Check for updates

AGU ADVANCING EARTH AND SPACE SCIENCE

Earth and Space Science

RESEARCH ARTICLE
10.1029/2022EA002338

Key Points:

- An end-to-end deep learning surrogate toward 3-D atmospheric transport process of pollutants was developed
- Continuous running of 400 and 1,000 steps were achieved for the 3-D consistency between the surrogate and the numerical benchmark
- Maximum of 164 speedup factors for computational efficiency were promoted via the neural network and Graphics Processing Unit architecture

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Z. Cheng,
chengz88@sjtu.edu.cn

Citation:

Xu, J. Z., Zhang, H. R., Cheng, Z., Liu, J. Y., Xu, Y. Y., & Wang, Y. C. (2022). Approximating three-dimensional (3-D) transport of atmospheric pollutants via deep learning. *Earth and Space Science*, *9*, e2022EA002338. https://doi.org/10.1029/2022EA002338

Received 22 MAR 2022
Accepted 16 JUN 2022

Author Contributions:

Conceptualization: J. Z. Xu, Z. Cheng, J. Y. Liu
Data curation: J. Z. Xu, H. R. Zhang, J. Y. Liu
Formal analysis: J. Z. Xu
Funding acquisition: Z. Cheng
Investigation: J. Z. Xu, Z. Cheng
Methodology: J. Z. Xu, H. R. Zhang, Z. Cheng, J. Y. Liu
Project Administration: Z. Cheng
Resources: J. Z. Xu, H. R. Zhang

© 2022 The Authors. Earth and Space Science published by Wiley Periodicals LLC on behalf of American Geophysical Union.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Approximating Three-Dimensional (3-D) Transport of Atmospheric Pollutants via Deep Learning

J. Z. Xu[1], H. R. Zhang[1], Z. Cheng[1], J. Y. Liu[1], Y. Y. Xu[2], and Y. C. Wang[3]

[1]Shanghai Environmental Protection Key Lab of Environmental Big Data and Intelligent Decision-making, School of Environmental Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, [2]MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai, China, [3]Network & Information Center, Shanghai Jiao Tong University, Shanghai, China

Abstract The physical transport process is the bottleneck of the computational efficiency in regional chemical transport modeling. The issue will be worse with the smaller time step due to increased iterations required with finer spatial resolution at scale. Reported surrogates of the transport process are usually unfeasible according to integrated assessment of efficiency promotion, long-term consistency, and spatial dimensions. This study intended to approximate the three-dimensional (3-D) transport process (including advection and diffusion) of a state-of-the-art chemical transport model, that is, Models 3/Community Multiscale Air Quality (CMAQ), via the U-Net structure of deep learning. Two temporal resolutions of models with 1-hr and 5-min were developed. Validation results indicated that single-step *R* squared of both models were higher than 0.9, and the lifetime for continuous running was 400 and 1,000 steps for 1-hr and 5-min model, respectively. Meanwhile, the computational efficiency can be promoted with the maximum of 164 times for 1-hr and 14 times for 5-min resolution on one GPU. The 1-hr deep learning surrogate could still achieve 12 times acceleration on the same CPU configurations of CMAQ, mainly through the end-to-end direct inferring rather than time step iterations. This study preliminarily proves the feasibility of the data-driven approach in approximating the 3-D complex transport process of atmospheric pollutants. Furthermore, computational efficiency can be efficiently improved while maintaining consistency and accuracy. Rapid transport simulation of different pollutants with wide concentration range can be expected, which will finally benefit the acceleration of whole chemical transport modeling.

1. Introduction

Chemical transport model (CTM) is a fundamental research tool for atmospheric environment and is widely used for air quality forecasting, source apportionment, and strategy design of pollution alleviation (Chuang et al., 2018; Shen et al., 2020; Zhang et al., 2014). Physical transport process is the driving force for primary pollutants, thus then affects the concentration of secondary pollutants (Byun & Schere, 2006; Tilt, 2019). It is usually numerically solved based on the Euler's mass continuity equation in CTM and occupies 57%~60% of the entire CTM's computation time (Colella & Woodward, 1984; Ying & Li, 2011; Zhang et al., 2013). Moreover, the exponential growth of computational cost is observed with finer spatial resolution due to the smaller time step and larger iteration steps required (Boffi et al., 2007; Kasim et al., 2020). The solution of physical transport process has become the critical bottleneck of the CTM's computation efficiency.

Machine learning, especially deep learning, has become a new research paradigm and acted as partial or whole replacement of complex geoscience models due to its excellent ability for non-linear fitting (Reichstein et al., 2019; Rolnick et al., 2019; Yuan et al., 2020). Specifically, for the surrogate of atmospheric transport process, Lauret et al. (2016) combined cellular automata and the artificial neural network (CA-ANN) to calculate the turbulence coefficient in horizontal two-dimensional (2-D) space. The R squared is over 0.7 for most testcases with the computation efficiency increases by ~1.5 times. Wang and Qian (2018) extended the study to three-dimensional (3-D) space with the same CA-ANN. *R* squared for a single time step vary between 0.2 and 0.95, and is reduced to almost zero after 100 multiple-time-step. The computation efficiency increased is still ~1.5 times. Vlasenko et al. (2021) emulated the entire CTM and the speed gain is 720 times of acceleration. Only 2-D horizontal distribution of daily average concentration, however, is concerned and the *R* squared values are only in the range of 0.38–0.67. The comprehensive performance with the aspects of efficiency promotion, long-term consistency, and spatial dimensions was still unsatisfied.

Software: J. Z. Xu, H. R. Zhang, J. Y. Liu
Supervision: Z. Cheng, Y. Y. Xu, Y. C. Wang
Validation: J. Z. Xu, H. R. Zhang
Visualization: J. Z. Xu, Z. Cheng
Writing – original draft: J. Z. Xu, Z. Cheng
Writing – review & editing: J. Z. Xu, Z. Cheng

This study proposes a deep learning surrogate for the 3-D atmospheric transport process, containing both advection and diffusion mechanisms. The data set was derived from the numerical solution for the transport process of a state-of-the-art CTM, and the features input into the U-Net neural network are determined accurately according to atmospheric transport theory. Validation results indicate that the data-driven deep learning surrogate has an excellent and stable ability under both 5-min and 1-hr time step, and the advantage of Graphics Processing Unit (GPU) hardware could help accelerate the numerical computational efficiency with 164 and 14 times for the two scenarios. The deep learning surrogate is expected to provide feasible ways for not only the forward solution for concentration of atmospheric pollutants, but also the inverse problems such as emission amount estimation derived from its natural property of automatic differentiation (Huang et al., 2021).

2. Model Development of the Deep Learning Surrogate

2.1. Collection and Pre-Processing of Data Set

The data set consisted of meteorological field, emission inventory, and concentrations of pollutant (shown in Figure 1). Meteorological field was estimated by the Weather Research and Forecasting (WRF) model (Skamarock & Klemp, 2008) with the initial conditions and observation data set from the National Centers for Environmental Prediction (NCEP) (Grell et al., 2005; NCEP, 2000, 2004a, 2004b; Otte & Pleim, 2010). Emission inventory was derived from the Multi-resolution Emission Inventory for China (MEIC v1.3, 0.25° × 0.25°, 2016, http://www.meicmodel.org) (Li et al., 2017; Zheng et al., 2018) and interpolated to the required 12 km *12 km horizontal resolution. A 3-D CTM numerical model of the Community Multiscale Air Quality (CMAQ) (U.S. EPA Office of Research and Development, 2019) was selected to simulate the concentration evolution with the meteorological field and emission inventory, as well as the benchmark for computation efficiency comparison for the deep learning surrogate. The CMAQ model was regarded as one of the state-of-the-art CTM and its scientific principles and algorithms could be referred to Binkowski and Roselle (2003). In this study, carbon monoxide (CO) was selected as the target pollutant because of its high level of concentration and negligible participation in atmospheric chemical reactions in urban atmosphere (shown in Figure S1 in Supporting Information S1). The modules of "CHEM" and "AERO," however, were both turned off during the running of the CMAQ model to assure only the transport process which included advection, diffusion and deposition are computed with the exclusion of chemistry process.

Two types of temporal resolution, that is, 1-hr and 5-min, were investigated, and the surrogate was developed and validated for each. The resolution of "1-hr" referred to the hourly average concentration and can be used for continuous simulation of inert pollutants only influenced by the transport process. Meanwhile, the resolution of "5-min" referred to the instantaneous concentration after 5 mins such that the surrogate of transport process could be coupled with the modules of chemical reaction and gas-particle partition smoothly for active pollutants. The geographical domain in both the CTM and the deep learning surrogate covered all regions of China and parts of southeast Asia with a 12 km grid resolution (512 rows × 512 columns, shown in Figure S1 in Supporting Information S1), as well as a total of 16 vertical layers with heights increasing from 0 to 19,308 m. A total of 24 days from 8 January to 1 February 2016 with the synchronization time step of 5-min was implemented for CMAQ simulation, and the hourly average concentrations and instantaneous concentrations every 5-min were then output and prepared for model training and evaluation. Therefore, there were 6,912 records for 5-min resolution and 576 records for 1-hr resolution in the data set. We selected 22 days as the training set and remaining 2 days as the validation set.

The input features for each record in the data set consisted of CO's initial concentration, emission amount of CO, wind speed in the horizontal direction and eddy diffusivity in vertical direction. The output target features were the concentration change of CO. The concentration change referred to the difference calculated by subtracting the initial concentration from output concentration after 5-min or 1-hr affected by the transport process. The concentration change was relatively low compared with the absolute concentration. Hence, the concentration alteration rather than the post-transport concentration was set as the output target feature to enable the deep learning surrogate to capture this change sensitively. Max-min normalization method was used for data preprocessing of meteorological features. The logarithmic function was used for data preprocessing of emission amount due to its extremely uneven spatial distribution (Ribeiro & Moniz, 2020).

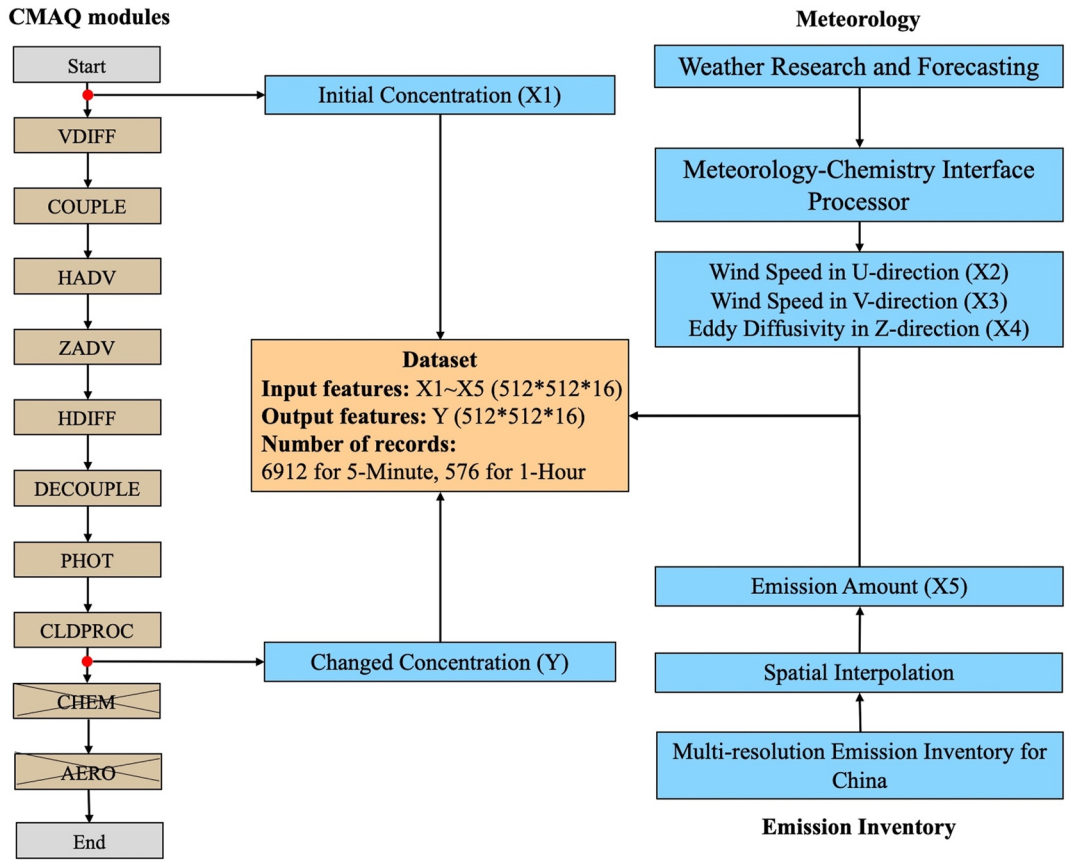

Figure 1. Flow chart of data set derivation for the deep learning surrogate. The brown part shows the science process sequence for the Community Multiscale Air Quality (CMAQ) chemical transport model (CCTM), the blue part shows the process of collection of concentration, meteorological data and emission, and the orange part is the composition of the data set. Two red points represent the breaking positions for data output.

2.2. Deep Learning Model Architecture

Since the physical transport process is mainly affected by instantaneous advection and diffusion conditions under single timestep and the integration of physical module and chemical module required to be conducted in one timestep, a deep U-shape convolutional neural network (U-Net) was selected to construct the deep learning model in this study (Dolz et al., 2018; Xing et al., 2022). U-Net used encoder, decoder and skip connections to extract image features while reducing information loss during encoding and decoding (Ronneberger et al., 2015). Figure 2 presents the specific configurations of U-Net model structures in this study. Five 3-D features with 16 vertical layers were expanded into 80 input channels, and the output was the spatial concentration change of 16 layers. The whole neural network included two sets of "down-sampling," "deconvolution," and "skip connection" operations. The configurations of internal model parameters were referred to the original U-Net structure (Ronneberger et al., 2015). A 1×1 convolution kernel was used in the last convolutional network layer before the output (Tompson et al., 2017), we excluded the activation function in the last layer as the change of concentration can be positive or negative in our case. Actually three U-Net structures with different depth were tested in this study (shown in Figure S2 in Supporting Information S1). The experimental results for three model structures indicated that the structure with two times of down-sampling obtained better performance than the other two structures according to the corresponding statistical index (shown in Figure S3 in Supporting Information S1). On the basis of U-Net-2, 3-D convolution kernel was also used for training, the R squared are 0.758 and 0.534 for 1-hr and 5-min model, and the root mean square error (RMSE) values are 6.081×10^{-3} and 0.833×10^{-4} for 1-hr and 5-min model, both worse than the case of 2-D convolution kernels (shown in Figure S4 in Supporting Information S1). The poor performance of the 3-D convolution kernel may be due to the characteristics of the convolution

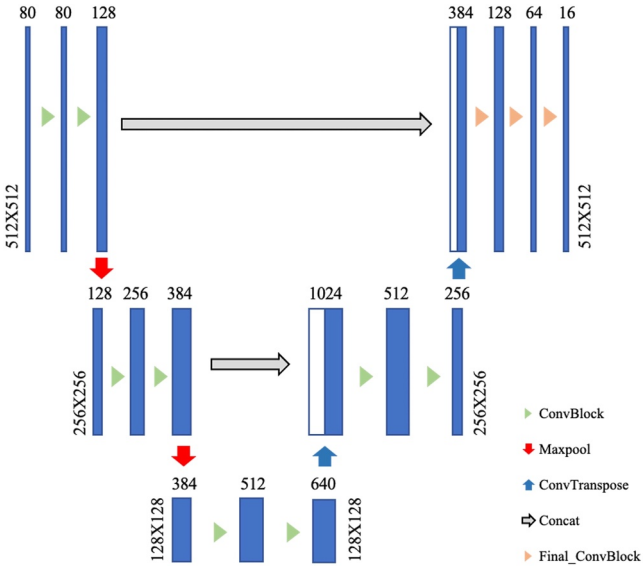

Figure 2. The architecture structure of the U-Net model used in this study. The U-Net neurol network has two times of down-sampling, deconvolution, and skip connection. The ConvBlock includes Conv2D 3 × 3, Batch normalization, ReLU, the Maxpool with a shape of 2x2, ConvTranspose includes ConvTranspose2D 3 × 3, and Final_ConvBlock have Conv2D 1 × 1 only.

kernel which will weaken the edge features and the current network structure of the 3-D convolution kernel. The detailed discussion can be found in Text S1 in Supporting Information S1.

A high-level neural network API of Pytorch (version 1.7.0) was used for the implementation of the U-Net model. All the model training and experiments were run in the Python (version 3.6.10) environment. The loss function and optimizer were Relative *L*2 error (R_{L2}) which could be calculated as Equation 1 (Li et al., 2020) and Adam (Kingma & Ba, 2014). The initial learning rate was 0.001, decreasing by a factor of 10 when the loss did not decrease. The batch size was set to 4, and the epoch size was set to 200. The remaining hyperparameters not described here were set to Pytorch default values. The loss trend for the surrogates and the numerator and denominator of the Relative *L*2 loss with the batch increases is shown in Figures S5 and S6 in Supporting Information S1.

$$R_{L2} = \frac{\|X_P - X_T\|_{L2}}{\|X_T\|_{L2}} \quad (1)$$

where X_T was the true value, X_P was the predicted value of the model.

To evaluate the acceleration ratios of the deep learning surrogate compared to the original CTM numerical model, the consuming clock time of the original CMAQ model and the U-Net model for the simulation of seven consecutive days on the CPU platform (4*40 cores) processor was compared. In addition, the computing time on the GPU platform was also collected for the deep learning surrogate. The specific hardware configurations in this study are listed in Table S1 in Supporting Information S1.

3. Model Validation of the Deep Learning Surrogate

3.1. Consistency Under a Single Time Step

The deep learning surrogate agreed well with the benchmark numerical model for the concentration change after a single time step. Concentration changes for the grids all over the domain were derived and compared (shown in Figure 3). *R* squared values of the 1-hr and 5-min models were both higher than 0.9. Meanwhile, the slope values of the fitting line for the two temporal resolutions reached as high as 0.9 with the intercepts close to zero. Several clusters of discrete points were noticed to be away from the 1:1 line for the case of 5-min, but not observed in the model of 1-hr. This phenomenon was reasonable as the concentration change was −0.3~0.3 ppm within 5 min, much lower than −2~2 ppm within 1 hr, making the deep learning surrogate more challenging to capture the tiny concentration changes of 5-min.

The consistency between the surrogate and benchmark was generally acceptable in vertical dimensional direction, especially for the height under the Planet Boundary Layer (PBL). *R* squared and RMSE were used for assessing consistency and plotted for each vertical layer (shown in Figure 4). Both two time-resolution models showed that *R* squared continues to decrease from >0.9 to <0.1 with the increase of vertical height, indicating that the surrogate has better performance in the surface and lower height. The *R* squared values at the twelfth layer in vertical dimension which corresponding to the top height of PBL (~3 km), however, could still stay at 0.5–0.6 for two resolutions. It then dropped dramatically to as low as 0.1 at the highest sixteenth layers with the height of 19 km. The pollutants inside the PBL are known to be well mixed while the pollutant's concentration above the PBL can be much lower which cause imbalance in the overall data (Kim et al., 2015; Lu et al., 2019), that will tremendously challenge the approximating ability of the deep learning surrogate (Buda et al., 2018; Liu et al., 2019). Furthermore, the mechanisms of the diffusion process above PBL such as the long-range transport were also more complex and different from those in the lower layers (Hu et al., 2010; Wagstrom & Pandis, 2011). It was interesting that the continuous decrease of RMSE value with higher layer seemed a "paradox" to the decrease of *R* squared. The trend of RMSE should be mainly attributed to the smaller range of concentration change with height increase rather than representing better performance under a higher layer. Overall, the deep

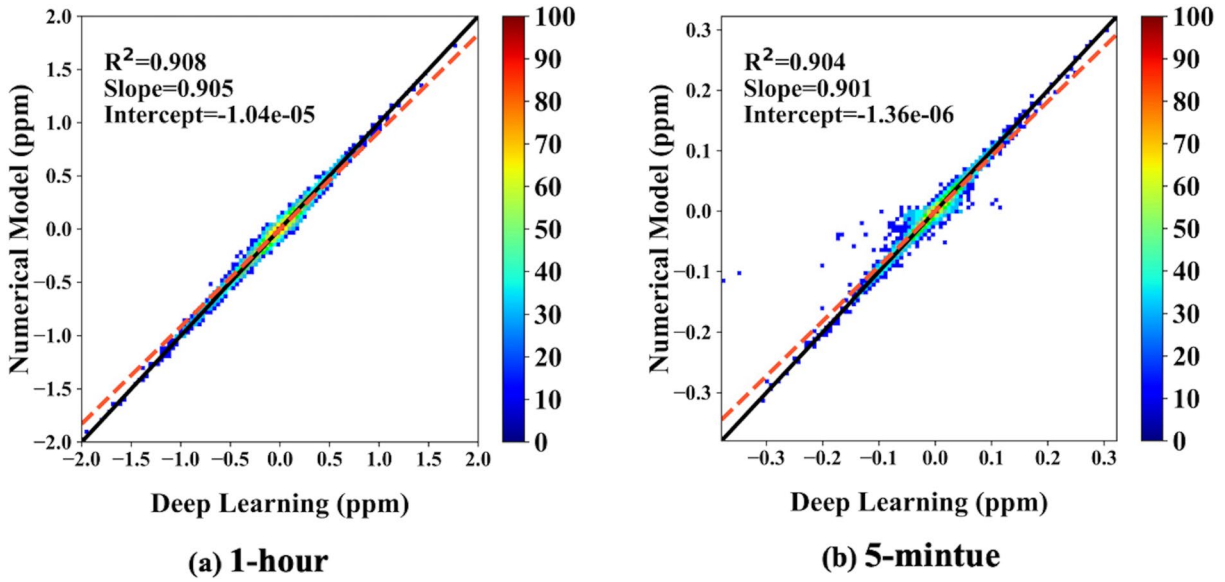

Figure 3. Density scatterplots of concentrations change predicted by the deep learning surrogate (*x*-axis) versus the benchmark provided by the numerical model (*y*-axis). The black solid line is the 1:1 line, and the red line is the linear best-fit line through the data. The total number of points in each plot is 83.88 million and the legend marked "100" represents 8.38 million points.

learning surrogate seemed more sensitive to the high concentration changes and the high emission amount in the lower layers. Accurate approximation under a single time step, especially for the height under PBL, provided the prerequisite for the deep learning surrogate to make reasonable continuous predictions.

Comparison with the reported index of previous studies manifested the better performance obtained in this study. Lauret et al. (2016) used the CA-ANN structure to simulate the 2-D atmospheric dispersion, with the *R* squared of 0.7 and slope of 0.84. Wang and Qian (2018) extend the mentioned model to 3-D application, with *R* squared ranging from 0.3 to 0.9 with the slope of 0.91–1.22 under different scenarios. Vlasenko et al. (2021) approximate the 2-D daily average concentration also derived from the CMAQ numerical model. Their *R* squared values were in the range of 0.38–0.67. The quality of the training data set and refinement of input features were regarded as the potential reasons for the better performance in this study. First, the data set was derived from the CMAQ numerical model, which is calculated according to strict theoretical equations without any noisy influence. Second, the

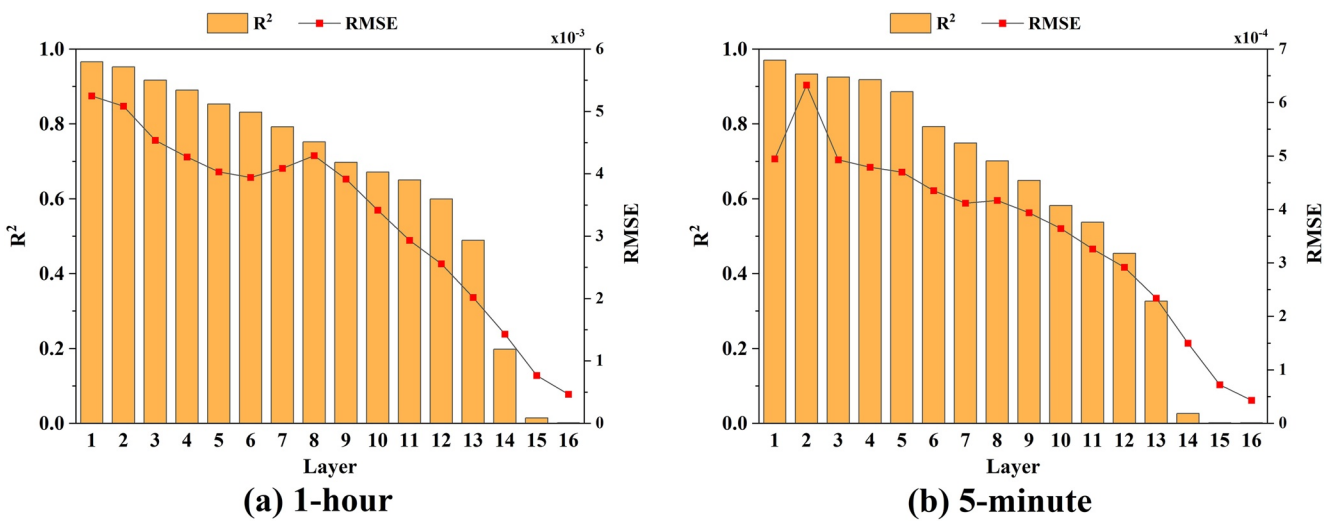

Figure 4. Statistical indicators (*R* squared and root mean square error [RMSE]) for a single time step between the deep learning surrogate and the numerical model at different vertical layers. The yellow bar represents *R* squared while the line with red rectangle represents RMSE value.

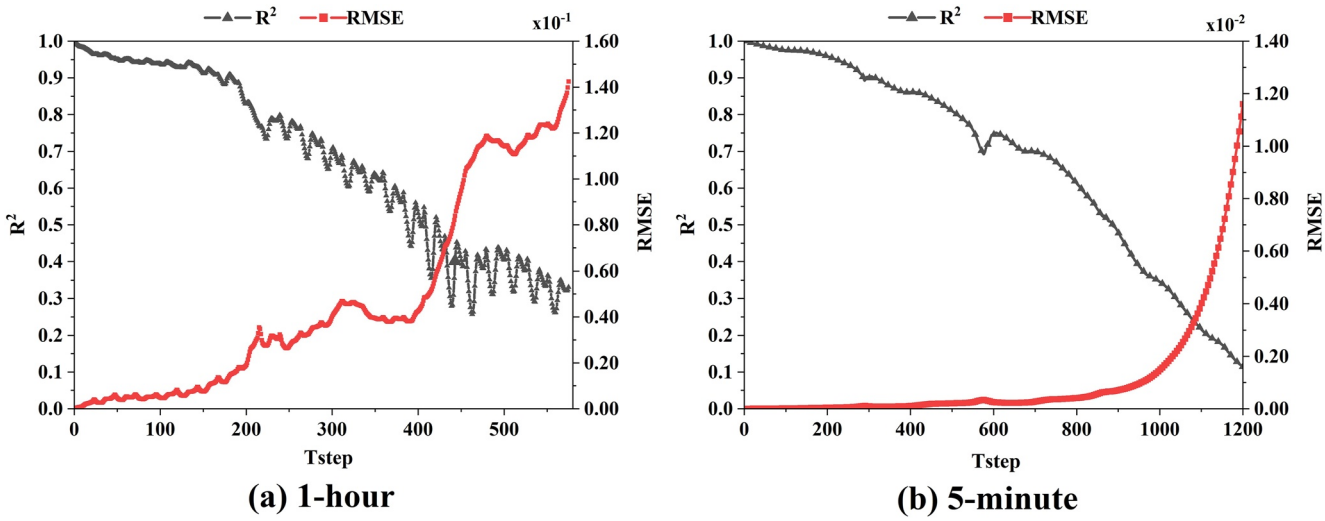

Figure 5. *R* squared and root mean square error (RMSE) evolution between the deep learning surrogate and the numerical benchmark for 1-hr and 5-min time resolutions. The line with gray triangle represents *R* squared while the line with red rectangle represents RMSE value.

vertical eddy diffusivity, which was a key parameter for the diffusion process (Marrouf et al., 2015), was set as the input features to help the U-Net model to learn the transport process accurately. This could be verified as we also compared the performance with the replacement of "eddy diffusivity" with more primary meteorological variables such as atmospheric pressure, temperature, humidity, etc., It was observed that the *R* squared values of primary meteorological variables were close or lower than those of single eddy diffusivity and the number of continuous running steps of primary meteorological variables were less than those of single eddy diffusivity (shown in Figure S7 in Supporting Information S1). At the same time, we consider the influence of boundary conditions and LULC on the model. Adding boundary conditions and adding LULC as an input channel has little effect on the overall model, while adding input channel will increase the computational burden definitely (shown in Figures S8 and S9 and Text S2 in Supporting Information S1).

3.2. Consistency Under Continuous Running

Model validation for multiple-time-step was essential for the long-term application of the deep learning surrogate. The initial concentrations at each time step are the sum of the initial concentration at the last time step and the concentration change generated by the deep learning surrogate. The continuous iterations of 400 (~17 days) and 1,000 (~83 hr) time step could be regarded as the lifetime for the resolution of 1-hr and 5-min, respectively, according to the comprehensive performance of two statistical indexes. Indicators of *R* squared and RMSE referred to the absolute ambient concentration for the grids all over the domain between the surrogate and the benchmark were estimated (shown in Figure 5). The values of *R* squared for the two resolutions decreased continuously due to the accumulation of errors in multiple-time-step. It reduced to 0.5 at the 400th step and 0.3 at the final 576th step for 1-hr resolution. Similarly, the *R* squared reduced to 0.4 at the 1,000th step and 0.1 at the final 1,200th step for 5-min resolution. Meanwhile, inflection points at the 400th step and 1,000th step for the 1-hr and 5-min resolution, respectively were observed with the evolution of RMSE values. The statistical result based on the concentration change during continuous running was also shown in Figure S10 in Supporting Information S1. The result for concentration change was generally agreed with that of the absolute concentration. The error of the 1-hr surrogate was dominantly originated from the internal bias of individual time step, while the error of the 5-min surrogate was significantly influenced by the bias accumulation of multiple steps.

The statistical performance of different vertical layers, however, presented significant diversities. Figure S11 in Supporting Information S1 selects five representative vertical layers corresponding to the height of below PBL (Layers 1, 4, and 8), PBL boundary (Layer 12) and above PBL (Layer 16). For the 1-hr deep learning surrogate, the *R* square values of the lower height (Layers 1 and 4) could maintain at ~0.6 even after 576 consecutive steps. In contrast, the statistical performance of upper height (Layers 12 and 16) was much worse than that of layers under PBL. The sudden change of RMSE from Layer 12 at the 400th step should be responsible for the inflection

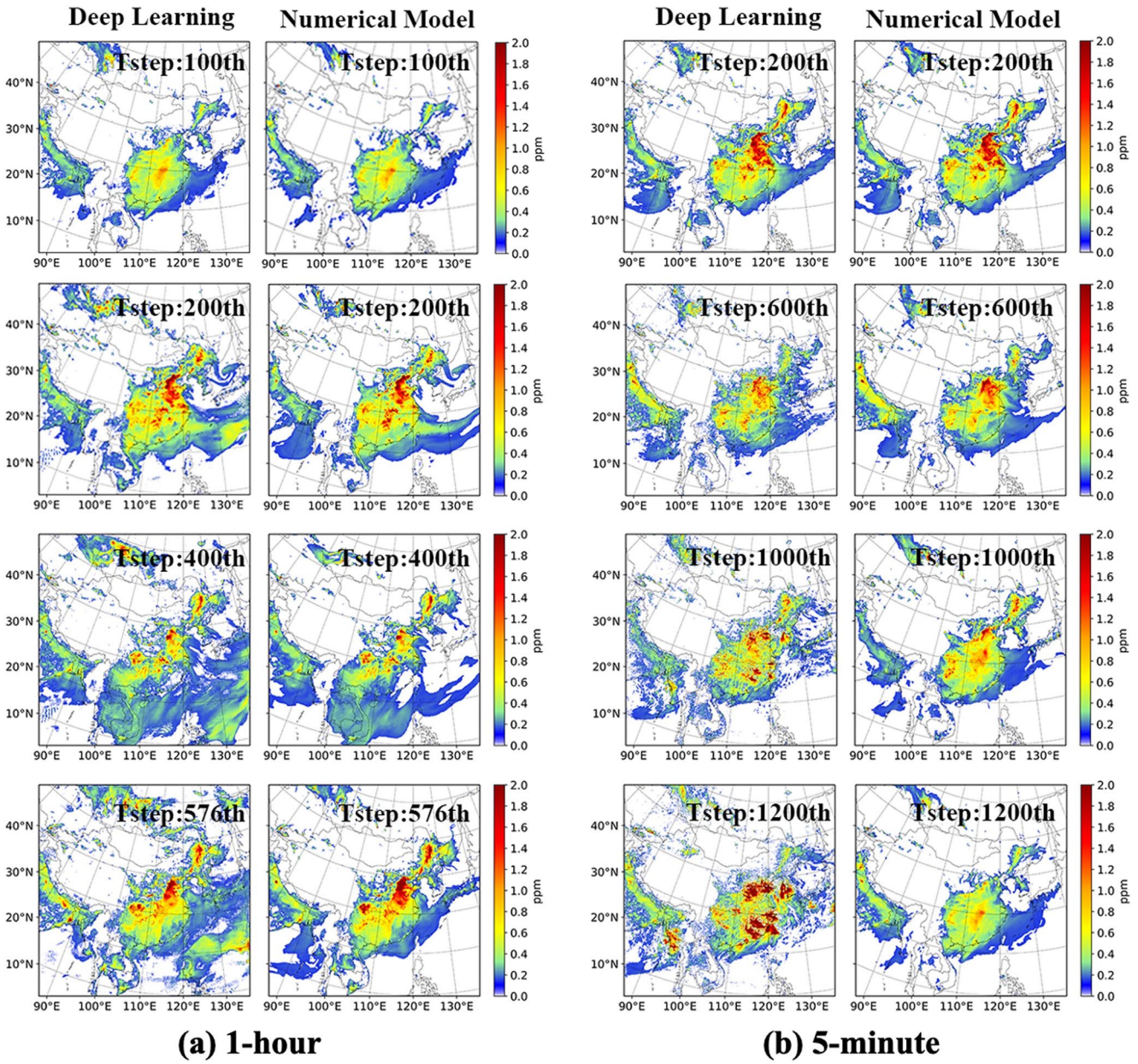

Figure 6. Concentration distribution of the vertical first layer for the deep learning surrogate and numerical model during continuous running. The 100th step, 200th step, 400th step, and 576th step were selected for the 1-hr time resolution, while the 200th step, 600th step, 1,000th step, and 1,200th step were selected for the 5-min time resolution.

point of Figure 5a as the RMSEs for other layers raised slowly and smoothly. The trend for the 5-min surrogate was similar to that of 1-hr, while the only difference was that the cause of RMSE inflection was attributed to all vertical layers rather than some single layer. The vertical performance of multiple-time-step was dominantly consistent with that of a single time step. However, the surface layer was what we cared most considering of health and ecosystem, and fortunately it had the best performance throughout all the vertical layers.

Spatial distribution could help to identify the deviation sources for the surrogates with two resolutions. Figure 6 shows the concentration comparison of spatial distribution in the surface (i.e., Layer 1) for 1-hr and 5-min resolutions, and the results of other vertical layers are presented in Figure S12 in Supporting Information S1. The spatial distribution of the surrogate could be observed to be highly consistent with the numerical benchmark until the 576th step for 1-hr resolution, especially for the hotspots area with high concentration. The deviation area mainly

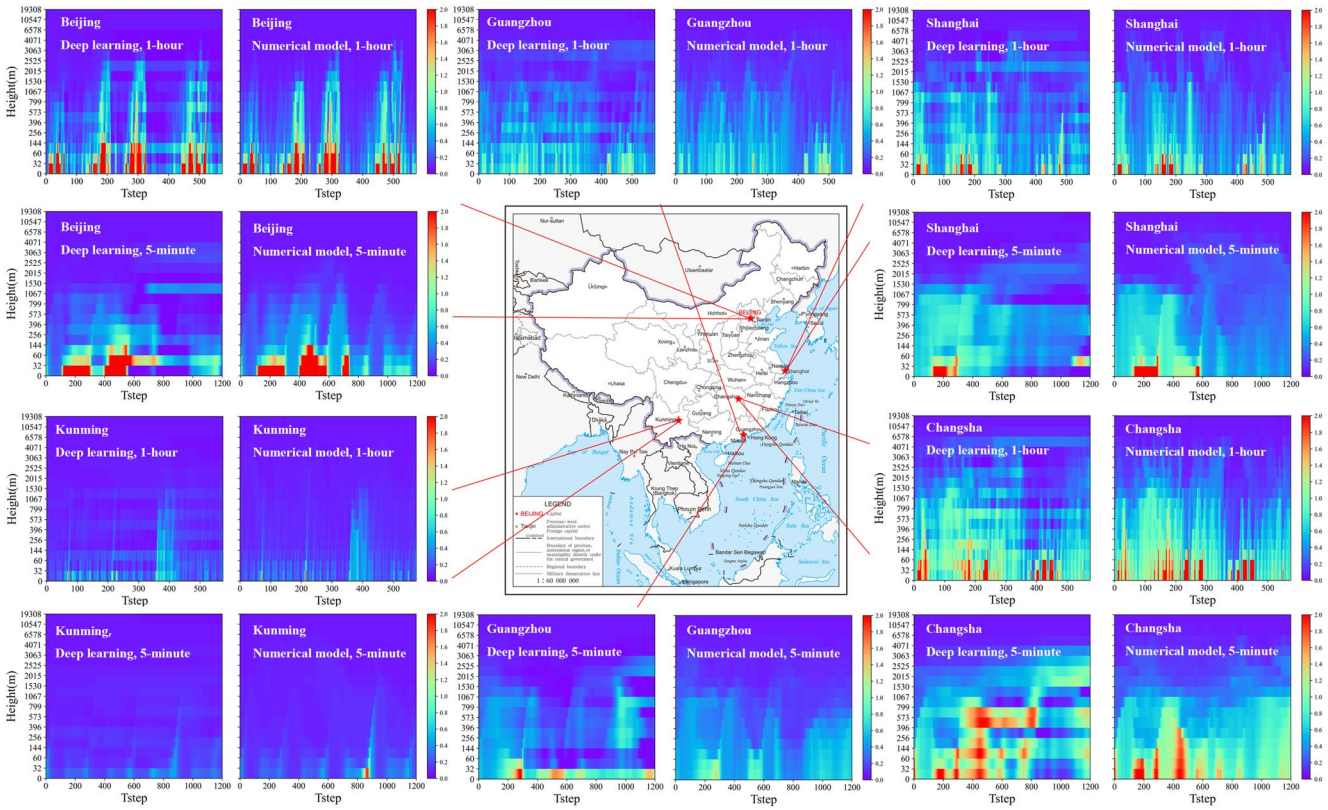

Figure 7. Concentration evolution of vertical distribution at different cities for the deep learning surrogate and numerical benchmark. The selected cities include Beijing, Shanghai, Guangzhou, Kunming, Changsha. The city name, model type and time resolution are labeled in each sub plot.

came from the offshore regions with low concentration, that is, the south-east part of the domain, which could be observed more clearly from the difference spatial distribution given by Figure S13 in Supporting Information S1. The surrogate might overestimate as high as 0.3 ppm at the sea area after the 200th step, but very limited underestimation was seen at the mainland until the final 576th step. In contrast, the surrogate for the 5-min resolution was mainly influenced by the deviation sources from the hotspots of mainland in China. Limited underestimation was seen until the 600th step. However, both significant overestimation and underestimation were existed at the hotspots of mainland in China from the 1,000th step, resulting in the overall worse performance compared with that of 1-hr resolution.

The temporal concentration evolution of vertical distribution was suitable for assessment of the 3-D surrogate in this study. It is difficult to present the vertical evolution for all grids, hence five cities all over China, that is, Beijing (north), Shanghai (east), Kunming (west), Changsha (middle) and Guangzhou (south) were selected. These five cities could also represent different pollution levels. The evolution of vertical concentrations was seen to be consistent well between the deep learning surrogate and numerical benchmark for both two resolutions (shown in Figure 7). The diurnal variation of concentration caused by the temporal change of PBL height, which was demonstrated by the numerical CMAQ model, could also be approximated almost perfectly by the deep learning surrogate for all cities. Relatively speaking, the performance of 1-hr resolution was better than that of 5-min resolution. For instance, there was a false hotspot pollution at the middle height during the 400th −800th step for Changsha city indicated by the surrogate, not seen in the benchmark numerical model.

The performance for multiple-time-step was always a bottleneck of deep learning surrogate from the reported studies. Wang and Qian (2018) tested one hundred different scenarios, and the *R* squared values dropped to zero at the 100th consecutive time steps for most scenarios, much smaller to the lifetime of 400 and 1,000 timesteps for the two resolutions in this study. Multiple-time-step validation has not even been investigated in the other two studies (Lauret et al., 2016; Vlasenko et al., 2021). In summary, the deep learning surrogate in this study could

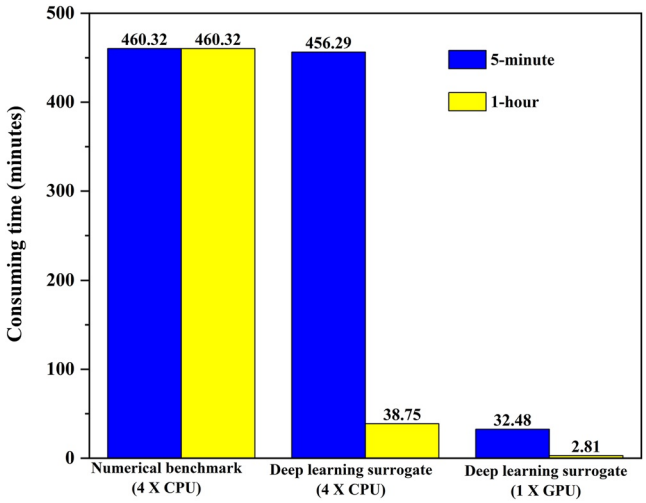

Figure 8. Comparison of consuming time for the numerical benchmark and deep learning surrogate under different time resolutions and hardware configurations. The numbers in the top of each bar represents the specific consuming minutes.

be generally implemented for a days-to-weeks application, especially for the area with notable anthropogenic emission and the height below PBL height such as the surface concentration.

3.3. Promotion of Computation Efficiency

Consuming clock time for the same period under different models and hardware configurations are compared and shown in Figure 8. It is worth noting that the consuming time of 1-hr was equal as that of 5-min for the CMAQ numerical benchmark because the results of 1-hr was averaged from the raw results of 5-min. Furthermore, the original CMAQ numerical model was written in Fortran framework and not suitable for the implementation on GPU platform, thus there was no computational time for the benchmark numerical model on GPU hardware.

The enlargement of the time step could notably reduce the total computation time under the same hardware. In this study, the consuming time was reduced from 460 min of the benchmark model to only 39 min of the deep learning surrogate for 1-hr resolution under the same CPU configuration (4*40 cores), achieving ∼12 times acceleration. The advantage of the deep learning surrogate over the numerical model was the end-to-end training method, which can directly predict the concentration 1 hr later without fixed iterations required by the numerical CMAQ model. However, the deep learning surrogate is not necessarily faster than the numerical benchmark under the same time step and CPU hardware. In this testcase, the deep learning surrogate with 5-min resolution consumed 456 min, almost the same as 460 min of the numerical CMAQ model using the same 4*40 CPU cores. If the deep learning architecture was more complicated or there were more inputs features, the consuming time of the surrogate could be even larger than the benchmark. This is explainable and reasonable because the large size of 3-D features input, complex architecture, and numerous neural nodes will bring large amount of big matrix operations. In contrast, the numerical operations of the traditional algorithm for physical transport process were simplified and optimized (Byun & Schere, 2006; Colella & Woodward, 1984).

Nevertheless, the deep learning architecture can efficiently accelerate the calculation relied on the GPU hardware. Compared with the time consuming (456 min for 5-min resolution and 39 min for 1-hr resolution) at the CPU platform of 4*40 cores, the same deep learning surrogate could reduce it to 32 and 3 min, respectively, at the platform of single one GPU. The accelerating ratio could reach as high as 13–14 for both two resolutions only attributed to the GPU hardware, which has the natural advantages in parallel tasks process.

In summary, the deep learning surrogate could accelerate the computation efficiency of the benchmark numerical model with the maximum speedup factor of 164 relying on the integrated benefit of end-to-end algorithm and GPU hardware. Reported acceleration work via deep learning also verified the above conclusions from this study. Only benefit from the end-to-end algorithm, Lauret et al. (2016) and Wang and Qian (2018) reported a speedup factor of 1.5 times through their neural networks. Liu et al. (2021) and Kelp et al. (2018) realized 10.6 times and 260 times acceleration, respectively via deep learning. Relying on both algorithm characteristics and GPU hardware, Vlasenko et al. (2021) reported 720 times speedup for a 1-day chemical transport simulation and Liu et al. (2021) reported 85.2 times speedup for a one-hour gas-phase chemistry solver simulation.

3.4. Limitations and Future Work

Although longer lifetime, better consistency, and significant acceleration for the deep learning surrogate of atmospheric transport process have been achieved, there are still several limitations in this study that should be investigated in the future. Currently, the study only focuses on carbon monoxide as the target pollutant, which was abundant in urban atmosphere. The applicability of the surrogate to other pollutants should be widely assessed as the emission intensity and concentration level vary in a large range for different pollutants. Additional normalization preprocessing for different species might be necessary to enhance the surrogate's applicability. Another

challenge is the applicability transfer to other meteorological seasons. The training data set and validation data set now were both in the winter season. It might be uncertain to apply the winter surrogate to other seasons such as summer because of significant different meteorological conditions. Inclusion of other seasons for the training data might relieve or solve this problem. After sufficiently validated, the deep learning surrogate for atmospheric transport process will be assembled with the deep learning surrogate for chemical process developed by our group Liu et al. (2021) to obtain a real chemical transport deep learning surrogate.

4. Conclusions

In this paper, a deep learning surrogate based on U-Net architecture was developed toward replacing the 3-D atmospheric transport process. Validation results showed that the deep learning model could well reproduce the horizontal advection and vertical diffusion processes. The R squared value was over 0.9 for a single time step, and the lifetimes for continuous running reached 400 and 1,000 steps for 1-hr and 5-min resolutions. The largest speedup factor of 164 was achieved for the deep learning surrogate via the advantage of end-to-end algorithm and GPU architecture.

This work preliminarily proved the feasibility of data-driven approach in approximating the 3-D diffusion process of pollutants. Meanwhile, computational efficiency can be notably improved while maintaining high consistency with the numerical model. The deep learning surrogate is expected to be a real "CTM" after coupling with the remaining chemical emulator. The natural property of automatic differentiation with the surrogate will also provide a convenient solution for inverse problems such as emission amount estimation.

Data Availability Statement

The source code of CMAQ can be retrieved from https://doi.org/10.5281/zenodo.3585898 (U.S. EPA Office of Research and Development, 2019) [Software]. The source code of WRF can be retrieved from https://www.mmm.ucar.edu/weather-research-and-forecasting-model (Skamarock & Klemp, 2008) [Software]. The MEIC Data can be retrieved from http://www.meicmodel.org (Li et al., 2017; Zheng et al., 2018) [Data set]. The NECP FNL data set is available to download from https://doi.org/10.5065/D6M043C6 (NCEP, 2000) [Data set]. The surface observations and vertical observations data can be found and downloaded from https://doi.org/10.5065/4F4P-E398 and https://doi.org/10.5065/39C5-Z211 (NCEP, 2004a; 2004b) [Data set]. The retrieved training and validation datasets are too large (>1T) to be uploaded for share. Anyone who is interested in them is welcomed to contact the corresponding author for a point-to-point transfer. The source code of the deep learning model could be found in Zenodo: https://doi.org/10.5281/zenodo.6556597 (Xu et al., 2022) [Software].

Acknowledgments
This work was supported by the National Natural Science Foundation of China (T2122022, 41975152). The CTM simulations and deep learning model training were completed on the "π 2.0" cluster system of the Center for High Performance Computing in Shanghai Jiao Tong University.

References

Binkowski, F. S., & Roselle, S. J. (2003). Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component 1. Model description. *Journal of Geophysical Research*, *108*(D6), 4183. https://doi.org/10.1029/2001JD001409

Boffi, D., Gastaldi, L., & Heltai, L. (2007). On the CFL condition for the finite element immersed boundary method. *Computers & Structures*, *85*(11–14), 775–783. https://doi.org/10.1016/j.compstruc.2007.01.009

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

Byun, D., & Schere, K. L. (2006). Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Reviews*, *59*(2), 51–77. https://doi.org/10.1115/1.2128636

Chuang, M. T., Lee, C. T., & Hsu, H. C. (2018). Quantifying PM2. 5 from long-range transport and local pollution in Taiwan during winter monsoon: An efficient estimation method. *Journal of Environmental Management*, *227*, 10–22. https://doi.org/10.1016/j.jenvman.2018.08.066

Colella, P., & Woodward, P. R. (1984). The piecewise parabolic method (PPM) for gas-dynamical simulations. *Journal of Computational Physics*, *54*(1), 174–201. https://doi.org/10.1016/0021-9991(84)90143-8

Dolz, J., Ben Ayed, I., & Desrosiers, C. (2018). Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities. *Lecture Notes in Computer Science*, *11383*, 271–282. https://doi.org/10.1007/978-3-030-11723-8_27

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment*, *39*(37), 6957–6975. https://doi.org/10.1016/j.atmosenv.2005.04.027

Hu, X. M., Nielsen-Gammon, J. W., & Zhang, F. (2010). Evaluation of three planetary boundary layer schemes in the WRF model. *Journal of Applied Meteorology and Climatology*, *49*(9), 1831–1844. https://doi.org/10.1175/2010JAMC2432.1

Huang, L., Liu, S., Yang, Z., Xing, J., Zhang, J., Bian, J., et al. (2021). Exploring deep learning for air pollutant emission estimation. *Geoscientific Model Development*, *14*(7), 4641–4654. https://doi.org/10.5194/gmd-14-4641-2021

Kasim, M., Watson-Parris, D., Deaconu, L., Oliver, S., Hatfield, P., Froula, D., et al. (2020). Up to two billion times acceleration of scientific simulations with deep neural architecture search. In *APS division of plasma physics meeting abstracts*. BO05-001. Retrieved from https://arxiv.org/abs/1201.0490

Kelp, M. M., Tessum, C. W., & Marshall, J. D. (2018). Orders-of-magnitude speedup in atmospheric chemistry modeling through neural network-based emulation. Retrieved from https://arxiv.org/abs/1808.03874

Kim, Y., Sartelet, K., Raut, J. C., & Chazette, P. (2015). Influence of an urban canopy model and PBL schemes on vertical mixing for air quality modeling over Greater Paris. *Atmospheric Environment*, *107*, 289–306. https://doi.org/10.1016/j.atmosenv.2015.02.011

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. Retrieved from https://arxiv.org/abs/1412.6980

Lauret, P., Heymes, F., Aprin, L., & Johannet, A. (2016). Atmospheric dispersion modeling using artificial neural network based cellular automata. *Environmental Modelling & Software*, *85*, 56–69. https://doi.org/10.1016/j.rse.2020.111716

Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., et al. (2017). Anthropogenic emission inventories in China: A review. National Science Review, *4*(6), 834–866. [Data set]. https://doi.org/10.1093/nsr/nwx150

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. Retrieved from https://arxiv.org/abs/2010.08895

Liu, C., Zhang, H., Cheng, Z., Shen, J., Zhao, J., Wang, Y., et al. (2021). Emulation of an atmospheric gas-phase chemistry solver through deep learning: Case study of Chinese Mainland. *Atmospheric Pollution Research*, *12*(6), 101079. https://doi.org/10.1016/j.apr.2021.101079

Liu, Z., Miao, Z., Zhan, X., Wang, J., Gong, B., & Yu, S. X. (2019). Large-scale long-tailed recognition in an open world. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2537–2546). https://doi.org/10.48550/arXiv.1904.05160

Lu, Y., Zhu, B., Huang, Y., Shi, S., Wang, H., An, J., & Yu, X. (2019). Vertical distributions of black carbon aerosols over rural areas of the Yangtze River Delta in winter. *Science of the Total Environment*, *661*, 1–9. https://doi.org/10.1016/j.scitotenv.2019.01.170

Marrouf, A. A., Essa, K. S., El-Otaify, M. S., Mohamed, A. S., & Ismail, G. (2015). The influence of eddy diffusivity variation on the atmospheric diffusion equation. *Open Journal of Air Pollution*, *4*(3), 109–118. https://doi.org/10.4236/ojap.2015.43011

National Centers for Environmental Prediction (NCEP). (2000) updated daily. NCEP FNL operational model global tropospheric analyses. [Data set]. Research Data Archive. continuing from July 1999. https://doi.org/10.5065/D6M043C6 .

National Centers for Environmental Prediction (NCEP). (2004a) updated daily. NCEP ADP global surface observational weather data. [Data set]. Research Data Archive. https://doi.org/10.5065/4F4P-E398

National Centers for Environmental Prediction (NCEP). (2004b) updated daily. NCEP ADP global upper air observational weather data. [Data set]. Research Data Archive. https://doi.org/10.5065/39C5-Z211

Otte, T. L., & Pleim, J. E. (2010). The meteorology-chemistry interface processor (MCIP) for the CMAQ modeling system: Updates through MCIPv3. 4.1. *Geoscientific Model Development*, *3*(1), 243–256. https://doi.org/10.5194/gmd-3-243-2010

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Ribeiro, R. P., & Moniz, N. (2020). Imbalanced regression and extreme value prediction. *Machine Learning*, *109*(9), 1803–1835. https://doi.org/10.1007/s10994-020-05900-9

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., et al. (2019). Tackling climate change with machine learning. Retrieved from https://arxiv.org/abs/1906.05433

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, *9351*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Shen, J., Zhao, Q., Cheng, Z., Huo, J., Zhu, W., Zhang, Y., et al. (2020). Evolution of source contributions during heavy fine particulate matter (PM2. 5) pollution episodes in eastern China through online measurements. *Atmospheric Environment*, *232*, 117569. https://doi.org/10.1016/j.atmosenv.2020.117569

Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. Journal of Computational Physics, *227*(7), 3465–3485. [Software]. https://doi.org/10.1016/j.jcp.2007.01.037

Tilt, B. (2019). China's air pollution crisis: Science and policy perspectives. *Environmental Science & Policy*, *92*, 275–280. https://doi.org/10.1016/j.envsci.2018.11.020

Tompson, J., Schlachter, K., Sprechmann, P., & Perlin, K. (2017). Accelerating Eulerian fluid simulation with convolutional networks. *Proceedings of Machine Learning Research*, *70*, 3424–3433. https://arxiv.org/abs/1607.03597

U.S. EPA Office of Research and Development. (2019). Cmaq. Zenodo. [Software]. https://doi.org/10.5281/zenodo.3585898

Vlasenko, A., Matthias, V., & Callies, U. (2021). Simulation of chemical transport model estimates by means of a neural network using meteorological data. *Atmospheric Environment*, *254*, 118236. https://doi.org/10.1016/j.atmosenv.2021.118236

Wagstrom, K. M., & Pandis, S. N. (2011). Contribution of long range transport to local fine particulate matter concerns. *Atmospheric Environment*, *45*(16), 2730–2735. https://doi.org/10.1016/j.atmosenv.2011.02.040

Wang, B., & Qian, F. (2018). Three dimensional gas dispersion modeling using cellular automata and artificial neural network in urban environment. *Process Safety and Environmental Protection*, *120*, 286–301. https://doi.org/10.1016/j.psep.2018.09.006

Xing, J., Zheng, S., Li, S., Huang, L., Wang, X., Kelly, J. T., et al. (2022). Mimicking atmospheric photochemical modeling with a deep neural network. *Atmospheric Research*, *265*, 105919. https://doi.org/10.1016/j.atmosres.2021.105919

Xu, J. Z., Zhang, H. R., Cheng, Z., Liu, J. Y., Xu, Y. Y., & Wang, Y. C. (2022). DL_3-DTransport. Zenodo. [Software]. https://doi.org/10.5281/zenodo.6556597 .

Ying, Q., & Li, J. (2011). Implementation and initial application of the near-explicit master chemical mechanism in the 3D community Multiscale Air quality (CMAQ) model. *Atmospheric Environment*, *45*(19), 3244–3256. https://doi.org/10.1016/j.atmosenv.2011.03.043

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., et al. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, *241*, 111716. https://doi.org/10.1016/j.rse.2020.111716

Zhang, H., Xie, B., Zhao, S. Y., & Chen, Q. (2014). PM2. 5 and tropospheric O_3 in China and an analysis of the impact of pollutant emission control. *Advances in Climate Change Research*, *5*(3), 136–141. https://doi.org/10.1016/j.accre.2014.11.005

Zhang, Y., Olsen, K. M., & Wang, K. (2013). Fine scale modeling of agricultural air quality over the Southeastern United States using two air quality models. Part I. Application and evaluation. *Aerosol and Air Quality Research*, *13*(4), 1231–1252. https://doi.org/10.4209/aaqr.2012.12.0346

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., et al. (2018). Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. Atmospheric Chemistry and Physics, *18*(19), 14095–14111. [Data set]. https://doi.org/10.5194/acp-18-14095-2018

References From the Supporting Information

Donnell, E. A., Fish, D. J., Dicks, E. M., & Thorpe, A. J. (2001). Mechanisms for pollutant transport between the boundary layer and the free troposphere. *Journal of Geophysical Research*, *106*(D8), 7847–7856. https://doi.org/10.1029/2000JD900730

Le, M., & Kayal, S. (2021). Revisiting edge detection in convolutional neural networks. *IEEE*, 1–9. In *2021 international joint conference on neural networks (IJCNN)*. https://doi.org/10.48550/arXiv.2012.13576

National Research Council. (2010). *Global sources of local pollution: An assessment of long-range transport of key air pollutants to and from the United States*. National Academies Press. Retrieved from https://nap.nationalacademies.org/read/12743/chapter/1