentrations in relation to human exposure.[8-9] Nevertheless, persistent challenges remain in the accurate estimation of ground concentrations from satellite retrievals. These challenges stem from difficulties in assigning column density to vertical layer-specific concentrations, whether through the application of statistic regressions,[10-11] machine learning techniques[12-20] or numerical model simulations.[21-23] The primary obstacle lies in the lack of accurate representation of vertical profiles, which exhibit strong spatial and temporal variations.[24] For instance, chemicals emitted from sources disperse into upper atmospheric layers, creating significant variations in vertical profiles in downwind areas.[25] Since most surface monitoring observations, often used as ground truth data, are situated in urban areas (such as China National Environmental Monitoring Center noted as CNEMC),[26] lacking information about rural regions. Consequently, many existing machine-learning models that heavily rely on ground-based observations may incorrectly assume vertical profiles in downwind areas, potentially leading to overestimations in rural locations. Numerical models are capable of explicitly representing the intricate processes of plume rise and dispersion from the emission sources following physical laws.[27] Some studies have introduced a column-to-surface ratio informed by numerical chemical transport models (CTM) to estimate satellite measurements at ground level,[21-23] or employ the simulated results of CTMs as a feature in their machine learning models.[14] However, this CTM simulation varies both in time and space and is subject to uncertainties stemming from various inputs and parameterization.[28-29] Additionally, the substantial computational demands of CTMs limit their application in near real-time scenarios. Effectively leveraging the strengths of CTMs in satellite retrievals and accurately interpolating satellite-derived column data into vertically resolved concentrations is an ongoing challenge.[30]

In principle, the satellite-measured column density represents a condensed value, derived from contributions at each vertical layer. Deciphering this column value into layer-specific concentrations is challenging, as it is an ill-posed problem when relying solely on detailed information about vertical structure. Previous studies attempted to address this challenge by incorporating multiple local features (target itself) into this decoding process. However, few of them have effectively leveraged information from surrounding grid cells and time steps (neighborhood data), likely because it can be too complex to integrate such neighborhood information. Machine learning, particularly deep-learning structures based on convolutional neural networks (CNNs), presents a promising solution,[13] as it can readily incorporate neighborhood data into the prediction process. Our previous studies have demonstrated its capacity in revealing the complicated atmospheric processes.[31-34]

To illustrate this approach, we present a novel method called DeepSAT4D, which using Deep-learning method to retrieve Satellite column measurement into Four-Dimensional chemical concentrations. More specifically, DeepSAT4D utilizes deep-learning models to regenerate the dynamic evolution of vertical structure stemming from complex atmospheric processes including plume rise and flow transport, thus to establish spatial-temporal correlations for both column and vertical layer-specific concentrations. We enriched the training data with multiple CTM model simulation under various of conditions to better represent the spatial and diurnal meteorological patterns, as well as the spatial distribution of column density. To customize the model for our specific task, we meticulously designed the model structure and framework, leveraging domain knowledge, as illustrated in Table S1.

The DeepSAT4D framework comprises three key modules:

ResNet: This module is used for vertical decomposition, transforming column density into a detailed vertical concentration structure. The ResNet model's architecture allows it to extract information from neighboring grid cells,[35] particularly for atmospheric system,[33] aiding in the determination of specific vertical structures influenced by complex atmospheric processes. The loss function is also carefully designed to enforce the mass conservation constraint by incorporating the predicted column density from the vertical-specific concentration.

ConvLSTM: Serving as the decoder for the Variational Autoencoder (VAE)[36] for temporal interpolation. This module structure[37] offers real-time predic-

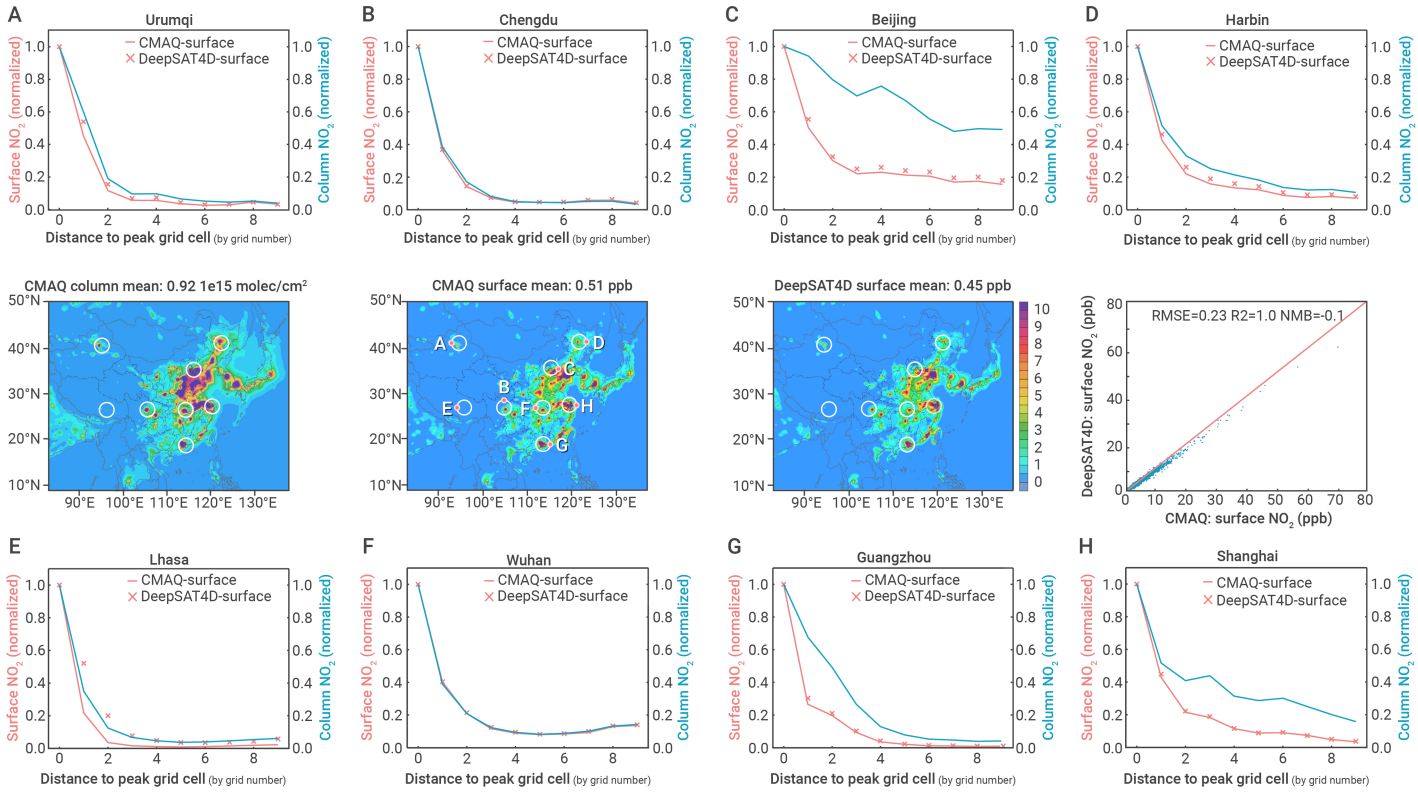


Figure 1. Comparison of NO_2 column density (CMAQ simulation, unit: 1×10 molec/cm^2) and surface NO_2 concentration (both simulated in CMAQ, and reproduced by DeepSAT4D, unit: ppb) across simulation domain (at 14:00 local time in 2017 baseline case) and in eight cities (selected the grid with highest surface concentration as the peak grid cell, and calculated the average concentrations in surrounding grid cells away from it by the distance up to 10 grid cells).

tions of time-series concentration in response to updated emissions. In adherence to physical laws, we feed the initial condition into the model twice to account for its role in atmospheric physical and chemical processes and its function as a modulatable baseline to facilitate the training process. Furthermore, we've designed a dual-model that incorporates both forward and backward models concurrently, considering satellite measurements at 24-hour intervals, to constrain the concentration and mitigate the issue of error accumulation.

UNet-LSTM: As the encoder for the VAE model,[32] this module estimates emission adjustments to align with the observed concentration. To mitigate potential ill-posed problems, we normalize emissions variability to a daily average level. Additionally, we train the model directly with the CTM data, eliminating the need for a priori emissions from bottom-up investigations, which are often impeded by data delays.

In this study, we employ the OMI satellite to retrieve nitrogen dioxide (NO_2) in China as an example, but it's worth noting that DeepSAT4D can be extended to other species and other regions with a similar model structure and corresponding training dataset. We initially trained using simulation data from the atmospheric chemistry transport numerical model (a commonly used physical model, WRF/CMAQ[38-39]) conducted over China domain with 27km by 27km spatial resolution (Figure S1). The trained models are then applied to OMI satellite products for the years 2017-2021. Ground measurement data (CNEMC) from the same period is used for comparison and evaluation of the surface concentration estimated from column density using DeepSAT4D. Further details about the machine learning modules and the numerical model used in this study can be found in the Method section.

As machine-learning based model, the DeepSAT4D model offers the advantage of real-time conversion from satellite column density to four-dimensional chemical concentration, eliminating the need for additional CTM model simulations. This approach significantly reduces the computational burden, essentially providing a CTM-free method. As illustrated in this study, we exclusively utilized simulation data for the baseline year of 2017 (simulation with 2017 baseline emissions and meteorology), along with two emission change scenarios and two future climate change scenarios in 2050 for training purposes (refer to Table S2). Subsequently, we applied this model to the years 2018-2021 without the need for any CTM simulations, as presented in the following Results section.

MATERIALS AND METHODS

Numerical model WRF/CMAQ

In this study, we employed the Community Multiscale Air Quality (CMAQ) model, version 5.2, utilizing the Carbon Bond 6 mechanism for gas-phase chemistry and the AERO6 mechanism for particulate matter chemistry. CMAQ, as one of the most widely used Chemical Transport Models (CTMs), has a well-established reputation for accurately simulating air pollutant concentrations.[38] Meteorological data were generated using the Weather Research and Forecasting (WRF) model,[39] version 3.8, configured identically to our previous studies.[58-59] Emission data were sourced from the ABaCAS-EI inventory, a high-resolution dataset developed by Tsinghua University with a spatial resolution as fine as 1km-by-1km and a temporal resolution of 1 hour.[60] Biogenic emissions were estimated using the Model for Emissions of Gases and Aerosols from Nature (MEGAN).[49] It is noteworthy that we have assessed the performance of WRF and CMAQ in simulating meteorological variables and air pollutant concentrations through extensive comparisons with observational data in our previous studies.[58-59]

The horizonal domain covers most of East Asia and is represented by grids of 182 rows and 232 columns with a horizontal resolution of 27 km by 27 km. The whole troposphere (from ground to 100mb) is defined by using 14 layers with sigma values as follows: 1.00, 0.995, 0.99, 0.98, 0.96, 0.94, 0.91, 0.86, 0.8, 0.74, 0.65, 0.55, 0.4, 0.2, and 0.00, corresponding to 19, 57, 114, 230, 386, 584, 910, 1375, 1908, 2618, 3598, 5061, 7620, 11944 meters above the ground at domain and annual averaged level.

In addition to national averages, our analysis also focuses on eight cities, namely Beijing, Shanghai, Guangzhou, Wuhan, Chengdu, Harbin, Lhasa, and Urumqi, as well as eight regions: North China Plain, Yangtze River Delta, Chuan-Yu, Liang-Hu, South China, Northeast China, Fen-Wei, and West China. The geographical distribution of the simulation domain, cities, and regions is provided in Figure S1.

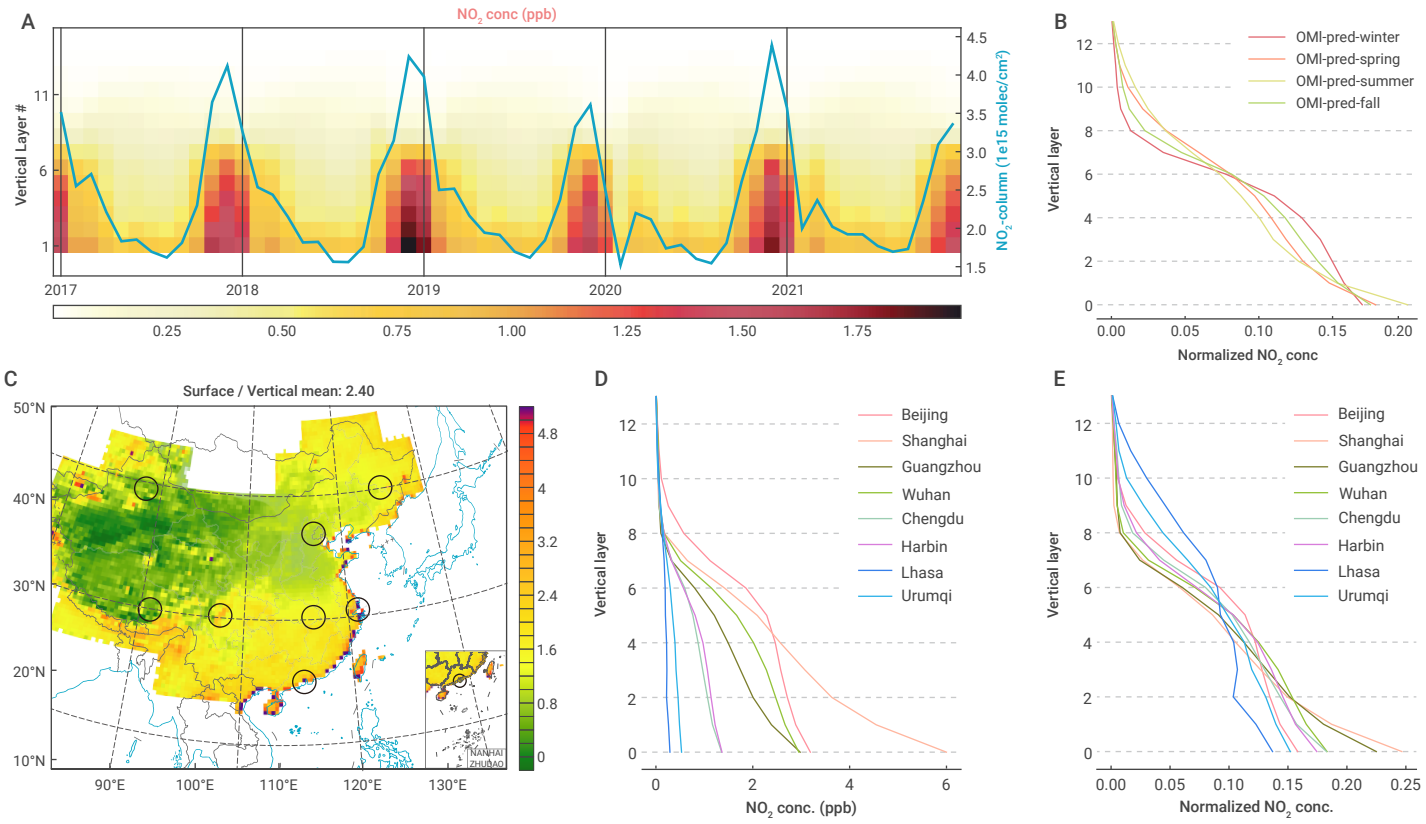


Figure 2. OMI satellite retrieved 14:00 NO_2 vertical structure profile time series during 2017-2021 (A), and seasonal 5-year averages (B), the ratio of surface to vertical mean concentration (C) and the vertical profiles in eight cities (averaging 10 grid cells surrounding the urban center) with absolute NO_2 concentration (unit: ppb) (D) and normalized values (sum =1) (E) (the top height of each vertical layer is at the following altitudes: 19, 57, 114, 230, 386, 584, 910, 1375, 1908, 2618, 3598, 5061, 7620, and11944meters. These values represent the annual mean domain average).

We conducted five simulation scenarios using the WRF/CMAQ models over a year (as summarized in Table S2). These scenarios including three meteorological conditions for the years of 2017, 2050-BCC-ssp126, and 2050-BCC-ssp585. Within each of these conditions, we considered different emission variations, including the 2017 baseline emissions, two random emission increases based on 2017 data, and emission levels corresponding to 2050-BCC-ssp126 and 2050-BCC-ssp585, as described in our previous paper.[61]

ResNet structure and training (from column density decompose to vertical layers)

The ResNet[35] model is trained to establish a connection between column density and vertical layer-specific concentrations under varying meteorological conditions, serving as an additional feature. The structure of the ResNet model is depicted in Figure S9.

Similar as our previous study,[33] we designed the ResNet model with 128 channels in the initial layer and incorporated 8 residue blocks. In addition to the column density at 14:00 local time, matching the OMI pass time, we included meteorological factors such as surface layer U- and V-direction winds (UV-wind), planetary boundary layer (PBL) height, 10-meter wind speed (WS), short-wave radiation (SWR), convective velocity scale (WSTAR), 2-meter temperature (T2), humidity (Q2), and Leaf Area Index (LAI) corresponding to the respective hour. To represent geographical information, we utilized a time-independent terrain height variable normalized to have a mean of 0 and a variance of 1. The prediction target includes NO_2 concentrations across all vertical layers. It's worth noting that further refinement of the vertical layer structure is possible, but it would result in a significant increase in computational resources, particularly in terms of the RAM required to store the training data.

For data augmentation, we introduced random cropping of the feature maps with dimensions of 60 rows by 60 columns. During training, we utilized the Mean Squared Error (MSE) loss function, running for a total of 2000 epochs. This number of epochs proved to be adequate for achieving good performance in both training and testing. Our learning rate commenced at 0.001 and was linearly decayed to zero at the end of the training process. We employed the Adam optimizer[62] to enhance model convergence.

The loss function consisted of two components: First, the loss of NO_2 concentration at each vertical layer was weighted by predefined ratios based on the average concentration level at each layer. This weighting was used to account for the larger scales of smaller concentrations at upper layers, making them comparable with surface layer concentrations. Second, the predicted concentration was used to estimate the column density, which was added as an additional component of the loss. This was done to optimize the model and ensure mass conservation across all vertical layers. It's important to note that this design also allows for retraining if column density calculations based on the satellite average-kernel[63] are considered, though it was not implemented in this study, as our OMI product lacked this information.

To provide better constraints on vertical properties, we included meteorological variables and column data for the current date as well as the previous and following dates. This allowed us to incorporate surrounding time-step information into the training process.

The training procedure involved ResNet (also applied to ConvLSTM) using data from the first 15 days of each month, totaling 720 days of records. Testing was conducted on the remaining days in each month across four different scenarios. Due to computational resource limitations, we adopted a 50-50 split between training and testing. It's important to highlight that we trained the model for all 24 hours to gain the benefits from the variability of features in the training data, even though the OMI data was applied specifically around local time 14:00. The model performance in training and test is very closed (RMSE<0.1 ppb, |NMB|<0.2, R^2=1), as presented in Figure S10.

ConvLSTM structure and training (from daily-averaged emission to concentration)

The ConvLSTM[37] model serves to estimate the concentration's temporal evolution by leveraging time-series meteorological factors, incorporating two

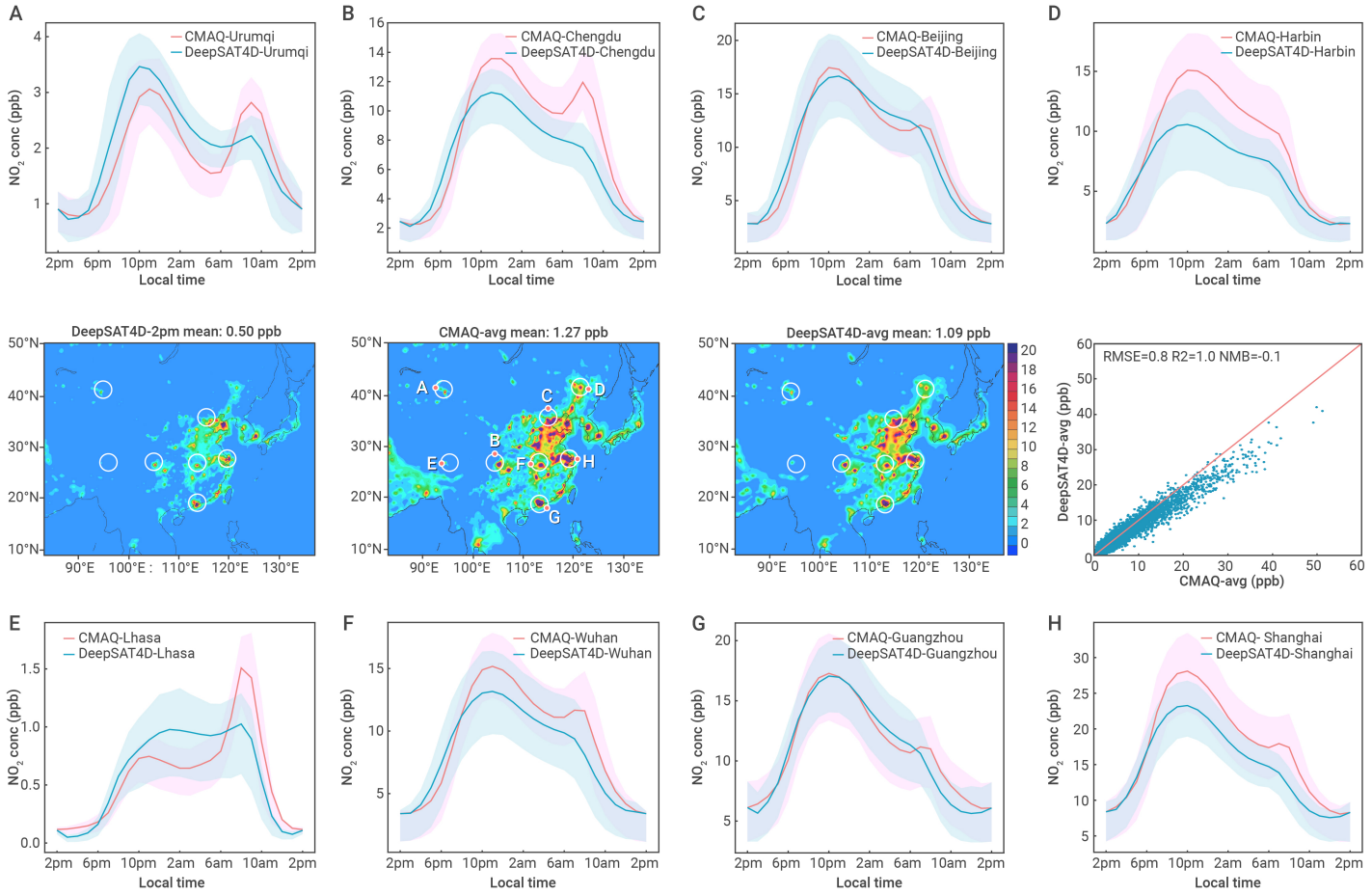


Figure 3. Comparison of 14:00 local time surface NO_2 and 24-hour averaged NO_2 across simulation domain and 24-hour diurnal variation in eight cities (simulated by CMAQ and reproduced by DeepSAT4D, annual mean in 2017 baseline case, unit: ppb).

concentrations with 24-hour step intervals. This data is interpolated into hourly values, providing a detailed temporal profile of concentration variations. The structural layout of the ConvLSTM model can be referenced in Figure S11.

We use time-series meteorological features to interpolate concentrations at specific hours into 24-hour variations. These features include the same set of meteorological factors observed over a 24-hour period, as well as time-independent normalized terrain height. Additionally, daily NOx emission data, utilized in CMAQ simulations, is integrated into the model. Notably, we refrain from using hourly NOx emission data to prevent ill-posed problems in subsequent tasks (i.e., avoid the possibility that multiple combinations of hourly NOx emissions can match to concentration estimates with satellite measurements at the satellite pass time). Instead, we opt to adjust emissions solely at the daily average level, effectively making simultaneous adjustments to emissions for each hour across a day.

Given that the measurement intervals occur every 24 hours, there are two temporal directions in which we predict time-series concentration. The first direction is forward, meaning we initiate the prediction at Hour 0 and forecast the subsequent hours (1, 2, ..., 24). While this approach is consistent with the time sequence in CMAQ, it's susceptible to error accumulation, especially as the prediction moves further away from the initialization. However, as the prediction hour approaches Hour 24, it draws nearer to the next initialization time, where constraints should be imposed.

To address this issue and make more accurate predictions, we have implemented an additional backward model. This model initiates the concentration at Hour 24 and proceeds to back-predict the reversed time-series concentrations for hours 23, 22, and so on until Hour 0. This dual-model design offers improved predictions for hours closer to Hour 24 but distant from Hour 0. By using both forward and backward models concurrently to constrain the concentration, we minimize the error accumulation problem. These models are trained together, with the loss function incorporating both models' outputs and comparing them with the CMAQ-simulated hourly NO_2 concentration as the ground truth.

The final prediction is the average of both models' predictions. In accordance with physical laws, the concentration previously predicted serves as the initial condition for predicting the next hour's concentration. This previous-hour concentration plays a dual role in the model: firstly, it participates in atmospheric physical and chemical processes and is combined with other factors such as meteorological variables within the ConvLSTM cell. Secondly, it acts as a baseline that can be modulated by perturbations. After passing through the ConvLSTM cell, the perturbation, resulting from the atmospheric processes, is added to the previous concentration status, creating the new concentration status. This design greatly facilitates the model's training process, as the prediction for the next hour's status closely resembles the previous hour's status.

The ConvLSTM model comprises four layers with varying channel numbers: 256, 128, 64, and 64, each equipped with a 3×3 kernel size. The hyperparameters (h, c) are automatically updated during the prediction, preserving historical information for long-term forecasting. The convolutional features facilitate the incorporation of neighboring grid information into the LSTM hyperparameters, recognizing the significant role of surrounding areas in atmospheric processes, in addition to local features.

Our model is trained to predict all 14 layers of NO_2 concentrations simultaneously, accounting for interactions among them (e.g., vertical diffusion). Similar to the ResNet, we trained the ConvLSTM using data from the first 15 days of each month, accumulating to a total of 720 days of records. Testing was performed on the remaining days of each month, across four different scenarios. The training initiates from the previous day's local time at 14:00,

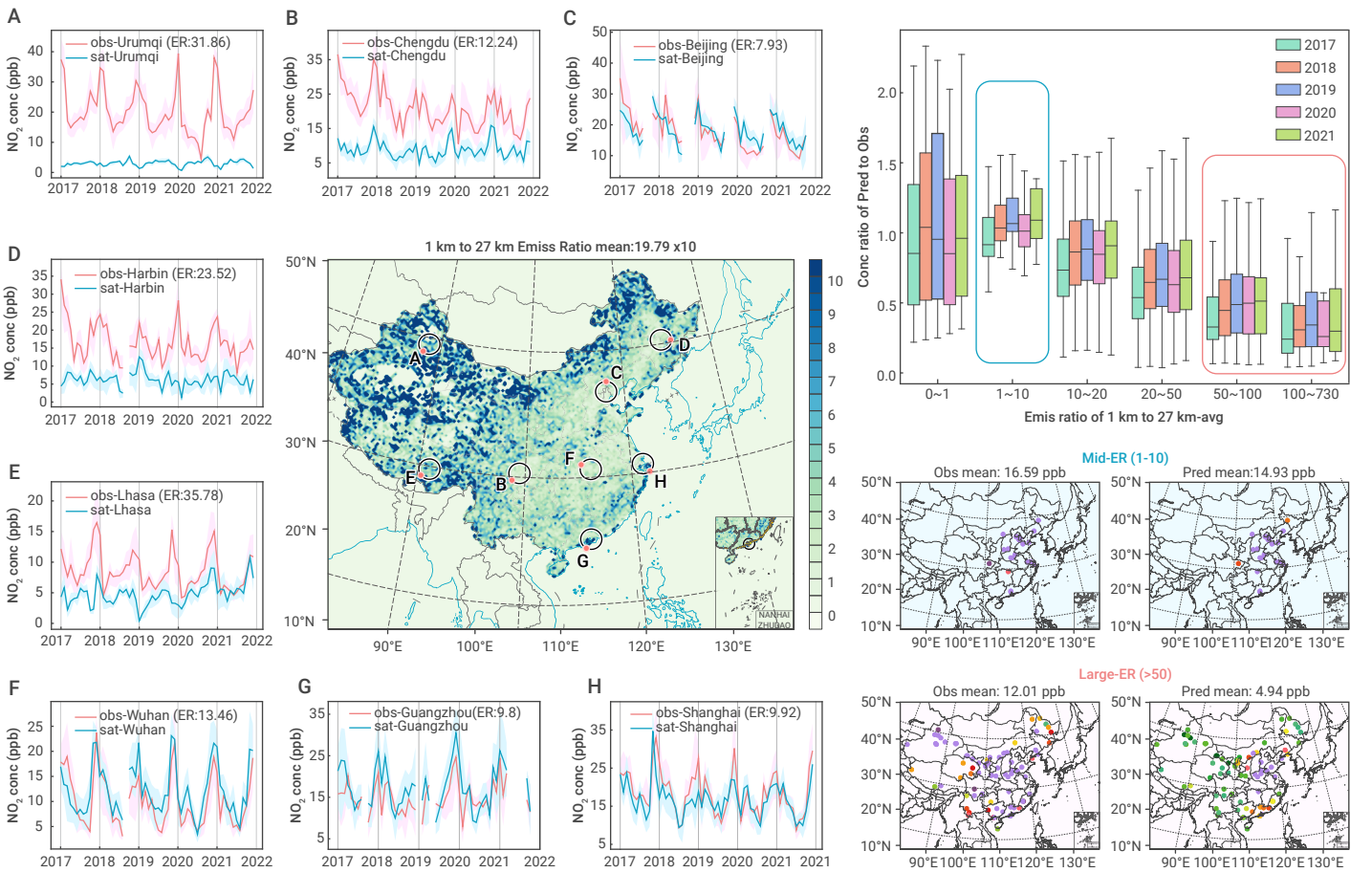


Figure 4. Trend of OMI-predict surface NO_2 concentration with DeepSAT4D during 2017-2021 comparing with ground observation in eight cities with the spatial distribution of 1km-to-27km emission ratio (ER) (left), and the comparison between OMI-prediction and observations by different group of ER values (right) (ER<1 indicate the sum of the emission in 27km grid cell are zero according to the 1km gridded emission file).

with a 24-hour time interval. The model's performance in testing achieves a score of RMSE<2.3 ppb, |NMB|<0.2, R^2=1, as depicted in Figure S12.

UNet-LSTM structure and training (from 2-hour concentrations to daily-averaged emission)

We employed the Variational Auto-Encoder (VAE) approach[35] to interpolate two specific hour concentrations measured by satellites into a complete 24-hour time series. Following the framework of our previous model,[31,34] the UNet-LSTM, we input all time-series features, including meteorological factors, into the LSTM layer to extract temporal features. These temporal features are then combined with time-independent features, such as terrain height, and the two-hour concentrations observed by satellites (at hour 0 and hour 24). This information is used to predict the 24-hour average emissions between these two time points. The model's structure can be found in Figure S13.

The model is trained as a generative machine-learning model, with emissions serving as the patent layer due to their substantial uncertainties, especially when compared to other factors and the absence of ground-truth data. In contrast to our previous study,[32] where we directly trained the VAE using satellite data, this study follows a different approach. We initially trained the VAE using CMAQ simulations and subsequently applied the trained VAE model with satellite products. This approach offers several advantages. Firstly, it enables us to leverage multiple CMAQ scenarios, combining different emission levels and meteorological conditions, as opposed to relying solely on the baseline data provided by satellites. Secondly, when applying the VAE with satellite data in subsequent years, it eliminates the need for a priori emissions, as all relevant conditions have been accounted for during the training process. The challenge of updating a priori emissions from bottom-up investigations is often hindered by data delays, leading to the use of outdated priori emission estimates that do not reflect current conditions. However, by training the VAE with CMAQ data, we can work with reliable prior data for emissions, covering a relatively wide range of variations. While it is acknowledged that uncertainties from the CMAQ model, including chemical and physical mechanisms, may also affect the VAE, we have carefully designed the loss function to minimize the impact of these uncertainties, as follows.

The loss function is similar as our previous designed,[32] consisted two parts. First is the adjustment of emission. Since in the decoder we only allow the change for the daily averaged emission level, it will have certain biases in prediction the time-series. Such loss can be helpful to minimize the discrepancy, to make better prediction of the emission adjustment matching with the two-time specific concentration. Consequently, the first part of the loss can be regarded as representing uncertainties in emissions. The updated emissions are integrated into the previously trained decoder, which encompasses the two-directional ConvLSTM models, to predict the hourly concentrations. The difference between these model-predicted concentrations and the CMAQ-simulated time-series concentrations serves as the other component of the loss function during training. While our objective is not to achieve a perfect match between the predicted concentration and the ground truth concentration, the second part of the loss function reflects uncertainties in input data, such as meteorology and satellite measurement, as well as errors in models. The balance between these two components of the loss is achieved through optimization during training, addressing all the uncertainties in emissions, potential errors associated with the CMAQ model, machine learning-model, and inputs, including OMI and meteorological variables.

Following the training process, the model's predictions for the 24-hour time-series concentrations, when supplied with satellite-retrieved two-hour concentrations, are considered as the spatially interpolated concentrations. The emission will be also predicted from the two-hour satellite measurement.

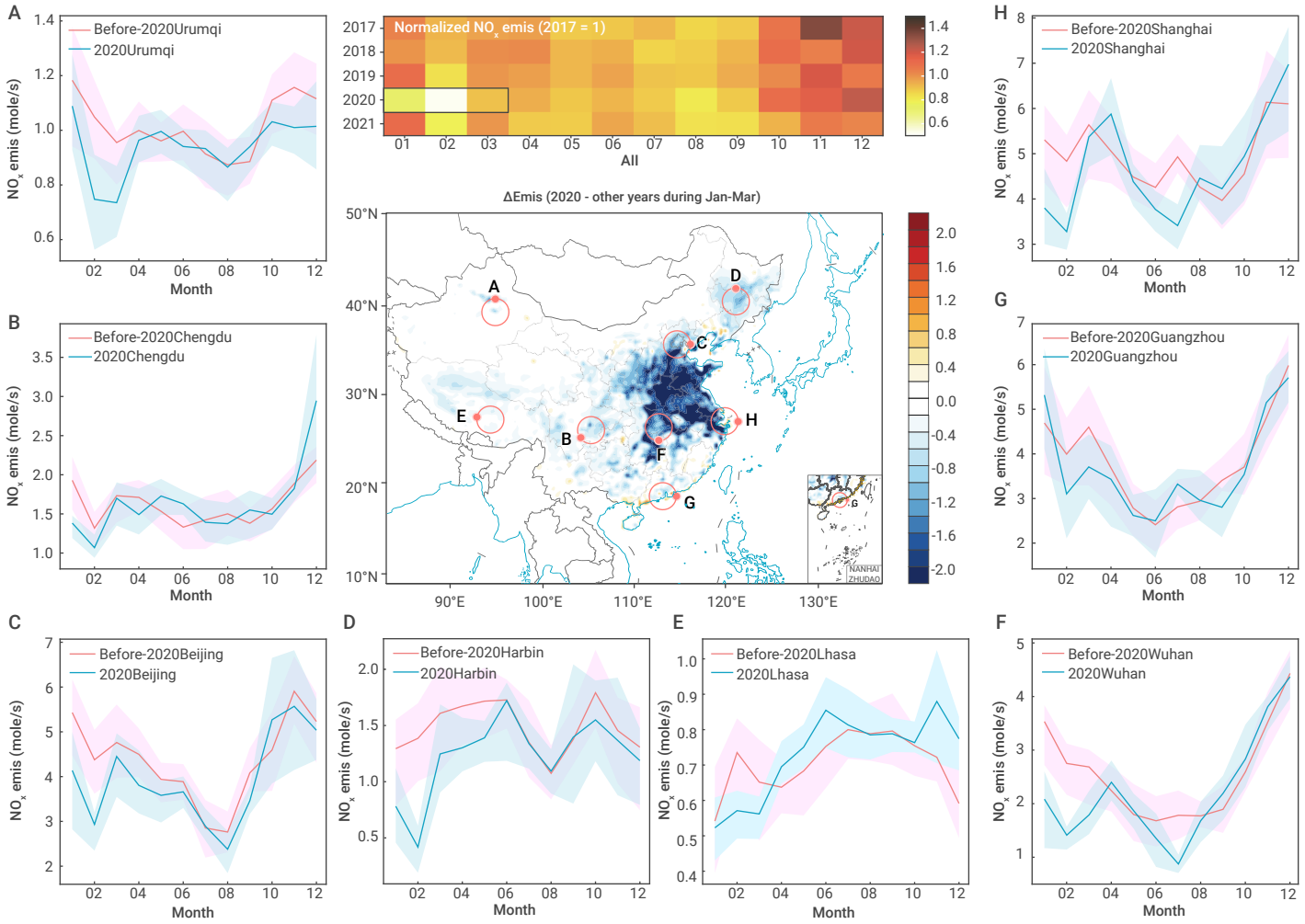


Figure 5. Estimated NOx emission variations by month, cites, and regions across 2017-2021 (unit: mol/s per 27km-by-27km grid cell).

We conducted comprehensive evaluations of the trained model to ensure that it maintains acceptable statistics (RMSE<1.3 mole/s, R^2>0.8, |NMB|<0.3), as depicted in Figure S14. Once the VAE is well-trained, it becomes a versatile tool that can be readily applied to other years without the need for any complex CTM simulations or priori emission estimations, rendering the DeepSAT4D model entirely CTM-free and highly efficient.

RESULTS

DeepSAT4D performance in retrieving vertically revolved NO_2 from column density

As described previously, DeepSAT4D can effectively utilize information from neighboring grid cells and temporal data to accurately reproduce the vertical structure of each grid cell, enabling the decoding of aggregated column data into concentrations at different vertical layers. This capability is particularly valuable when it comes to ground-level NO_2, which is closely related to human exposure and emission sources.

We first use CMAQ simulation data to validate the performance of DeepSAT4D in retrieving surface NO_2 from the CMAQ column density. Results suggest that DeepSAT4D excels in reflecting the spatial gradients of ground-level NO_2 concentrations (the lowest layer in CMAQ, approximately 19 meters above the ground) as presented in Figure 1.

Ground-level NO_2 concentrations (annual mean at 14:00 local time) often exhibit more pronounced gradients than their column densities, as indicated by the more significant decline in ground-level NO_2 concentration from the peak grid cell to the surrounding grid cells when compared to column density. This is particularly evident in megacities like Beijing, Shanghai, and Guangzhou, where emission sources are prevalent. This discrepancy occurs because NO_2 is transported aloft more than through the surface, and aloft concentrations also contribute to the NO_2 column. The discrepancy is less pronounced in Chengdu, likely attributable to low dispersion to the surrounding area facilitated by its location in a basin.[40] The gradients in ground-level NO_2 concentrations and column densities in Wuhan exhibit nearly no difference. This is mainly attributed to the significant emissions released at high altitudes from point sources with elevated stacks, a characteristic more pronounced in Wuhan compared to other cities (see Figure S2). The noticeable distinction between the two is anticipated to become evident when efforts to control aloft emissions are implemented in the future (see Figure S3). Evidently, DeepSAT4D adeptly captures the spatial gradient discrepancy between surface NO_2 and column density. It accurately predicts a more substantial reduction in surface NO_2 compared to the reduction in NO_2 column as one moves away from the peak NO_2 grid cells, though its performance in Lhasa, a relatively clean area, is limited, likely due to the restricted variation of emissions in remote areas within the current training dataset. DeepSAT4D accomplishes this by leveraging information from neighboring grid cells and time steps, enabling precise predictions of surface NO_2 concentration gradients that may not necessarily align with column data. As expected (refer to Figure S4), at the grid cell with the highest ground-level NO_2 concentration, typically situated closer to emission sources, the vertical profiles indicate a higher concentration near the surface. In contrast, the surrounding areas exhibit a more significant presence of NO_2 aloft due to dispersion and transport processes. DeepSAT4D adeptly captures this pattern, utilizing the spatial distribution of column density in conjunction with the corresponding meteorological conditions.

The NO_2 vertical profiles, as revealed by multiple years of OMI satellite retrievals with DeepSAT4D (see Figure 2), exhibit pronounced spatial-temporal variations. In general, during winter, NO_2 displays the smallest vertical

gradients compared to other seasons. This phenomenon can be attributed to weaker sink processes, including low dispersion and limited photochemical reactions which resulted in longer lifetime of NO_2, allowing it to disperse and remain present across vertical layers for extended periods. In contrast, the summer season exhibits the opposite trend. Strong vertical mixing and increased solar radiation lead to a shorter lifetime of NO_2, resulting in both lower surface and column densities, as well as the most significant vertical gradients observed throughout the year.

On a horizontal scale, NO_2 vertical profiles generally exhibit the most significant gradients in polluted cities with better dispersion conditions, such as Shanghai and Guangzhou, situated in coastal areas. In comparison to inland cities like Beijing, Wuhan, and Chengdu, coastal cities benefit from meteorological conditions that facilitate the transport of pollution to their surrounding regions. Inland cities tend to have more localized pollution aloft, resulting in reduced vertical gradients for NO_2. Cities in remote areas, such as Lhasa and Urumqi, exhibit the smallest vertical gradients due to their lower emissions, primarily representing rural areas in the analysis. These results align with observed NO_2 vertical profiles in general, indicating relatively larger vertical gradients in urban and industrial areas[41] compared to rural or remote areas.[42] Moreover, the gradients are more pronounced in the morning compared to noon and afternoon time periods (see Table S3).[43] In essence, the NO_2 vertical profiles obtained through DeepSAT4D align with a reasonable spatial-temporal pattern that is consistent with physical laws and actual environmental conditions.

DeepSAT4D performance in interpolating 24-hour NO_2 from OMI pass time

By incorporating additional features such as emission and meteorological factors that influence the diurnal variation of NO_2 concentrations, DeepSAT4D can effectively interpolate and provide a comprehensive 24-hour estimation of NO_2 concentrations starting from the original OMI pass time (at around 13:30 local time, we choose the 14:00 model data to align with it, as the CMAQ is conducted on an hourly basis) per day. This interpolation enhances the temporal resolution of NO_2 data, allowing for a more complete and detailed representation of NO_2 concentration throughout the day.

By comparison against the CMAQ simulation, the DeepSAT4D exhibits the capability to predict the complete diurnal pattern of surface NO_2 concentrations (as shown in Figure 3), with elevated concentrations during the night and lower levels during the day. Notably, the OMI pass time typically falls during the period of the lowest NO_2 concentrations throughout the day (in the afternoon), potentially leading to high biases when estimating overall daily-averaged concentrations without accounting for the hourly variations driven by meteorological factors as did in most previous studies.

The daily peak concentration of NO_2 can be up to twice as high, domain-wide, compared to the measurement time by OMI satellite, and this variation can range from 3 to 8 times across different spatial locations. DeepSAT4D effectively exhibits its ability to interpolate the complete 24-hour variation in NO_2 concentrations, which is generally consistent with CMAQ simulations. However, some uncertainties may be present, particularly around 10:00 am, corresponding to the traffic rush hour in the morning.

We further estimated daily mean surface NO_2 concentrations from OMI-retrieved column data using DeepSAT4D for the 2017-2021 period, as depicted in Figure 4, to compare with ground-based observations from CNEMC. It's important to note the fundamental differences in measurement nature between OMI satellite data (averaging over a 27 km by 27 km area) and ground-based monitors (measuring at specific locations). These two methods measure NO_2 differently, and this difference can be significant, particularly for NO_2, a species with a relatively short lifetime that's greatly influenced by emission sources and exhibits large spatial gradient at sub-grid scale.[44]

To demonstrate this issue, we introduce the 1km-to-27 km Emission Ratio (ER) as an indicator. ER is calculated by taking the ratio of the maximum emissions within 729 1-by-1 km grid cells within one 27-by-27 km grid cell to the average emissions of all 729 1-by-1km grid cells. These emissions are based on 1 km spatially resolved emission data and can range from 1 to 729. High ER values signify substantial spatial heterogeneity in emissions within a given 27km grid cell and the potential for significant spatial gradients at sub-grid scales.

Results suggest a good agreement between OMI predictions and ground-based measurements for sites with small ER values (ranging from 1 to 10). These sites are primarily located in eastern China and exhibit high NO_2 concentrations, such as in the cities of Beijing, Shanghai, and Guangzhou. However, as ER values increase, indicating a stronger impact from spatial heterogeneity in emissions, the discrepancy between OMI predictions and ground-based measurements becomes more substantial. Significant differences emerge in areas with strong spatial heterogeneity (ER > 20), and OMI-predicted NO_2 concentrations are notably lower than the measurements obtained from ground monitors. This discrepancy arises from the significant spatial heterogeneity in emissions, impeding ground-based monitor sites from accurately representing average levels for the entire 27km grid cell observed by OMI. This challenge extends to diurnal variation predictions when compared against ground-based measurements (see Figure S5). While DeepSAT4D effectively captures high concentrations during the night and low concentrations during the day, OMI predictions tend to be higher than ground-based measurements at night but lower during the day. The substantial dispersion during the daytime results in larger spatial gradients at the sub-grid scale, leading to a more significant underestimation during daytime. Given that satellite products lack observations during nighttime, the assumption completely relies on the CMAQ model, which is also subject to uncertainties associated with emissions or model mechanisms.[45] Leveraging more advanced geostationary satellite products with hourly temporal resolution measurements could prove beneficial in improving the prediction of diurnal variations in the future. It's also important to note that the 1 km grid emissions are established by downscaling data based on factors such as population, road maps, and point source locations. The ER data may also suffer uncertainties, particularly when dealing with missing point sources, which can substantially elevate ER values, especially in rural areas where emissions are scarce (represent by the ER<1 in Figure 4).

These results also imply that traditional data fusion or machine learning methods that use all available ground sites as ground truth for interpolating NO_2 concentrations may lead to significant overestimation, particularly in rural and western areas. Higher spatial resolution satellite data products can help mitigate this resolution problem, but it's equally crucial to have more detailed and widespread coverage of ground sites to provide valid data with good representativeness at regional scale.

Fast inference of NOx emission variations from OMI-satellite with DeepSAT4D

By applying temporal interpolation with DeepSAT4D, the emissions at a daily averaged scale can be assimilated to align with the two time slots measured by the OMI satellite. This process involves a machine-learning-based inversion optimized method, specifically the Variational Autoencoder (VAE) detailed in the method section. This approach enables DeepSAT4D to rapidly estimate changes in NOx emissions based on observed concentration variations, as illustrated in Figure S6.

The application of DeepSAT4D for estimating NOx emission changes during the period from 2017 to 2021, based on OMI-satellite data, is presented in Figure 5. The results indicate higher emissions during winter and lower emissions in summer, primarily driven by winter heating activities in northern China. Interestingly, the estimated NOx emissions in early 2020 are significantly lower, by more than 40%, in comparison to previous years, suggesting the impact of COVID-19 on emissions.[46-48] DeepSAT4D effectively captures these changes across all eight cities in China, demonstrating its ability to monitor and analyze variations in emissions in near-real time.

DISCUSSION

Satellite products hold tremendous potential for measuring atmospheric chemicals on a broad spatial-temporal scale. With the expansion of satellite products offering increased spatial and temporal resolution, satellite data is expected to play an increasingly vital role in enhancing our understanding of atmospheric chemistry and its associated impacts. However, their utility is restricted by the absence of vertical structure and diurnal variations. To address these limitations, this study developed an innovative deep learning model (DeepSAT4D) which is capable of converting satellite-measured

column density into comprehensive four-dimensional concentration fields. The DeepSAT4D has been effectively deployed for the retrieval of OMI-NO_2 data from 2017-2021 in China, clearly demonstrating its proficiency in decomposing vertically resolved NO_2 information from column density data, accurately interpolating 24-hour NO_2 concentrations based on OMI measurements, and efficiently estimating NO_x emission variations. These capabilities provide invaluable support for research endeavours related to epidemiology, human exposure, policy design, and air quality forecasting.

The success of the newly developed DeepSAT4D relies on three crucial aspects that have been identified as key challenging issues as follows.

First, DeepSAT4D effectively addresses the limitation on vertical structure by harnessing information from surrounding grid cells, utilizing a deep-learning architecture based on CNN. This is particularly significant because high values across the spatial domain often signify emission sources or urban centers, which exhibit a typical vertical profile characterized by surface-level abundance. Conversely, pollution edges or diminishing values along with airflow often indicate rural areas, where the vertical profile displays more contributions from higher altitudes due to the transport of air into the upper layers. It is important to incorporate such information serves as an additional feature in establishing the relationship between column and surface concentrations. Moreover, considering the time period surrounding the data with variations in meteorological conditions proves to be beneficial in identifying the vertical structure. This is especially crucial for pinpointing point sources like power plants and industry combustions, as they are most affected by meteorological variations during the plume rising process. While the OMI-satellite used in this study only provides data for one specific local time hour (14:00) once a day, the implementation of geostationary satellites with hourly data will significantly enhance the accuracy of vertical structure estimation in the future.

Secondly, DeepSAT4D tackles the challenge of filling in missing hourly data and estimating daily means from a single daily measurement taken at 14:00 local time from the OMI satellite. This is achieved by incorporating additional features, encompassing both emission and meteorological factors that drive the diurnal variation of NO_2 concentration throughout the day. Recognizing that meteorological variations tend to be more accurate and readily available compared to emissions data, current DeepSAT4D is designed to focus on reducing emissions variability to a daily average level. Such design helps to avoid potential ill-posed problems that can arise when attempting to retroactively adjust emissions based on concentration data, given the often complex and non-linear relationship between emissions and concentrations. This approach is also well-founded as it aligns with the notion that the diurnal variation of emissions is often correlated with changes in meteorological conditions. This concept finds support in models like the meteorology driven emission modeling for both biogenic (e.g., MEGAN[49]) and anthropogenic emissions (Met-Emis[50]). With this in mind, DeepSAT4D effectively interpolates NO_2 concentrations for the other 23 hours between two OMI-retrieved time slots. This interpolation is made possible by incorporating hourly meteorological factors and drawing upon previously established knowledge of emission patterns acquired during the training phase, utilizing multiple CMAQ simulations and advanced ConvLSTM structure. Consequently, this approach enables the estimation of NO_2 concentrations throughout the day. The comparison between DeepSAT4D predictions and ground-based measurements also underscores a notable shortcoming in the integration of ground-based data at the grid cell level. This limitation is accentuated by the considerable spatial heterogeneity observed at the sub-grid level, which can substantially undermine the representativeness of ground-based measurements. Furthermore, it's crucial to acknowledge that these ground-based measurements are primarily acquired in urban centers and exhibit distinct vertical profiles when contrasted with the prediction grids, as depicted in Figure S7. These challenges may result in a substantial overestimation of ground-level NO_2 concentration in rural areas when employing traditional machine-learning methods to integrate ground-based measurements directly into the grid cell, as depicted in Figure S8. In contrast, DeepSAT4D predictions align closely with CTM-based measurements, presenting a similar level of ground-level NO_2 concentration while exhibiting a spatial distribution much closer to OMI-column data than the original CTM. This highlights DeepSAT4D's ability to mitigate uncertainties associated with the CTM. However, limitations persist, especially in Tibet, where some unusually high values require further investigation in future studies, possibly benefiting from an increased abundance of training data.

Thirdly, our results demonstrate that DeepSAT4D is capable of accurately estimating surface NO_2 concentrations while efficiently updating emissions. Through the VAE structure, emissions are adjusted as a by-product to align with specified concentration points at 24-hour intervals. Notably, our model was trained using data from 2017 and effectively applied for the years 2018-2021, all without the need for additional emission estimates or complex CTM simulations, thus rendering it completely CTM-free. These findings underscore the tremendous potential of DeepSAT4D for real-time emission updates to enhance the accuracy and efficiency for air quality forecasting in the future.[51-52]

The study also highlights the natural connection between the spatial-temporal relationship in NO_2 concentrations, which originate from the same emission sources and evolve through atmospheric processes like diffusion and advection driven by meteorological factors. It demonstrates the feasibility of decoding the satellite-measured column density back into chemical concentrations at each layer and hour when a proper linkage is established. This process can be considered an efficient means of fusing simulation data with satellite observations since it incorporates modeling information as features when developing the linkage.

This study focuses on the OMI-NO_2 products, serving as a demonstration of the DeepSAT4D method. It is highly recommended to utilize more recent satellite products with higher spatial and temporal resolution, such as Tropospheric Monitoring Instrument (TROPOMI),[53] Geostationary Environment Monitoring Spectrometer (GEMS),[54] and Himawari-8[55-57] in the future, while it's important to note that training DeepSAT4D with CTM simulations requires corresponding high resolution (about 4km). This poses a challenge for conducting large-scale CTM simulations at the national or global level. Furthermore, the results might be affected by the uncertainties inherent in the CTM itself and the absence of a comprehensive training datasets, particularly for the western area (as indicated by relatively poor performance in Lhasa). Nevertheless, with the advancement in computational abilities and machine learning, the DeepSAT4D method introduced in this study can be readily extended to other pollutants and regions. This study introduces an effective approach for training machine-learning models, leveraging physical model (e.g., CMAQ) results to generate diverse datasets. This methodology enhances the performance of data-driven models, such as DeepSAT4D. While the physical model may encounter uncertainties in input data (e.g., meteorological variables) and mechanisms, it undergoes continuous improvement benefiting from the ongoing growth of scientific knowledge in atmospheric chemistry and physics. These uncertainties can be further mitigated, allowing DeepSAT4D to offer more accurate and reliable predictions for chemical concentrations and emissions derived from satellite products, which can be useful to support the further analysis on atmospheric chemical impacts on human health and Earth's ecosystem.

REFERENCES

1. Lerdau, M. T., Munger, J. W., and Jacob, D. J. (2000). The NO_2 flux conundrum. Science **289**(5488): 2291–2293. DOI: 10.1126/science.289.5488.2291.
2. Xue, T., Tong, M., Wang, M., et al. (2023). Health impacts of long-term NO_2 exposure and inequalities among the Chinese population from 2013 to 2020. Environmental Science & Technology **57**(13):5349-5357. DOI: 10.1021/acs.est.2c08022.
3. Dong, Z., Wang, S., Jiang, Y., et al. (2023). An acid rain–friendly NH_3 control strategy to maximize benefits toward human health and nitrogen deposition. Science of The Total Environment **859**: 160116. DOI: 10.1016/j.scitotenv.2022.160116.
4. Callies, J., Corpaccioli, E., Eisinger, M., et al. (2000). GOME-2-Metop's second-generation sensor for operational ozone monitoring. ESA bulletin **102**: 28–36. https://www.esa.int/esapub/bulletin/bullet102/Callies102.pdf
5. Bovensmann, H., Burrows, J. P., Buchwitz, M., et al. (1999). SCIAMACHY: Mission objectives and measurement modes. Journal of the atmospheric sciences **56**(2): 127–150. DOI: 10.1175/1520-0469(1999)056<0127:SMOAMM>2.0.CO;2.
6. Celarier, E. A., Brinksma, E. J., Gleason, J. F., et al. (2008). Validation of Ozone Monitoring Instrument nitrogen dioxide columns. Journal of Geophysical Research: Atmospheres **113**(D15): 1–2. DOI: 10.1029/2007JD008908.
7. Waters, J. W., Froidevaux, L., Harwood, R. S., et al. (2006). The earth observing system microwave limb sounder (EOS MLS) on the Aura satellite. IEEE transactions on geoscience and remote sensing **44**(5): 1075–1092. DOI: 10.1109/TGRS.2006.873771.

8. Achakulwisut, P., Brauer, M., Hystad, P., et al. (2019). Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO_2 pollution: Estimates from global datasets. The Lancet Planetary Health **3**(4): e166–e178. DOI: 10.1016/S2542-5196(19)30046-4.
9. Geddes, J. A., Martin, R. V., Boys, B. L., et al. (2016). Long-term trends worldwide in ambient NO_2 concentrations inferred from satellite observations. Environmental health perspectives **124**(3): 281–289. DOI: 10.1289/ehp.1409567.
10. Anand, J. S., and Monks, P. S. (2017). Estimating daily surface NO_2 concentrations from satellite data–a case study over Hong Kong using land use regression models. Atmospheric Chemistry and Physics **17**(13): 8211–8230. DOI: 10.5194/acp-17-8211-2017.
11. Chi, Y., Fan, M., Zhao, C., et al. (2021). Ground-level NO2 concentration estimation based on OMI tropospheric NO2 and its spatiotemporal characteristics in typical regions of China. Atmospheric Research **264**: 105821. DOI: 10.1016/j.atmosres.2021.105821.
12. Yu, M., and Liu, Q. (2021). Deep learning-based downscaling of tropospheric nitrogen dioxide using ground-level and satellite observations. Science of the Total Environment **773**: 145145. DOI: 10.1016/j.scitotenv.2021.145145.
13. Ghahremanloo, M., Lops, Y., Choi, Y., et al. (2021). Deep learning estimation of daily ground - level NO2 concentrations from remote sensing data. Journal of Geophysical Research: Atmospheres **126**(21): e2021JD034925. DOI: 10.1029/2021JD034925.
14. Wei, J., Liu, S., Li, Z., et al. (2022). Ground-level NO_2 surveillance from space across China for high resolution using interpretable spatiotemporally weighted artificial intelligence. Environmental Science & Technology **56**(14): 9988–9998. DOI: 10.1021/acs.est.2c03834.
15. Long, S., Wei, X., Zhang, F., et al. (2022). Estimating daily ground-level NO_2 concentrations over China based on TROPOMI observations and machine learning approach. Atmospheric Environment **289**: 119310. DOI: 10.1016/j.atmosenv.2022.119310.
16. Grzybowski, P. T., Markowicz, K. M., and Musiał, J. P. (2023). Estimations of the ground-level NO_2 concentrations based on the Sentinel-5P NO_2 tropospheric column number density product. Remote Sensing **15**(2): 378. DOI: 10.3390/rs15020378.
17. Li, T., and Cheng, X. (2021). Estimating daily full-coverage surface ozone concentration using satellite observations and a spatiotemporally embedded deep learning approach. International Journal of Applied Earth Observation and Geoinformation **101**: 102356. DOI: 10.1016/j.jag.2021.102356.
18. Wang, Y., Yuan, Q., Li, T., et al. (2021). Estimating daily full-coverage near surface O_3, CO, and NO_2 concentrations at a high spatial resolution over China based on S5P-TROPOMI and GEOS-FP. ISPRS Journal of Photogrammetry and Remote Sensing **175**: 311–325. DOI: 10.1016/j.isprsjprs.2021.03.018.
19. Zhu, S., Xu, J., Fan, M., et al. (2023). Estimating near-surface concentrations of major air pollutants from space: A universal estimation framework LAPSO. IEEE Transactions on Geoscience and Remote Sensing **61**: 1–11. DOI: 10.1109/TGRS.2023.3248180.
20. Zhu, S., Xu, J., Zeng, J., et al. (2023). LESO: A ten-year ensemble of satellite-derived intercontinental hourly surface ozone concentrations. Scientific Data **10**(1): 741. DOI: 10.1038/s41597-023-02656-4.
21. Lamsal, L.N., Martin, R.V., Van Donkelaar, A., et al. (2008). Ground - level nitrogen dioxide concentrations inferred from the satellite - borne Ozone Monitoring Instrument. Journal of Geophysical Research: Atmospheres, **113**(D16):15. DOI: 10.1029/2007jd009235.
22. Bechle, M. J., Millet, D. B., and Marshall, J. D. (2013). Remote sensing of exposure to NO2: Satellite versus ground-based measurement in a large urban area. Atmospheric Environment **69**: 345–353. DOI: 10.1016/j.atmosenv.2012.11.046.
23. Lin, J.-T., Martin, R. V., Boersma, K. F., et al. (2014). Retrieving tropospheric nitrogen dioxide from the Ozone Monitoring Instrument: Effects of aerosols, surface reflectance anisotropy, and vertical profile of nitrogen dioxide. Atmos. Chem. Phys., **14**(3):1441–1461. DOI: 10.5194/acp-14-1441-2014.
24. Lin, J. T., and McElroy, M. B. (2010). Impacts of boundary layer mixing on pollutant vertical profiles in the lower troposphere: Implications to satellite remote sensing. Atmospheric Environment **44**(14): 1726–1739. DOI: 10.1016/j.atmosenv.2010.02.009.
25. Wang, Y., Dörner, S., Donner, S. , et al. (2019). Vertical profiles of NO_2, SO_2, HONO, HCHO, CHOCHO and aerosols derived from MAX-DOAS measurements at a rural site in the central western North China Plain and their relation to emission sources and effects of regional transport. Atmospheric Chemistry and Physics **19**(8): 5417–5449. DOI: 10.5194/acp-19-5417-2019.
26. Kong, L., Tang, X., Zhu, J., et al. (2021). A 6-year-long (2013–2018) high-resolution air quality reanalysis dataset in China based on the assimilation of surface observations from CNEMC. Earth System Science Data **13**(2): 529–570. DOI: 10.5194/essd-13-529-2021.
27. Houyoux, M. R., and Vukovich, J. M. (1999). Updates to the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system and integration with Models-3. The Emission Inventory: Regional Strategies for the Future **1461**: 1–11.
28. Han, K. M., Lee, S., Chang, L. S., et al. (2015). A comparison study between CMAQ-simulated and OMI-retrieved NO_2 columns over East Asia for evaluation of NOx emission fluxes of INTEX-B, CAPSS, and REAS inventories. Atmospheric Chemistry and Physics **15**(4): 1913–1938. DOI: 10.5194/acp-15-1913-2015.
29. Kuhlmann, G., Lam, Y. F., Cheung, H. M., et al. (2015). Development of a custom OMI NO_2 data product for evaluating biases in a regional chemistry transport model. Atmospheric Chemistry and Physics **15**(10): 5627–5644. DOI: 10.5194/acp-15-5627-2015.
30. Liu, L., Zhang, X., Xu, W., et al. (2020). Reviewing global estimates of surface reactive nitrogen concentration and deposition using satellite retrievals. Atmospheric Chemistry and Physics **20**(14): 8641–8658. DOI: 10.5194/acp-20-8641-2020.
31. Huang, L., Liu, S., Yang, Z., et al. (2021). Exploring deep learning for air pollutant emission estimation. Geoscientific Model Development **14**(7): 4641–4654. DOI: 10.5194/gmd-14-4641-2021.
32. Xing, J., Li, S., Zheng, S., et al. (2022). Rapid Inference of Nitrogen Oxide Emissions Based on a Top-Down Method with a Physically Informed Variational Autoencoder. Environmental Science & Technology **56**(14): 9903–9914. DOI: 10.1021/acs.est.1c08337.
33. Xing, J., Zheng, S., Ding, D., et al. (2020). Deep learning for prediction of the air quality response to emission changes. Environmental science & technology **54**(14): 8589–8600. DOI: 10.1021/acs.est.0c02923.
34. Xing, J., Zheng, S., Li, S., et al. (2022). Mimicking atmospheric photochemical modeling with a deep neural network. Atmospheric research **265**: 105919. DOI: 10.1016/j.atmosres.2021.105919.
35. He, K., Zhang, X., Ren, S., et al. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). DOI: 10.1109/CVPR.2016.90.
36. Kingma, D. P., and Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114. DOI: 10.48550/arXiv.1312.6114
37. Shi, X., Chen, Z., Wang, H., et al. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems **28**. DOI: 10.1007/978-3-319-21233-3_6.
38. Appel, K., Pouliot, G., Simon, H., et al. (2013). Evaluation of dust and trace metal estimates from the Community Multiscale Air Quality (CMAQ) model version 5.0. Geoscientific Model Development **6**(4):883-899. DOI: 10.5194/gmd-6-883-2013.
39. Skamarock, W. C., Klemp, J. B., Dudhia, J., et al. (2008). A Description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-475+STR **113**. DOI: 10.13140/RG.2.1.2310.6645.
40. Deng, Y., Li, J., Li, Y., et al. (2019). Characteristics of volatile organic compounds, NO_2, and effects on ozone formation at a site with high ozone level in Chengdu. Journal of Environmental Sciences **75**: 334–345. DOI: 10.1016/j.jes.2018.05.004.
41. Kang, Y., Tang, G., Li, Q., et al. (2021). Evaluation and evolution of MAX-DOAS-observed vertical NO_2 profiles in urban Beijing. Advances in Atmospheric Sciences, **38**(7): 1188–1196. DOI: 10.1007/s00376-021-0370-1.
42. Liu, S., Cheng, S., Ma, J., et al. (2023). MAX-DOAS measurements of tropospheric NO_2 and HCHO vertical profiles at the longfengshan regional background station in northeastern China. Sensors **23**(6): 3269. DOI: 10.3390/s23063269.
43. Chen, L., Pang, X., Li, J., et al. (2022). Vertical profiles of O_3, NO_2 and PM in a major fine chemical industry park in the Yangtze River Delta of China detected by a sensor package on an unmanned aerial vehicle. Science of the Total Environment **845**: 157113. DOI: 10.1016/j.scitotenv.2022.157113.
44. Tao, H., Xing, J., Zhou, H., et al. (2020). Impacts of improved modeling resolution on the simulation of meteorology, air quality, and human exposure to $PM_{2.5}$, O_3 in Beijing, China. Journal of Cleaner Production **243**:118574. DOI: 10.1016/j.jclepro.2019.118574.
45. Toro, C., Foley, K., Simon, H., et al. (2021). Evaluation of 15 years of modeled atmospheric oxidized nitrogen compounds across the contiguous United States. Elem Sci Anth **9**(1): 00158. DOI: 10.1525/elementa.2020.00158.
46. Xing, J., Li, S., Jiang, Y., et al. (2020). Quantifying the emission changes and associated air quality impacts during the COVID-19 pandemic on the North China Plain: A response modeling study. Atmospheric Chemistry and Physics **20**(22): 14347–14359. DOI: 10.5194/acp-20-14347-2020.
47. Fan, C., Li, Z., Li, Y., et al. (2021). Variability of NO_2 concentrations over China and effect on air quality derived from satellite and ground-based observations. Atmospheric Chemistry and Physics **21**(10): 7723–7748. DOI: 10.5194/acp-21-7723-2021.
48. Zheng, B., Zhang, Q., Geng, G., et al. (2021). Changes in China's anthropogenic emissions and air quality during the COVID-19 pandemic in 2020. Earth System Science Data **13**(6): 2895–2907. DOI: 10.5194/essd-13-2895-2021.
49. Guenther, A.B., Jiang, X., Heald, C.L., et al. (2012). The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2. 1): An extended and updated framework for modeling biogenic emissions. Geosci. Model Dev. **5**(6):1471–1492. DOI: 10.5194/gmd-5-1471-2012.
50. Baek, B. H., Coats, C., Ma, S., et al. (2023). Dynamic Meteorology-induced Emissions Coupler (MetEmis) development in the Community Multiscale Air Quality (CMAQ): CMAQ-MetEmis. Geoscientific Model Development **16**(16): 4659–4676. DOI: 10.5194/gmd-16-4659-2023.
51. Demerjian, K., Beauharnois, M., Ku J., et al. (2013). Developing real-time emissions estimates for enhanced air quality forecasting. EM: Air and Waste Management Associations Magazine for Environmental Managers. Air &Waste Management Association, Pittsburgh, PA 11:22-27. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NERL&dirEntrId=265045
52. Wu, H., Kong, L., Tang, X., et al. (2023). Air Quality Forecasting with Inversely Updated

Emissions for China. Environmental Science &Technology Letters 10(8):655-661. DOI: 10.1021/acs.estlett.3c00266.
53. Van Geffen, J., Boersma, K. F., Eskes, et al. (2020). S5P TROPOMI NO_2 slant column retrieval: Method, stability, uncertainties and comparisons with OMI. Atmospheric Measurement Techniques, **13**(3): 1315–1335. DOI: 10.5194/amt-13-1315-2020.
54. Park, S. S., Kim, S. W., Song, C. K., et al. (2020). Spatio-temporal variability of aerosol optical depth, total ozone and NO_2 over East Asia: Strategy for the validation to the GEMS Scientific Products. Remote Sensing **12**(14): 2256. DOI: 10.3390/rs12142256.
55. Liu, J., and Chen, W. (2022). First satellite-based regional hourly NO2 estimations using a space-time ensemble learning model: A case study for Beijing-Tianjin-Hebei Region, China. Science of The Total Environment **820**: 153289. DOI: 10.1016/j.scitotenv.2022.153289.
56. Ri, X., Tana, G., Shi, C., et al. (2022). Cloud, atmospheric radiation and renewal energy application (CARE) version 1.0 cloud top property product from Himawari-8/AHI: Algorithm development and preliminary validation. IEEE Transactions on Geoscience and Remote Sensing **60**:1-11. DOI: 10.1109/TGRS.2022.3172228.
57. Tana, G., Ri, X., Shi, C., et al. (2023). Retrieval of cloud microphysical properties from Himawari-8/AHI infrared channels and its application in surface shortwave downward radiation estimation in the sun glint region. Remote Sensing of Environment **290**: 113548. DOI: 10.1016/j.rse.2023.113548.
58. Ding, D., Xing, J., Wang, S., et al. (2019). Impacts of emissions and meteorological changes on China's ozone pollution in the warm seasons of 2013 and 2017. Frontiers of Environmental Science & Engineering **13**:1-9. DOI: 10.1007/s11783-019-1160-1.
59. Ding, D., Xing, J., Wang, S., et al. (2019). Estimated contributions of emissions controls, meteorological factors, population growth, and changes in baseline mortality to reductions in ambient $PM_{2.5}$ and $PM_{2.5}$-related mortality in China, 2013–2017. Environmental health perspectives **127**(6):067009. DOI: 10.1289/EHP4157.
60. Zheng, H., Zhao, B., Wang, S., et al. (2019). Transition in source contributions of $PM_{2.5}$ exposure and associated premature mortality in China during 2005–2015. Environment international **132**:105111. DOI: 10.1016/j.envint.2019.105111.
61. Liu, S., Xing, J., Wang, S., et al. (2021). Health benefits of emission reduction under 1.5°C pathways far outweigh climate-related variations in China. Environmental Science & Technology **55**(16):10957-10966. DOI: 10.1021/acs.est.1c01583.
62. Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. DOI: 10.48550/arXiv.1412.6980.
63. Eskes, H. J., and Boersma, K. F. (2003). Averaging kernels for DOAS total-column satellite retrievals. Atmospheric Chemistry and Physics **3**(5): 1285–1291. DOI: 10.5194/acp-3-1285-2003.

ACKNOWLEDGMENTS

This work was supported by the Open Research Program of the International Research Center of Big Data for Sustainable Development Goals (CBAS2022ORP01), and Microsoft Climate Research Initiative program.

AUTHOR CONTRIBUTIONS

SL and JX designed the methodology and conducted the WRF/CMAQ, satellite retrieval, and machine-learning experiment and write the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.59717/j.xinn-geo.2024.100061

LEAD CONTACT WEBSITE

http://jszy.whu.edu.cn/lisiwei

The Innovation Geoscicence, *Volume 2*

DeepSAT4D: Deep learning empowers four-dimensional atmospheric chemical concentration and emission retrieval from satellite

DOI: https://doi.org/ 10.59717/j.xinn-geo.2024.100061

Siwei Li and Jia Xing

Table of Contents

Figure S12 performance of ConvLSTM.

Figure S13 Variational auto-encoder (VAE) model structure.

Figure S14 performance of UNet-LSTM in reproducing the emissions.

Table S1 Summary of input / output for DeepSAT4D

Data	ResNet	Conv-LSTM	Unet-LSTM
Column density	***Input***	-	-
	Hour 0/24*	-	-
Vertical conc.	***Output***	***Input***	***Input***
	Hour 0/24	Hour 0/24	Hour 0+24
	-	***Output***	***Output***
	-	Hour 1-23	Hour 1-23
Met.	***Input***	***Input***	***Input***
	Hour 0/24	Hour 0-24	Hour 0-24
Emis.	-	***Input***	***Output***
	-	daily	daily

*decoding the one-hour NO_2 column measured at satellite pass time into the vertical layer-specific hourly NO_2 concentrations; Hours 0 and 24 correspond to the hour that aligns with the satellite pass time, while Hours 1-23 represent the intervening hours

Table S2 Summary of WRF-CMAQ model simulation

Scenario	Name	Meteorology	Anthropogenic Emission
1	2017-base	2017	2017
2	2017-v2	2017	Increased randomly
3	2017-v3	2017	Increased randomly
4	BCC-ssp126	2050 ssp126	2050-low
5	BCC-ssp585	2050 ssp585	2050-high

*all scenarios are obtained from our previous study (Xing et al., 2022), First 15 days of each month of Scenario 1,3,4,5 was used for training; rest days of Scenario 1,3,4,5 and all days of Scenario 2 was used for testing

Table S3 Comparison of DeepSAT4D predicted NO_2 vertical profiles with observations

Location	Time	Height range AGL (m)	Variations (ppb per 100m)		Reference
			Observation	DeepSAT4D	
An industry park in YRD (30.16ºN, 120.90ºE)	Aug 19-21, 2020 (9:00, 11:00, 13:00, 15:00, 17:00 LT)	0-100	-1.88	-2.03	Chen et al., 2022
		100-300	-1.19	-0.42	
		300-500	-1.30	-0.09	
	Oct 21,24,25, 2020 (9:00, 11:00, 13:00, 15:00, 17:00 LT)	0-100	-2.39	-2.48	
		100-300	-1.64	-1.49	
		300-500	-1.56	-0.07	
	Dec 4, 6, 8, 2020 (9:00, 11:00, 13:00, 15:00, 17:00 LT)	0-100	-6.85	-1.26	
		100-300	-1.43	-0.78	
		300-500	-1.53	-0.49	
	Jan 28-30, 2021 (9:00, 11:00, 13:00, 15:00, 17:00 LT)	0-100	-2.37	-1.68	
		100-300	-2.14	-1.12	
		300-500	-1.89	-0.56	
	Mar 23-25, 2021 (9:00, 11:00, 13:00, 15:00, 17:00 LT)	0-100	-5.81	-1.87	
		100-300	-0.96	-0.57	
		300-500	3.77	-0.19	
An urban residential area in Beijing (39.97ºN, 116.37ºE)	Apr 1 to May 31, 2019 (8:00-10:00 LT)	0-500	-2.1	-3.00	Kang et al., 2021
		500-1300	-0.50	-0.38	
		1300-3000	-0.18	-0.20	
	Apr 1 to May 31, 2019 (11:00-13:00 LT)	0-500	-0.20	-0.30	
		500-1300	-0.35	-0.31	
		1300-3000	-0.12	-0.09	
	Apr 1 to May 31, 2019 (14:00-17:00 LT)	0-500	-0.30	-0.20	
		500-1300	-0.20	-0.20	
		1300-3000	-0.17	-0.09	
Remote area, Longfengshan in north eastern China (44.73ºN, 127.60ºE)	Oct 24, 2020-Oct 13, 2021 (daytime)	0-500	-0.11	-0.27	Liu et al., 2023
		500-4000	-0.016	-0.019	

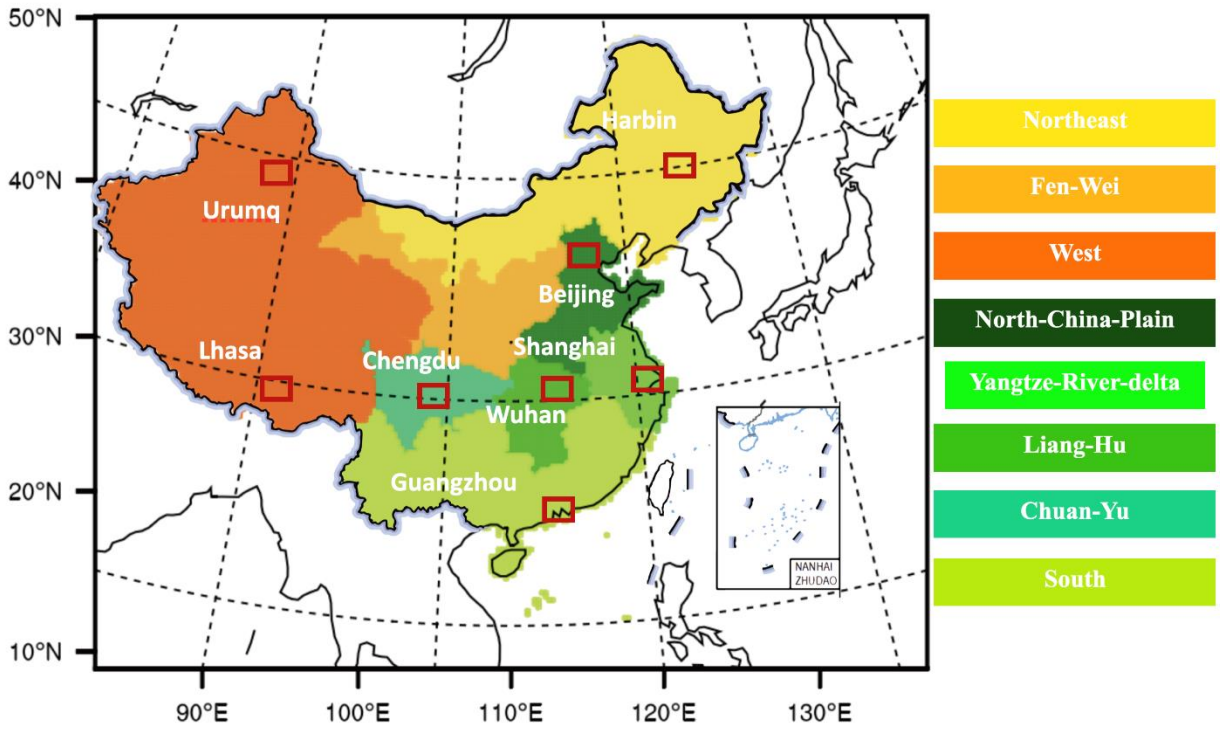
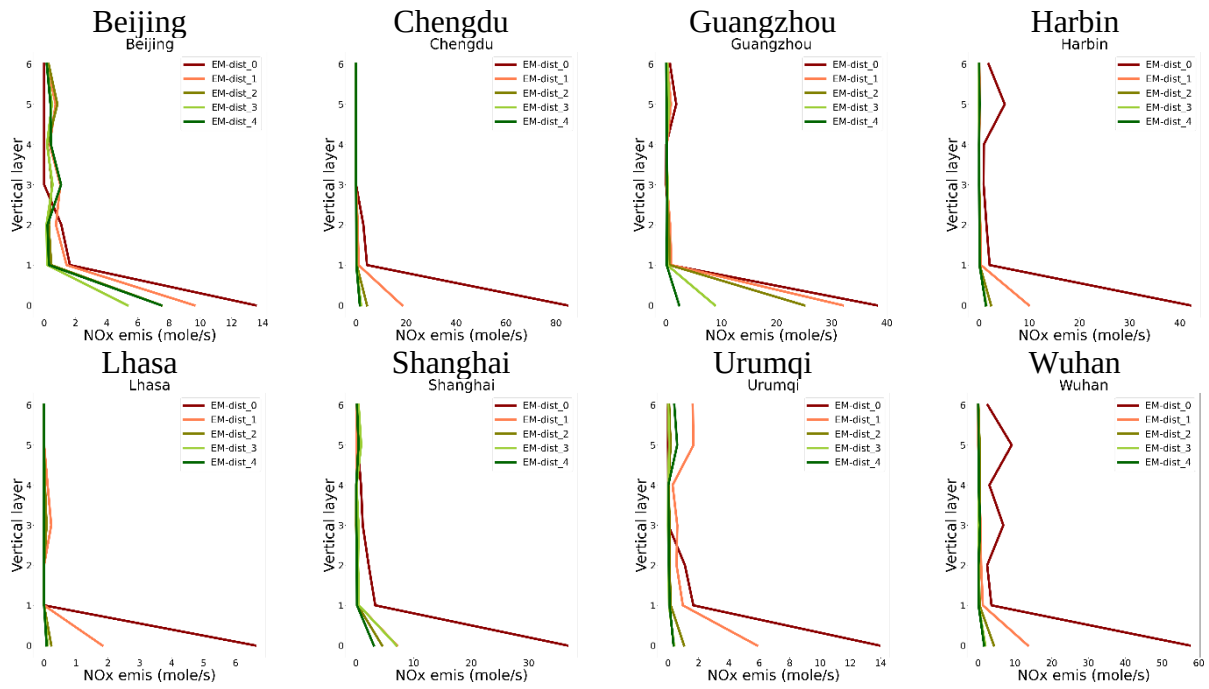
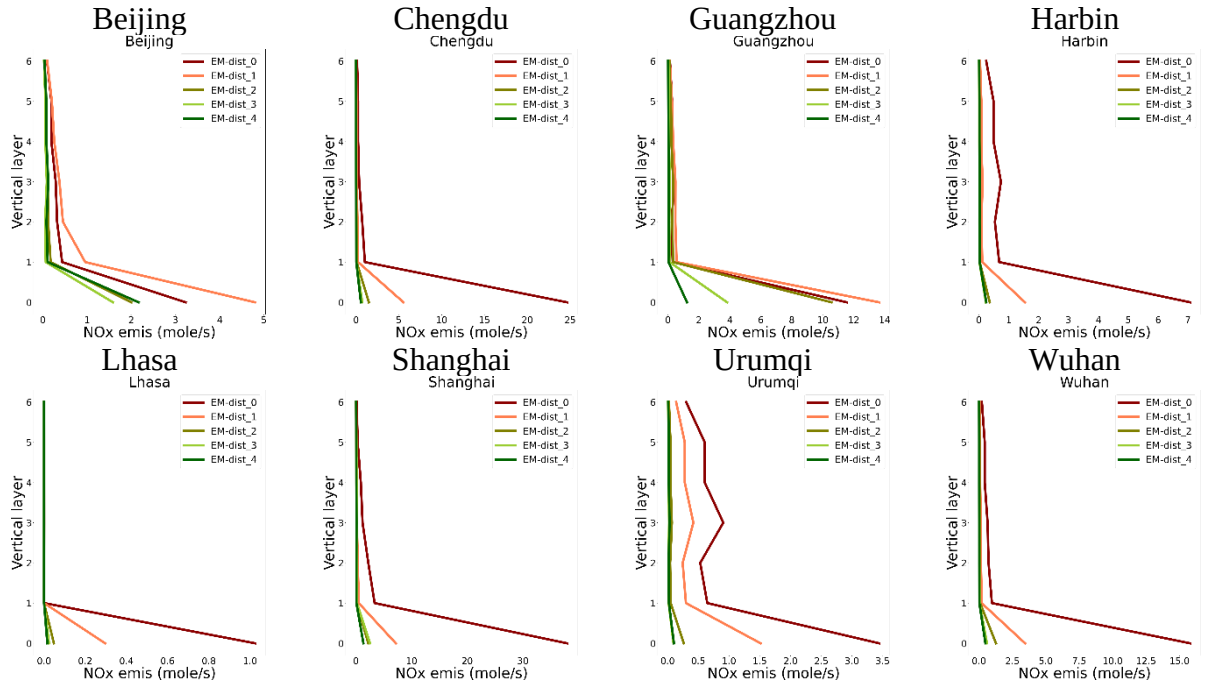


Figure S1 Simulation domain and location of cities and regions.



(a) 2017-base



(b) BCC-ssp126

Figure S2 Comparison of vertical profiles of emissions among eight cities and surrounding area (data used in CMAQ simulations, local time 14:00 at annual mean level)

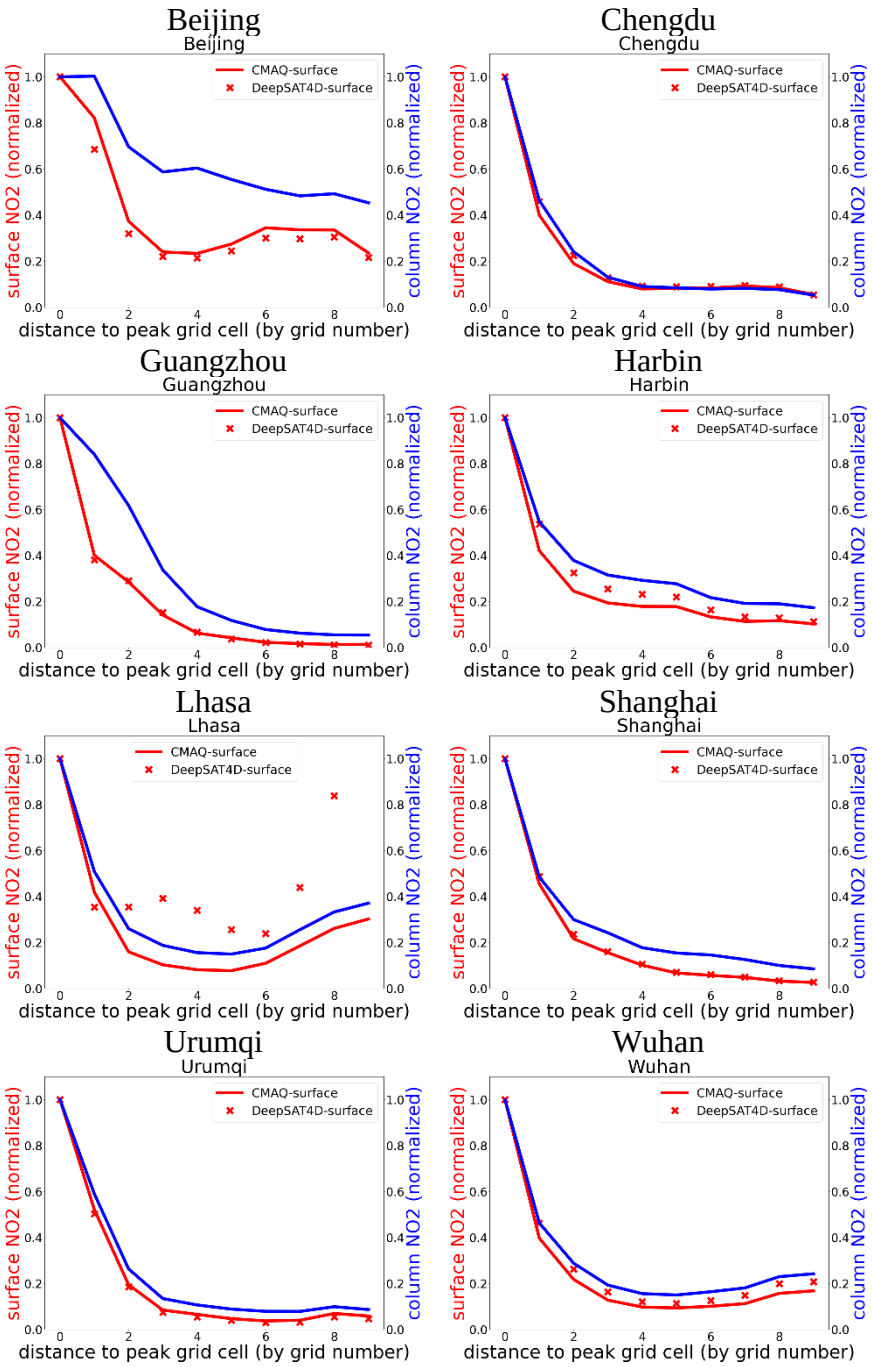


Figure S3. Spatial gradients of NO_2 column density and surface NO_2 concentration and in eight cities in BCC-ssp126 low emission scenario (selected the grid with highest surface concentration as the peak grid cell, and calculated the average concentrations in surrounding grid cells away from it by the distance up to 10 grid cells)

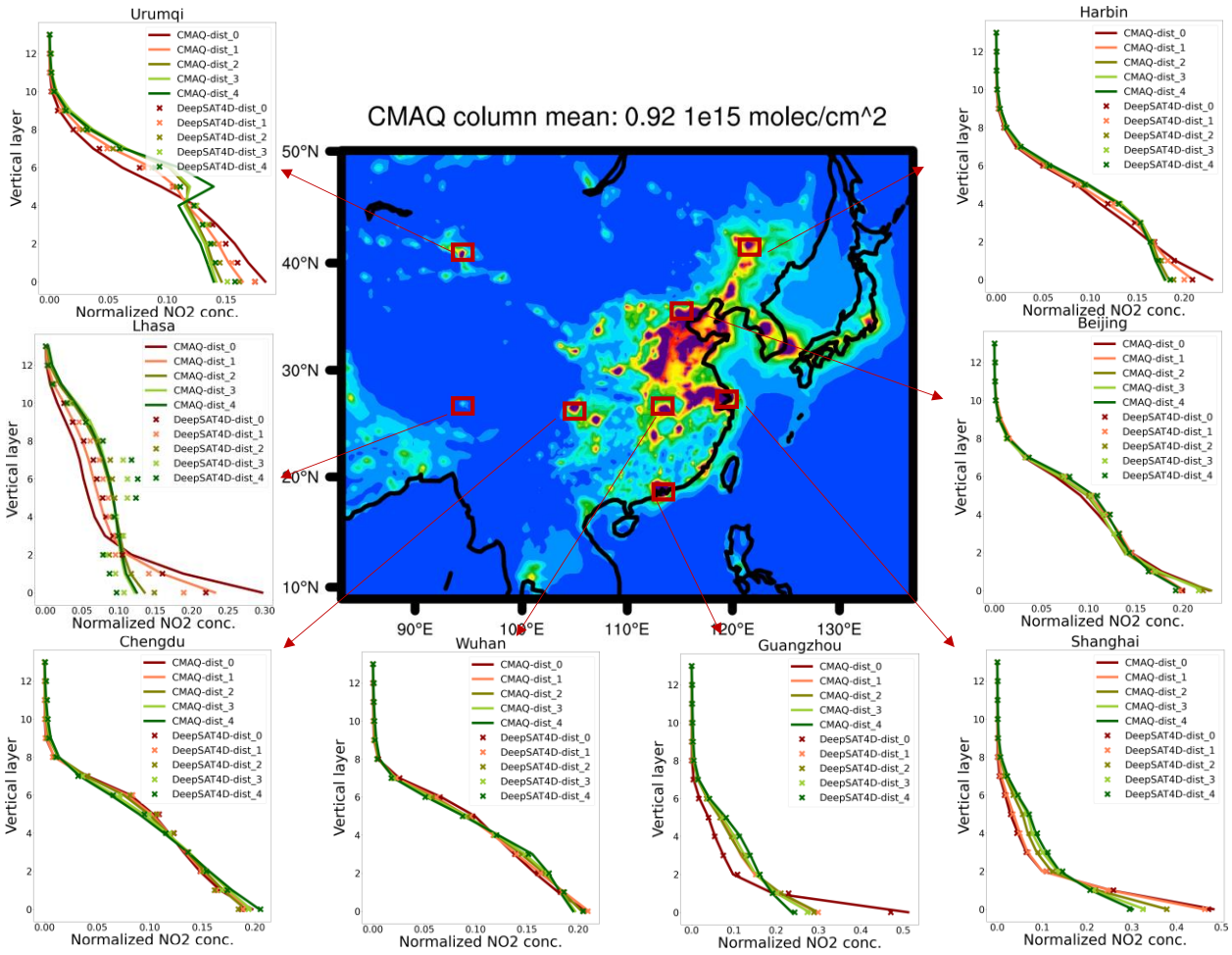


Figure S4 Comparison of NO_2 vertical profiles between CMAQ simulation and DeepSAT4D prediction.

(Notably, the highest surface NO_2 concentrations generally result in a greater concentration near the surface, while the surrounding areas exhibit higher NO_2 levels aloft due to dispersion and transport processes. The DeepSAT4D effectively captures this pattern, utilizing spatial column density distribution and corresponding meteorological conditions.)

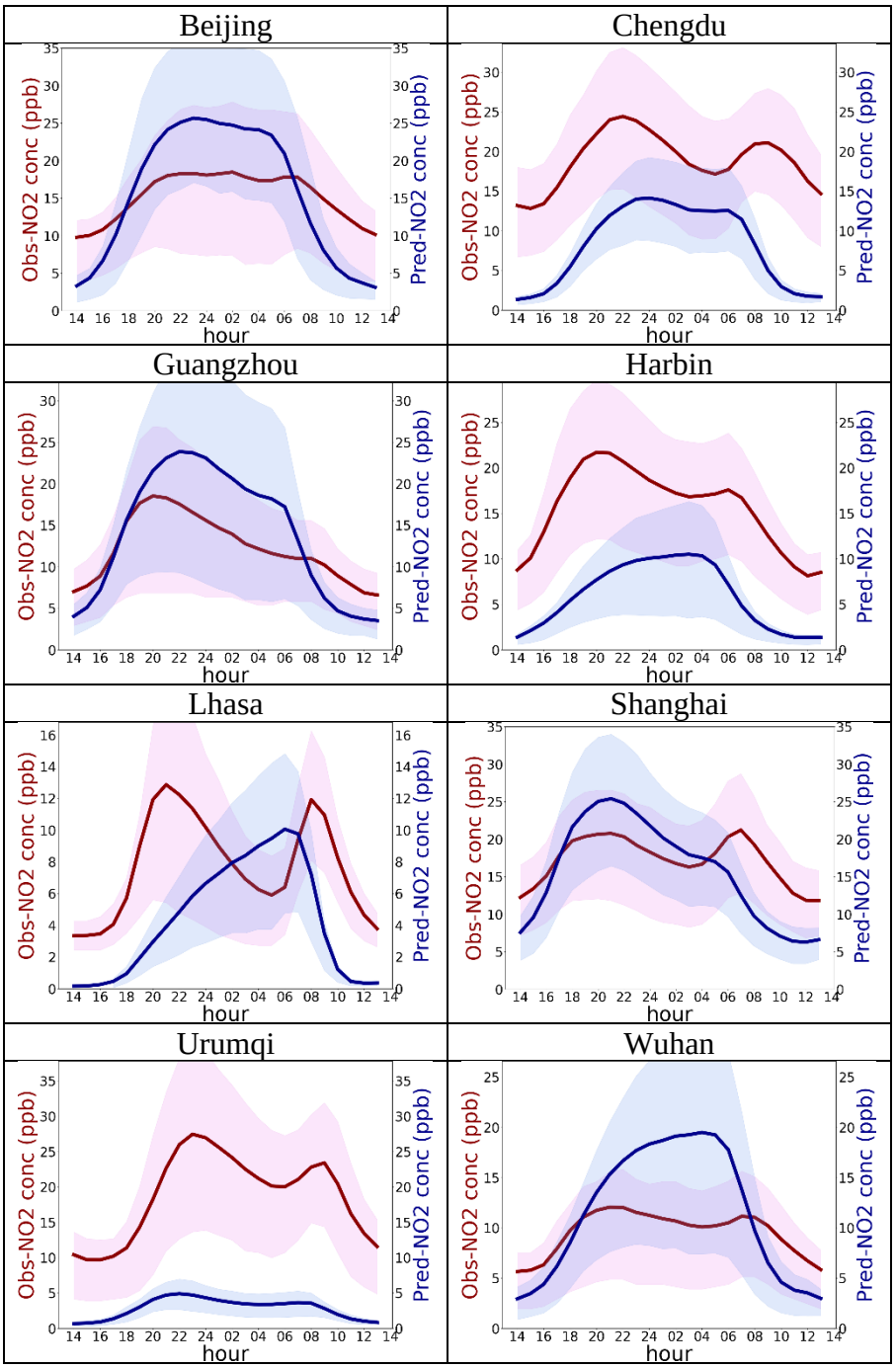


Figure S5. Comparison of diurnal variation of NO_2 in eight cities predicted with DeepSAT4D with the ground measurement (unit: ppb)

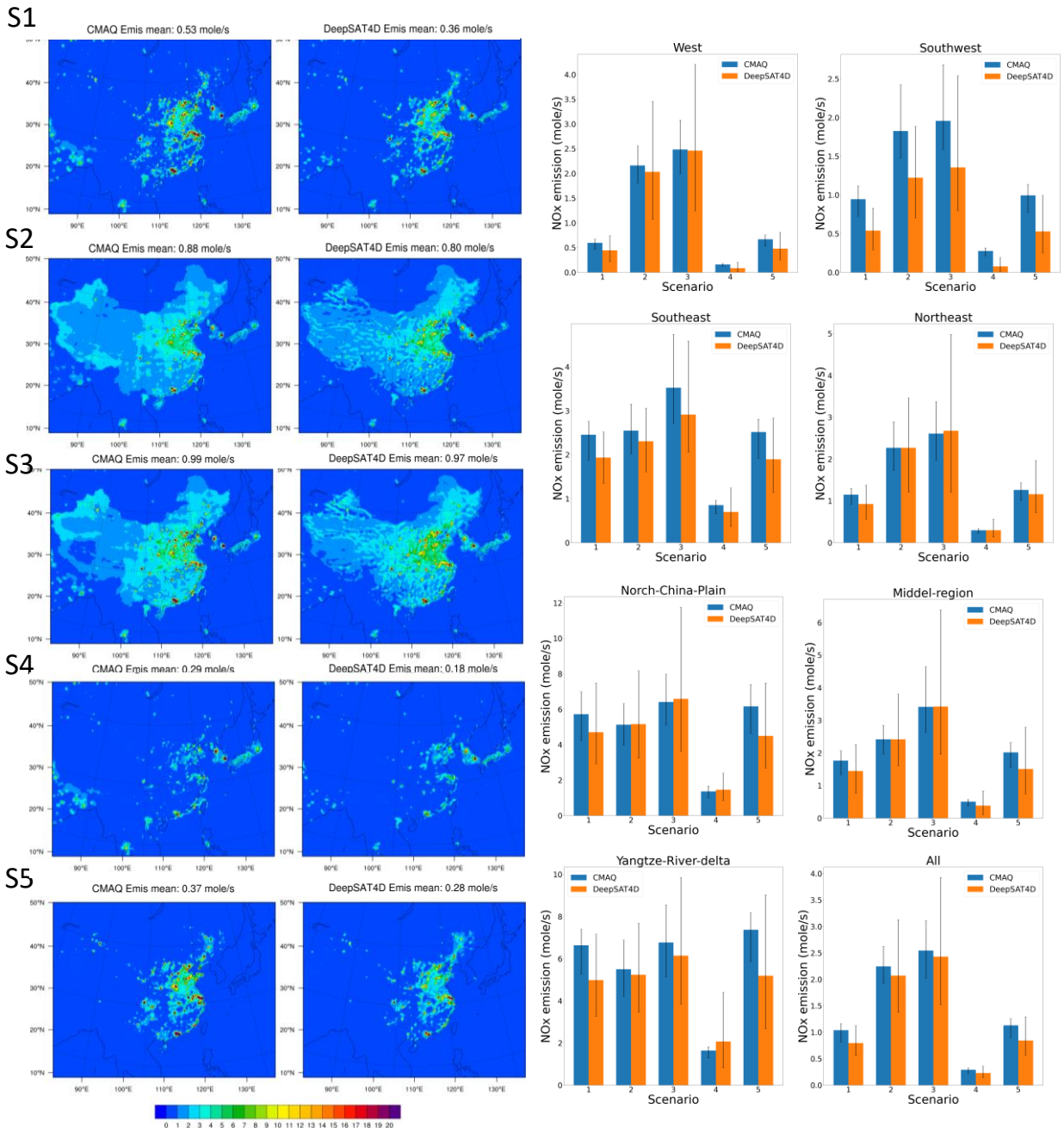


Figure S6 Comparison of DeepSAT4D reproduced emissions against with those used in CMAQ simulation across different scenarios (annual mean) with error bars representing the 5% and 95% percentiles across days

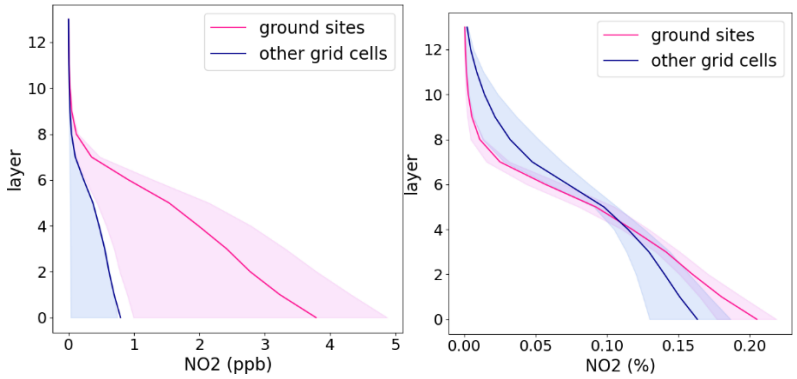


Figure S7 Comparison of vertical profiles of NO_2 concentration (left: absolute value; right: normalized) between CNEMC site location grid cells and others

(the presence of imbalanced samples from CNEMC grids may result in significant high biases when interpolating to rural areas, primarily due to the lack of vertical structure information.)

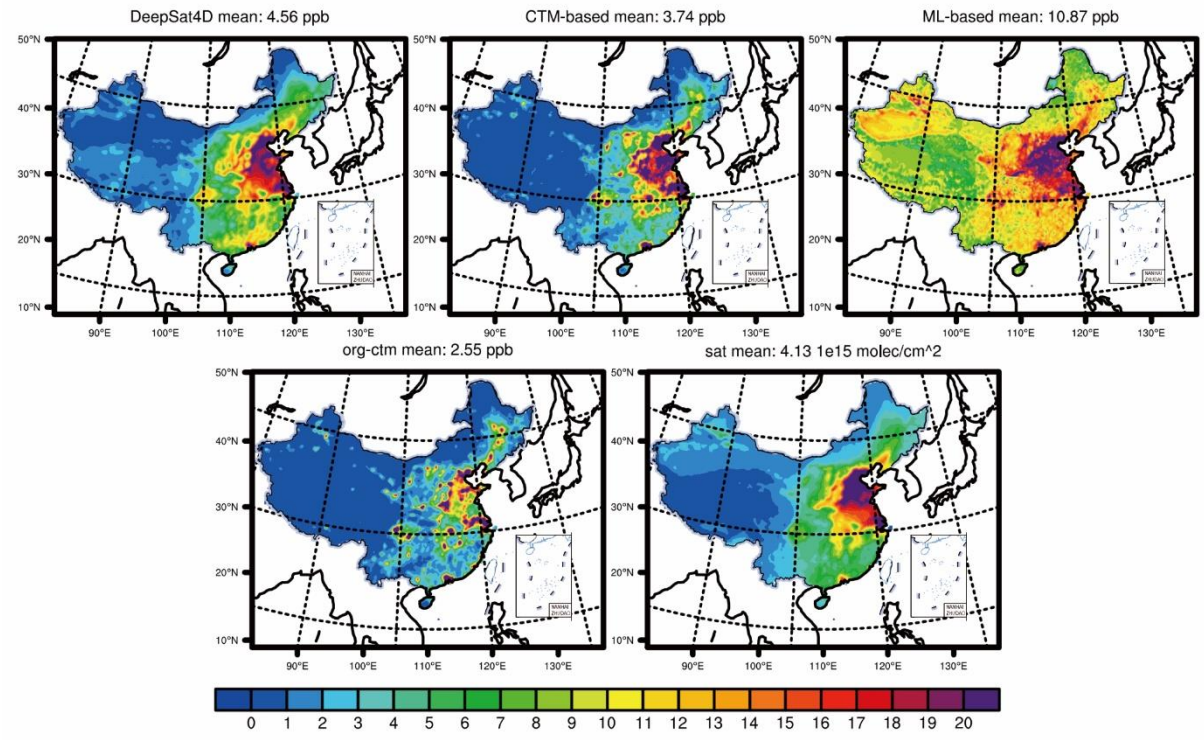


Figure S8 Comparison of ground-level NO_2 concentration between prediction from DeepSAT4D, traditional CTM-based (using CTM simulated ratio of xNO2 to ground NO2 applied to the OMI column for each grid cell) and ML-based method (lightGBM method as described in our previous study, Xing et al., 2020), as well as the original CMAQ simulation and OMI-NO_2 column density.

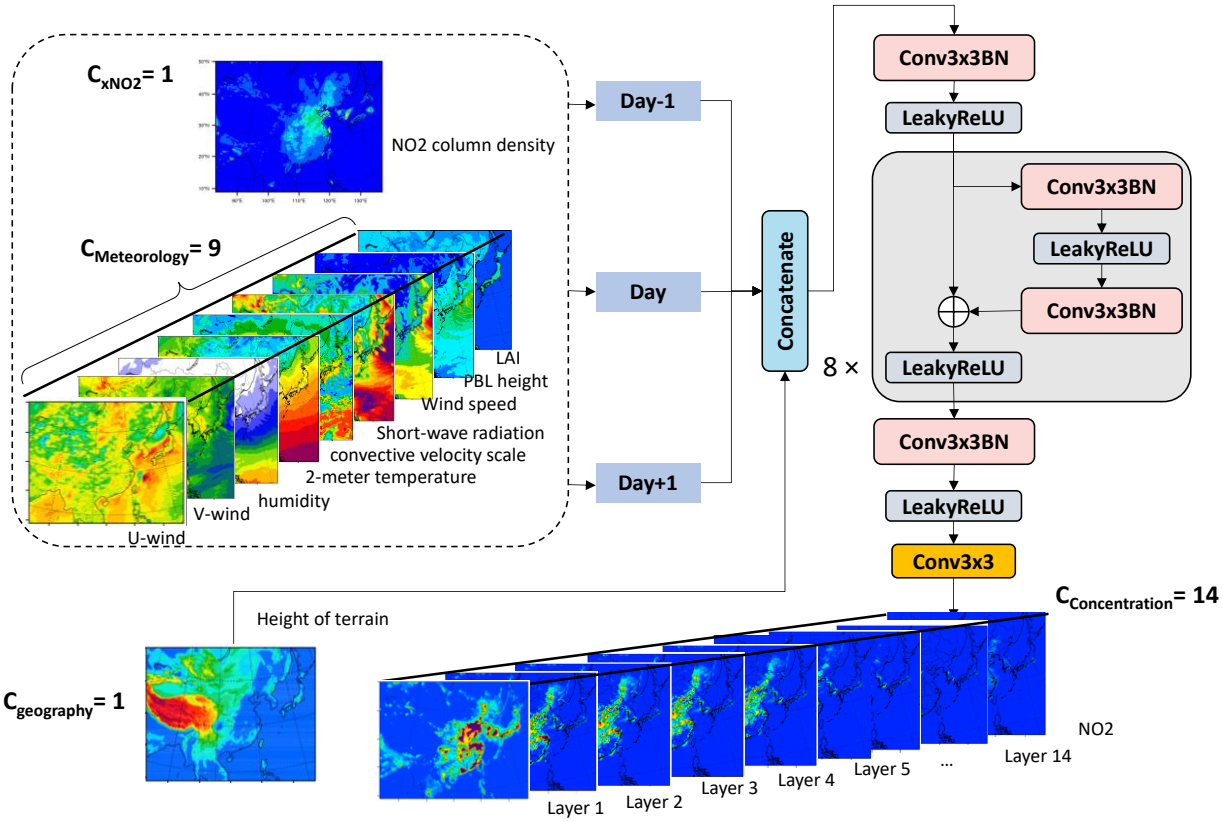
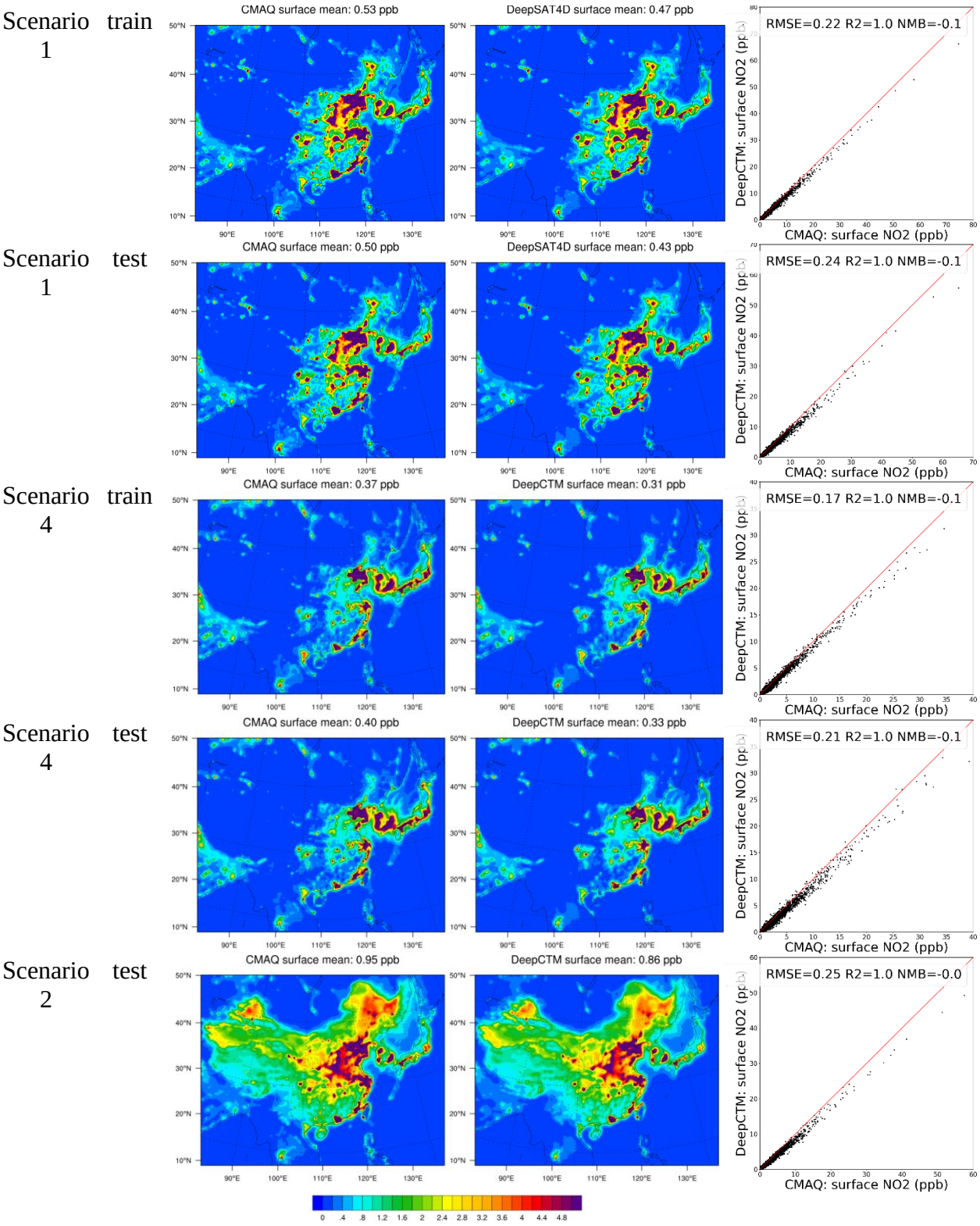
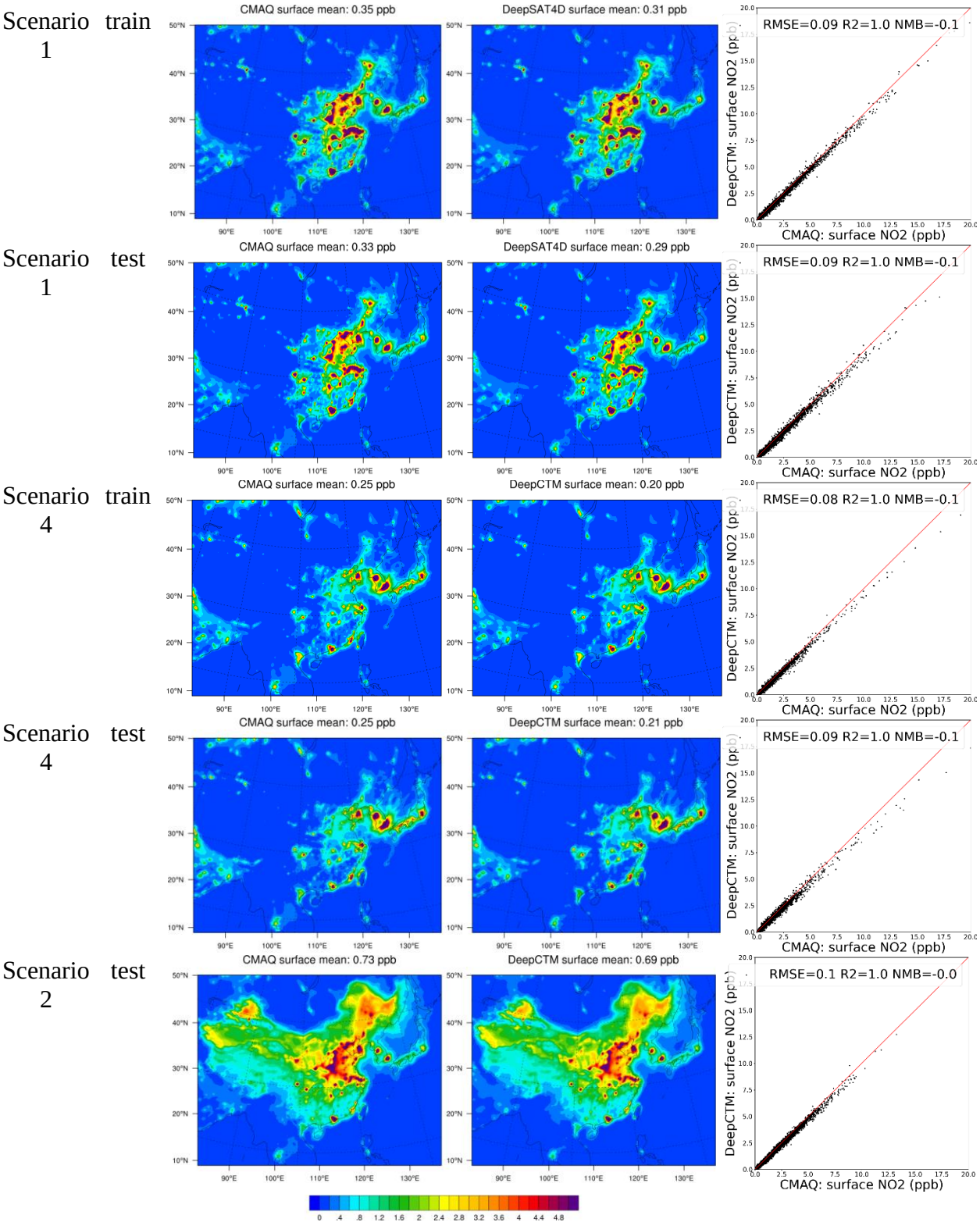


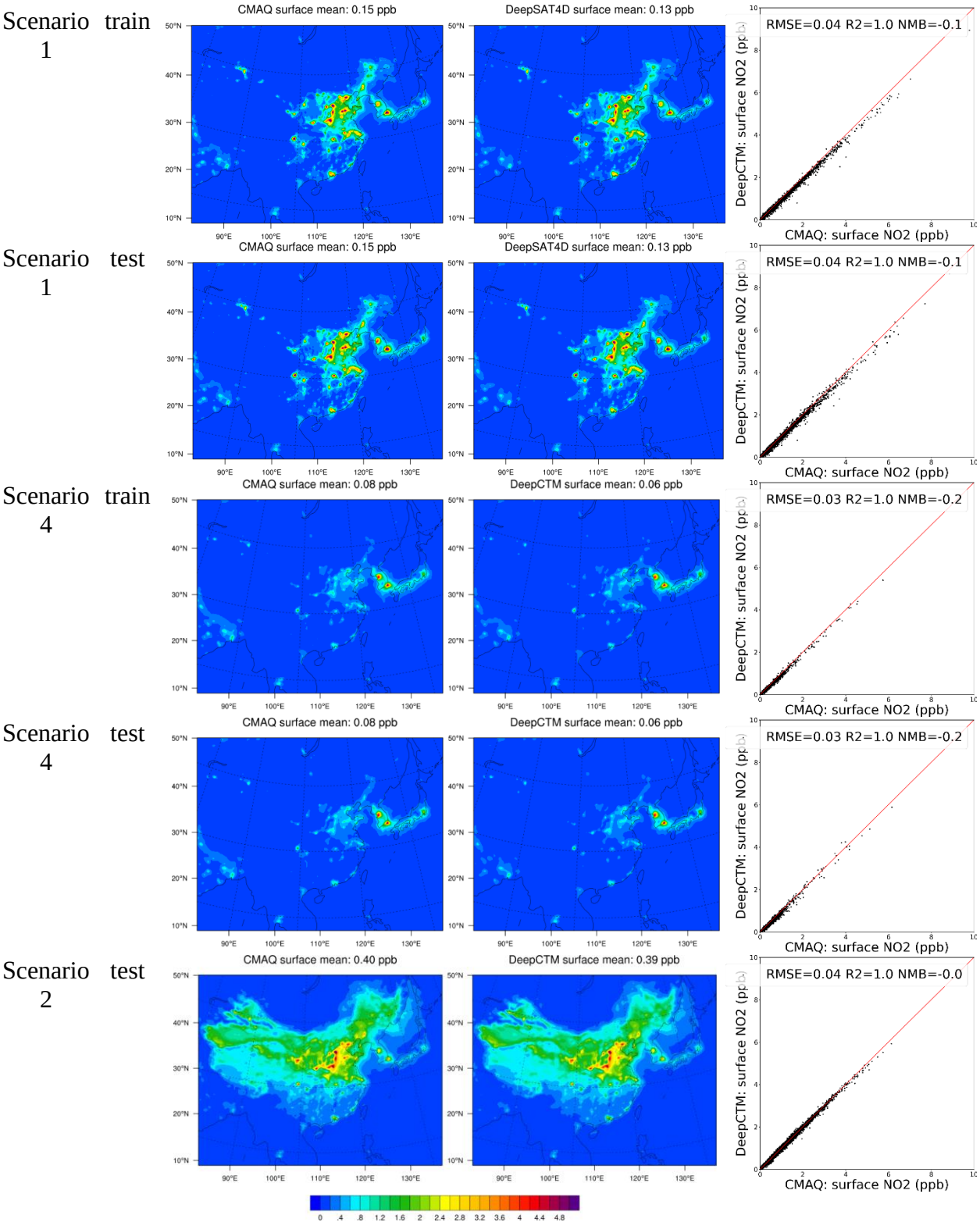
Figure S9 ResNet Model structure



(a) surface layer


Scenario train 1
CMAQ surface mean: 0.35 ppb
DeepSAT4D surface mean: 0.31 ppb
RMSE=0.09 R2=1.0 NMB=-0.1
Scenario test 1
CMAQ surface mean: 0.33 ppb
DeepSAT4D surface mean: 0.29 ppb
RMSE=0.09 R2=1.0 NMB=-0.1
Scenario train 4
CMAQ surface mean: 0.25 ppb
DeepCTM surface mean: 0.20 ppb
RMSE=0.08 R2=1.0 NMB=-0.1
Scenario test 4
CMAQ surface mean: 0.25 ppb
DeepCTM surface mean: 0.21 ppb
RMSE=0.09 R2=1.0 NMB=-0.1
Scenario test 2
CMAQ surface mean: 0.73 ppb
DeepCTM surface mean: 0.69 ppb
RMSE=0.1 R2=1.0 NMB=-0.0
DeepCTM: surface NO2 (ppb)
CMAQ: surface NO2 (ppb)
0 .4 .8 1.2 1.6 2 2.4 2.8 3.2 3.6 4 4.4 4.8

(b) Layer 3



(c) Layer 7

Figure S10 Performance of ResNet (all day averages)

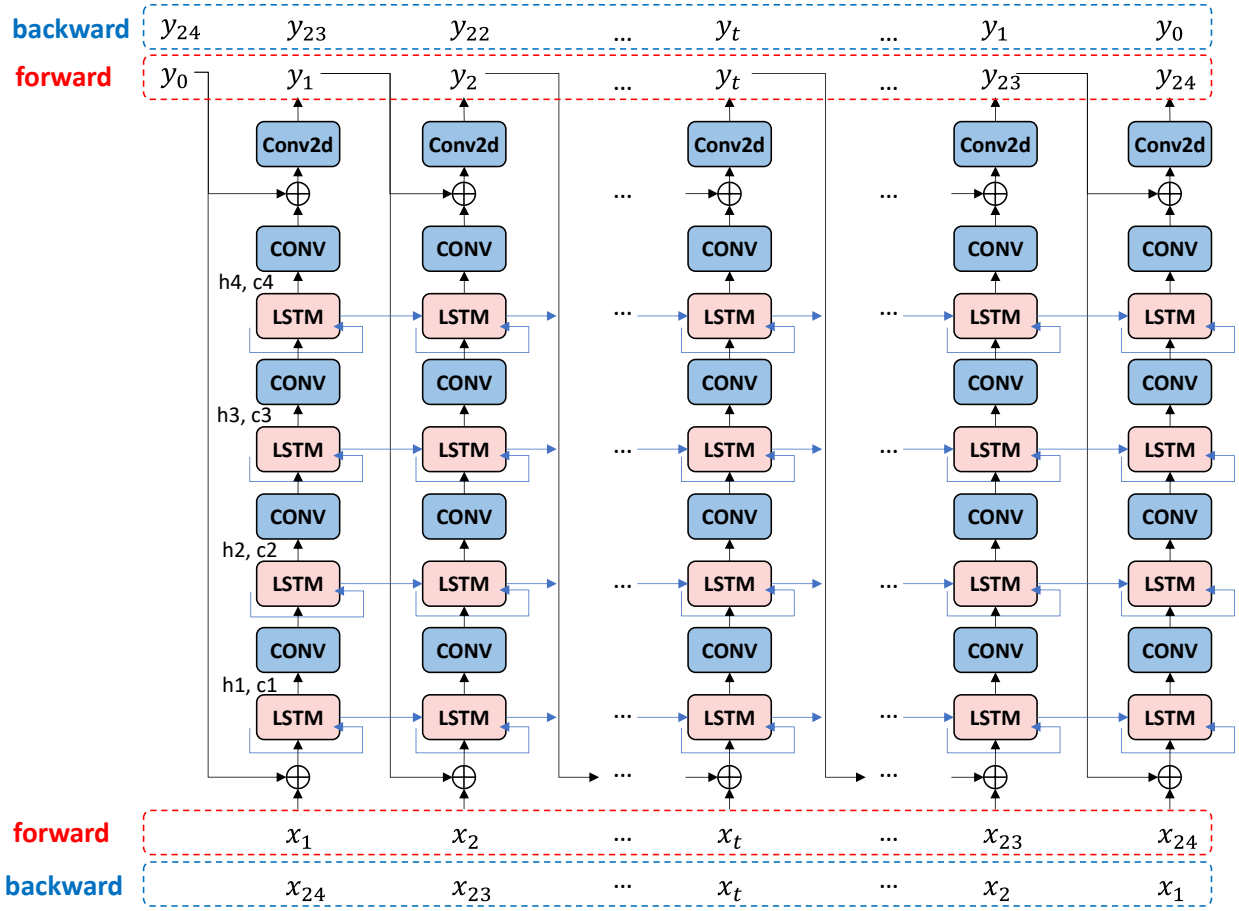
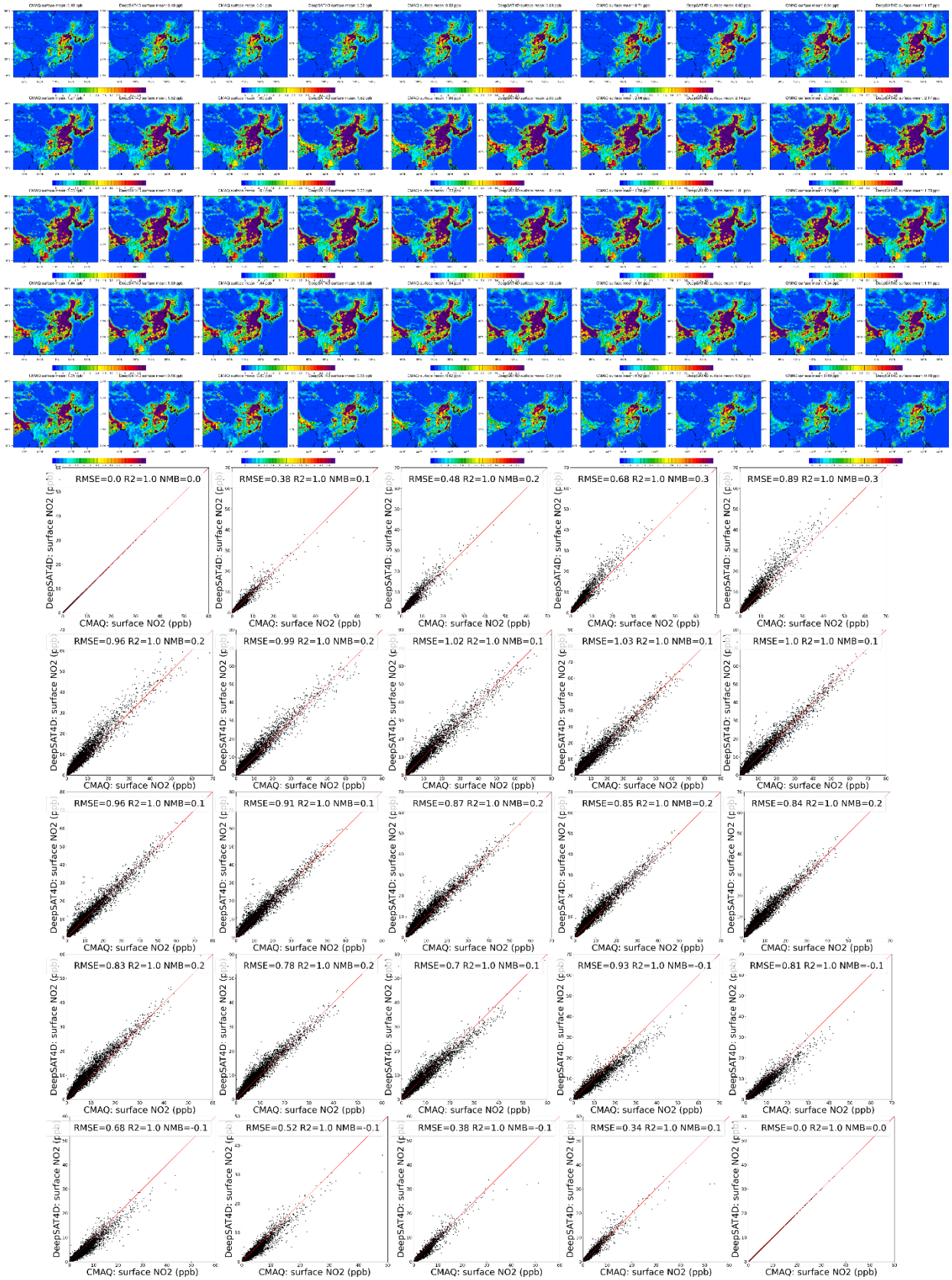
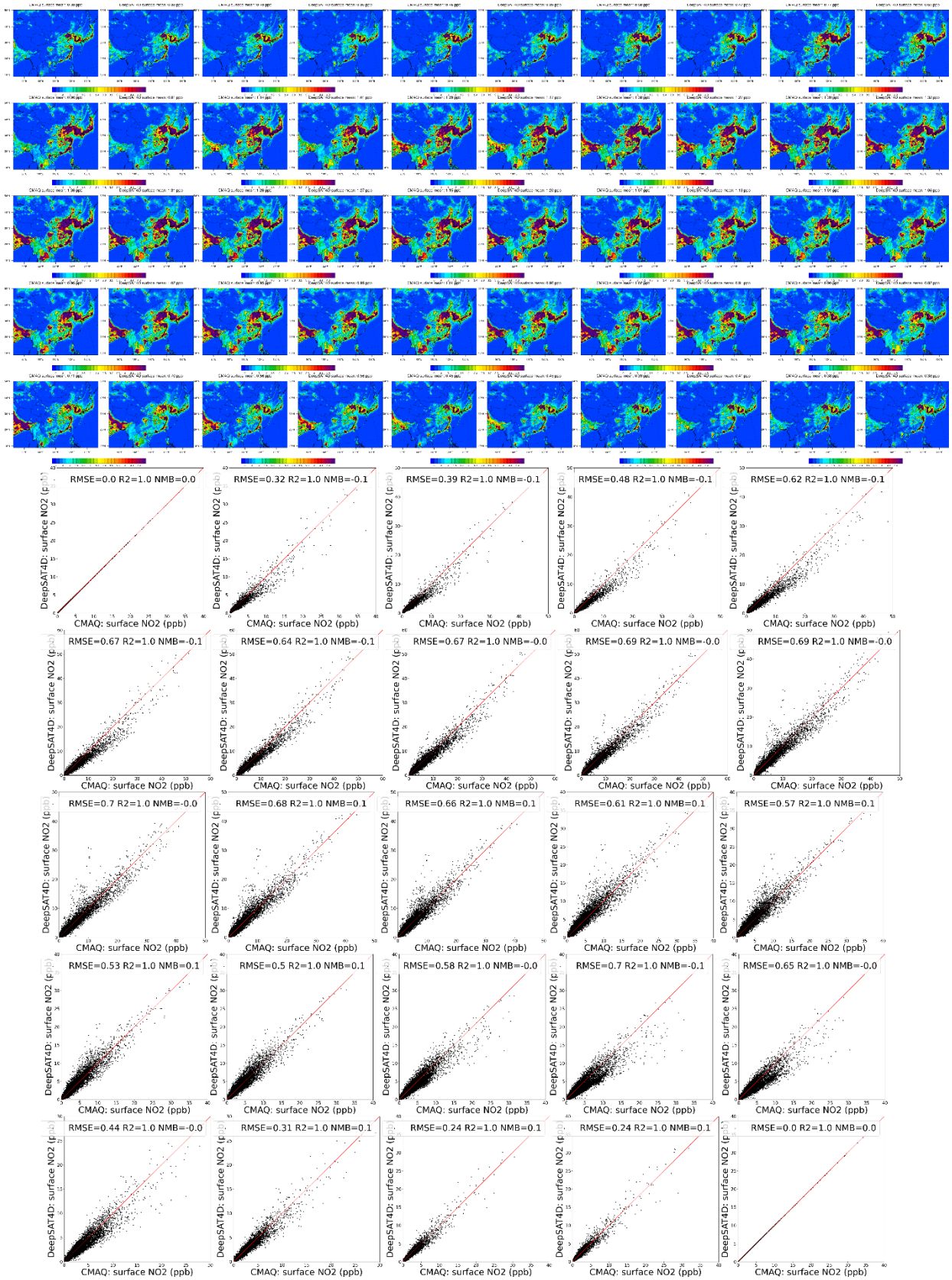


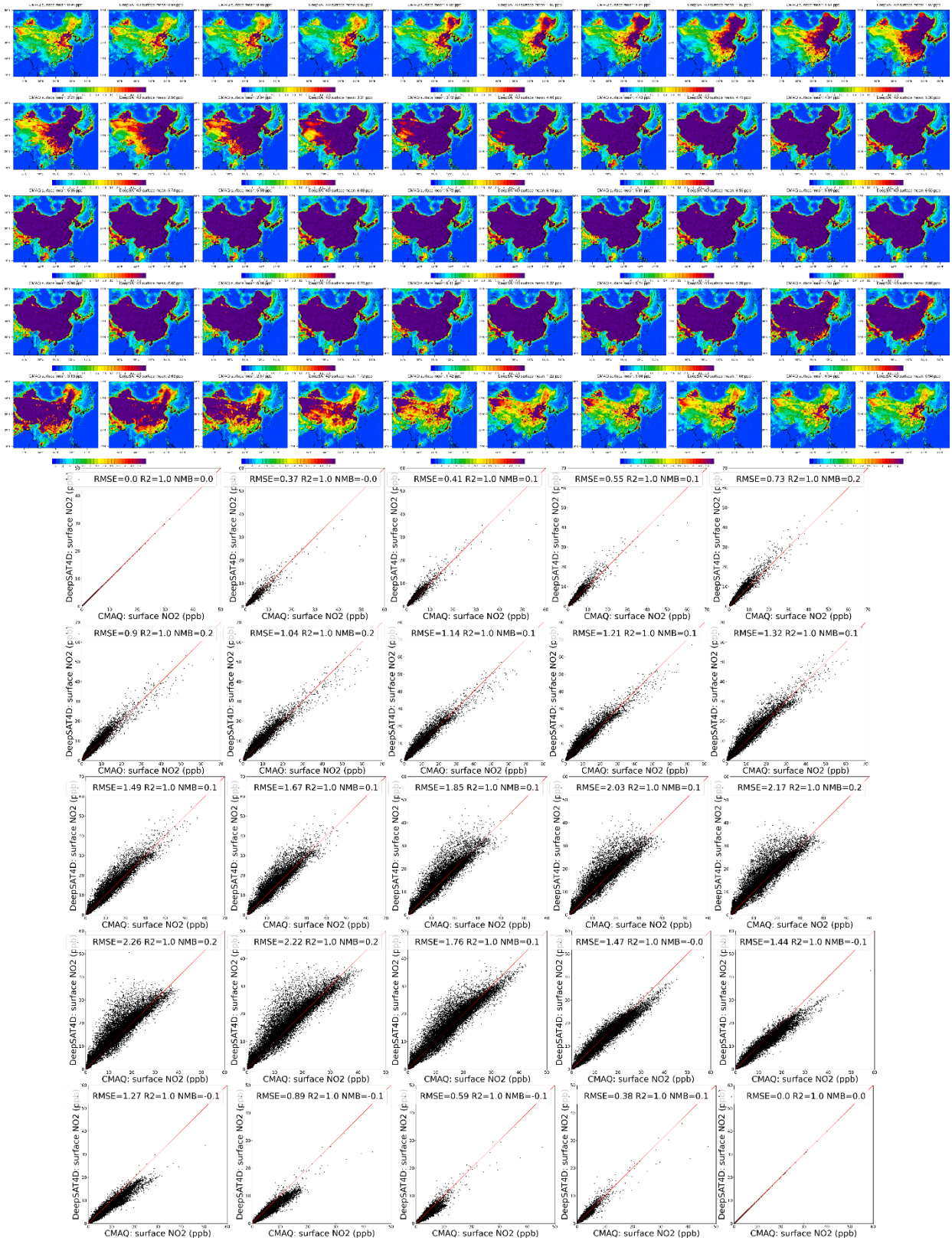
Figure S11 two-direction ConvLSTM model structure



(a) Scenario 1-test



(b) Scenario 4-test



(c) Scenario 2-test

Figure S12 performance of ConvLSTM

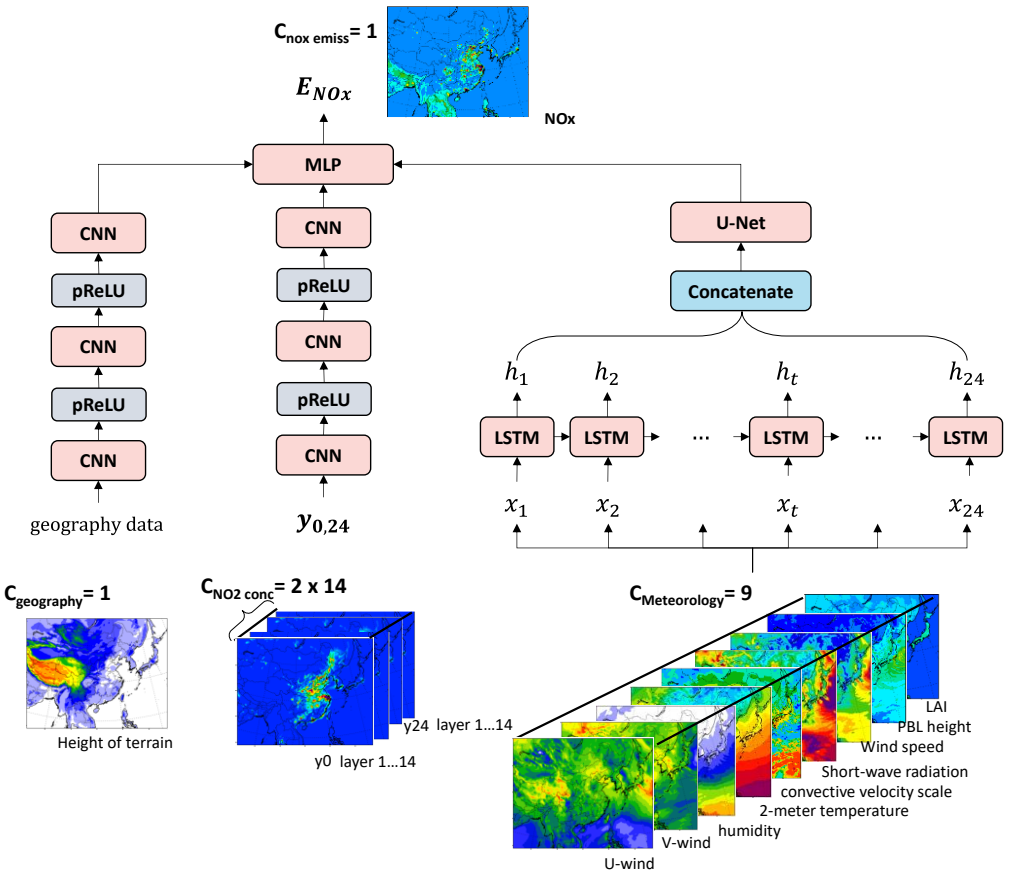


Figure S13 Variational auto-encoder (VAE) model structure

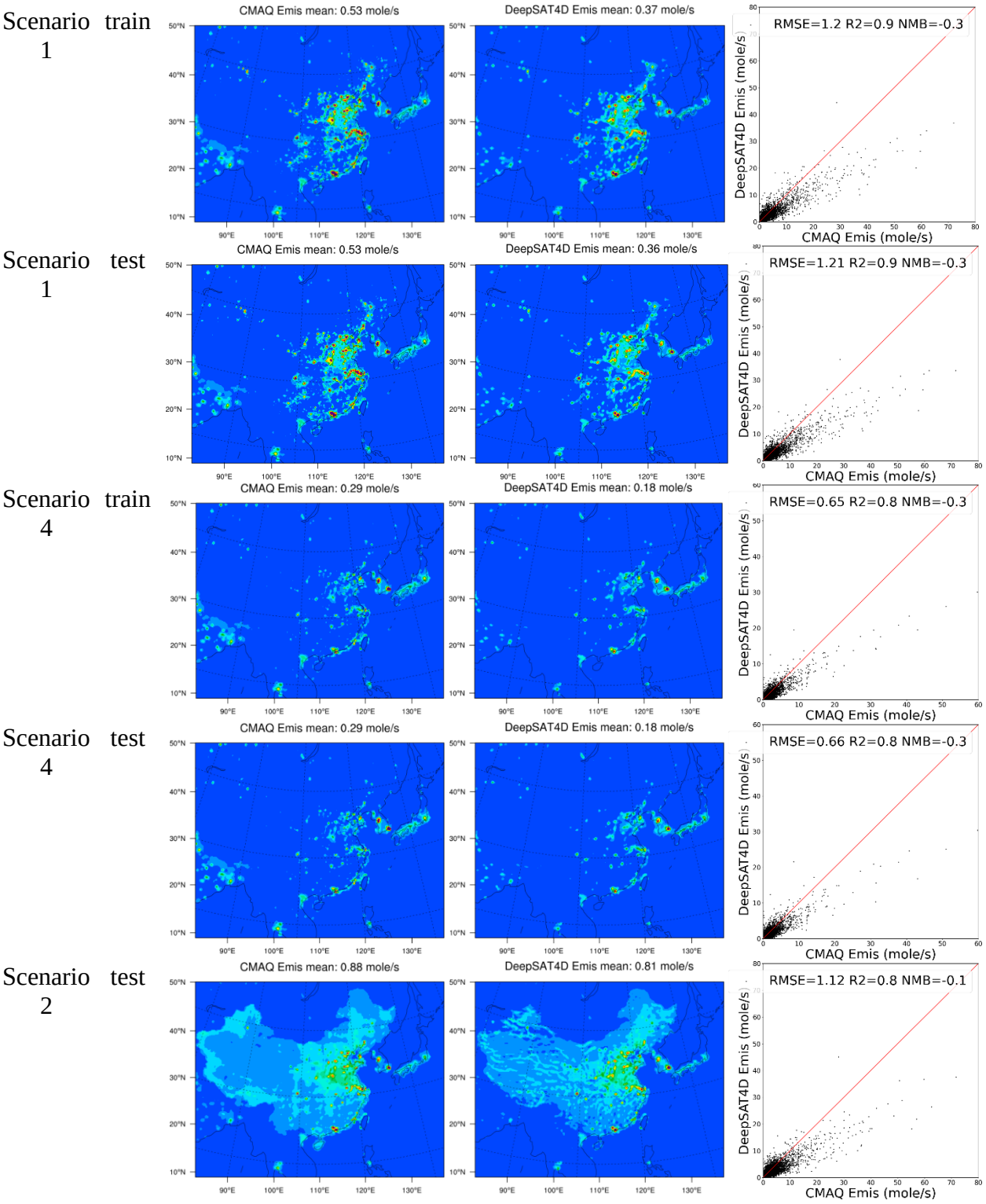


Figure S14 performance of UNet-LSTM in reproducing the emissions