JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2021MS002806

Special Section:
Machine learning application to Earth system modeling

Spatio-Temporal Hourly and Daily Ozone Forecasting in China Using a Hybrid Machine Learning Model: Autoencoder and Generative Adversarial Networks

Meiling Cheng[1], Fangxin Fang[1], Ionel M. Navon[2], Jie Zheng[3], Xiao Tang[4], Jiang Zhu[4], and Christopher Pain[1]

[1]Department of Earth Science and Engineering, Applied Modelling and Computation Group, Imperial College London, London, UK, [2]Department of Scientific Computing, Florida State University, Tallahassee, FL, USA, [3]Center for Excellence in Regional Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, China, [4]International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Abstract Efficient and accurate real-time forecasting of national spatial ozone distribution is critical to the provision of effective early warning. Traditional numerical air quality models require a high computational cost associated with running large-scale numerical simulations. In this work, we introduce a hybrid model (VAE-GAN) combining a generative adversarial network (GAN) with a variational autoencoder (VAE) to learn the dynamic ozone distributions in spatial and temporal spaces. The VAE-GAN model can not only decipher the complex nonlinear relationship between the inputs (the past states/ozone and meteorological factors) and outputs (ozone), but also provide ozone forecasts for a long lead-time beyond the training period. The performance of VAE-GAN is demonstrated in hourly and daily spatio-temporal ozone forecasts over China. The training datasets from 2013 to 2017 and validation datasets from 2018 to 2019 are the collection of data from the air quality reanalysis datasets. With the use of VAE, large dataset sizes are decreased by three orders of magnitude, enabling hourly and daily forecasts to be computed in seconds. Results show that the VAE-GAN achieves a reasonable accuracy in the prediction of both the spatial and temporal evolution patterns of hourly and daily ozone fields, as compared to the Nested Air Quality Prediction Modeling System (commonly used in China), the reanalysis data and observations during the validation period. Thus, the VAE-GAN is a cost-effective tool for large data-driven predictions, which can potentially reinforce air pollution prediction efforts in providing risk assessment and management in a timely manner.

Plain Language Summary This work presents a hybrid machine learning model for hourly and daily ozone forecasting in spatial and temporal spaces. Our goal is to use the machine learning model for exploring the complex nonlinear relationship between the meteorological factors and ozone concentration, and to perform long lead-time ozone forecasts accurately and efficiently. The reanalysis ozone datasets from 2013 to 2019 over China are used for processing different training and prediction scenarios. Our results show that the proposed machine learning model can predict the spatio-temporal evolution patterns of the hourly and daily ozone concentration accurately and much more efficiently in comparison to the Nested Air Quality Prediction Modeling System. Such a data-driven model is promising in different applications, for example, providing early warnings of high ozone concentrations in densely populated areas.

Key Points:

- A hybrid deep learning model is proposed for efficient and accurate real-time ozone forecasting with a high spatial resolution
- The VAE-GAN enables hourly and daily forecast to be achieved in seconds while the same level of accuracy is reached compared to Nested Air Quality Prediction Modelling System
- The performance of the VAE-GAN model is evaluated in hourly and daily ozone forecasts over China from the year 2018–2019

Correspondence to:
F. Fang,
f.fang@imperial.ac.uk

Citation:
Cheng, M., Fang, F., Navon, I. M., Zheng, J., Tang, X., Zhu, J., & Pain, C. (2022). Spatio-temporal hourly and daily ozone forecasting in China using a hybrid machine learning model: Autoencoder and generative adversarial networks. *Journal of Advances in Modeling Earth Systems*, *14*, e2021MS002806. https://doi.org/10.1029/2021MS002806

Received 14 SEP 2021
Accepted 14 FEB 2022

Author Contributions:
Conceptualization: Meiling Cheng
Data curation: Meiling Cheng, Ionel M. Navon, Jie Zheng, Xiao Tang, Jiang Zhu
Funding acquisition: Meiling Cheng, Fangxin Fang, Christopher Pain
Investigation: Meiling Cheng, Fangxin Fang

1. Introduction

The presence of ozone at the ground level has detrimental effects on the health of humans and the ecological environment (H. Wang, Li, et al., 2020; Ezimand & Kakroodi, 2019). Ozone concentration involves tremendous spatio-temporal variability based on photochemical interactions between nitrogen oxides (NO_x) and volatile organic compounds under certain meteorological circumstances (Jenkin & Clemitshaw, 2000). In recent years, the tropospheric ozone has turned into a primary atmospheric pollutant due to rapid growth of industrialization and urbanization, which results in approximately 80 thousand premature deaths every year (Feng et al., 2019). Therefore, it is necessary to make accurate and efficient ozone concentration forecasts in spatial and temporal spaces, thus providing guides for early warning management and control.

Methodology: Meiling Cheng, Fangxin Fang
Resources: Ionel M. Navon, Jie Zheng, Xiao Tang, Jiang Zhu
Software: Meiling Cheng
Supervision: Fangxin Fang, Christopher Pain
Validation: Meiling Cheng
Visualization: Meiling Cheng, Fangxin Fang
Writing – original draft: Meiling Cheng, Fangxin Fang
Writing – review & editing: Meiling Cheng, Fangxin Fang, Ionel M. Navon

Traditional chemistry transport models (CTMs) for ozone concentration forecasts consider the physical and chemical properties of an atmosphere in the processes of diffusion, transfer and air pollution (Ezimand & Kakroodi, 2019; Hu et al., 2016; Kumar et al., 2012). They are able to predict the spatio-temporal distributions of ambient ozone on a large scale (Mo et al., 2020). For example, Wang et al. (2013) used a global CTM (GEOS-Chem) to investigate the effects of 2000–2050 global changes in climate and anthropogenic emissions on surface ozone over China. However, the ozone concentration forecast involves nonlinear, strong coupling and multivariate problems, thus increasing the difficulty of long lead-time prediction (Stern et al., 2008; Su et al., 2020). In addition, the accuracy of CTMs is often limited by relatively coarse resolution, high uncertainty in emission inventory, model assumptions and parameters, as well as high computational cost (R. Liu, Ma, et al., 2020; Sharma et al., 2017).

These challenges have motivated researchers to exploit hybrid efficient and accurate models, for example, a variety of machine learning algorithms (Chattopadhyay et al., 2020; Weyn et al., 2020). Machine learning models are capable of efficiently handling big data and easily identifying trends and patterns (Hoshyaripour et al., 2016; Jia et al., 2020; Kumar et al., 2017). Several previous works have utilized different machine learning models to predict ambient ozone concentrations directly from observations and post-processing dynamical modeling outputs (Watson, 2019). For example, the MultiLayer Perceptron (MLP) was utilized by Dutot et al. (2007) to carry out hourly maximum ozone forecast in the next day in the center of France. Similarly, Eslami et al. (2019) employed a deep convolutional neural network (CNN) to predict the hourly ozone concentration by using atmospheric data at the previous day along with in-situ ozone and NO_2 concentrations as inputs. In the work of AlOmar et al. (2020), the hybrid model (W-ANN) combining an artificial neural network (ANN) with a wavelet transform (WT) approach, could provide ozone concentration forecast at multiple time scales (2-, 3-, 4 and 5-ahead hours). While these models can predict the future behavior of ozone concentration given knowledge of its past and present states at meteorological stations, they rarely explore spatio-temporal dependencies exhaustively. It was also found that the model performance was more severely degraded with a long lead-time (Weyn et al., 2019). Recently, H. Wang, Li, et al. (2020) presented the spatial and temporal (monthly) correlation in the monitoring network for regional ozone prediction. Furthermore, Zhan et al. (2018) used the random forest (RF) and Kriging approaches to interpolate the monitoring data on uniform grids and then investigated the spatial and temporal (daily) relationship of ozone concentrations. These models presented in (H. Wang, Li, et al., 2020; Zhan et al., 2018) are strongly dependent on the availability (locations and period) of monitoring data. The predictive accuracy is usually restricted due to a lack of sufficiently dense observations.

In this study, a hybrid physical-informed deep learning model is proposed for real-time ozone forecasting with a high spatial resolution. In hybrid deep learning modelling, the training and validation datasets originate from the reanalysis data which optimally combines the numerical solutions from the physical air quality model (here, the Nested Air Quality Prediction Modelling System - NAQPMS) and observed data in monitoring stations. Physical modelling results from the governing equations can be used to improve understanding of complex ozone transport processes and provide the details of the spatial distribution of ozone concentrations, while monitoring data can be used for correcting physical modelling results, thus improving predictive accuracy.

A newly developed generative adversarial network (GAN) model (Cheng et al., 2020) is introduced for spatial-temporal (hourly/daily) ozone forecasting in this work. GANs proposed by Goodfellow et al. (2014), have been attractive for producing high-resolution samples (e.g., images). Various versions of GANs have been applied for generating different types of high-dimensional data, such as geostatistical maps (Laloy et al., 2018), seismic waves (Li et al., 2018), weather maps (Gagne et al., 2020), total electron content map (Chen et al., 2019). GANs have the capability of learning hierarchical feature representations in various image analysis problems (Zhong et al., 2019). Recently, Cheng et al. (2020) first introduced a variational autoencoder (VAE) into GAN to explore spatial and temporal flow dynamics with a reduction of the computational cost. The original GAN consists of a generative model (generator) and a discriminative model (discriminator). The generator and discriminator contest through an adversarial process to optimize the generative model. Here, the VAE-GAN is a game theoretic framework that involves three modules, an encoder, a decoder and a discriminator (Makhzani et al., 2015). The role of the encoder is to transform real and fake samples in a high-dimensional space into a low-dimensional latent space. The low-dimensional representations, several orders of magnitude smaller than the dimensional size of the original datasets, are then applied to the adversarial network for representation learning and parameter optimization, thus alleviating computational burdens (Cheng et al., 2020). After that, the decoder continues to transfer the

latent states to high-dimensional samples. In essence, the VAE-GAN model includes a self-training generative and discriminative mechanism, providing an attractive way for learning the real sample data distribution in the absence of any prior physical or extensive statistical knowledge.

In this work, we have developed a spatio-temporal VAE-GAN for hourly and daily ozone forecasting. We use convolutional networks in the VAE-GAN model to improve its scalability for large data-driven modelling and implement a recursively forecasting strategy in the model architecture for achieving the long lead-time forecast. Overall, the VAE-GAN model developed in this study can: (a) learn and classify the dynamic ozone features using high-dimensional datasets; (b) decipher the complex nonlinear relationship between the present states and future events among all training samples; (c) provide accurate hourly and daily ozone forecast in a long lead-time beyond the training period. By using high-resolution reanalysis data, we show that the VAE-GAN can represent both the spatial and temporal variability of ozone concentration. To our best knowledge, this is the first proposal of the VAE-GAN for hourly and daily ozone forecasting using a combination of dynamic learning and real-time forecasting strategy.

In this work, the VAE-GAN model has been applied to hourly and daily forecasting of ozone concentrations with a high spatial resolution (15 × 15 km) across China. The datasets are collected from the air quality reanalysis datasets during 2013–2019. The accuracy of the VAE-GAN model has been evaluated by comparing the results with those obtained from both the physical air quality model (NAQPMS) and observations at monitoring stations. The remainder of this paper is organized as follows. Section 2 provides the details of datasets and methods used in this study, the VAE-GAN architecture and forecasting workflow. This is followed by Section 3 with model performance analysis of ozone forecasting capabilities in spatial and temporal spaces. Finally in Section 4, conclusions are presented.

2. Data and Methods

Ozone forecasting involves nonlinear, strong coupling and heterogeneous problems (Su et al., 2020). As investigated by Zhan et al. (2018), the key ozone-related meteorological variables are the temperature, relative humidity, wind speed, evaporation and sunshine duration. Among these meteorological conditions, the daytime surface temperature is an important driver of ozone episodes. A high temperature can accelerate the chemical kinetic reaction and emissions of the natural components of ozone increase (Bloomer et al., 2009; Gu et al., 2020; Yin et al., 2019). This work thus focuses on developing an ozone forecasting model depending on temperature. The impact of other meteorological factors (such as humidity and wind speed) is also explored in Section 3.4.

Given a series of ozone observations $\{O_{t-p}, \ldots, O_t\}$ (where p is the length of historical time steps) and the temperature data T_{t+1} at the predictive time level $(t + 1)$, an one-step-ahead ozone forecast can be expressed mathematically as:

$$\widehat{O}_{t+1} = \mathcal{F}\left(O_{t-p}, \ldots, O_t, T_{t+1}\right), \tag{1}$$

where $\widehat{O}_{t+1} = \left(\widehat{O}^1_{t+1}, \ldots, \widehat{O}^k_{t+1}, \ldots, \widehat{O}^{M_k}_{t+1}\right)$ $(k \in (1, M_k)$, M_k is the number of points in the study area) represents the predictive zone concentration at time level $(t + 1)$ and $\mathcal{F}$ is a forecasting model. In this work, the VAE-GAN-based ozone forecasting model is selected as the forecasting model $\mathcal{F}$. A schematic diagram of the proposed VAE-GAN framework is presented in Figure 1.

2.1. Data

2.1.1. Reanalysis Data

Ground-level ozone has become a severe pollutant in major urban areas of China (Zhan et al., 2018). Our objective is to develop an efficient and accurate machine learning tool for providing an interactive ozone map over China, which shows real-time (hourly/daily) ozone concentrations with a high spatial resolution across China. However, there are only 1436 monitoring stations over China, which are sparse and heterogeneous. It is obvious that the monitoring data is not enough to represent the spatial distribution of ozone on a high spatial scale. Therefore, we select the 7-year Chinese air quality reanalysis (CAQRA) data set for our training and validation purposes (Kong et al., 2021). The CAQRA dataset from 2013 to 2019 was produced by assimilating surface observations at 1436 monitoring stations into the Nested Air Quality Prediction Modeling System (NAQPMS).

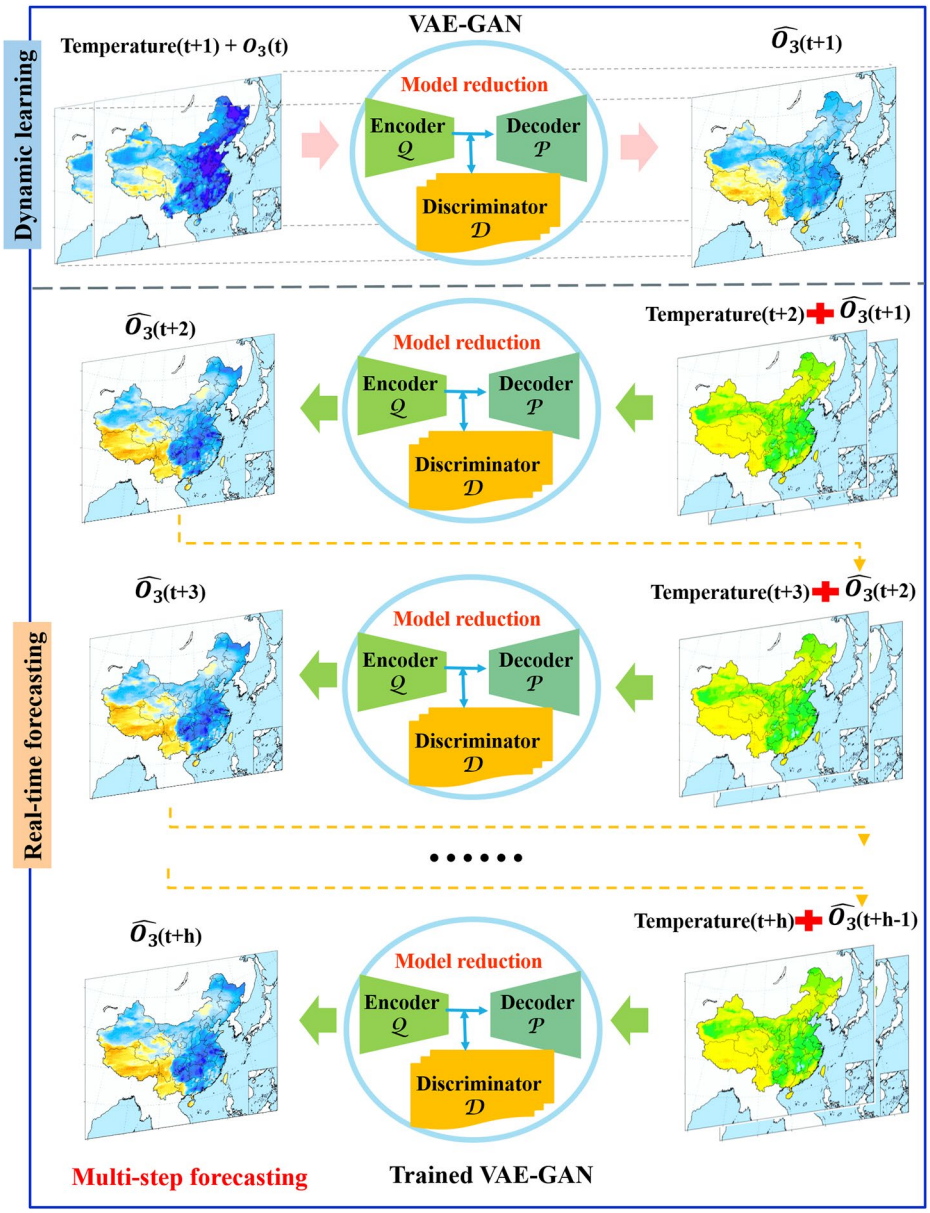

Figure 1. Schematic of real-time ozone forecasting. The architecture consists of two processes: Dynamic learning and real-time forecasting. In real-time forecasting process, one-step ozone forecast can be obtained by a direct prediction for one-step (hourly/daily) lead-time. To get forecasts for a long lead-time, a recursive strategy is adopted, which uses the last one-step ozone forecast as a new input to predict the next-step ozone concentration.

Such combination could be the optimal way by which physical modeling results from the NAQPMS can provide dynamic understanding from the governing equations, while monitoring data may find unexpected patterns not provided by physical modeling.

The CAQRA dataset contains six pollutants ($PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and ozone) and meteorological factors (temperature, humidity and wind speed) at the high spatial (15 × 15 km) and temporal (hourly) resolution, which are publicly accessible at Science Data Bank (ScienceDB). Each ozone map or meteorological map across China has a high spatial resolution with a total node number of 47,232. The ozone concentration over China varies in both spatial and temporal spaces. For example, as shown in Figure 2, it can be noted that the distributions of average hourly ozone concentration in June and December of 2018 are heterogeneous, exhibiting significant seasonal variations. This can be explained by the fact that ozone production is tightly correlated with meteorological

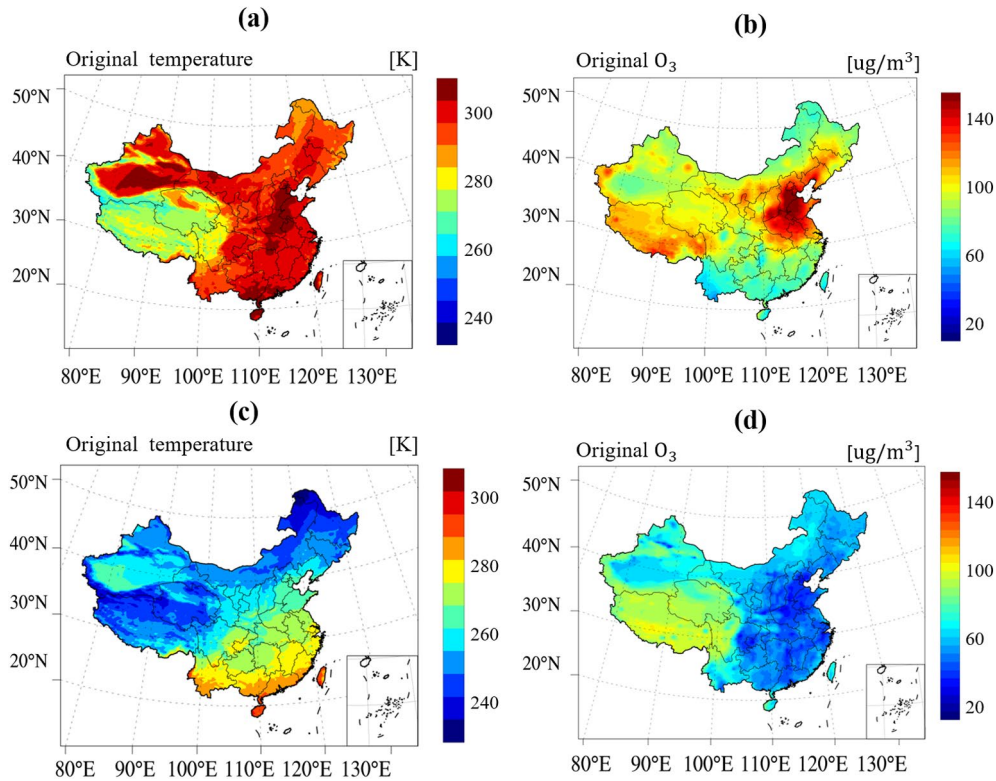

Figure 2. Spatial distribution of the average hourly temperature and ozone concentration in China. (a) and (b) June; (c) and (d) December in 2018.

conditions, such as ambient temperature, sunshine radiation and humidity, causing the variation of emitted intensity of atmospheric pollutants in different seasons (Bloomer et al., 2010). The high ozone concentration in June is closely related to the enhanced photochemistry reaction due to high temperature and strong solar radiance in Summer (Kong et al., 2021). To demonstrate the performance of our proposed model with seasonal variations, we choose the months of June and December as representing Summer and Winter while exhibiting hourly spatio-temporal ozone forecasts in the following Section 3.1.

2.1.2. Monitoring and Physical Modeling Data

The objective of this study is to develop a machine learning-based ozone forecasting model, which will be accurate and more efficient than existing air quality models. For comparison purposes, the NAQPMS (commonly used for real-time pollution forecasting in China, here referred to as the physical model) is chosen. The NAQPMS model is driven by the hourly meteorological fields produced by the Weather Research and Forecasting (WRF) model, and the aqueous-phase chemistry and wet deposition are simulated based on the Regional Acid Deposition Model (RADM) mechanism in the Community Multi-scale Air Quality (CMAQ) model. The emissions of air pollutants considered in the NAQPMS model include the monthly anthropogenic, biomass burning, biogenic volatile organic compound (BVOC), marine VOC, soil NO_x and lightning NO_x emissions. More details of the NAQPMS configuration are introduced in Kong et al. (2021). The hourly physical modeling dataset at the high spatial (15×15 km) and temporal (hourly) resolutions is generated from the NAQPMS in June of 2018, which has the same spatial and temporal resolutions as the reanalysis data (CAQRA data set). We compare the VAE-GAN results with those from the NAQPMS, where the monitoring measurements are used as the reference data. Hourly monitoring of ozone data is published by the China National Environmental Monitoring Centre (CNEMC). The hourly ozone observations are collected from 1436 stations in June of 2018 (Kong et al., 2021).

2.1.3. Data Processing

During the training and validation processes, the input and output datasets are pre-processed by a Standard Scaler normalization method (Buitinck et al., 2013). The ozone concentration and meteorological factors are scaled to a

Table 1
Summary of Training and Validation Datasets Used in Hourly and Daily Ozone Forecasting (Where the Model Inputs are the Meteorological Conditions (Temperature, Humidity and Wind Speed) and the Ozone Concentrations at the Last Time Level While the Targeted Outputs are the Ozone Concentrations at the Predictive Time Level)

Scale	Training dataset	Validation dataset
Hourly forecasting	Reanalysis data (June and December of 2017)	Reanalysis data (June and December of 2018)
		Monitoring data (June of 2018)
		Physical modeling data (June of 2018)
Daily forecasting	Reanalysis data (2013–2017)	Reanalysis data (2018–2019)

specific interval prior to training the VAE-GAN model. In this study, the description of the training and validation datasets in hourly and daily ozone forecasting is provided in Table 1.

As shown in Table 1, we select the reanalysis datasets in 2017 for model training and 2018 for validation in hourly ozone forecasting (in Section 3.1 and Section 3.2). During the validation period, the reanalysis data is used for initializing ozone forecasting. We compare the forecasting results with the reanalysis data, monitoring observations and results from the physical model-NAQPMS in the six regions of whole China (North, Northeast, Southwest, Southeast, Northwest, and Central) as well as highly populated regions (Jing-Jin-Ji (JJJ) Yangtze River Delta (YRD) and Pearl River Delta; in Section 3.1 and Section 3.2). In daily ozone forecasting, we use the reanalysis data from the year 2013–2017 for model training and the reanalysis data from the year 2018–2019 for model validation (in Section 3.3 and Section 3.4).

2.2. Dynamic Learning Algorithm: VAE-GAN

Given a large historical dataset $O = \{O_{t-p}, \ldots, O_t\} \in \Re^{(p+1)\times M_k}$ (where $p+1$ is the historical time length and M_k is the number of nodes) over the whole China area, it is challenging to efficiently predict the ozone concentration in spatial and temporal spaces. Here the VAE machine learning technique, used in a wide variety of applications (Gonzalez & Balajewicz, 2018), is introduced with the scope of dimensionality reduction. The architecture of VAE (including an encoder $\mathcal{Q}$ and a decoder $\mathcal{P}$) in the VAE-GAN model is used to identify low-dimensional representations of ozone concentration maps as they evolve in time. In comparison to traditional reduced order models (ROMs) using proper orthogonal decomposition (POD; Xiao et al., 2019), the VAE has a nonlinear activation function for its nonlinearly weighted input, thus better representing the nonlinear physical processes (Cheng et al., 2020).

A GAN consists of two modules: a generator $\mathcal{G}$ and a discriminator $\mathcal{D}$. The generator is designed to produce solutions from a random dataset while the discriminator is tasked to discriminate the real samples from the generated solutions. The VAE-GAN is a probabilistic autoencoder that uses the GAN framework as a variational inference algorithm (Makhzani et al., 2015). In the hybrid VAE-GAN (as shown in the middle column of Figure 1), the inputs μ and the real solutions O_d (the targeted outputs) are fed into the encoder $\mathcal{Q}$, which transforms the high-dimensional datasets into low-dimensional representations ζ and ζ_d respectively. The discriminator $\mathcal{D}$ then discriminates the low-dimensional representations ζ and ζ_d in an efficient way. The latent states ζ_d and ζ are then transformed into the reconstructed samples $\widehat{O_d}$ (the reconstructed outputs) and real solutions $\widehat{O}$ (the predicted outputs) respectively in the decoder $\mathcal{P}$.

For the encoder, we use a deep CNN where the model inputs are the meteorological conditions (temperature T_{t+1} at the predictive time level) and historical ozone dataset O while the targeted outputs are the ozone concentrations O_{t+1} at the predictive time level. The CNNs exhibit superior performance in extracting the high nonlinearity and chaotic nature from the inputs with a two-dimensional structure (Bolton & Zanna, 2019). The encoder has four convolutional layers followed by one densely connected layer. Each hidden layer is followed by a Leaky rectified linear unit (Leaky ReLU) activation function to extract feature maps from inputs. To avoid dimensionality loss of the edges of images, the padding operation is applied in convolutional layers. After the hidden layers, the fully connected layer is applied to further reduce the dimension of features from its previous layer and find the most task-relevant features for inference. This hierarchical convolutional feature learning acts as nonlinear model reduction, which finally transfers the inputs μ and targeted output O_d into the low dimensional latent states ζ

Table 2
Architecture of Encoder, Decoder and Discriminator in the VAE-GAN

Module	Layer
Encoder $\mathcal{Q}$	Input (47,232)
	Conv2D, leaky ReLU ($\alpha = 1$)
	Conv2D, leaky ReLU ($\alpha = 1$)
	Conv2D, leaky ReLU ($\alpha = 1$)
	Full connected
	Output (64)
Decoder $\mathcal{P}$	Input (64)
	Full connected, leaky ReLU ($\alpha = 1$)
	Conv2D, leaky ReLU ($\alpha = 1$)
	Conv2D, leaky ReLU ($\alpha = 1$)
	Conv2D, leaky ReLU ($\alpha = 1$)
	Conv2D, Tanh
	Output (47,232)
Discriminator $\mathcal{D}$	Input (64)
	Full connected, leaky ReLU ($\alpha = 1$)
	Full connected, leaky ReLU ($\alpha = 1$)
	Full connected, leaky ReLU ($\alpha = 1$)
	Full connected, leaky ReLU ($\alpha = 1$)
	Full connected, leaky ReLU ($\alpha = 1$)
	Full connected, Sigmoid
	Output (1)

and ζ_d respectively. It is shown in Table 2 that the size of datasets on ozone spatial distributions is reduced from 47,232 to 264 in the latent space, that is, a decrease of three orders of magnitude.

The discriminator is then used for distinguishing the real latent state ζ_d from the generated latent state ζ at the reduced space. The latent states ζ and ζ_d are processed using five hidden layers, each using a densely connected layer with a Leaky ReLU activation function. The output layer is densely connected to the final hidden layer with a sigmoid activation function (Lee & You, 2019; Mosser et al., 2017), to yield a number between 0 and 1 representing the probability that the ζ is a fake sample created from inputs μ (as opposed to a targeted output O_d).

For the decoder, it is a reverse process for reconstructing the latent states ζ and ζ_d to the high-dimensional states $\widehat{O}$ and $\widehat{O_d}$. The decoder has four hidden layers including one densely connected layer and three convolutional layers, each layer followed by a Leaky ReLU activation function. The final layer uses a tanh activation (Lee & You, 2019; Mosser et al., 2017) with outputs $\widehat{O}$ and $\widehat{O_d}$ shown in Table 2.

Now the question is how can we train a forecasting model? The VAE-GAN is intuitively similar to VAEs, with the key difference that the VAE-GAN replaces the Kullback-Leibler divergence penalty of VAEs with the adversarial loss described below (Makhzani et al., 2015). The training process involves the minimization of reconstruction loss between the reconstructed outputs $\widehat{O}$ and targeted ones O_d, as well as the maximization of adversarial loss in distinguishing between the latent states ζ and ζ_d. The reconstruction loss jointing the encoder and the decoder is expressed as:

$$\mathcal{L}_{rec}(\theta, \phi) = \| O_d - \mathcal{P}\left(\mathcal{Q}\left(O_d, \theta\right), \phi\right) \|, \tag{2}$$

and the adversarial loss jointing the encoder, decoder and discriminator is calculated as:

$$\mathcal{L}_{adv}(\theta, \phi, \psi) = E_{\zeta \sim p_\zeta(\zeta)}(\log \mathcal{D}(\mathcal{Q}(\mu))) + E_{\zeta_d \sim p_{\zeta_d}(\zeta_d)}\left(\log\left(1 - \mathcal{D}\left(\mathcal{Q}\left(O_d\right)\right)\right)\right), \tag{3}$$

where θ, ψ, ϕ represent the parameters in the encoder $\mathcal{Q}$, discriminator $\mathcal{D}$ and decoder $\mathcal{P}$, respectively.

Typically, the VAE-GAN uses the hybrid loss ($\mathcal{L}_{rec} + \mathcal{L}_{adv}$) as the objective minimization-maximization function as follows:

$$\min_{\theta,\phi} \max_{\psi} E_{\zeta \sim p_\zeta(\zeta)}(\log \mathcal{D}(\mathcal{Q}(\mu))) + E_{\zeta_d \sim p_{\zeta_d}(\zeta_d)}\left(\log\left(1 - \mathcal{D}\left(\mathcal{Q}\left(O_d\right)\right)\right)\right) + \| O_d - \mathcal{P}\left(\mathcal{Q}\left(O_d, \theta\right), \phi\right) \|. \tag{4}$$

the model can be minimized and maximized by stochastic gradient descent with backpropagation. This means that the loss function derivation is propagated throughout the network using the chain rule to update the model parameters θ, ϕ, ψ. In this work, we choose the adaptive moment estimation (Adam) as an adaptive stochastic gradient descent (SGD) algorithm, which has been proven to be efficient for different types of deep networks (Kingma & Ba, 2014). Once the training process is completed, the model parameters (including weights and bias) are saved for real-time forecasting.

2.3. Real-Time Forecasting

The multi-step-ahead time series forecasting can be described as an estimation of future time series $\{O_{t+1}, \ldots, O_{t+H}\}$. Referring to the work of Cheng et al. (2021), the multi-step-ahead ozone forecasting process can be summarized as consisting of the following algorithmic steps:

1. The forecasting model of VAE-GAN is trained to perform a new one-step-ahead forecast $\widehat{O}_{t+1}$ during the predictive period ($t + 1 \in (t_f, t_N)$).

2. Selecting the new input temperature T_{t+2} at predictive time level $(t+2)$.
3. Using the one-step predicted $\hat{O}_{t+1}$ and temperature T_{t+2} as new inputs, to forecast the next-step-ahead ozone concentration at predictive time level $(t+2)$ by the trained model.
4. Repeating steps (1–3) for H times at the multi-step forecasting strategy as shown in Figure 1.
5. Finally, the length-H timesteps of forecasts at point k ($k \in [0, M_k]$) are obtained as:

$$\hat{O}^k_{t+h} = \begin{cases} \{\mathcal{Q}, \mathcal{P}\}\left(O^k_{t-p+h-1}, \ldots, \hat{O}^k_{t+1}, \ldots, \hat{O}^k_{t+h-1}, T^k_{t+h}\right), & (h \in \{2, \ldots, p+1\}) \\ \{\mathcal{Q}, \mathcal{P}\}\left(\hat{O}^k_{t-p+h-1}, \ldots, \hat{O}^k_{t+h-1}, T^k_{t+h}\right), & (h \in \{p+2, \ldots, H\}) \end{cases}.$$

we use the Python library Keras (Gulli & Pal, 2017) with the TensorFlow 1.1.0 (Abadi, 2016) backend for all experiments of the proposed model. The performance of the VAE-GAN model developed in this study is assessed using the root mean squared error (RMSE), coefficient of correlation R, normalized mean bias (NMB) and normalized mean error (NME) (Emery et al., 2017) which are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N}\left(\hat{O}_j - O_{d,j}\right)^2}{N}}, \tag{5}$$

$$R = \frac{\sum_{j=1}^{N}\left(\hat{O}_j - \overline{\hat{O}}\right)\left(O_{d,j} - \overline{O}_d\right)}{\sqrt{\sum_{j=1}^{N}\left(\hat{O}_j - \overline{\hat{O}}\right)^2}\sqrt{\sum_{j=1}^{N}\left(O_{d,j} - \overline{O}_d\right)^2}}, \tag{6}$$

$$NMB = \frac{\sum_{j=1}^{N}\left(\hat{O}_j - O_{d,j}\right)}{\sum_{j=1}^{N} O_{d,j}} \times 100, \tag{7}$$

$$NME = \frac{\sum_{j=1}^{N}\left|\hat{O}_j - O_{d,j}\right|}{\sum_{j=1}^{N} O_{d,j}} \times 100, \tag{8}$$

where the subscript j represents the pairing of N targeted ozone concentration O_d and predictions $\hat{O}$ by points ($[0, M_k]$) and time ($[t_f, t_N]$), and the overbars signify means over points and/or time.

3. Results and Discussion

3.1. Hourly Ozone Forecasting Analysis Over China in Summer and Winter

In this section, we demonstrate the trained VAE-GAN model has the capability of capturing the details of the spatial distribution of ozone concentrations over China. For this purpose, the post-processed reanalysis dataset with a high spatial resolution (15×15 km) is thus chosen as the 'true' reference during the validation period (see Table 1). The trained VAE-GAN is used for hourly ozone forecasting in June and December 2018. A comparison between the predicted ozone concentrations and reanalysis data is provided in Section 3.1.1, as well as the corresponding error estimation is carried out in Section 3.1.2.

3.1.1. Spatio-Temporal Hourly Forecasting of Ozone Concentration

The spatial distribution of hourly ozone concentrations in June and December in 2018 is shown in Figure 3. It can be seen that the VAE-GAN model captures a large part of the spatial distribution of ozone concentration at lead-times $t = 30$, 60 hr in June and December, respectively. It is noted that the ozone concentration in June is higher than that in December. This can be explained by the fact that a high temperature leads to increased convection,

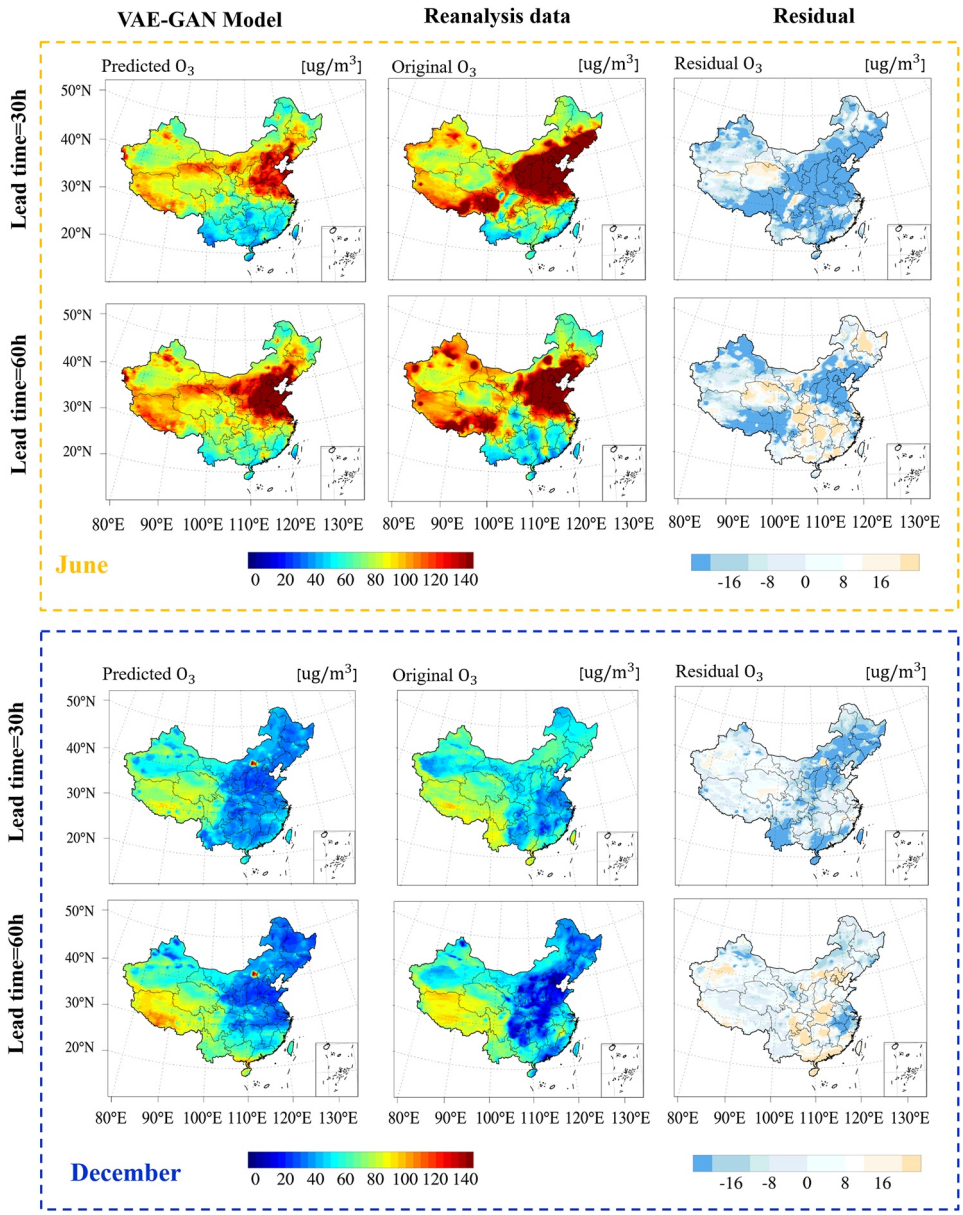

Figure 3. Comparison of the spatial distribution of ozone fields obtained from the VAE-GAN model (left) and the reanalysis data (middle) at lead-times of 30 and 60 *hr* in June and December in 2018.

unstable atmosphere stratification, and increased precipitation, which are beneficial to the diffusion and deposition of ozone (Yan et al., 2021). By contrast, the low temperature in December causes a surface inversion, and the height of the atmospheric mixed layer is low, which is not conductive to vertical convection. Compared to the reanalysis data, in June, the predicted ozone concentration is somewhat underestimated over the central of China at different lead-times, while slightly overestimated over the North-east and South-east area close to the North Pacific Ocean.

The last column of Figure 3 shows the residual maps between the ozone distributions predicted by the VAE-GAN and reanalysis data. Table 3 indicates that the mean residual varies from $-12.4\ \mu g/m^3$ (June) and $-6.91\ \mu g/m^3$ (December) at the lead-time of $t = 30\ hr$ to $-5.88\ \mu g/m^3$ (June) and $-1.34\ \mu g/m^3$ (December) at the lead-time of $t = 60\ hr$ respectively. It is found that 10% and 90% residuals in June are higher than those in December, that is, the predicted results using the VAE-GAN have lower accuracy in June in comparison to that in December. This may be caused by the complex and high ozone heterogeneity in the spatial distribution in June. Overall,

Table 3
Residuals of Ozone Fields Between the VAE-GAN Model and the Reanalysis Data at the Lead-Times t = 30, 60 hr in June and December in 2018

	June		December	
Indicators	Lead-time = 30hr	Lead-time = 60hr	Lead-time = 30hr	Lead-time = 60hr
Mean residual (ug/m^3)	−12.4	−5.88	−6.91	−1.34
90th percentile (ug/m^3)	13.95	23.44	8.92	15.09
10th percentile (ug/m^3)	−40.81	−32.30	−24.58	−18.39

the trained VAE-GAN has captured the spatial characteristics of ozone concentrations for given new inputs at a lead-time beyond the training period. Although a discrepancy between the predicted ozone concentrations and reanalysis data, the VAE-GAN is able to provide the same level of accuracy as that of the physical model-NAQPMS (see Section 3.2). To improve the predicted accuracy during the high ozone episode (≥160 $\mu g/m^3$), the convolutional long short-term memory (ConvLSTM) approach is introduced to the VAE-GAN model. For details, see Appendix A.

3.1.2. Error Analysis of Hourly Ozone Forecasting

Further error analysis of hourly ozone forecasting is carried out through the RMSE and correlation coefficient. The temporal variations of spatial-averaged RMSEs and correlation coefficient of hourly predictive results from the VAE-GAN are shown in Figure 4. It can be noticed that the spatial-averaged RMSE values in June and December in 2018 are generally around 22 $\mu g/m^3$ and 13 $\mu g/m^3$ during the entire forecasting period. The corresponding correlation coefficient values are shown in Figure 4b. We can see that the ozone concentration predicted by the VAE-GAN model shows a good correlation with the reanalysis data over a long lead-time. Most of correlation

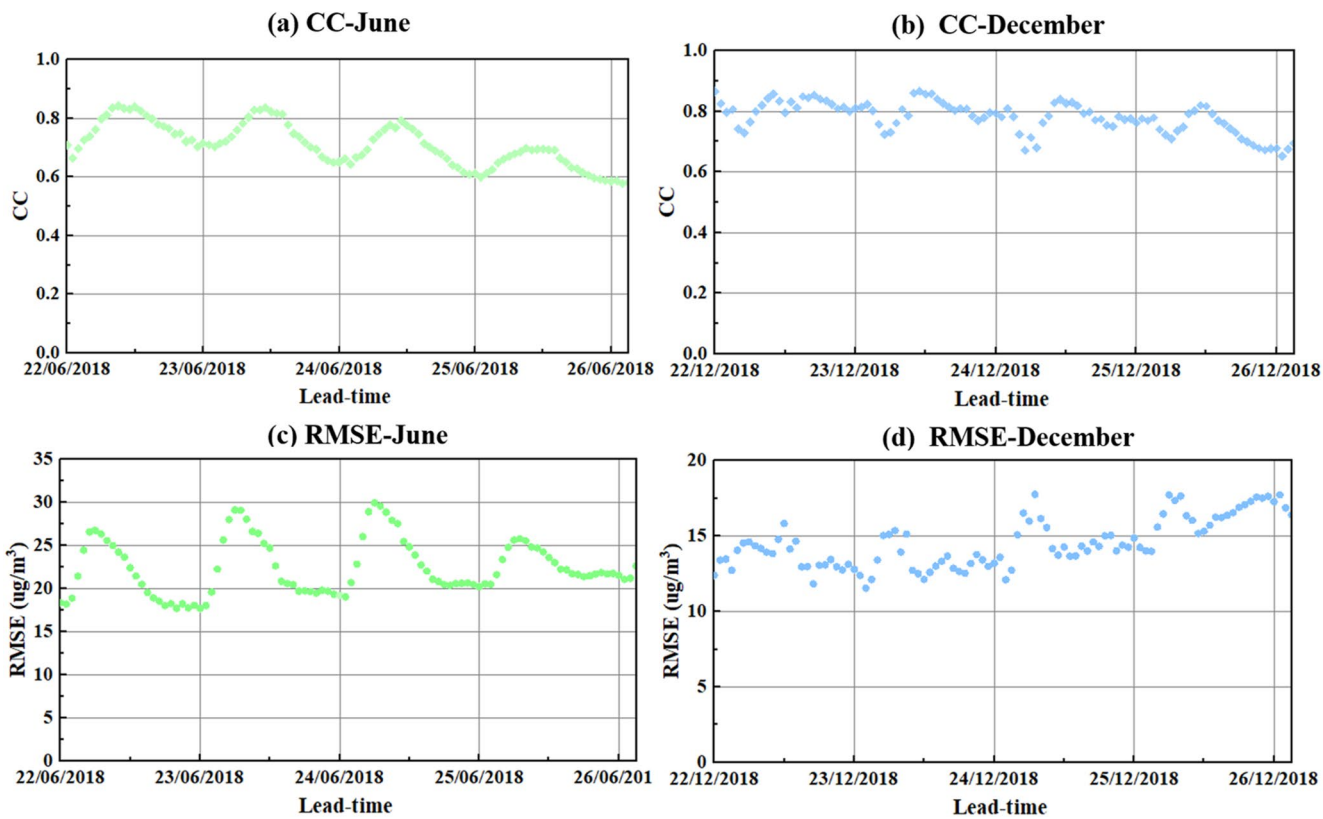

Figure 4. The spatial-averaged RMSEs and correlation coefficients of ozone concentration between the VAE-GAN and the reanalysis data in June and December in 2018.

coefficients are around or beyond 0.6 during the forecasting period in June. In addition, it is worth noting that the magnitude of correlation coefficients in December is close to 0.9. This indicates that the VAE-GAN performs better in hourly ozone forecasting in December.

Figure 5 illustrates the spatial distribution of temporal-averaged RMSEs. It can be noted that most of the temporal-averaged RMSEs are below 20 $\mu g/m^3$ in June and 16 $\mu g/m^3$ in December, except for the North-east areas where the largest errors exist due to complex factors including human activities, meteorology and industries influencing the ozone concentration. From the cumulative distribution of temporal-averaged RMSEs in both months (Figure 5c), it is shown that the temporal-averaged RMSEs of 80% points are smaller than 25 $\mu g/m^3$. In addition, it is clear that the mean values of temporal-averaged RMSEs are 17.23 $\mu g/m^3$ and 9.88 $\mu g/m^3$ for June and December respectively.

3.2. Comparison of Hourly Ozone Forecasting Between VAE-GAN and the Physical Model-NAQPMS

In this section, to further demonstrate the forecasting performance of the VAE-GAN, the predictive accuracy of the VAE-GAN is evaluated with an existing air quality physical model-NAQPMS. The comparison of ozone results obtained from the VAE-GAN and NAQPMS has been undertaken through (a) the spatial distribution and temporal variation of ozone concentrations over China (Section 3.2.1), where the reanalysis data is considered as the 'true' reference; and (b) ozone concentration at monitoring stations (Section 3.2.2), where the monitoring measurement is used as the 'true' reference.

3.2.1. Spatio-Temporal Distribution of Ozone Forecasting Compared to Physical Modeling

The spatial distribution of hourly ozone forecasting results obtained from the VAE-GAN and physical model-NAQPMS at lead-times $t = 40, 60\ hr$ is plotted in Figure 6. We can see that the VAE-GAN and physical model-NAQPMS exhibit almost the same level of acceptable performance for hourly ozone forecasting. Compared with the physical model-NAQPMS, the VAE-GAN can better capture the spatial characteristics of high ozone concentrations in densely populated areas in China.

From Figure 6, we note that the spatio-temporal distributions of ozone concentrations have a high spatial heterogeneity across China. For further analysis of temporal variation, the whole country is thus divided into six

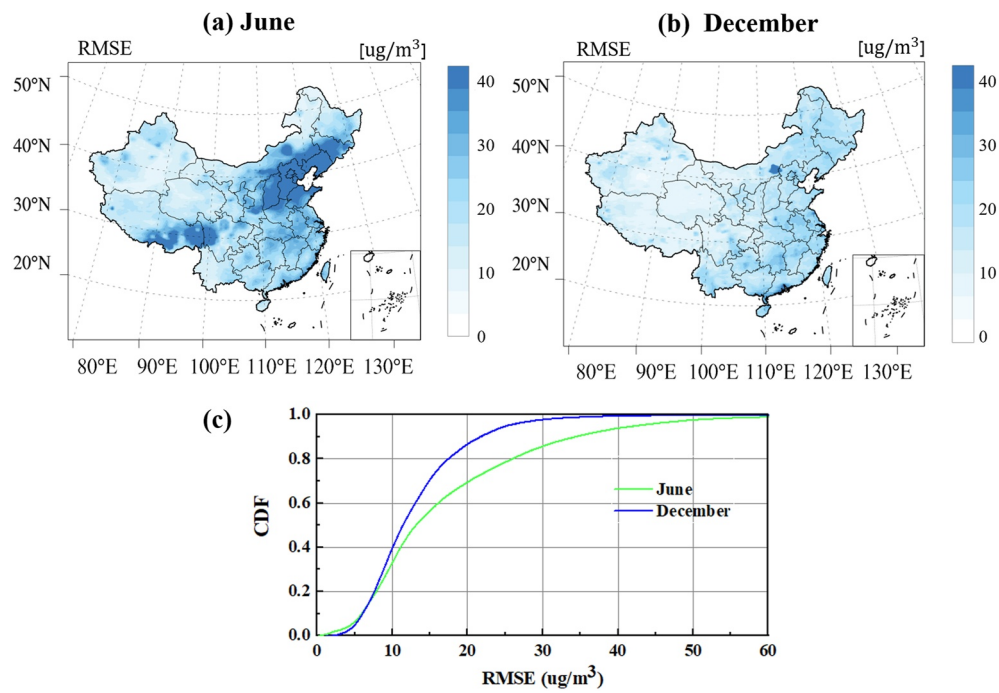

Figure 5. The spatial distribution of temporal-averaged RMSEs and the comparison of the cumulative density functions for the temporal-averaged RMSEs in June and December in 2018.

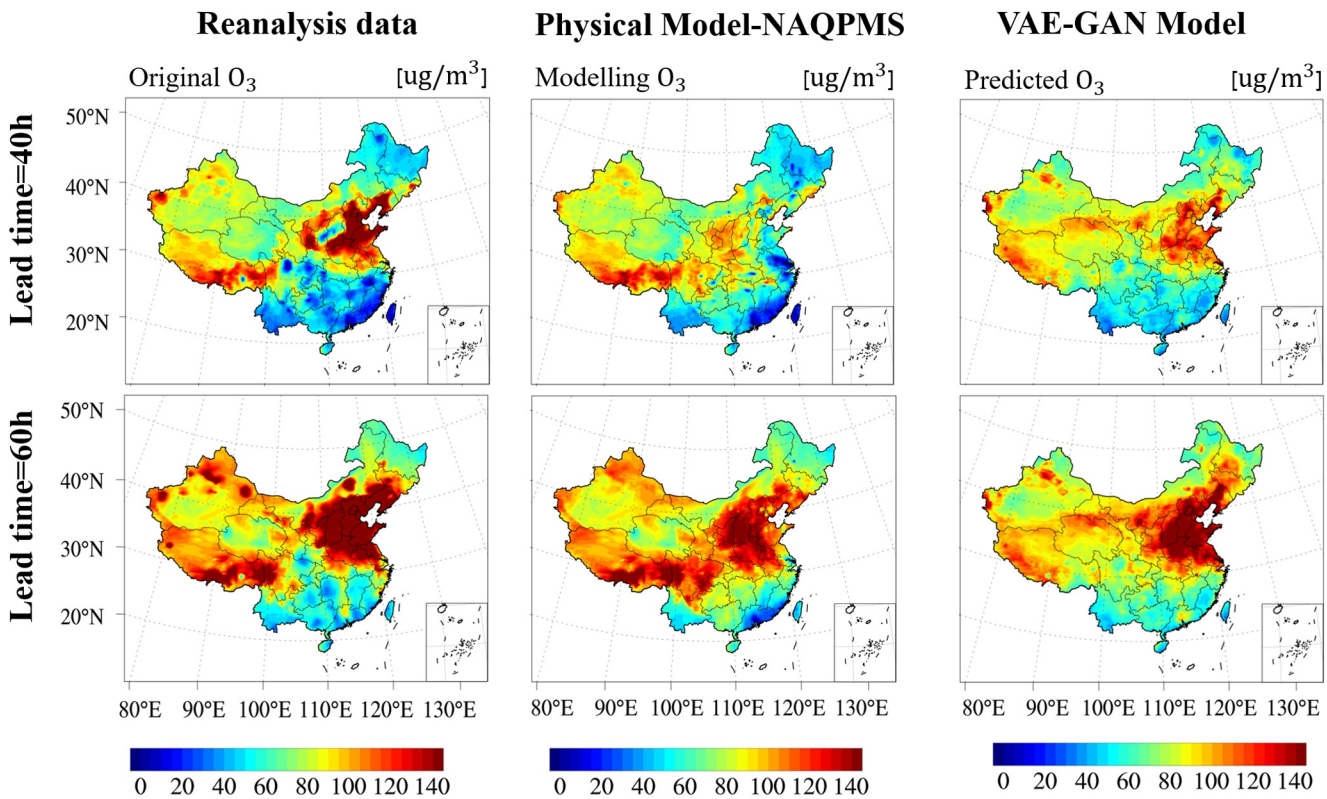

Figure 6. Comparison of the spatial distribution of ozone fields obtained from the VAE-GAN model (right), physical model-NAQPMS (middle) and the reanalysis data (left) at lead-times $t = 40, 60\ hr$ in June of 2018 respectively.

regions: the North China Plain (NCP), Northeast China (NE), Southwest China (SW), Southeast China (SE), Northwest China (NW), and Central China (shown in Figure 7). Figure 7 shows the spatial-averaged RMSEs of predicted results in six regions during the predictive period of June in 2018. It is noted that the predicted results from both the physical model and VAE-GAN have a low value of RMSE (below 30 $\mu g/m^3$) in the NW region. The highest RMSE of predicted ozone concentration occurs in the NCP region (Figure 7c), which is consistent with the findings in Zhan et al. (2018). The spatial heterogeneity of ozone concentrations in summer is associated with the spatial pattern of humidity conditions. The drier condition in the NCP region leads to a high-level accumulation of ambient ozone. The predicted ozone concentrations from both the physical model-NAQPMS and VAE-GAN are underestimated in the NCP, thus resulting in the highest RMSE. However, the VAE-GAN model generally performs better than the physical model-NAQPMS in the NCP region, which can be seen in Figure 6. In other regions (including NE, Central, SE and SW regions), the performance of the VAE-GAN model is comparable with the physical model-NAQPMS.

3.2.2. VAE-GAN-Based Ozone Forecasting Compared to Physical Modeling and Observations at Monitoring Stations

The performance of the VAE-GAN in spatio-temporal forecasting is further evaluated by comparing the predicted ozone concentrations with those from the physical model-NAQPMS and the observations at monitoring stations. Here we choose three study regions in China: JJJ, YRD and Pearl River Delta region (PRD; shown in Figure 8), which are densely populated regions in China. Figure 8 shows the time series of hourly predicted and monitoring ozone concentrations during 22–30 June of 2018 in the major cities located in the three regions. It is noted that the ozone concentration is gradually increasing in the morning and reaches a peak at noon, and has a low value during the night time. This can be explained as follows. The formation of ozone is involved in the processes of photochemical reactions, advection, entrainment, and deposition. The entrainment of ozone into the atmospheric boundary layer greatly contributes to the increase of ozone concentration at day time (Freire et al., 2017), while

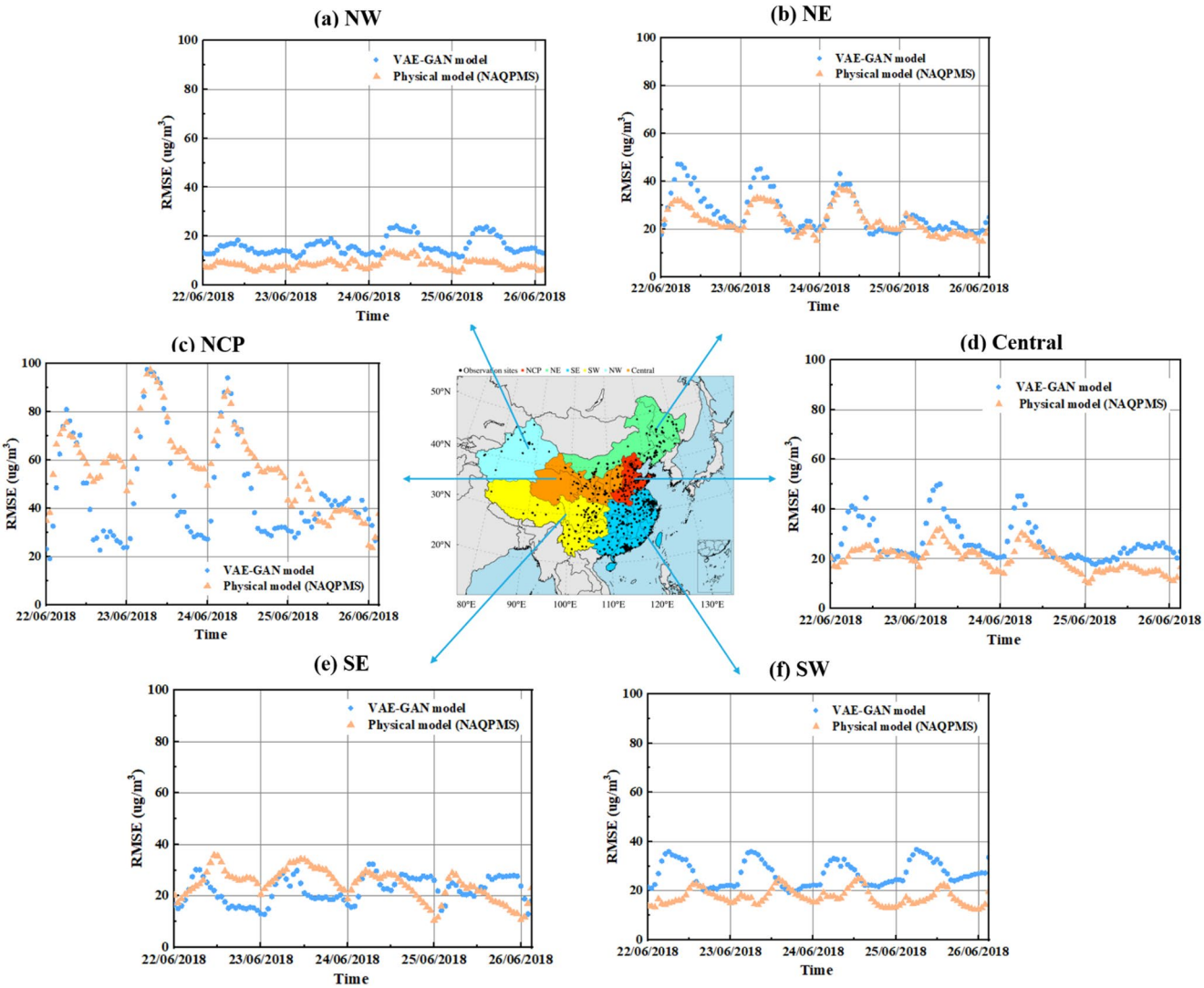

Figure 7. The spatial-averaged RMSE of predictive results from the VAE-GAN and physical model-NAQPMS during predictive period of June in 2018. There are six regions: The North China Plain, Northeast China, Southwest China, Southeast China, Northwest China, and Central China (refer to Kong et al. (2021)). The reanalysis data is used as 'true' reference.

due to the stable nocturnal boundary layer caused by radiative cooling at night time, the surface ozone concentrations typically reach a minimum level by nitrogen oxide titration and dry deposition (H. Liu, J. Liu, et al., 2020).

We can see that the predicted ozone by the VAE-GAN model is largely in good agreement with observations during the predictive period (22-30 June in 2018). It is noticed the VAE-GAN performs better than the physical model-NAQPMS in Shanghai, Nanjing, Hangzhou, Hefei in the YRD region. The discrepancy between the VAE-GAN predicted results and observations occurs when the peak values of ozone drastically increase at day time due to large uncertainties in traffic, emission inventory and meteorological conditions. To better capture the peak, the VAE-GAN-LSTM model is introduced in Appendix A. In Figure A1, the predicted ozone shows a reduction of underestimation in the ozone peaks compared to that using the VAE-GAN model, and exhibits a good match with the observations in cities with high ozone concentrations, such as Beijing, Tianjin, Shijiazhuang, Nanjing and Hefei.

The performance of the machine learning-based modeling and the physical model has further been evaluated using the statistical error indicators in Equation 5 (Emery et al., 2017) and the corresponding results are shown in Table 4. The ozone concentrations predicted by the VAE-based models mostly have a smaller value of RMSE,

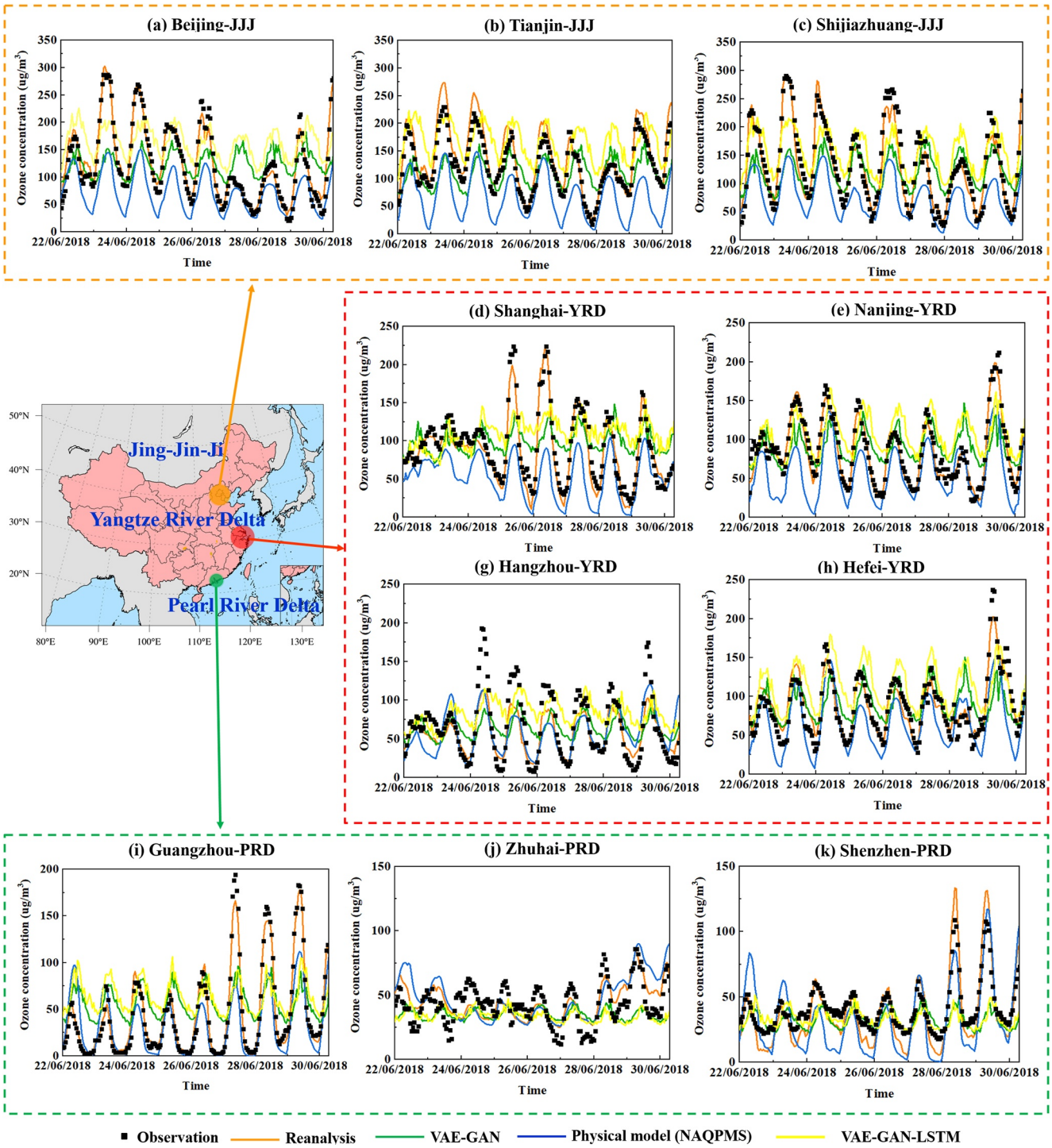

Figure 8. Ozone time series at the cities in three densely populated regions (Jing-Jin-Ji, Yangtze River Delta, and Pearl River Delta) using the observations, reanalysis data, physical model-NAQPMS, VAE-GAN and VAE-GAN-LSTM models.

NMB and NME than that from the physical model-NAQPMS in most of regions in China. The NMB and NME values of predicted results in the JJJ and YRD regions largely lie within ±22% and 25% respectively for the VAE-GAN and VAE-GAN-LSTM models, while those are beyond ±35% (NMB) and 35% (NME) for the physical model-NAQPMS. The performance of the VAE-GAN model is thus comparable or better than that of the physical model-NAQPMS in the JJJ and YRD regions. In the PRD region, the zone concentration mostly

Table 4
Error Analysis of Results From the Reanalysis Data, the Physical Model-NAQPMS, VAE-GAN and VAE-GAN-LSTM Models Where the Observations in June of 2018 are Taken as References (True Values)

		Correlation Coefficient				RMSE (ug/m^3)			
Zone		Reanalysis	Physical model (NAQPMS)	VAE-GAN	VAE-GAN-LSTM	Reanalysis	Physical model (NAQPMS)	VAE-GAN	VAE-GAN-LSTM
Jing-Jin-Ji	Beijing	0.98	0.83	0.52	0.57	20.02	63.38	43.34	57.98
	Tianjin	0.97	0.91	0.68	0.76	18.62	72.33	56.93	53.55
	Shijiazhuang	0.97	0.75	0.50	0.34	13.09	50.39	40.81	46.28
Yangtze River Delta	Shanghai	0.97	0.83	0.55	0.56	9.68	38.56	34.97	39.52
	Nanjing	0.95	0.74	0.58	0.37	22.12	29.64	35.89	42.68
	Hangzhou	0.95	0.79	0.54	0.58	13.24	32.41	33.05	40.29
	Hefei	0.99	0.83	0.58	0.48	6.98	30.21	42.25	46.38
Pearl River Delta	Guangzhou	0.98	0.82	0.42	0.36	10.90	18.18	18.66	18.64
	Shenzhen	0.68	0.46	0.27	0.23	11.82	17.73	17.05	17.54
	Zhuhai	0.49	0.37	0.33	0.52	42.13	52.11	21.59	19.31
		NMB (%)				NME (%)			
Zone		Reanalysis	Physical model (NAQPMS)	VAE-GAN	VAE-GAN-LSTM	Reanalysis	Physical model (NAQPMS)	VAE-GAN	VAE-GAN-LSTM
Jing-Jin-Ji	Beijing	−12.18	−45.58	−18.30	23.29	12.88	45.58	23.95	29.23
	Tianjin	−5.65	−46.85	−22.62	0.87	9.38	46.85	28.29	22.76
	Shijiazhuang	−6.23	−44.04	−18.88	−11.96	8.69	44.04	21.96	19.25
Yangtze River Delta	Shanghai	−3.70	−34.30	−18.68	−3.07	6.94	34.69	23.86	19.60
	Nanjing	−26.95	−29.85	−37.45	−22.40	26.95	30.44	37.78	25.66
	Hangzhou	−3.75	−24.21	−14.98	7.24	10.09	26.27	21.77	24.10
	Hefei	−4.87	−42.93	−48.18	−46.01	5.45	42.93	48.18	46.01
Pearl River Delta	Guangzhou	21.88	−4.26	−59.48	−60.34	21.88	14.50	59.48	60.34
	Shenzhen	−2.91	0.52	−56.91	−63.20	2.91	2.42	56.91	63.20
	Zhuhai	22.42	27.09	−13.13	−7.46	23.99	28.83	14.63	11.82

has a low value of <100 $\mu g/m^3$ compared to that in the JJJ and YRD regions. The variation of ozone in the PRD region is linked to summer monsoon (Shao et al., 2009) which has not currently been considered in the machine learning-based models, but in physical modeling. This can explain why the performance of the physical model is better than the VAN-based models in the PRD region.

Generally speaking, the VAE-GAN model developed in this study is relatively reliable when used for ozone forecasting during a long lead-time. Our results suggest that the well-trained VAE-GAN has captured the spatio-temporal characteristics of the ozone concentration variation. In comparison to the physical model-NAQPMS, the VAE-GAN model can provide a high efficient ozone forecast while the same level of accuracy is achieved. This highlights the potential of the VAE-GAN model to be applied in conjunction with many real-time applications, for example, early warning for air pollution emergencies.

3.3. Daily Ozone Forecasting Analysis Over China

In this section, the capability of the VAE-GAN is further demonstrated in daily ozone forecasting of the year 2018 in JJJ, YRD and PRD regions. The reanalysis datasets are split into the training dataset from the year 2013–2017 and the predictive dataset in 2018.

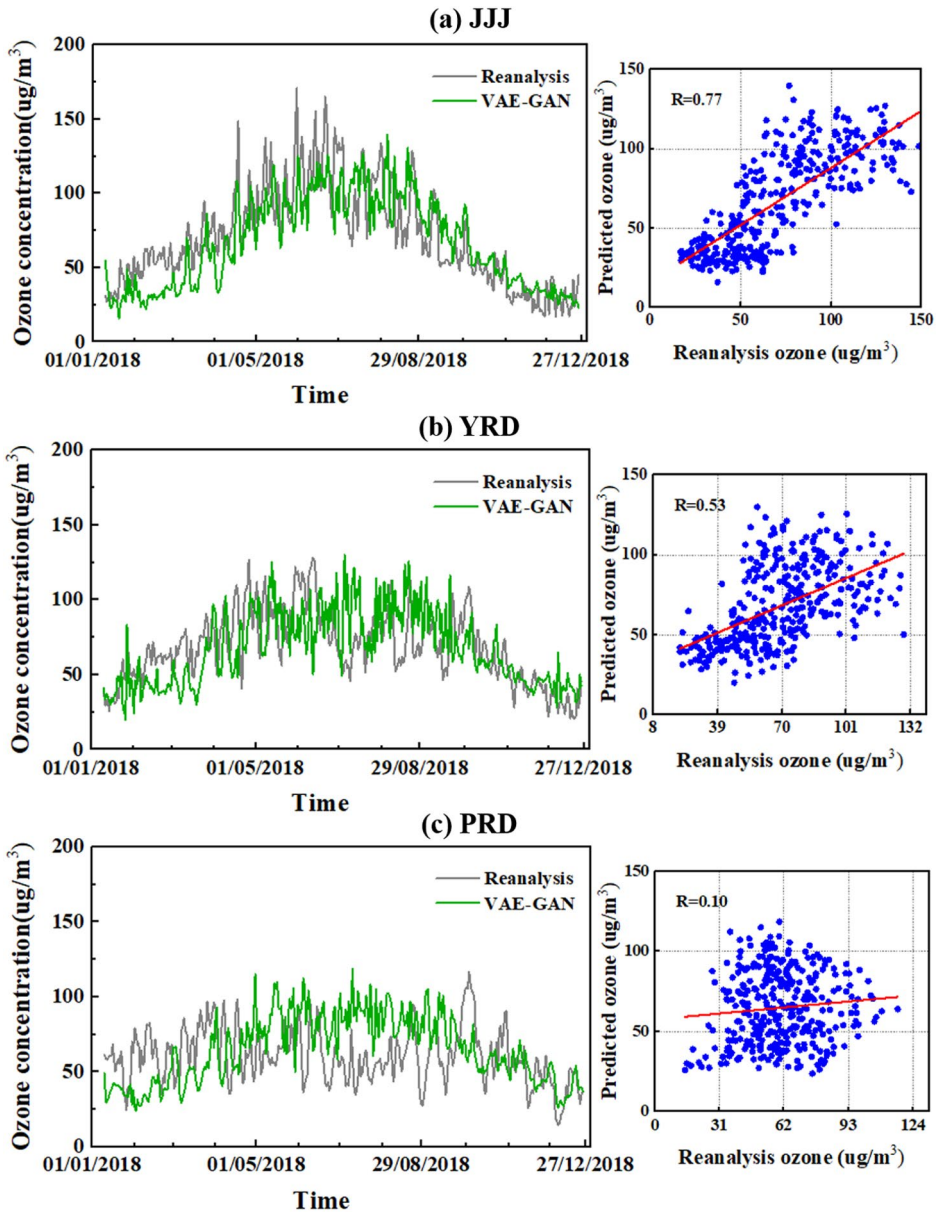

Figure 9. A comparison of the temporal variation of spatial-average daily ozone concentrations from the VAE-GAN model and the reanalysis data in 2018.

Figure 9 shows the temporal variation of spatial-average daily ozone concentrations over the JJJ, YRD, and PRD regions during the 2018 predictive period. As shown in Figure 9, it can be noticed that the predicted ozone concentration exhibits good agreement with the reanalysis data in the JJJ and YPR regions. In the PRD region, the VAE-GAN model captures the daily variability, but overestimates the ozone in the summer. This could be explained by the responses between the meteorological conditions and ozone. In the regions of JJJ and YRD, the daily variation of ozone concentration shows the seasonal variation (an inverse U-shaped trend, which is highly related to the temperature; Yang et al., 2020), where the highest and lowest ozone concentrations occur in Summer and Winter respectively. However, the trend of the ozone seasonal variation is weak in South China, including the PRD region. This is due to the warmer temperatures throughout the year and the Asian Summer monsoon, which brings cloudy and raining weather, marine air and strong deep convection and other unfavorable factors for ozone formation and accumulation (Lu et al., 2018; Ma & Yin, 2021; Yang et al., 2020). It is a challenge in existing air quality models when taking into account these complex meteorological factors. This can

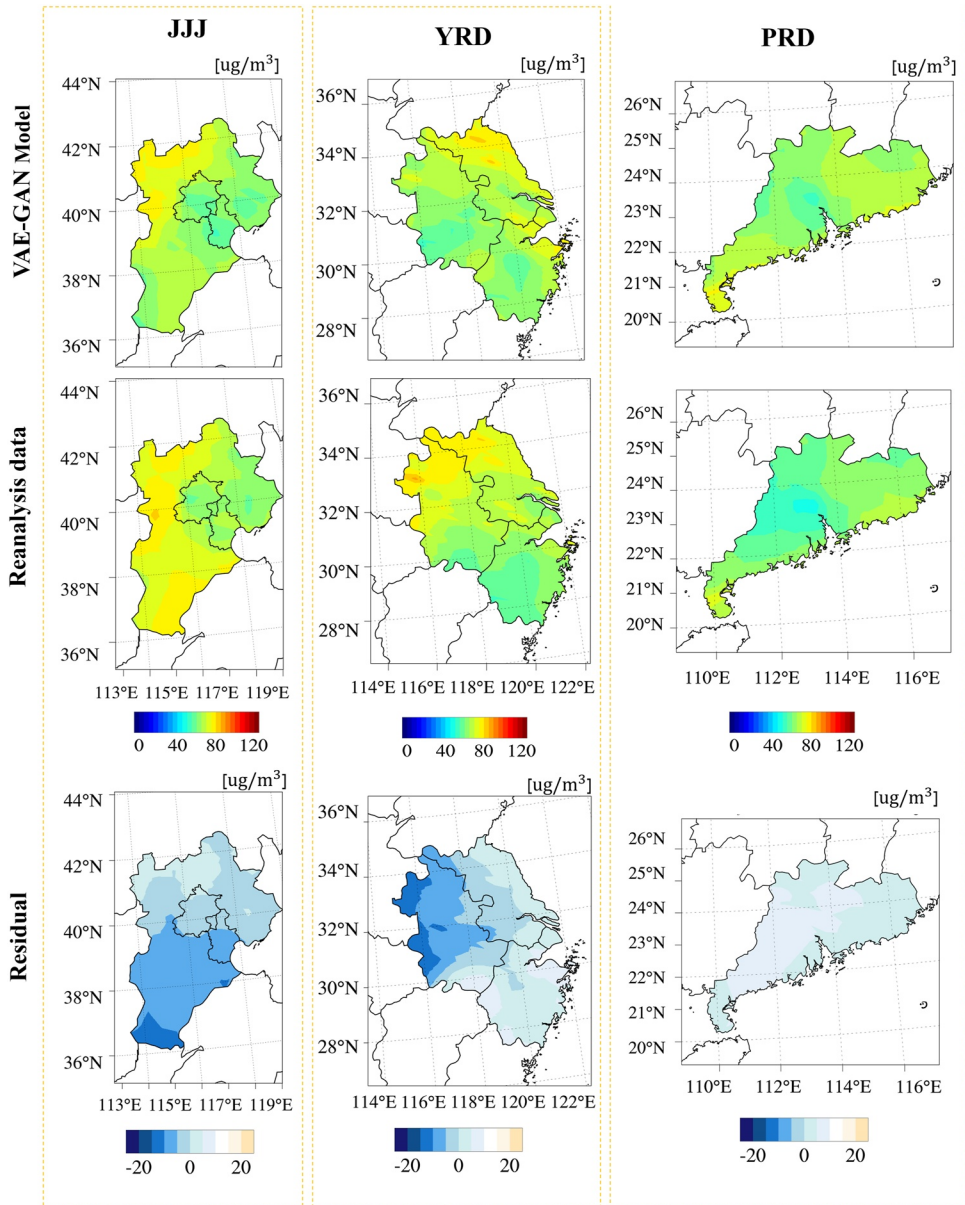

Figure 10. A comparison of the spatial distribution of temporal-average daily ozone concentrations from the VAE-GAN model and the reanalysis data in the Jing-Jin-Ji, Yangtze River Delta and Pearl River Delta regions.

be seen in Table 4 that a poor correlation-prevails (average 40%) between the predicted results and monitoring data in the PRD exists for both the proposed VAE-GAN and physical model-NAQPMS models. Coupling of the complex meteorological conditions (e.g., monsoon) into the VAE-GAN training process will be explored to improve predictive accuracy in future studies.

Figure 10 shows the spatial distribution of temporal-average daily ozone concentrations over the predictive period (the whole year of 2018). The predicted results from the VAE-GAN are generally comparable to the reanalysis data in JJJ, YRD and PRD regions. It can be seen that the VAE-GAN model captures a large part of the spatial distribution of ozone concentrations. To further evaluate the VAE-GAN results, the relative errors of the predicted results over the three regions are presented in Figure 11. Reasonable predictive accuracy is largely achieved with a low relative error of around ±10%. Overall, the VAE-GAN model skilfully captures the spatial and temporal variation of ozone concentration during the predictive period.

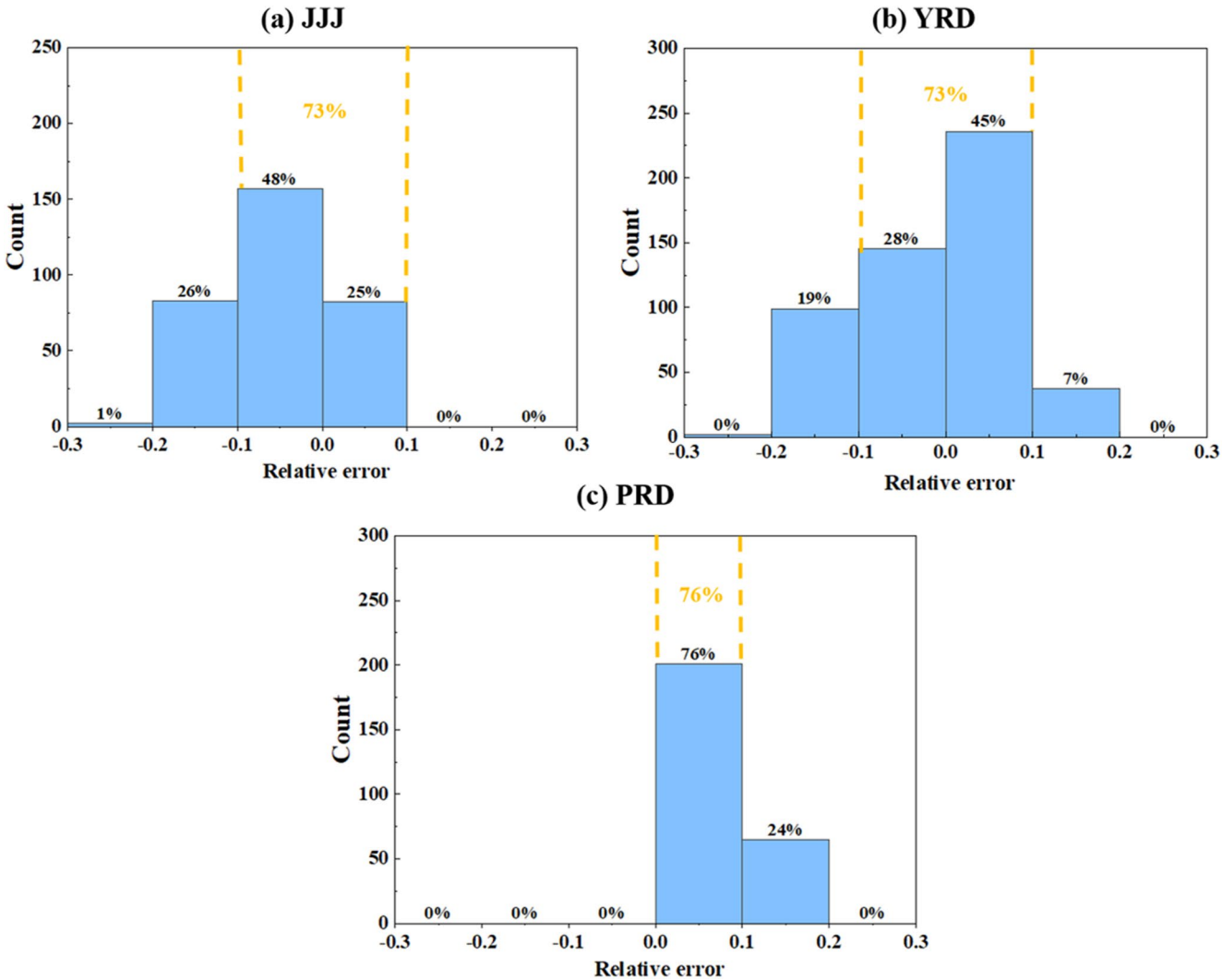

Figure 11. A relative error histograms of the temporal-average daily ozone concentrations between the VAE-GAN model and the reanalysis data over the whole year of 2018. The blue bins aggregate the respective grid points over three regions. Numbers on top of the bins are percentage values with respect to the total number of grid points. Yellow numbers indicate the area proportion of the globe, where values of relative error are within the ±10% interval.

3.4. Impact of Meteorological Factors

In our study cases described above, the temperature is considered as one of model inputs since it is the top ozone-related meteorological condition. As known, other meteorological factors (relative humidity and wind speed) are also important in ozone formation and accumulation. For example, low humidity causes a reduction of cloud cover which contributes to photochemical reaction (Chen et al., 2020). A high wind speed has a negative impact on ozone concentration by transporting and cleaning air pollutants out of the study areas (Z. B. Wang, Li, 2020). In this section, two deep learning models (the VAE-GAN and the deep convolutional GAN (DCGAN)) are used for quantifying the impact of the meteorological factors on daily ozone forecasting during the validation period of 2018 and 2019. The model inputs are the meteorological conditions (temperature, relative humidity and wind speed) at the predictive time level and historical ozone concentrations while the targeted output is the ozone concentration at the predictive time level. The reanalysis datasets are used for training and validation purposes: the training datasets compassing the time period between years 2013 and 2017 and validation datasets from 2018 to 2019. The densely populated JJJ region in China is selected as the study area. We estimate the performance of the VAE-GAN and DCGAN models in daily forecasting with consideration of: (a) no meteorological factors (no met); (b) temperature (temp); (c) temp + relative humidity (rh), and (d) temp + rh + wind).

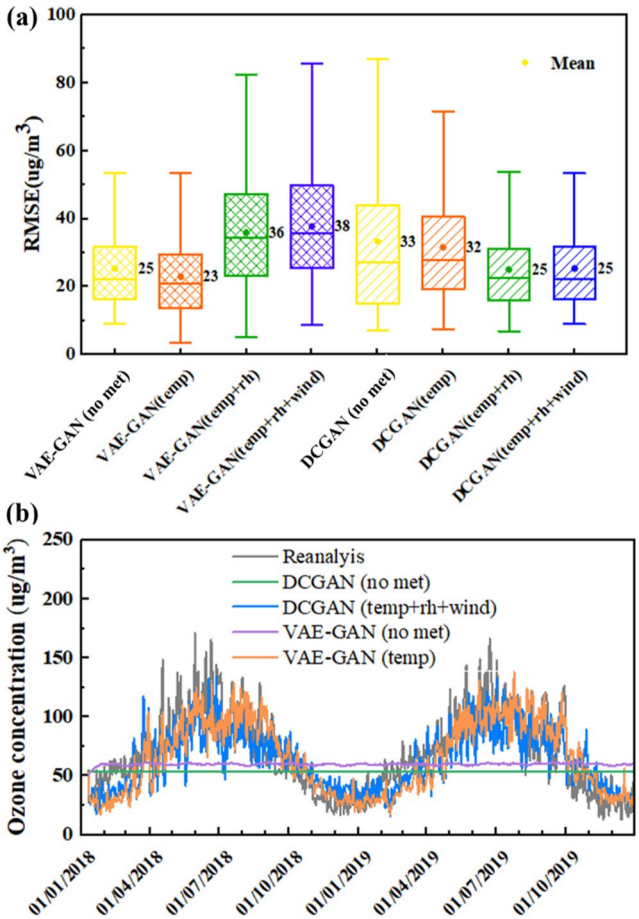

Figure 12. Predictive daily results from the cases of VAE-GAN and deep convolutional GAN model with/without the introduction of the meteorological factors into the model inputs, where the densely populated Jing-Jin-Ji region in China is selected as the study area and the predictive period is from the year 2018–2019. *Note.* That in the case of no meteorological factors the models only take the last-timestep ozone as the model input.

The spatial-averaged RMSEs of predicted ozone concentrations are plotted in Figure 12a while the predicted daily ozone concentrations are shown in Figure 12b. It is shown in Figure 12b that without the introduction of the meteorological factors, both of the VAE-GAN and DCGAN models fail to capture the temporal variation of daily ozone concentrations from the year 2018 and 2019.

In DCGAN modeling, it is clearly seen in Figure 12a that the more meteorological factors are included in the model inputs, the smaller the spatial-averaged RMSEs of daily ozone results. The DCGAN reaches its best performance when the impact of temperature, relative humidity and wind speed is considered.

In VAE-GAN modeling, the introduction of temperature plays a crucial role in improving the accuracy of ozone forecasting. The VAE-GAN with the consideration of temperature performs best among all the study cases. The VAE-GAN achieves the lowest RMSE value of 23 $\mu g/m^3$ in daily ozone concentration prediction in the JJJ region from the year 2018–2019. In Figure 12b, it can be noted that by comparing with the reanalysis data, the VAE-GAN model can better represent the peaks of ozone concentrations than the DCGAN model. However, it is also noticed that in contrast to the DCGAN model, the RMSE of predicted results is decreased when more meteorological factors (relative humidity and wind speed) are considered in the model inputs. This is due to the limitation of the VAE-GAN model. For multiple variable inputs, the quality of generated outputs suffers from the generator (decoder $\mathcal{P}$) attempting to cover all data manifolds in the data space, thus leading to a poor model performance (Pandeva & Schubert, 2019). Coupling of multiple meteorological inputs with the VAE-GAN will be investigated in our future work.

Overall, the temperature or other meteorological factors play a dominant role in improving the accuracy of ozone forecasting, for example, reducing uncertainties and accumulative error with every step of ozone forecasting.

4. Conclusions

In this study, a machine learning-based forecasting framework is implemented for efficient and accurate real-time (hourly and daily) ozone forecasting in China. A hybrid deep learning VAE-GAN approach is proposed to better explore the spatial and temporal features of ozone concentrations, where GAN has the capability of learning hierarchical feature representations of ozone concentrations while VAE is used for dimensionality reduction and avoiding model collapse. The hybrid VAE-GAN model consists of both adversarial training and real-time forecasting processes, which narrow the gaps between the generative and targeted ozone fields from the past to future states and provide accurate ozone forecasts for a long lead-time beyond the training period in a recursive way.

In our hybrid VAE-GAN ozone modeling, the training and validation datasets consist of hourly reanalysis datasets at a high spatial resolution of 15 × 15 km across China from 2013 to 2019. By using VAE, the high-dimensional size of input datasets on ozone and temperature spatial distributions is reduced by three orders of magnitude. It is evident that the VAE-GAN is an efficient tool for large data-driven predictions. Our results illustrate that the VAE-GAN model has captured the spatial and temporal evolution patterns of ozone concentrations during the predictive period of 2018 and 2019 in comparison to the reanalysis data, observations and physical model-NAQPMS. The VAE-GAN model can provide a high efficient ozone forecast while the same level of accuracy is achieved compared to the physical model-NAQPMS. However, it is also noticed that a low predictive accuracy of ozone concentrations occurs in some local regions where the daily variation of ozone concentrations exhibits

volatility. In future studies, it will be necessary to develop a more comprehensive forecasting scheme to improve the predictive accuracy by combining the VAE-GAN with data assimilation techniques for uncertainty reduction.

Appendix A: Model Comparison

In the section, the performance of our VAE-GAN model is evaluated against a VAE-GAN-LSTM model, which introduces the convolutional long short-term memory (ConvLSTM) layer (Xingjian et al., 2015) into the encoder and decoder of the VAE-GAN. The comparative plots of hourly ozone forecasting results obtained from the VAE-GAN and VAE-GAN-LSTM models during the predictive period of June in 2018 are shown in Figure A1. It can be observed in Figure A1(a) that a relatively good agreement between the observed and forecasting hourly ozone concentration is achieved, while the arrival time of forecasting ozone peaks is slightly delayed. We can see that during the validation period of days 22–26 in June 2018, the VAE-GAN-LSTM model can better predict the ozone peaks than the VAE-GAN model, where the reanalysis data is considered as the reference. However, the VAE-GAN-LSTM model fails to capture the low ozone concentration at night, thus having higher the RMSEs than that of the VAE-GAN model in Figure A1(b). This situation also reflects on the spatial distribution of ozone forecasts as shown in Figure A2. Generally speaking, the use of LSTM in the VAE-GAN can well capture the high ozone concentration although the RMSE of results from the VAE-GAN-LSTM is slightly higher than that of the VAE-GAN. It is thus suggested that the VAE-GAN-LSTM model has the potential for hourly high ozone forecasting in China.

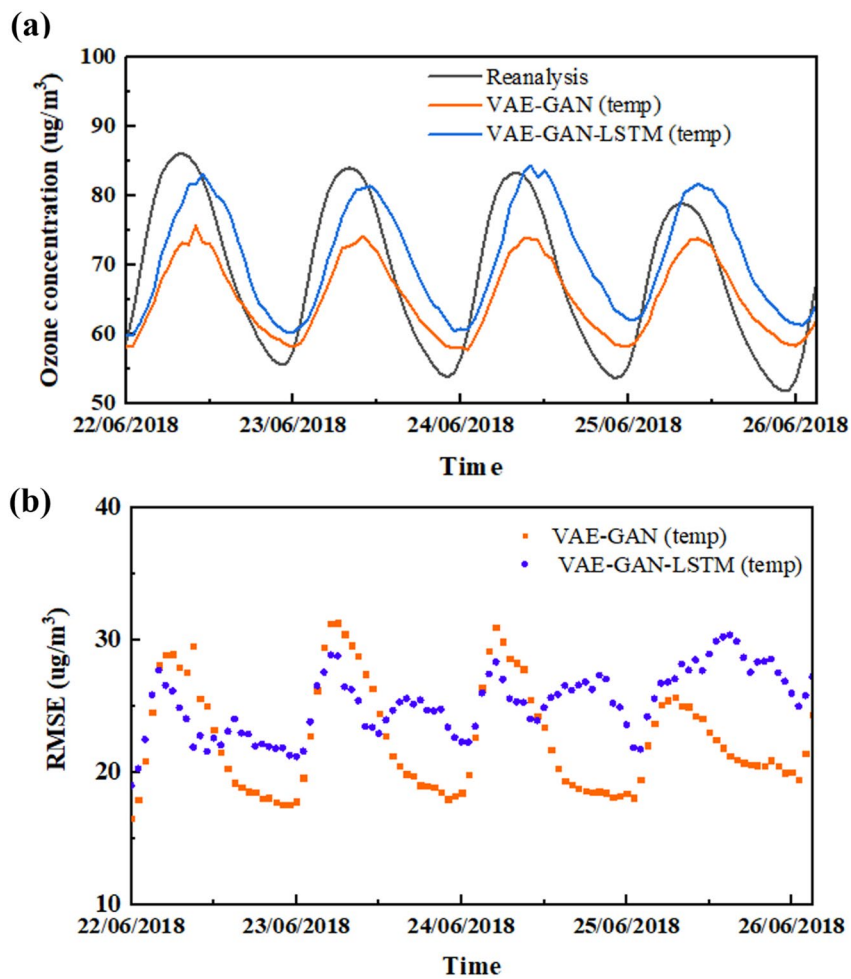

Figure A1. Spatial-averaged ozone concentration variation during predictive period of June in 2018. The VAE-GAN and VAE-GAN-LSTM models take both the temperature and last-step predictive ozone as the model inputs. The VAE-GAN-LSTM model adopts the convolutional long short-term memory (ConvLSTM) layer into the encoder and decoder of the VAE-GAN.

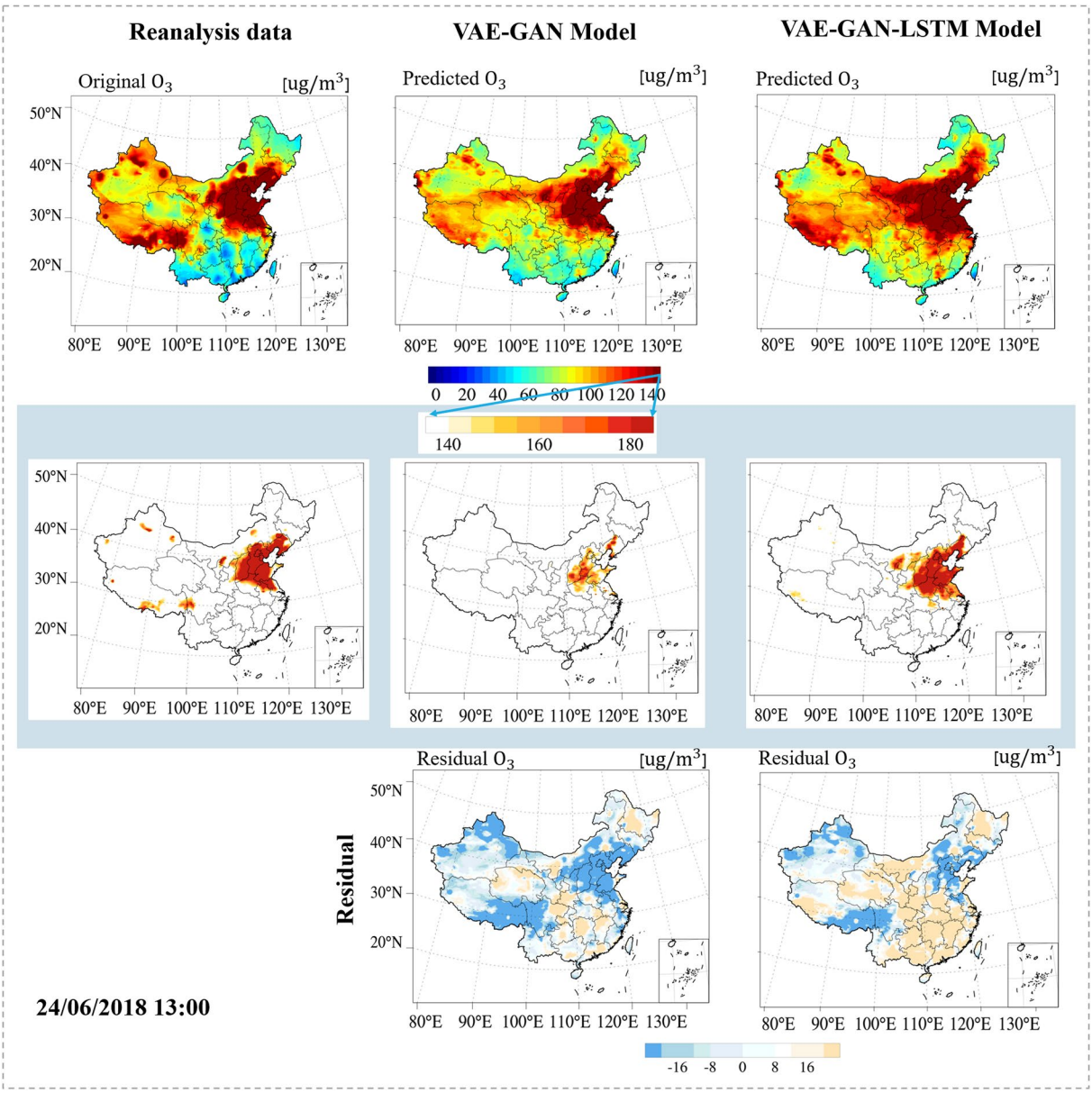

Figure A2. Comparison of the spatial distribution of ozone fields obtained from the VAE-GAN model and the VAE-GAN-LSTM at the lead-time of 60 *hr* in June of 2018.

Appendix B: Ozone Forecasting Using NO_2

NO_2 is the primary element involved in the process of forming ozone. When sunlight interacts with NO_2, the gas molecules break down and release monoatomic oxygen (O) atoms which then react to diatomic oxygen (O_2) molecules and form the trioxygen molecule known as ozone (Ezimand & Kakroodi, 2019). Here the impacts of the precursors (here, NO_2) on ozone concentrations have been explored using the proposed VAE-GAN model.

B1. Hourly Ozone Forecasting Using NO_2 From Reanalysis Datasets

In real-time ozone forecasting in realistic scenarios, both NO_2 and ozone are the unknown variables (outputs) to be predicted/solved using either the air quality model or the AI-based model (here VAE-GAN). Thus it is not suitable to consider NO_2 as the model input in real-time ozone forecasting. Even so, it is worth exploring the

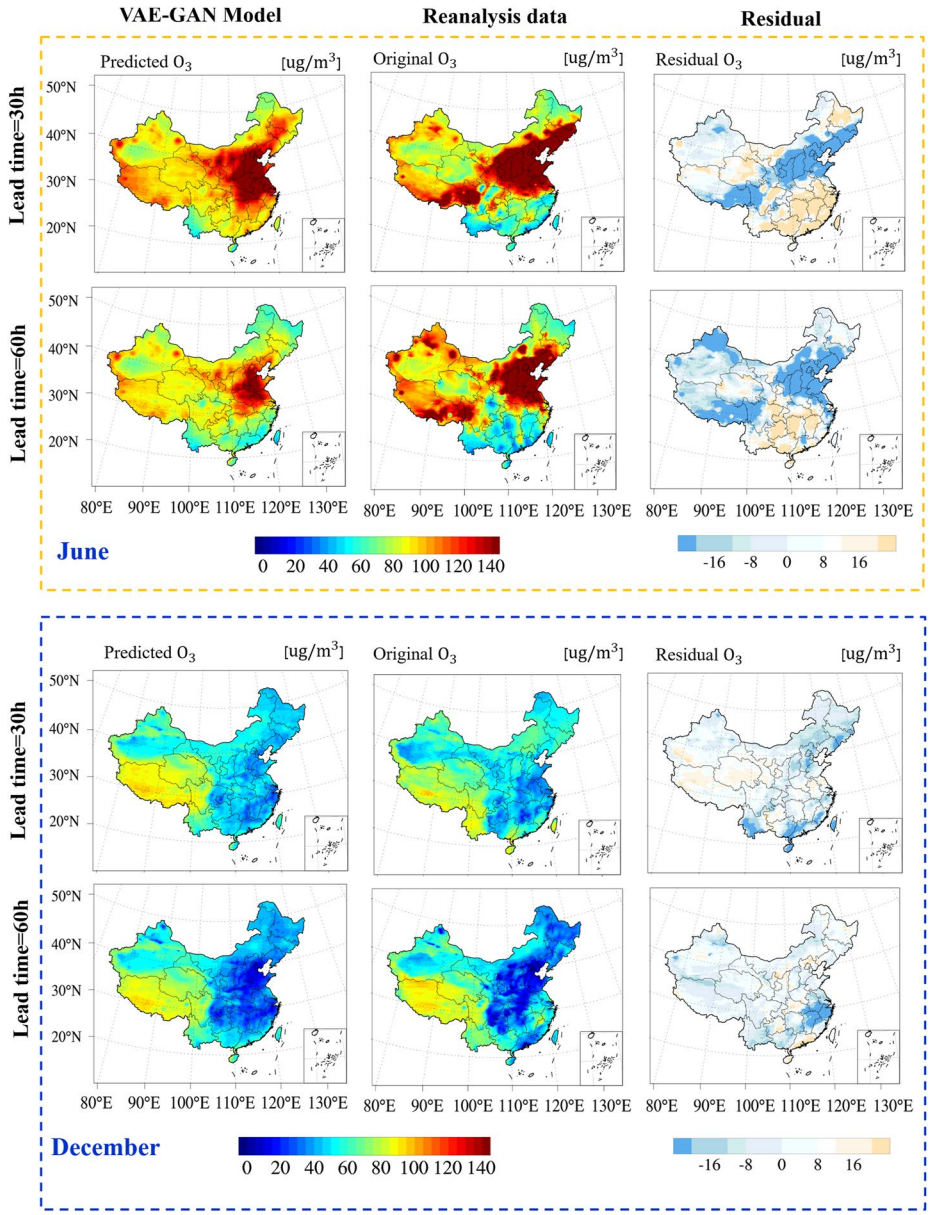

Figure B1. Comparison of the spatial distribution of ozone fields obtained from the VAE-GAN model (left) and the reanalysis data (middle) at the lead-times of 30 and 60 *hr* in June and December in 2018 respectively.

relationship between NO_2 and ozone using the VAE-GAN model based on history ozone and NO_2 datasets, where the NOx concentration is the model input while the ozone concentration is the model output.

As introduced in Section 2.1.1, the hourly NO_2 and ozone datasets are collected from the reanalysis CAQRA datasets. We used the hourly datasets in 2017 for training, and 2018 for validation. The spatial performance of the VAE-GAN model for hourly ozone forecasting in June and December of 2018 is shown in Figure B1. It can be seen that the VAE-GAN model captures a large part of the spatial distribution of ozone concentration at lead-times $t = 30, 60hr$ in December. The last row of Figure B1 shows the residual maps between ozone maps predicted by the VAE-GAN and reanalysis data. In June, the predicted ozone concentration is slightly overestimated over the south of China at different lead-times, while slightly underestimated over the NCP region.

The RMSE and correlation coefficient of predictive results from the VAE-GAN are shown in Figure B2. It can be noticed that the RMSE values in June and December in 2018 are generally below 17 $\mu g/m^3$ during the whole

forecasting period. The average RMSE values are 14.56 $\mu g/m^3$ and 8.17 $\mu g/m^3$ for June and December respectively. The corresponding correlation coefficient values are shown in Figure B2(b). We note that in particular, models which accurately capture the regime behavior will also show good correlation statistics when averaged over a long lead-time. Most of the correlation coefficients are around or beyond 0.8 during the forecasting period. It is worth noting that the magnitude of the correlation coefficient in December is close to 0.9. This means the VAE-GAN performs better in hourly ozone forecasting in December.

B2. Impact of NO_2 Emissions on Ozone Forecasting

In realistic scenarios, the local NOx emission from massive anthropogenic emissions, industries, or transportation is another factor impacting ozone forecasting. In this work, the NOx and VOC emissions were considered as inputs, but removed during the selection of driving variables. There are two reasons for this. First, the available NOx emission data (source) are averaged monthly or daily. The daily variation of NOx emission remains the same during the same month. The NOx emissions are constant at a daily temporal scale. Thus, the hourly temporal variation of ozone concentrations will not be affected by a constant NOx emission during the training process. However, the emission of NOx varies over space, for example, a high emission in densely populated cities. The impact of the spatial variation of NOx emission is implicitly considered during the training process. The reason for this is that the training datasets (the reanalysis datasets) combine the physical modeling results, where NOx and VOC emissions have been considered in NAQPMS model (Kong et al., 2021) and observations. We speculate the predictive results can reflect the contribution of other driven factors (including NOx and VOC) on ozone concentration. To validate this hypothesis, we have undertaken one modeling experiment, where the average temperature ($27°C \pm 3°C$) in June is used as inputs everywhere over China. Can we expect the VAE-GAN models are able to predict the spatial distribution of ozone concentrations over China? The right column in Figure B3 shows the ozone concentration variation where the input temperature is around 300K ($27°C \pm 3°C$). We can see that the predictive ozone varies in space with the same temperature as model inputs. It is noted that the ozone variation over the regions can match well with the NO_2 emission (right column in Figure B3) in June, a high ozone concentration in the densely populated cities, mainly in the east of China. Therefore, the predictive ozone

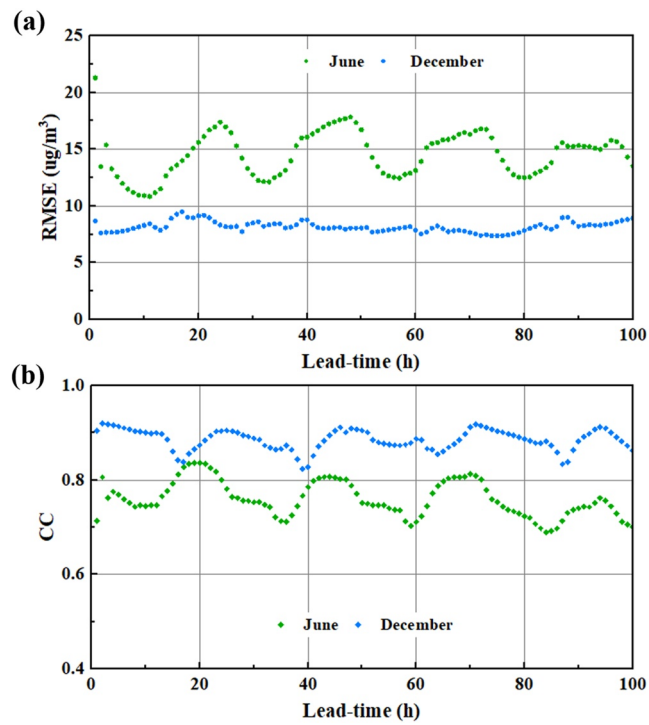

Figure B2. The RMSE and correlation coefficients of ozone concentration obtained from the VAE-GAN and the reanalysis data in two months in 2018.

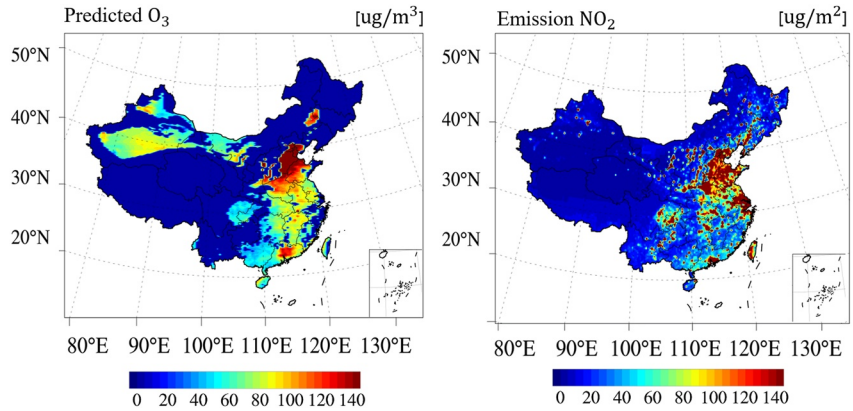

Figure B3. The predictive ozone (left) and emission of NO_2 (right) distribution of China in June of 2018.

in space has indicated the patterns of ozone formation with emission intensities associated with industrial and agricultural activities.

Second, in ozone forecasting, the importance of NOx and NMVOC emissions is lower than expected although the ozone formation is associated with NOx emissions. Zhan et al. (2018) pointed out that this may be due to regional transport or titration effects. NOx and NMVOC may be transported from urban to rural areas via advection, causing the location of ozone formation to occur some distance away from the emission source. NOx titration tends to counteract the importance of NOx emissions to ozone levels. In the work of Ma et al. (2021), they also found that ozone precursors such as NO_2 and satellite-based formaldehyde (HCHO) observations were found to play relatively weak roles in modulating ozone variations, though these factors were crucial to the ozone formation. In addition, Zhan et al. (2018) found that the inclusion of NOx and NMVOC into the machine learning model can introduce a noise which tends to be higher than the gained information, which is consistent with what occurred in our study.

Finally, we would like to further discuss the importance of the choice of driving variables in ozone forecasting. For machine learning-based ozone forecasting at stations, the model performance may be improved with increasing inputs (Su et al., 2020; Yafouz et al., 2021). In contrast, for spatio-temporal ozone forecasting, with the increase of the number of driving factors, the number of modeling parameters to be optimially estimated will significantly increase, especially on a large spatial scale. It would thus compromise all the drivers during the training process, thus weakening the relationship between the most important driver (e.g., temperature) and ozone concentrations. It is therefore suggested that only key-driven variables should be chosen for model prediction.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The VAE-GAN is a public model and its releases are accessible through this GitHub repository: https://github.com/eriklindernoren/Keras-GAN/tree/master/aae. The CAQRA dataset is available from Kong et al. (2021) (https://doi.org/10.11922/sciencedb.00053). The monitoring air pollution data used in this study was accessed through the Ministry of Ecology and Environment of the People's Republic of China (http://www.mee.gov.cn/).

Acknowledgments
This work was supported by the China Scholarship Council (No. 201806270238) and Engineering and Physical Sciences Research Council (EPSRC; MAGIC (EP/N010221/1), INHALE (EP/T003189/1), MUFFINS (EP/P033148/1) and PREMIERE (EP/T000414/1)) in the UK. The authors are grateful for the support of the Imperial College ICT service. We would like to acknowledge the reviewers and Editor for their in-depth perspicacious comments that contributed to improving the presentation of this paper.

References

Abadi, M. (2016). Tensorflow: Learning functions at scale. In *Proceedings of the 21st acm sigplan international conference on functional programming* (Vol. 51). https://doi.org/10.1145/3022670.2976746 SIGPLAN Not.

AlOmar, M. K., Hameed, M. M., & AlSaadi, M. A. (2020). Multi hours ahead prediction of surface ozone gas concentration: Robust artificial intelligence approach. *Atmospheric Pollution Research*, *11*(9), 1572–1587. https://doi.org/10.1016/j.apr.2020.06.024

Bloomer, B. J., Stehr, J. W., Piety, C. A., Salawitch, R. J., & Dickerson, R. R. (2009). Observed relationships of ozone air pollution with temperature and emissions. *Geophysical Research Letters*, *36*(9). https://doi.org/10.1029/2009gl037308

Bloomer, B. J., Vinnikov, K. Y., & Dickerson, R. R. (2010). Changes in seasonal and diurnal cycles of ozone and temperature in the eastern US. *Atmospheric Environment*, *44*(21–22), 2543–2551. https://doi.org/10.1016/j.atmosenv.2010.04.031

Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, *11*(1), 376–399. https://doi.org/10.1029/2018ms001472

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., et al. (2013). *API design for machine learning software: Experiences from the scikit-learn project*. arXiv preprint arXiv:1309.0238.

Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2020). Data-driven super-parameterization using deep learning: Experimentation with multiscale lorenz 96 systems and transfer learning. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2020MS002084. https://doi.org/10.1029/2020ms002084

Chen, Z., Jin, M., Deng, Y., Wang, J. S., & Huang, H. (2019). Improvement of a deep learning algorithm for total electron content maps: Image completion. *Journal of Geophysical Research: Space Physics*, *124*(1), 790–800. https://doi.org/10.1029/2018ja026167

Chen, Z., Li, R., Chen, D., Zhuang, Y., Gao, B., Yang, L., et al. (2020). Understanding the causal influence of major meteorological factors on ground ozone concentrations across China. *Journal of Cleaner Production*, *242*, 118498. https://doi.org/10.1016/j.jclepro.2019.118498

Cheng, M., Fang, F., Navon, I. M., & Pain, C. (2021). A real-time flow forecasting with deep convolutional generative adversarial network: Application to flooding event in Denmark. *Physics of Fluids*, *33*(5), 056602. https://doi.org/10.1063/5.0051213

Cheng, M., Fang, F., Pain, C., & Navon, I. M. (2020). An advanced hybrid deep adversarial autoencoder for parameterized nonlinear fluid flow modelling. *Computer Methods in Applied Mechanics and Engineering*, *372*, 113375. https://doi.org/10.1016/j.cma.2020.113375

Dutot, A. L., Rynkiewicz, J., Steiner, F. E., & Rude, J. (2007). A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions. *Environmental Modelling & Software*, *22*(9), 1261–1269. https://doi.org/10.1016/j.envsoft.2006.08.002

Emery, C., Liu, Z., Russell, A. G., Odman, M. T., Yarwood, G., & Kumar, N. (2017). Recommendations on statistics and benchmarks to assess photochemical model performance. *Journal of the Air & Waste Management Association*, *67*(5), 582–598. https://doi.org/10.1080/10962247.2016.1265027

Eslami, E., Choi, Y., Lops, Y., & Sayeed, A. (2019). A real-time hourly ozone prediction system using deep convolutional neural network. *Neural Computing & Applications*, 1–15. https://doi.org/10.1007/s00521-019-04282-x

Ezimand, K., & Kakroodi, A. (2019). Prediction and spatio-temporal analysis of ozone concentration in a metropolitan area. *Ecological Indicators*, *103*, 589–598. https://doi.org/10.1016/j.ecolind.2019.04.059

Feng, R., Zheng, H., Zhang, A., Huang, C., Gao, H., & Ma, Y. (2019). Unveiling tropospheric ozone by the traditional atmospheric model and machine learning, and their comparison: A case study in hangzhou, China. *Environmental pollution*, *252*, 366–378. https://doi.org/10.1016/j.envpol.2019.05.101

Freire, L., Gerken, T., Ruiz-Plancarte, J., Wei, D., Fuentes, J., Katul, G., et al. (2017). Turbulent mixing and removal of ozone within an amazon rainforest canopy. *Journal of Geophysical Research: Atmospheres*, *122*(5), 2791–2811. https://doi.org/10.1002/2016jd026009

Gagne, D. J., Christensen, H. M., Subramanian, A. C., & Monahan, A. H. (2020). Machine learning for stochastic parameterization: Generative adversarial networks in the Lorenz'96 model. *Journal of Advances in Modeling Earth Systems*, *12*(3), e2019MS001896. https://doi.org/10.1029/2019ms001896

Gonzalez, F. J., & Balajewicz, M. (2018). *Deep convolutional recurrent autoencoders for learning low-dimensional feature dynamics of fluid systems*. arXiv preprint arXiv:1808.01346.

Goodfellow, I., Pouget Abadie, J., Mirza, M., Xu, B., Warde Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).

Gu, Y., Li, K., Xu, J., Liao, H., & Zhou, G. (2020). Observed dependence of surface ozone on increasing temperature in Shanghai, China. *Atmospheric Environment*, *221*, 117108. https://doi.org/10.1016/j.atmosenv.2019.117108

Gulli, A., & Pal, S. (2017). *Deep learning with keras*. Packt Publishing Ltd.

Hoshyaripour, G., Brasseur, G., Andrade, M., Gavidia Calderón, M., Bouarar, I., & Ynoue, R. Y. (2016). Prediction of ground-level ozone concentration in são paulo, Brazil: Deterministic versus statistic models. *Atmospheric environment*, *145*, 365–375. https://doi.org/10.1016/j.atmosenv.2016.09.061

Hu, J., Chen, J., Ying, Q., & Zhang, H. (2016). *One-year simulation of ozone and particulate matter in China using WRF/CMAQ modeling system*. Atmospheric Chemistry and Physics, *16*(16), 10333–10350. https://doi.org/10.5194/acp-16-10333-2016

Jenkin, M. E., & Clemitshaw, K. C. (2000). Ozone and other secondary photochemical pollutants: Chemical processes governing their formation in the planetary boundary layer. *Atmospheric Environment*, *34*(16), 2499–2527. https://doi.org/10.1016/s1352-2310(99)00478-1

Jia, B., Dong, R., & Du, J. (2020). Ozone concentrations prediction in Lanzhou, China, using chaotic artificial neural network. *Chemometrics and Intelligent Laboratory Systems*, *204*, 104098. https://doi.org/10.1016/j.chemolab.2020.104098

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980.

Kong, L., Tang, X., Zhu, J., Wang, Z., Li, J., Wu, H., 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. https://doi.org/10.5194/essd-13-529-2021

Kumar, N., Middey, A., & Rao, P. S. (2017). Prediction and examination of seasonal variation of ozone with meteorological parameter through artificial neural network at NEERI, Nagpur, India. *Urban Climate*, *20*, 148–167. https://doi.org/10.1016/j.uclim.2017.04.003

Kumar, R., Naja, M., Pfister, G., Barth, M., Wiedinmyer, C., & Brasseur, G. (2012). Simulations over South Asia using the weather Research and forecasting model with chemistry (WRF-chem): Chemistry evaluation and initial results. *Geoscientific Model Development*, *5*(3), 619. https://doi.org/10.5194/gmd-5-619-2012

Laloy, E., Hérault, R., Jacques, D., & Linde, N. (2018). Training-image based geostatistical inversion using a spatial generative adversarial neural network. *Water Resources Research*, *54*(1), 381–406. https://doi.org/10.1002/2017wr022148

Lee, S., & You, D. (2019). Data-driven prediction of unsteady flow over a circular cylinder using deep learning. *Journal of Fluid Mechanics*, *879*, 217–254. https://doi.org/10.1017/jfm.2019.700

Li, Z., Meier, M. A., Hauksson, E., Zhan, Z., & Andrews, J. (2018). Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical Research Letters*, *45*(10), 4773–4779. https://doi.org/10.1029/2018gl077870

Liu, H., Liu, J., Liu, Y., Ouyang, B., Xiang, S., et al. (2020). Analysis of wintertime O_3 variability using a random forest model and high-frequency observations in Zhangjiakou—An area with background pollution level of the North China Plain. *Environmental Pollution*, *262*, 114191. https://doi.org/10.1016/j.envpol.2020.114191

Liu, R., Ma, Z., Liu, Y., Shao, Y., Zhao, W., & Bi, J. (2020). Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: A machine learning approach. *Environment International*, *142*, 105823. https://doi.org/10.1016/j.envint.2020.105823

Lu, X., Hong, J., Zhang, L., Cooper, O. R., Schultz, M. G., Xu, X., et al. (2018). Severe surface ozone pollution in China: A global perspective. *Environmental Science and Technology Letters*, *5*(8), 487–494. https://doi.org/10.1021/acs.estlett.8b00366

Ma, M., Yao, G., Guo, J., & Bai, K. (2021). Distinct spatiotemporal variation patterns of surface ozone in China due to diverse influential factors. *Journal of Environmental Management*, *288*, 112368. https://doi.org/10.1016/j.jenvman.2021.112368

Ma, X., & Yin, Z. (2021). Dipole pattern of summer ozone pollution in the east of China and its connection with climate variability. *Atmospheric Chemistry and Physics Discussions*, 1–23. https://doi.org/10.5194/acp-21-16349-2021

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., & Frey, B. (2015). *Adversarial autoencoders*. arXiv preprint arXiv:1511.05644.

Mo, Y., Li, Q., Karimian, H., Fang, S., Tang, B., Chen, G., & Sachdeva, S. (2020). A novel framework for daily forecasting of ozone mass concentrations based on cycle reservoir with regular jumps neural networks. *Atmospheric Environment*, *220*, 117072. https://doi.org/10.1016/j.atmosenv.2019.117072

Mosser, L., Dubrule, O., & Blunt, M. J. (2017). Reconstruction of three-dimensional porous media using generative adversarial neural networks. *Physical Review E*, *96*(4), 043309. https://doi.org/10.1103/physreve.96.043309

Pandeva, T., & Schubert, M. (2019). *Mmgan: Generative adversarial networks for multi-modal distributions*. arXiv preprint arXiv:1911.06663.

Shao, M., Zhang, Y., Zeng, L., Tang, X., Zhang, J., Zhong, L., & Wang, B. (2009). Ground-level ozone in the Pearl River Delta and the roles of VOC and NO_x in its production. *Journal of Environmental Management*, *90*(1), 512–518. https://doi.org/10.1016/j.jenvman.2007.12.008

Sharma, S., Sharma, P., & Khare, M. (2017). Photo-chemical transport modelling of tropospheric ozone: A review. *Atmospheric Environment*, *159*, 34–54. https://doi.org/10.1016/j.atmosenv.2017.03.047

Stern, R., Builtjes, P., Schaap, M., Timmermans, R., Vautard, R., Hodzic, A., et al. (2008). A model inter-comparison study focussing on episodes with elevated PM_{10} concentrations. *Atmospheric Environment*, *42*(19), 4567–4588. https://doi.org/10.1016/j.atmosenv.2008.01.068

Su, X., An, J., Zhang, Y., Zhu, P., & Zhu, B. (2020). Prediction of ozone hourly concentrations by support vector machine and kernel extreme learning machine using wavelet transformation and partial least squares methods. *Atmospheric Pollution Research*, *11*(6), 51–60. https://doi.org/10.1016/j.apr.2020.02.024

Wang, H., Li, X., Wang, D., Zhao, J., Peng, Z., He, H., et al. (2020). Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach. *Journal of Cleaner Production*, *253*, 119841. https://doi.org/10.1016/j.jclepro.2019.119841

Wang, Y., Shen, L., Wu, S., Mickley, L., He, J., & Hao, J. (2013). Sensitivity of surface ozone over China to 2000-2050 global changes of climate and emissions. *Atmospheric Environment*, *75*, 374–382. https://doi.org/10.1016/j.atmosenv.2013.04.045

Wang, Z. B., Li, J. X., & Liang, L. W. (2020). Spatio-temporal evolution of ozone pollution and its influencing factors in the Beijing-Tianjin-Hebei Urban Agglomeration. *Environmental Pollution*, *256*, 113419. https://doi.org/10.1016/j.envpol.2019.113419

Watson, P. A. (2019). Applying machine learning to improve simulations of a chaotic dynamical system using empirical error correction. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1402–1417. https://doi.org/10.1029/2018ms001597

Weyn, J. A., Durran, D. R., & Caruana, R. (2019). Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2680–2693. https://doi.org/10.1029/2019ms001705

Weyn, J. A., Durran, D. R., & Caruana, R. (2020). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2020MS002109. https://doi.org/10.1029/2020ms002109

Xiao, D., Fang, F., Heaney, C. E., Navon, I. M., & Pain, C. (2019). A domain decomposition method for the non-intrusive reduced order modelling of fluid flow. *Computer Methods in Applied Mechanics and Engineering*, *354*, 307–330. https://doi.org/10.1016/j.cma.2019.05.039

Xingjian, S., Chen, Z., Wang, H., Yeung, D., Wong, W., & Woo, W. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems* (pp. 802–810).

Yafouz, A., Ahmed, A. N., Zaini, N., Sherif, M., Sefelnasr, A., & El-Shafie, A. (2021). Hybrid deep learning model for ozone concentration prediction: Comprehensive evaluation and comparison with various machine and deep learning algorithms. *Engineering Applications of Computational Fluid Mechanics*, *15*(1), 902–933. https://doi.org/10.1080/19942060.2021.1926328

Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., & Li, F. (2021). Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Systems with Applications*, *169*, 114513. https://doi.org/10.1016/j.eswa.2020.114513

Yang, G., Liu, Y., & Li, X. (2020). Spatiotemporal distribution of ground-level ozone in China at a city level. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-64111-3

Yin, Z., Cao, B., & Wang, H. (2019). Dominant patterns of summer ozone pollution in eastern China and associated atmospheric circulations. *Atmospheric Chemistry and Physics*, *19*(22), 13933–13943. https://doi.org/10.5194/acp-19-13933-2019

Zhan, Y., Luo, Y., Deng, X., Grieneisen, M. L., Zhang, M., & Di, B. (2018). Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment. *Environmental Pollution*, *233*, 464–473. https://doi.org/10.1016/j.envpol.2017.10.029

Zhong, Z., Sun, A. Y., & Jeong, H. (2019). Predicting CO_2 plume migration in heterogeneous formations using conditional deep convolutional generative adversarial network. *Water Resources Research*, *55*(7), 5830–5851. https://doi.org/10.1029/2018wr024592