

Article

Embedded Generative Air Pollution Model with Variational Autoencoder and Environmental Factor Effect in Ulaanbaatar City

Bulgansaikhan Baldorj¹, Munkherdene Tsagaan^{2,*}, Lodoysamba Sereeter³ and Amanjol Bulkhbai⁴

¹ Department of Physics, Mongolian University of Science and Technology, Ulaanbaatar 14191, Mongolia; bulgansaikhan@must.edu.mn

² Department of Mathematics, Mongolian University of Science and Technology, Ulaanbaatar 14191, Mongolia

³ Department of Engineering, German-Mongolian Institute for Resource and Technology, Ulaanbaatar 14191, Mongolia; lodoysamba@gmit.edu.mn

⁴ Department of Remote Sensing, Information and Research Institute of Meteorology, Hydrology and Environment, Ulaanbaatar 15160, Mongolia; b.amanjol@yahoo.com

* Correspondence: ts.munkherdene@must.edu.mn; Tel.: +976-90098001

Abstract: Air pollution is one of the most pressing modern-day issues in cities around the world. However, most cities have adopted air quality measurement devices that only measure the past pollution levels without paying attention to the influencing factors. To obtain preliminary pollution information with regard to environmental factors, we developed a variational autoencoder and feedforward neural network-based embedded generative model to examine the relationship between air quality and the effects of environmental factors. In the model, actual SO_2 , NO_2 , $PM_{2.5}$, PM_{10} , and CO measurements from 2016 to 2020 were used, which were assembled from 15 differently located ground monitoring stations in Ulaanbaatar city. A wide range of weather and fuel measurements were used as the data for the influencing factors, and were collected over the same period as the air pollution data were recorded. The prediction results concerned all measurement stations, and the results were visualized as a spatial–temporal distribution of pollution and the performance of individual stations. A cross-validated R^2 was used to estimate the entire pollution distribution through the regions as SO_2 : 0.81, $PM_{2.5}$: 0.76, PM_{10} : 0.89, and CO : 0.83. Pearson’s chi-squared tests were used for assessing each measurement station, and the contingency tables represent a high correlation between the actual and model results. The model can be applied to perform specific analysis of the interdependencies between pollution and environmental factors, and the performance of the model improves with long-range data.

Keywords: air pollution; embedded generative model; environmental factor effect; latent variable



Citation: Baldorj, B.; Tsagaan, M.; Sereeter, L.; Bulkhbai, A. Embedded Generative Air Pollution Model with Variational Autoencoder and Environmental Factor Effect in Ulaanbaatar City. *Atmosphere* **2022**, *13*, 71. <https://doi.org/10.3390/atmos13010071>

Academic Editors: Lianfa Li, Meredith Franklin, Jun Wu and Haidong Kan

Received: 15 September 2021

Accepted: 28 December 2021

Published: 31 December 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Serious environmental issues are common in densely populated large cities in developing countries. There are many factors contributing to this, starting with climate change; however, urbanization is a global problem [1]. In the past two decades, the number of inhabitants in Ulaanbaatar city has increased more than three times. The capital city of Mongolia sits on the Tuul River, the third-longest river in Mongolia, in a valley surrounded by large mountains, the geographical map of which is shown in Figure 1a. As the city was built in large river watershed lowland between mountains, the temperature is often inverse during the wintertime [2]. Moreover, the city’s geographical and climate features, as well as its poor heating system infrastructure, make the city one of the most air-polluted cities in the world today. The standard sewerage and pure water supply systems were initially designed for 400,000 inhabitants 46 years ago, but now around one 1.4 million people live in Ulaanbaatar, [3]. Some of these 1.4 million citizens are living in the boundary areas of

the city, called ger districts in some reports, which are suburbs that are not connected to the main heating supply. In these suburbs, every family has its stoves and brick stoves as their heat source. Figure 1 shows two neighboring districts: one comprises standard heating-supplied buildings, and the other is a suburb with traditional heating stoves. Unfortunately, the suburban area increases every year [4]. Figure 1 also illustrates the suburban area and the area of standard heated buildings, with the ger district area being 6–7 times larger than the area with apartments. In the warm season, wood is the most commonly used material in traditional stoves, but in the cold season, coal is used. Furthermore, in the outer boundary areas of the ger districts, other materials are sometimes used to reach the minimum warm temperature for living, such as tires, bones, old clothes, etc.

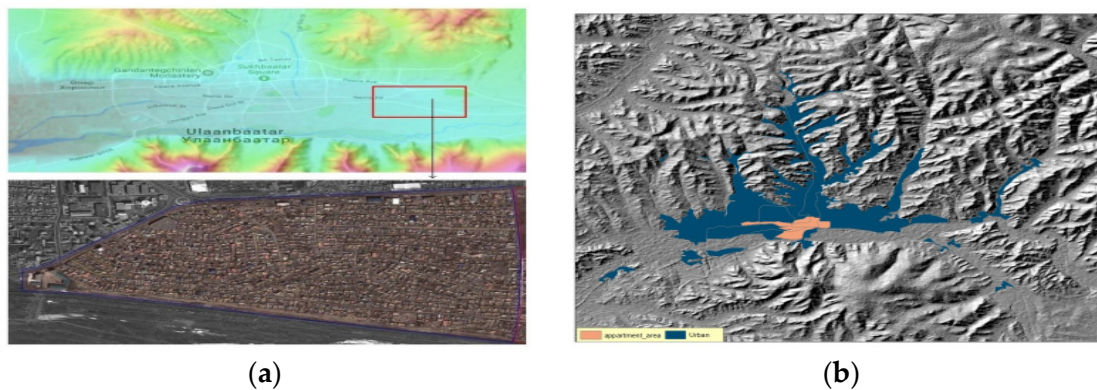


Figure 1. Infrastructure of Ulaanbaatar (UB). (a) Geographical map of Ulaanbaatar (UB) and suburban areas. (b) Apartment area vs. ger district area.

Hence, the smoke from coal and other materials resulted in a horrifying situation in winter, as reported in some earlier works [5,6]. Before the government standardized fuel in Ulaanbaatar, in one example, the harmful dust level was 6–7 times higher than the most lenient World Health Organization standards, according to the World Bank Report in 2012 [7].

From 15 May 2018, the Government of Mongolia strictly banned the import and use of raw coal in Ulaanbaatar, and issued an order for the use of improved fuel instead. Hence, the air-quality index (AQI) has improved, and there was a big difference in the AQI between 2019 and 2020, especially in the wintertime. Consequently, the AQI is influential in addressing air-quality improvement because climatology contributed to improving fuels. However, the role of the AQI measurement network needs to be clarified. To be precise, even devices that accurately measure air quality are only informed by values of measurement from the past, with very few cases in real-time. Alternatively, the tools are simply about listening, i.e., how harmful the AQI was. In this situation, a predictive method may be the solution regarding the historical data information of the AQI, climatology, and improved fuel. Thus, we developed an embedded generative model for the AQI that depends on the environmental factor effect (EFE) influencing the AQI, which is a result of the city's geographical features, such as temperature, wind, atmospheric pressure, etc., including improved and raw fuel measurement quantities.

Therefore, studying the interdependence between AQI and EFE with fuel measurement quantities is the purpose of this work. The variational autoencoder (VAE) and multilayer perceptron methods were used to develop an embedded generative approach. From the VAE learning, we produced new data based on its latent variables. Latent variables play a crucial role in VAE generation, highlighting data and reducing its dimensionality in a shallow space. The EFE with fuel measurement quantities was measured simultaneously with the AQI. The dimensions of the latent variables and the EFE with fuel measurement quantities were the same size. Hence, to discover the relationship between the AQI and the EFE data's latent variable, a multilayer perceptron was used. Thirty-two variables were

gathered in the EFE data, such as ambient temperature, pressure, and wind speed, as well as the impact of raw coal and improved fuel information. For the VAE learning process, a four-year period from 11 of Ulaanbaatar’s (UB’s) air-quality measurement networks were used. The proposed method’s results were evaluated by a Pearson chi-squared test and compared with a non-linear regression prediction in the final step.

The measurement quantities of raw coal and improved fuels in three different laboratories are shown in Table 1. The items in Table 1 are as follows: **Q1**—received basis refers to the moisture content of the coal (maximum percentage of water); **Q2**—dry basis ash content (maximum percentage of dry); **Q3**—dry basis volatile matter (maximum percentage of dry); **Q4**—total dry basis sulfur content (maximum percentage of dry); **Q5**—lower calorific value (kcal/kg min); **Q6**—percentage of carbon; **Q7**—percentage of hydrogen; **Q8**—percentage of nitrogen; and **Q9**—percentage of oxygen.

Table 1. Fuel quality comparison.

Fuel Type	Measured Laboratory	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Improved Fuel	Southwest region, Russia	2.4	22.2	19.6	0.89	6374	67.63	3.81	1.83	3.68
	Central Geological Laboratory	1.9	22.7	19.8	0.87	6984	-	-	-	-
	Laboratory of Mineral Resources and Petroleum Authority	2.7	23.3	19.3	0.86	6334	65.1	5.4	1.32	7.57
	Institute of Chemistry and Chemical Technology	0.8	22.9	18.68	0.89	5918	63.86	3.76	1.69	-
	National standard MNS 5679:2019	≤10	≤29	≤22	≤1.0	≤4200	-	-	-	-
Raw Coal	National standard MNS 3818:2011	37.5	17.5	44.8	0.38	3360	-	-	-	-
	National standard MNS 6226:2011	6	28	26	1.3	5500	-	-	-	-
	CRRRI China	-	-	-	-	-	67.3	4.2	0.95	17.2
	CRRRI China	-	-	-	-	-	69.3	3.8	3.8	3.1

Related Work on ANN and Deep Learning for Ambient Air Pollution Estimation

The application of a neural network (NN) for air pollution is associated with the development of an artificial neural network (ANN), with mainly machine learning and VAE being applied in prediction works. For instance, in early works, the ANN technique was applied for pollution prediction: Reich et al. used 24 h SO₂ data for simple ANN architecture and discussed some of the limitations of the approach [8]. Moustiris et al. used much larger meteorological and air pollution data to predict the pollution index of NO₂, CO, SO₂, and O₃ three days ahead in seven different places in Greece [9]. For urban air pollution prediction, an ANN was applied and compared with multiple time series methods, such as ARIMA, fuzzy time series, and principal component analysis (PCA), using 7–10 years of Malaysian air pollution measurement data [10–12]. A more expanded description of the ANN-based approach can be found in the review work in [13–15]. Furthermore, an ANN was used for the integrated assessment modeling of global climate change and ecological research [16,17].

Machine learning (ML) approaches have been applied for more automatic and accurate air pollution predictions from a large amount of input data with numerous outputs. Zhang et al. studied the superior predictive ability of ML methods based on six years of Hong Kong air-quality index measurements [18]. Moreover, the special gradient-boosting ML approach was applied for predicting the PM_{2.5} concentration in China [19]. Delavar et al. used an autoregressive non-linear neural network with an external input for improving air pollution prediction in the case of Tehran city [20]. The research of Ly et al. considered the dependence between the pollution index and the sources obtained from multisensory measurements and weather data in air pollution predictive analysis-based ML approaches [21]. Similar work was reported by Arnaudo et al. in the urban area of Milan city using different configurations employing machine and deep learning models, such as a linear regressor, an artificial neural network using Bayesian regularization, a random forest regressor, and a long short-term memory network [22].

Lee et al. presented a gradient boosting-based machine learning approach for predicting $PM_{2.5}$ concentration based on a large-scale data set in Taiwan city [23]. For UB city, $PM_{2.5}$ concentration prediction has been studied using two different ML approaches, and high-performance results have shown the spatiotemporal variations in the high-emission areas [24]. Moreover, Meredith et al. compared several machine learning methods for ground-level pollutant mixtures of particulate matter with a diameter smaller than $PM_{2.5}$ for air-quality assessment in Mongolia [25]. The survey works of ML methods for ambient air pollution assessment are summarized in [26–30].

Among the generative models for air contamination, deep learning methods are a more recently developed approach. Tien et al. worked on air-quality inference based on a graph-based matrix completion problem, and applied a variational model on graph convolutional autoencoders [31]. Additionally, they used a mobile Internet of Things sensor for fine-grained air quality in deep learning inference [32]. The spatiotemporal prediction of air pollution based on deep learning approaches has been studied widely for many different purposes. Li et al. studied the spatiotemporal deep learning-based air-quality prediction method that inherently considers the spatial and temporal correlations and prediction results of all stations compared with traditional time series prediction models [33].

Li et al. introduced the advantage of using two-stage models in geographically weighted machine learning [34], and deep learning to robustly impute a long time series of multi-angle implementation of atmospheric correction aerosol optical depth [35]. The achievements of embedded methods with deep learning approaches were considered in [36,37], and the high performance of autoencoder-based residual networks consisting of learning models is analyzed in [38].

Fan et al. studied the framework and feasibility of an idea based on deep RNN [39], whereas improved feature analysis was considered in [40] based on fine-grained air quality.

2. Methodology

2.1. Area of Study and Data Collection

The capital city UB has a history spanning more than 360 years and is currently home to 50% of the country's population. This population is continuously growing; therefore, the most significant problem is in an urban ecological context, with environmental pollution also dramatically increasing. In addition, the city has a harsh and unique climate, and its geographical location behavior is the main cause of this air pollution. The city is located in a large valley surrounded by mountains, at an elevation of approximately 1580 m above sea level. Due to the elevation and being located far from the sea, the city has a very long-lasting winter, making UB one of the coldest capital cities in the world. Ulaanbaatar receives an average of 239 mm of precipitation per year, of which 1 mm or more falls in 39 days, and 5 mm or more of rain falls in 15 days. Generally, this area is arid, with the humidity sometimes decreasing to 4% per year. The wind comes mainly along the valley, but the strength of the wind is low throughout the year, with 10% of the wind observed in the spring months.

The average annual surface air temperature is $-0.7\text{ }^{\circ}\text{C}$, dropping to $-45\text{ }^{\circ}\text{C}$ in the winter and warming up to $+59\text{ }^{\circ}\text{C}$ in the summer. Sudden frosts occur on the surface of the soil in all months except July. Therefore, the ground freezes to a considerable depth of 4–8 m. The average wind speed for the country is less than 2 m/s, which affects the air pollution levels, making them more constant for an extended time. Additionally, the average wind speed of the capital city is primarily 0.5 m/s from the northern and northwestern areas, which means that there is not much air circulation. Thus, air pollution remains stable for long periods around the center of the city. Figure 2 shows the wind information maps across four years.

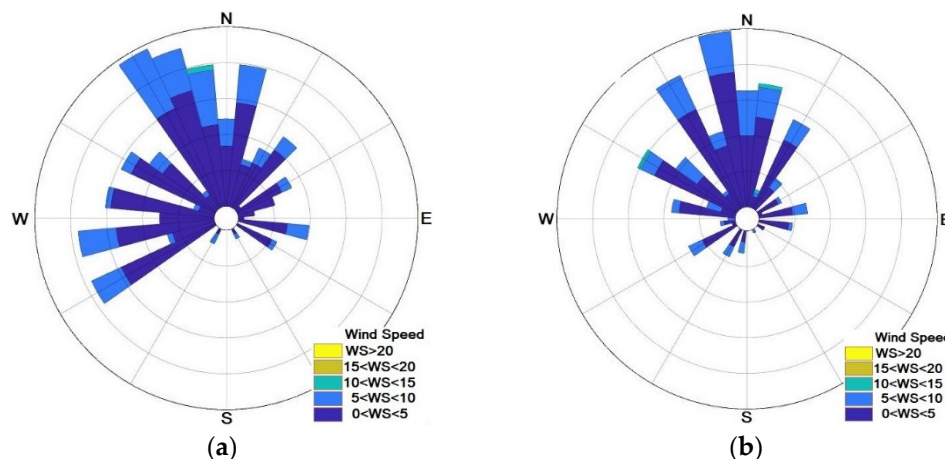


Figure 2. The wind direction and speed map of UB city. (a) The 2016 wind rose map. (b) The 2019 wind rose map.

The National Agency for Meteorology and Environmental Monitoring (NAMEM) established the first air-quality monitoring station in Ulaanbaatar in 1977. This was the foundation for the establishment of an air-quality monitoring network in Mongolia. Until 2008, there were four permanent air-quality monitoring stations in Ulaanbaatar, which measured only two indicators, sulfur dioxide and nitrogen dioxide. There was no continuous air-monitoring system built on the outskirts of Ulaanbaatar city.

Due to Ulaanbaatar’s air pollution issues, there is a need to expand the network by equipping it with automatic, continuous measurement capabilities and improving its capacity. The air-quality monitoring network consists of the Air Pollution Reducing Department of Capital City (APRD) Network and the NAMEM Network. A German grant established the APRD Network with five stations in 2008, and the NAMEM Network, with 10 stations, was established by means of a French loan in 2010. Currently, 12 automatic stations are measuring sulfur dioxide, nitrogen oxides, carbon monoxide, ozone, PM_{10} , and $PM_{2.5}$ every 15–30 min, and 3 non-automatic controls of sulfur dioxide and nitrogen dioxide (Figure 3). The concentration is determined using a chemical solution.

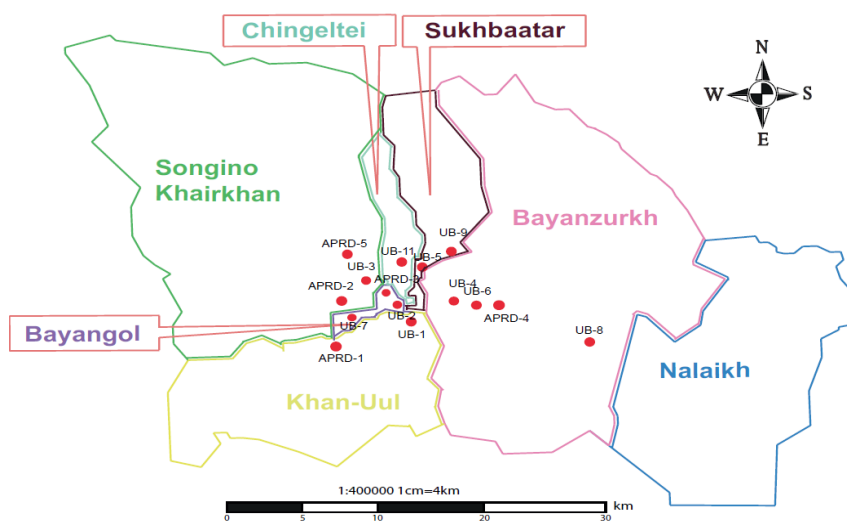


Figure 3. State map of UB and air-quality index (AQI) measurement networks.

2.2. VAE for Embedded Generative Method

Let $y = \{y_1, y_2 \dots y_n\}$ be the AQI measurements collected from measurement networks at different locations in the city. Furthermore, $x = \{x_1, x_2 \dots x_n\}$ represents the EFE of the air quality, which is measured at the same time as y . Next, the purpose is to

understand the relationship between those two quantities; $F : x \rightarrow y$. In the case of air quality, the relationship is completely unknown; generally, it is assumed to be a system of highly non-linear maps. Moreover, data of x and y raw and sampling methods are very different in each measurement. In this situation, a variational autoencoder (VAE) has the benefit of representing the data in low dimensional continuous latent variables with intractable posterior distributions [41].

The VAE method optimizes the weight parameters \mathbb{W} and bias \mathbb{b} of the function: $\bar{F}(\mathbb{W}, \mathbb{b}, x)$, which is a non-linear approximation of the relation map, as follows:

$$F = \operatorname{argmin}_{\mathbb{W}, \mathbb{b}} \frac{1}{N} \sum_{i=1}^N \| \mathbf{y}_i - \bar{F}(\mathbb{W}, \mathbb{b}, \mathbf{x}_i) \|^2 \tag{1}$$

where $\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^N$ is the data set provided and stored from the relation map F . The optimization process of $\bar{F}(\mathbb{W}, \mathbb{b}, \mathbf{x}_i)$ consists of two maps, encoder Φ and decoder Ψ . The encoder reduces the dimension of y into dimensionally reduced space, $\Phi: \mathbf{y} \rightarrow \mathbf{h}$, and then the decoder map generates new $\hat{\mathbf{y}}$ values from the reduced space, $\Psi: \mathbf{h} \rightarrow \hat{\mathbf{y}}$. Thus, when an optimal parameter is reached, we obtain the pair of (Φ, Ψ) , which satisfies the following minimization:

$$(\Phi, \Psi) = \operatorname{argmin}_{\mathbb{W}, \mathbb{b}} \frac{1}{N} \left[\sum_{i=1}^N \| \mathbf{y}_i - \Psi(\Phi(\mathbf{y}_i)) \|^2 + KL(N(\mu, \sigma), N(0, I)) \right] \tag{2}$$

where KL is the Kullback–Leibler (KL) divergence loss between $N(\mu, \sigma)$ and $N(0, I)$ (along with this formulation, we refer [42,43]). KL measures how the distribution of $N(\mu, \sigma)$ associated with data \mathbf{y} is different to the normal distribution, defined as:

$$KL(N(\mu, \sigma), N(0, I)) = \frac{1}{2} \sum_{s=1}^l (\mu_s^2 + \sigma_s^2 - 1 - \ln(\sigma_s^2)) \tag{3}$$

where l is the dimension of dimensionally reduced space \mathbf{h} , called latent space. To obtain the latent space from the data, the multilayer perceptron (MLP) is the advised method. The map of reducing dimension consists of several hidden layers, and every layer has finite nodes. In the encoder process, the number of nodes in the steps of layer is conventional in reducing the order. w_{ij}^k is denoted as the weight parameters of the layers between k th and $(k + 1)$ th, whereas i and j are node positions, respectively. Then, the k th layer weight parameters can be represented in the following matrix form:

$$W^{(k)} = \begin{bmatrix} w_{1,1}^k & \cdots & w_{m,1}^k \\ \vdots & \ddots & \vdots \\ w_{1,n}^k & \cdots & w_{m,n}^k \end{bmatrix} \tag{4}$$

where m is the number of nodes in the k th layer, and n is the number of nodes in the $(k + 1)$ th layer. To curtail the dimension of input that the number of nodes in the layer satisfies:

$$N > \cdots > \underbrace{\cdots}_{k-1} > \underbrace{m}_{k-1} > \underbrace{n}_{k+1} > \underbrace{\cdots}_{k+1} > l$$

In recurrent form, this can be described as:

$$\mathbf{h}^{(1)} = W^{(1)} \mathbf{y} + b^{(1)} \tag{5}$$

$$\mathbf{h}^{(k)} = W^{(k)} \mathbf{h}^{(k-1)} + b^{(k)} \tag{6}$$

where $b^{(k)} = [b_1^{(k)}, b_2^{(k)}, \dots, b_n^{(k)}]^\top$ refers to the bias parameters and \top is the matrix transpose operation. Every $h^{(k)} = \{h_1^{(k)}, h_2^{(k)}, \dots, h_n^{(k)}\}$ retains the key feature of the previous steps, and the final space h is considered to be the feature space. Then, the encoder function is represented as:

$$\Phi(y) = \zeta(W^{(k)} \dots \zeta(W^{(2)} \zeta(W^{(1)}y + b^{(1)}) + b^{(2)}) + \dots + b^{(k)}) \tag{7}$$

where ζ is the sigmoid function. The main feature that distinguishes a VAE from other generating models is its latent space. If we look at latent space alone, its actual value is a quantity that cannot be measured directly. However, it contains the core of the essential information in the data, determining in the latent space the KL-divergence loss and reparameterization trick is vital. In the process of optimization, Equation (3) provides a close distribution to normal with μ and σ expressed as:

$$\mu = W^\mu \Phi(y) + b^\mu \tag{8}$$

$$\sigma = W^\sigma \Phi(y) + b^\sigma \tag{9}$$

where $W^\mu, W^\sigma, b^\mu,$ and b^σ are the weight and bias for μ and σ . Then, the latent variable is described as:

$$h = \mu + \sigma \odot \epsilon \tag{10}$$

where ϵ is the normally distributed auxiliary noise variable and \odot is the element-wise product. Figure 4 shows the reparameterization trick on h , the difference with and without ϵ is intractable posterior distributions.

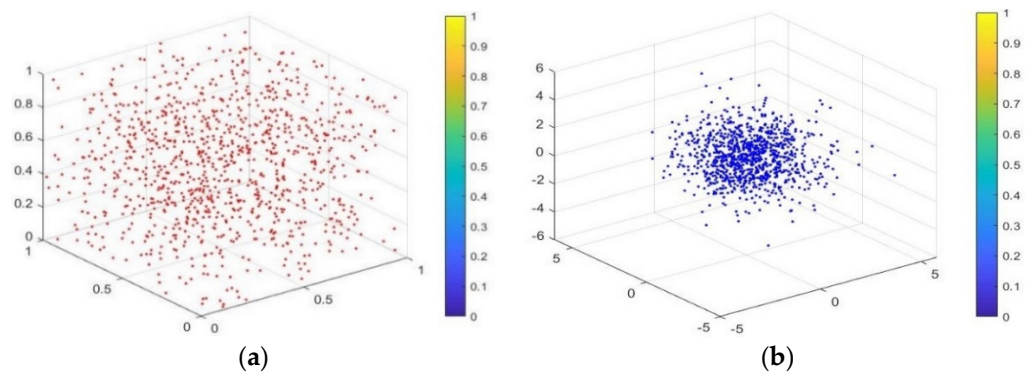


Figure 4. The reparameterization trick on h . (a) The random sample of μ and σ . (b) The random sample of μ and σ with ϵ .

More precisely, the latent space is centered at the coordinate and distributed in a bounded interval in any direction; thus, any linear combination between two points of the space belongs to its space (for more detailed properties of this phenomena, refer to Section 3.4 of [42]).

When h variables are found, the decoder function is defined as:

$$\Psi(h) = \zeta(W^{2k} \dots \zeta(W^{k+2} \zeta(W^{(k+1)}h + b^{(k+1)}) + b^{(k+2)}) + \dots + b^{(2k)}) \tag{11}$$

Then, the entire process of VAE is to determine the parameters of W and b , where:

$$\mathbb{W} = \{W^{(1)}, \dots, W^{(k)}, W^{(\mu)}, W^{(\sigma)}, W^{(k+1)}, \dots, W^{(2k)}\} \tag{12}$$

$$\mathbb{b} = \{b^{(1)}, \dots, b^{(k)}, b^{(\mu)}, b^{(\sigma)}, b^{(k+1)}, \dots, b^{(2k)}\} \tag{13}$$

The process of optimizing those parameters is based on backpropagation and a gradient descent algorithm. Initially, it starts with random choice in \mathbb{W} and \mathbb{b} , then the sensitiveness of training errors in little changes in those parameters, and the derivative of the training error leads to the chain rule and backpropagation. The network architecture is represented in Figure 5, and the architecture of the VAE, and the algorithm of optimizing the parameters, are described in the next section, along with the embedded network.

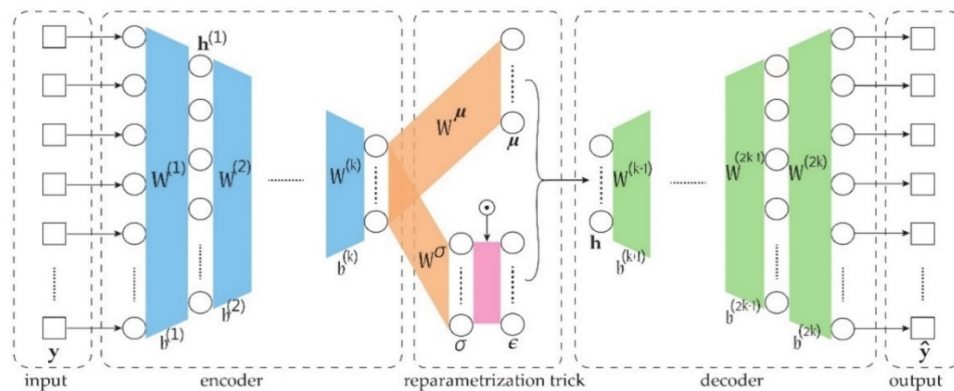


Figure 5. The architecture of the variational autoencoder (VAE).

2.3. Embedding Network for the Relationship between EFE and AQI

The formulas described in Equations (2)–(13) are basic VAE methods for generating new AQI values from the latent variables obtained in the learning process. In order to approximate the function F , whereby \mathbb{W} , \mathbb{b} , and x are as described in Equation (1), we need different approaches to the typical VAE approach. However, the method defined in Equations (2)–(13) determines the vital space of the AQI, which is the latent variable h . To find the relationship between latent space data and the EFE, we use the feedforward neural network approach, as the period of sampling time is equal.

Expressly, the first column of Equation (14) represents n daily measurements of AQI, and the next column is the latent variables obtained through the algorithm applied to daily measurements of the AQI. Additionally, EFEs were also measured on the same day with y :

$$\begin{array}{ccc}
 \mathbf{y}_1 = \{y_1^{(1)}, \dots, y_m^{(1)}\} & \mathbf{h}_1 = \{h_1^{(1)}, \dots, h_l^{(1)}\} & \mathbf{x}_1 = \{x_1^{(1)}, \dots, x_l^{(1)}\} \\
 \mathbf{y}_2 = \{y_1^{(2)}, \dots, y_m^{(2)}\} & \mathbf{h}_2 = \{h_1^{(2)}, \dots, h_l^{(2)}\} & \mathbf{x}_2 = \{x_1^{(2)}, \dots, x_l^{(2)}\} \\
 \dots & \dots & \dots \\
 \mathbf{y}_n = \{y_1^{(n)}, \dots, y_m^{(n)}\} & \mathbf{h}_n = \{h_1^{(n)}, \dots, h_l^{(n)}\} & \mathbf{x}_n = \{x_1^{(n)}, \dots, x_l^{(n)}\}
 \end{array} \tag{14}$$

Hence, the data are available to learn the relationship between h and x . Apply the feedforward neural network to approximate the function $f : h \rightarrow x$; indeed, it is the universal approximation theorem for f . Thus, it gives the $f(000000, 33333333; x)$ function, which is an approximation function as $f \sim f(000000, 33333333; x)$ with weight and bias parameters corresponding to network nodes. Then, the new generating AQI values can be described through the embedding as:

$$\begin{cases} f : x \rightarrow h \\ \Psi : h \rightarrow \hat{y} \end{cases} \tag{15}$$

The architecture of the two combined networks is represented between y , h , and x in Figure 6.

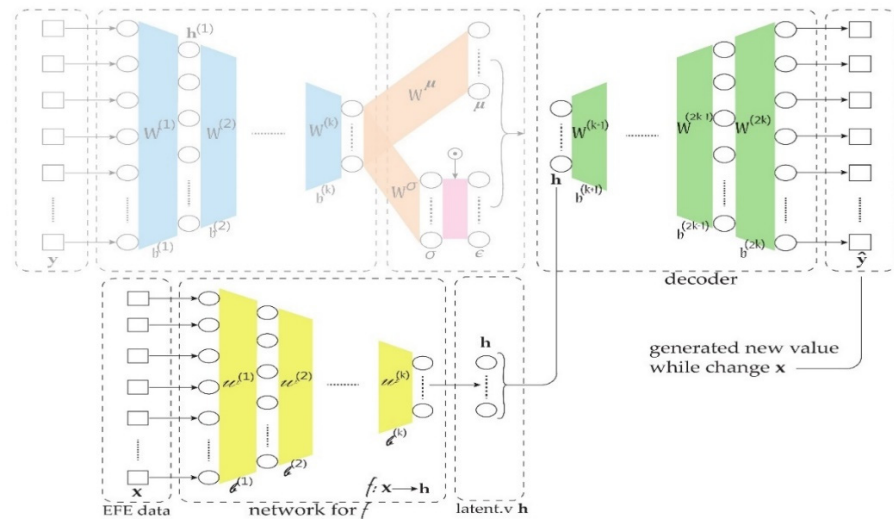


Figure 6. The architecture of the embedded network.

The Algorithm 1 for the embedded generative model:

Algorithm 1: embedded modeling algorithm

Result: find h latent space, and save model

(W, b) -initialization; η -step size hyperparameter; set error estimation;

while not done **do**

τ -sample batch;

for all τ **do**

Evaluate $\nabla (W, b) \sum_{i=1}^N (\|y_i - \bar{F}(W, b, x_i)\|^2) \tau$ by gradient decent method;

If error not small enough **then**

Update $(W, b) \leftarrow (W, b) - \eta \sum_{i=1}^N (\|y_i - \bar{F}(W, b, x_i)\|^2) \tau$

end

end

end

Save the above model, which is Ψ ;

Result: find the relation map f ;

(w, b) -initialization; ζ -step size hyperparameter; set error estimation;

set algorithm for (w, b) same with (W, b) ;

Result: embedding of Ψ , f ;

Generate new AQI values based on the embedding;

$\Psi: (f: x \rightarrow h) \rightarrow \hat{y}$

3. Experiment

3.1. Data Expression

To validate the proposed method of illustrating the interdependency between AQI and EFE, we used four years of air pollution concentration measurements and four years of weather data. The correlation between AQI and pollution concentration is defined as:

$$AQI = \frac{(I_{high} - I_{low})}{(C_{high} - C_{low})} (C - I_{low}) + I_{low} \tag{16}$$

where C is the amount of pollutant concentration per cubic meter of air, C_{low} and C_{high} are the lowest and highest breakpoints of C , and I_{low} and I_{high} are the index breakpoints corresponding to C_{low} and C_{high} . However, the AQI is designed to disseminate air-quality information using color coding as it is accessible to the general public, normalizing the measurement into world standards, with the relationship between AQI and C being direct one-to-one. Furthermore, the pollutant concentration is available from measurement stations, with most of them even working automatically, and we used C instead of AQI. In addition, the pollution concentration data that were assembled between 1 January 2016 and 1 June 2020 from 15 pollution measurement stations and positions are shown in Figure 3. Three of them determined the pollution concentration based on active sampling tests, and the remaining stations automatically measured the concentration. The automatic stations used the light source to obtain the signal difference between the reference and the incoming air in sampling manifolds. This working principle analyzed the amount of different pollution concentrations by micrograms per cubic meter in a minute and eliminated rotation for optic adjustment. Related to the working principle and purpose, each station had different measuring capabilities, namely SO_2 measured from 15 stations, NO_2 measured from 13 stations, PM_{10} measured from 14 stations, $PM_{2.5}$ measured from 8 stations, and CO measured from 10 stations, with a total of 60 pollution concentrations gathered for the experiment. As in the example, the daily concentration of SO_2 in 2016 is expressed in Figure 7, which was measured by UB-2, with the tolerable content level being $50 \mu\text{g}/\text{m}^3$.

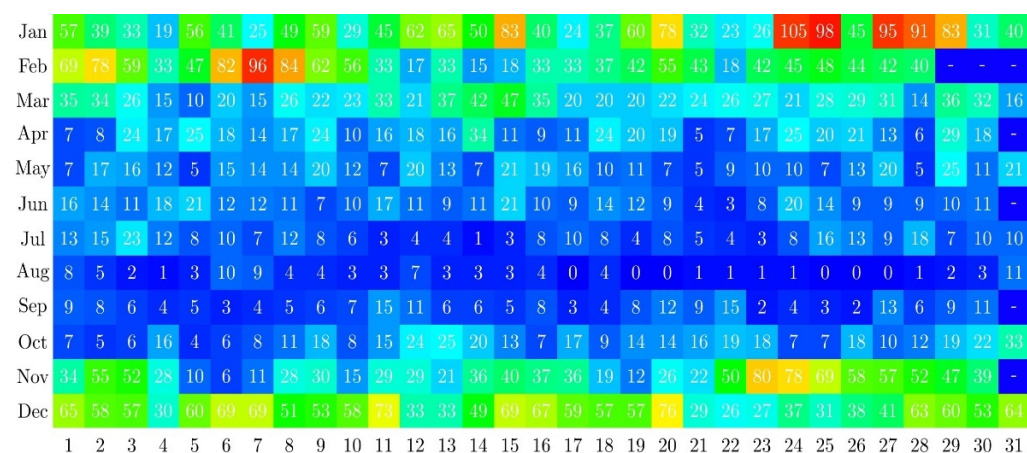


Figure 7. The daily concentration of SO_2 in 2016, measured by UB-2, tolerable content level is $50 \mu\text{g}/\text{m}^3$.

The EFE data were composed of measurements of a comprehensive range of weather data sampled from five automatic weather measurement stations in UB. Weather data involve various levels of wind speed, wind direction, temperature, relative humidity, and the most crucial information for the fuel period. Furthermore, weather information was collected within the same period as the air pollution concentration measurement.

Surface wind regimes depend on local climatic conditions and atmospheric circulation. Each season, the average wind speeds are very different in UB, 1.8–6.3 m/s in spring,

–2.5 m/s in winter, and 1.2–5.4 m/s in other seasons. In winter, due to the strong anti-cyclones, the prevailing wind speed slows down, whereas in other seasons, the west, northwest, and north winds prevail along the prevailing air flow. Depending on the location of the weather stations in UB, the wind direction, speed, and frequency are quite varied. For example, in the station at the airport, the east wind is predominant in all seasons except spring, different to the station at the university. The four years of the probability distribution of the direction and wind speed are visualized in Figure 8. The degree of wind direction is defined in cardinal degrees as compass points.

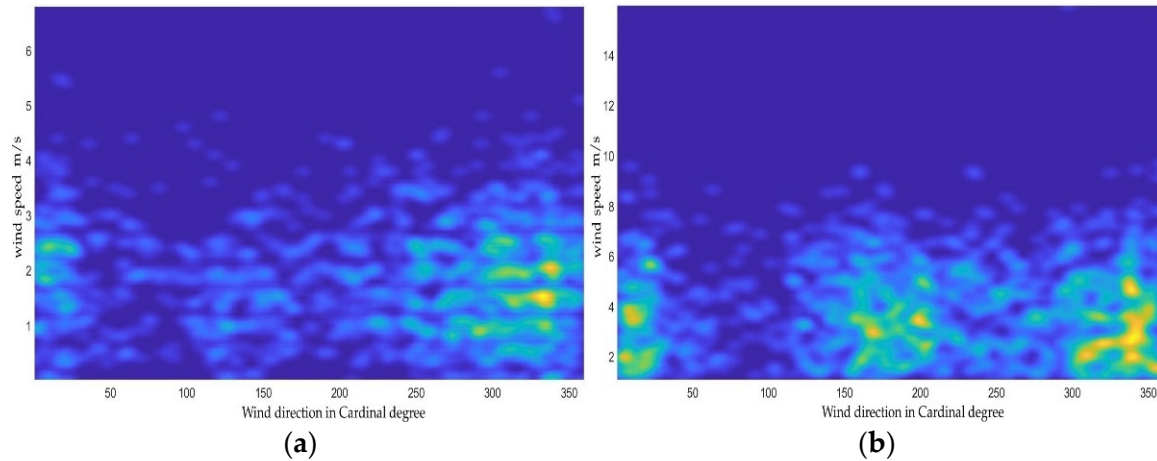


Figure 8. The distribution of wind with respect to direction and speed in 2016–2019. (a) The station at the university. (b) The station at the airport.

Furthermore, four years of temperature and relative humidity information are expressed in Figure 9. For both quantities, the frequency is counted with respect to amounts of years, and the absolute measurements are represented in a bivariate histogram. In the illustration of temperature, the higher frequencies appear more often in the cold season over the four years of data than in the warm season. The amount of relative humidity is inversely related to air temperature and depends on cloud precipitation. The average annual relative humidity around Ulaanbaatar is 55–76%.

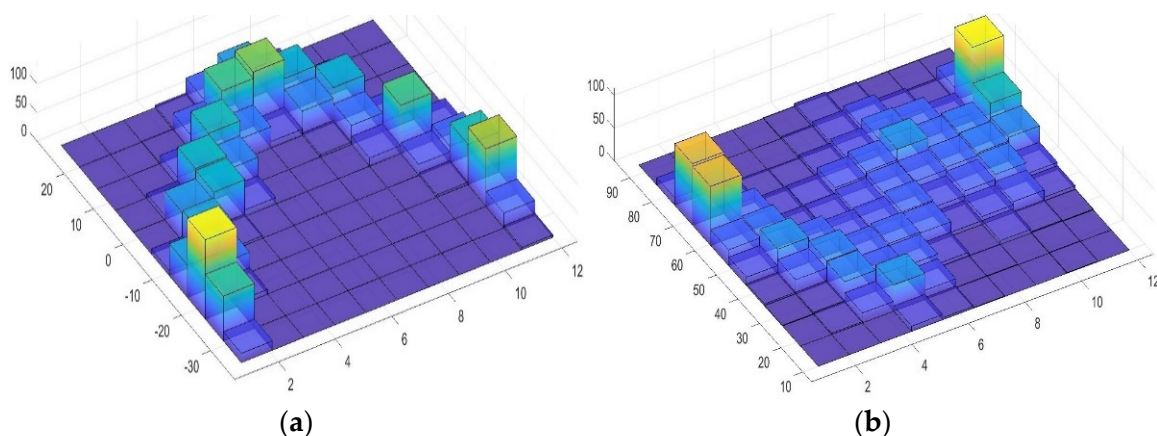


Figure 9. The bivariate histogram expression of temperature and humidity in 2016–2019. (a) Temperature. (b) Humidity.

3.2. Visualization of the Results

In the VAE-based learning process, to obtain the feature space h of the AQI data, we need to separate data into two parts, i.e., training and testing. The data size was

$365 \times 60 = 87,600$ from 8–15 air pollution concentration measuring stations, and the data were separated into 60,000 and 27,600 for training and testing, respectively. The computation architecture for the VAE consisted of three hidden layers with nodes 40, 30, and 20, and one input channel had a length of 60. The training epoch was 200,000, the batch size was 1000, and the learning rate was 0.001. We used an RmsProp optimizer with the *ReLU* activation function in the hidden layer and the sigmoid activation function in the output layer.

Python and TensorFlow were used for the VAE implementation, and a PC equipped with an Intel(R) Core (TM) i7-2600-CPU@3.40GHz, with an NVIDIA GeForce GTX 1060ti 6GB GPU was used for the computation. The Python plotting tool, MATLAB, and LaTeX pgfplot packages were used in the data and result visualization. In Figures A1–A3 in the Appendix A, the typical air pollution measurement of SO_2 and the convergence rate of the VAE computation at a particular epoch are shown.

With this latent space computation, we completed the VAE using the algorithm illustrated in pseudocode 1. Next, for the relationship between the EFE data and the latent variables of the VAE, the development of a feedforward NN for learning was required, which is observed in the second step of pseudocode 1. The daily measurement of the EFEs consisted of 32 quantities. Hence, the feedforward NN's input channel was 32, and it had one hidden layer with 25 nodes. Along with a training epoch of 10,000, a batch size of 20, and a learning rate of 0.001, we used the Adam optimizer with the *ReLU* activation function in the hidden layer and the sigmoid activation function in the output layer, whereas 70% and 30% of the data were used for the training and testing sets, respectively.

Finally, to validate our suggested method, the EFE data of the first three months of 2020 were selected for generating the predicted air pollution concentration. With respect to the first 90 days and the pollution measurement stations, the model provides $90 \times 60 = 5400$ as the number of results. This large number of results requires an expression of the outcome in an optimal way, and we elected a few specific depictions. For this reason, the most frequently polluted days of each pollution index were chosen to illustrate the results. Table 2 represents the most frequently polluted dates in the first three months of our four years of data.

Table 2. The most frequently polluted air pollution dates in 2016–2019.

	SO_2	NO_2	PM_{10}	$PM_{2.5}$	CO
Jan	12	11	11	11	12
Feb	1	1	3	4	1
Mar	1	1	31	2	1

The coldest days of the year directly affected the reason for representing the same behavior for gaseous pollution in the first three months of integration in Table 2. By long-standing average, the coldest days of the year are between 8 and 15 January. Therefore, in January, SO_2 , NO_2 , $PM_{2.5}$ and CO are formed from fuels burned in home stoves for heating in the ger districts with high atmospheric pressure and temperature inversion. However, particular pollution is related to poor infrastructure. When the warm season starts, particular pollution from the ger districts increases due to the mobility of a large number of vehicles; consequently, at the end of March, the measurement of PM_{10} greatly increases.

Furthermore, the inverse distance weight (IDW) interpolation method was used for the pollution concentration map. The inverse weight is defined as:

$$\omega_{ij} = \frac{1/d_{ij}^2}{\sum_{i=1}^n 1/d_{ij}^2} \quad (17)$$

where n is the number of points for interpolation, and d_{ij} is distance between the i th interpolation point and the j th air pollution measurement station. Then, based on Equations (14) and (15), the concentration values at the i th interpolation point is estimated as:

$$y_i = \omega_{ij}\Psi(f(x_j)) \tag{18}$$

The results of Equation (18) are expressed in Figures 10–12, corresponding to the integrated values in Table 2; 1 February and March for SO_2 and CO , and 11 January for $PM_{2.5}$ were selected, respectively.

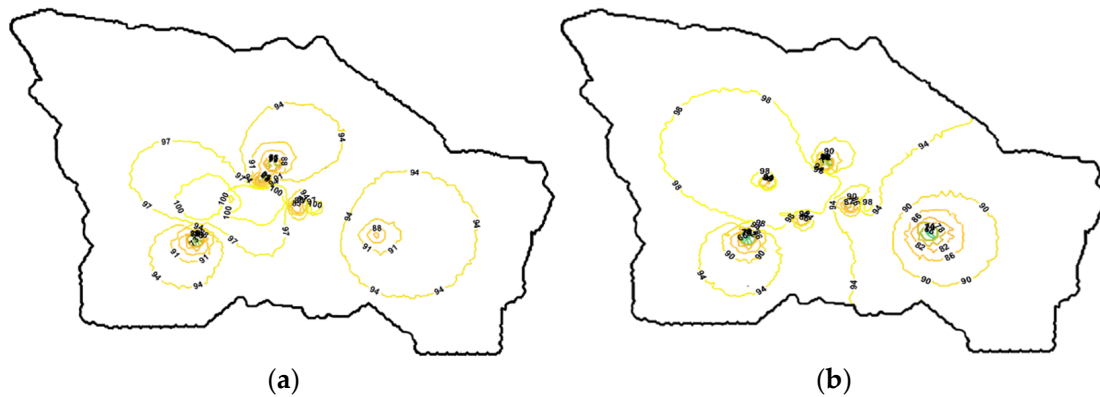


Figure 10. The distribution map of SO_2 on 1 February 2020. (a) The distribution of actual SO_2 . (b) The distribution of predicted SO_2 .

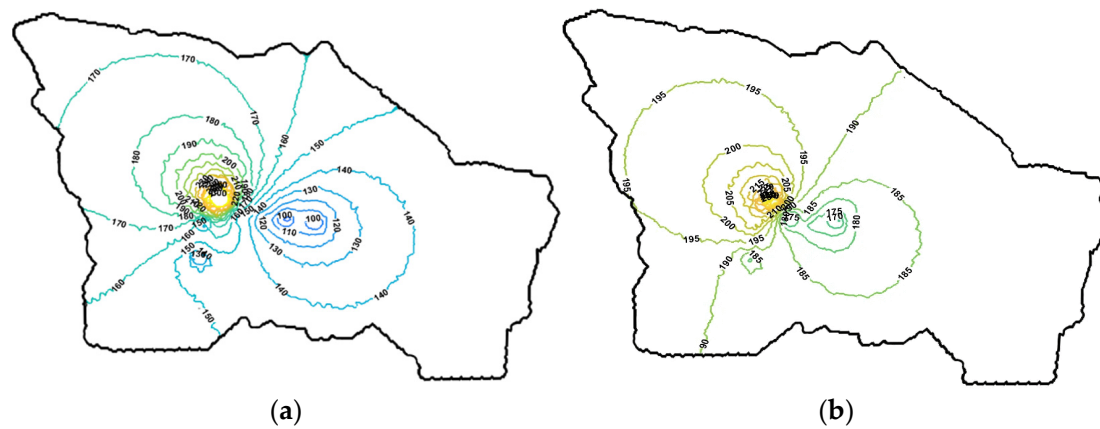


Figure 11. The distribution map of $PM_{2.5}$ on 11 January 2020. (a) The distribution of actual $PM_{2.5}$. (b) The distribution of predicted $PM_{2.5}$.

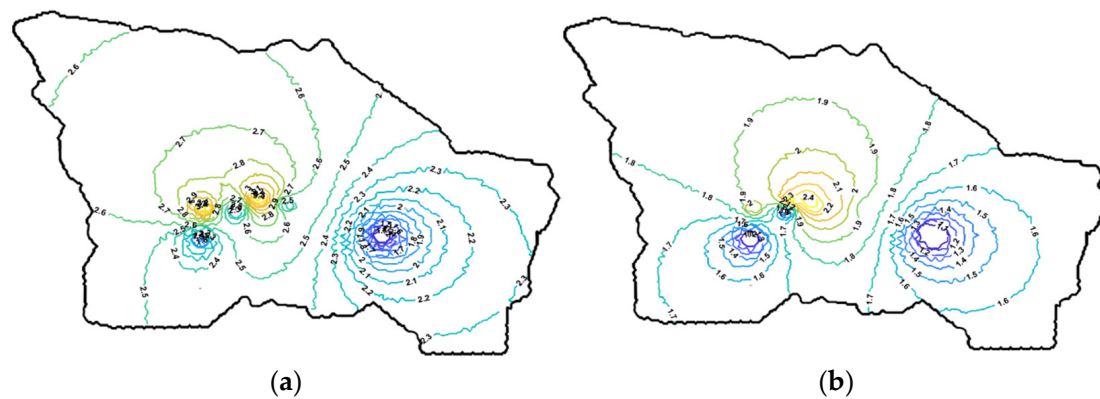


Figure 12. The distribution map of CO on 1 March 2020. (a) The distribution of actual CO . (b) The distribution of predicted CO .

For more results, we randomly selected one of the pollution measuring stations for every five air pollutions, and 90 days of both actual and model prediction results are compared in Appendix A. To evaluate the performance of the proposed model, the most commonly used correlation coefficient and Pearson’s chi-squared methods were used. The positive correlation coefficient indicates higher correlation between testing values. As in the example, Figures 13 and 14 represent the correlation coefficients between the actual and predicted values of SO_2 , $PM_{2.5}$, PM_{10} , and CO obtained from all measurement stations, respectively. In the figure, the corresponding correlation coefficients are highlighted in red in the correlation matrix plot. For more validation purpose, we randomly selected one measurement station and computed the contingency tables for 90 days of actual and predicted values, which expressed in Figures A4–A8 in Appendix A. Two of the corresponding contingency tables for NO_2 and $PM_{2.5}$ are shown in Figure A9.

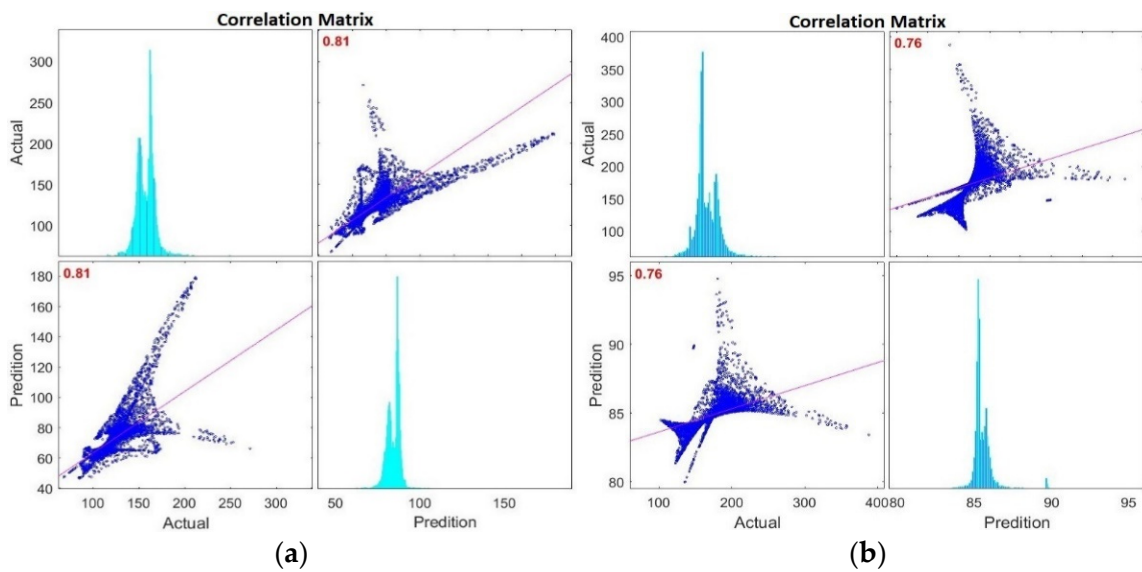


Figure 13. The correlation matrix of actual and predicted values. (a) SO_2 , (b) $PM_{2.5}$.

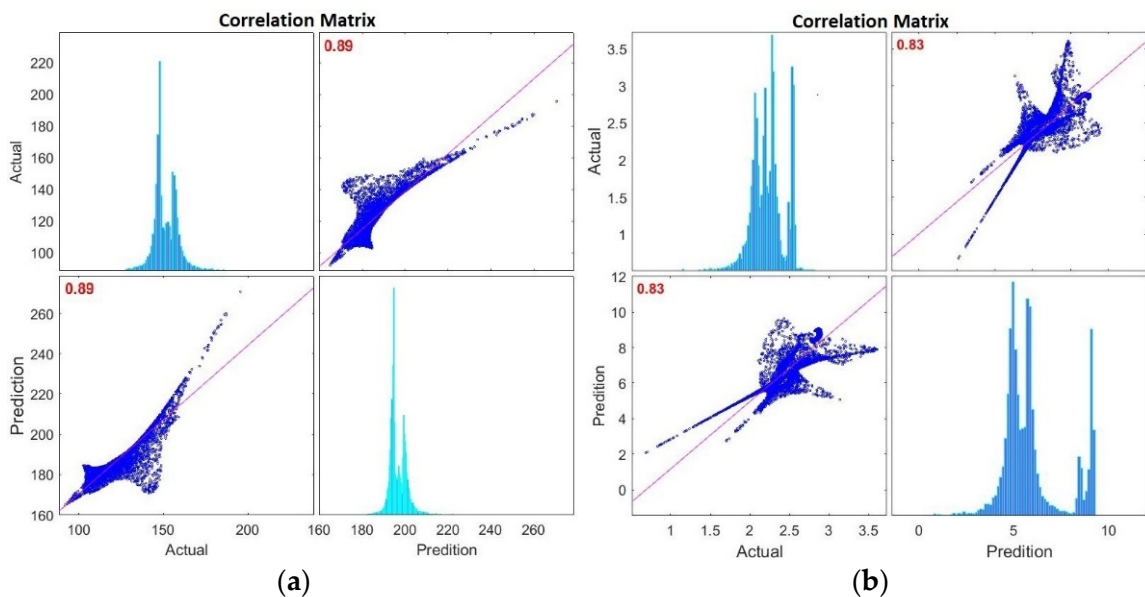


Figure 14. The correlation matrix of actual and predicted values. (a) PM_{10} , (b) CO .

For the remaining three air pollution concentrations, the contingency table was similar, whereas the average chi-squared result was $\chi_{ave.}^2 = 115.14$. Thus, the statistics for the Pearson's chi-squared test were highly dependent on each other.

4. Conclusions

In this study, a variational autoencoder-based embedded generative air pollution model was developed. The proposed approach represented substantial progress in the reliable prediction of five air pollution concentrations. The embedded generative model consists of a VAE and a feedforward NN to discover the relationship between the air pollution concentration and environmental factor effects. The convergence of learning networks based on the gradient descent method with RmsProp and Adam optimizer techniques were applied, and the convergence rate was 1×10^{-3} , consequentially. To construct the proposed model, we simultaneously assembled four years (2016–2020) of air pollution concentration measurements and weather data. In addition, information on improved and raw fuel was involved in the environment factor data. The prediction values of the model corresponded to all the pollution measurement stations, with a total of $15 \times 90 = 1350$ prediction values obtained for the test with respect to 15 stations in the first 90 days of 2020. Thus, the results can be visualized in two ways. First, a pollution map can be created by using the inverse distance weight interpolation throughout the regions of UB, which expresses a daily pollution map for estimating the pollution concentration for local coordinates of the city. For instance, the difference between the local measurement and the prediction results for PM_{10} in ger district areas located in UB suburbs in four directions was $21.24 \mu\text{g}/\text{m}^3$, $-9.32 \mu\text{g}/\text{m}^3$, $11.05 \mu\text{g}/\text{m}^3$, and $-11.87 \mu\text{g}/\text{m}^3$, respectively, in the Airport, Zuragt, Tolgoit, and Amgalan districts on 16 January 2020. The correlation coefficient method was applied while constructing the pollution concentration map of the city for each pollution index, and the average of these indicates the entire performance of the model as 0.8280. The Pearson's chi-squared test method was used for testing the second visualization performance of the individual pollution measuring stations and expressed a high correlation between actual and predicted. A further aim of considering the VAE and NN embedding to investigate inversely the interdependencies between the weather and the fuel conditions, by sampling normal to good days of air pollution, could be considered for future study.

Author Contributions: Conceptualization, B.B. and M.T.; methodology, B.B., M.T. and L.S.; investigation, B.B., M.T. and L.S.; writing—original draft preparation, B.B., M.T., L.S. and A.B.; writing—review and editing, M.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The convergence rate of the VAE computation at a particular epoch. We designed two methods to compare the results of the embedded model with real measurements. The first way was represented in Figures 10–12, in which we gathered results belonging to all measurement stations and used the IDW method to construct a concentration map for one day. For the second way, we chose one station and sampled the results concerning the days that EFE was applied to the model. In this way, we applied the first 90 days' data of 2020 for EFE to the (15), the model provided 59×90 values.

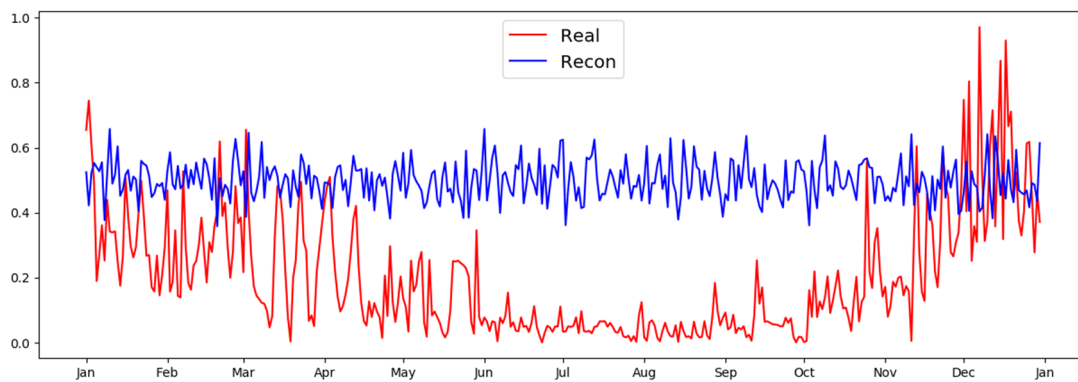


Figure A1. Convergence of the VAE computation at the first epoch.

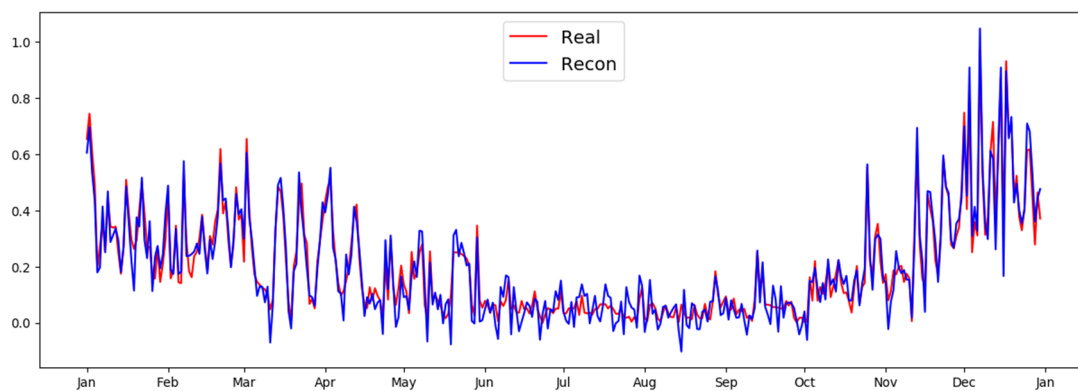


Figure A2. Convergence of the VAE computation at an epoch of 400.

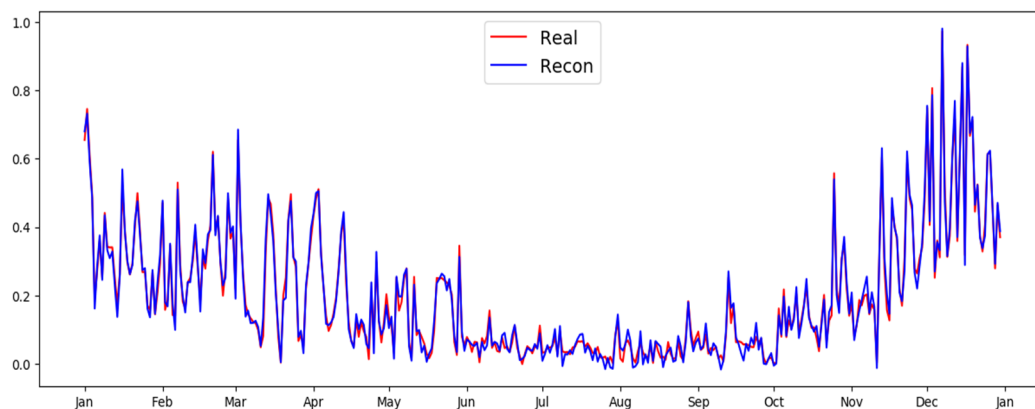


Figure A3. Convergence of the VAE computation at an epoch of 800.

From the results of 59 stations, we randomly chose the station and compared the 90 days of results to the actual measurements (Figures A4–A6). In Figure A4, the real and predictive values of NO_2 for the UB-5 station are presented. Meanwhile, Figure A5 shows the graphics of the actual measurement and the predictive results of the embedded model of SO_2 for the UB-7 station. Additionally, in Figure A6, the graphics of the real measurement and the predictive values of the method for PM_{10} for the APRD-2 station are expressed.

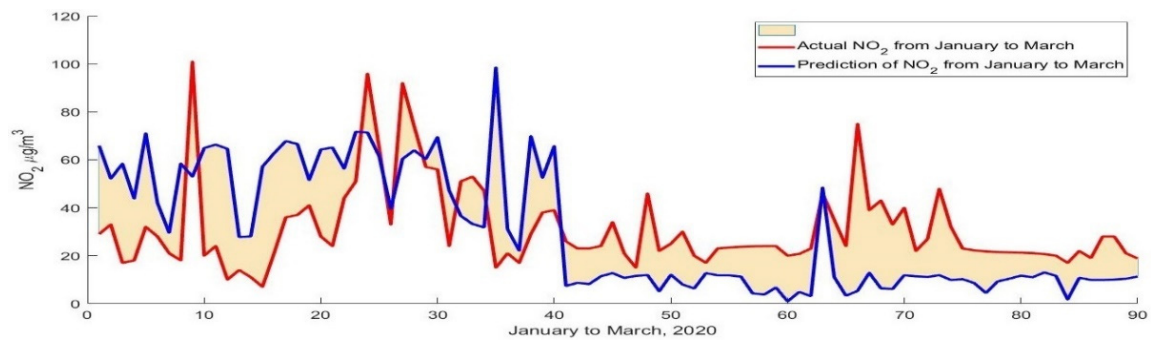


Figure A4. The first three months of 2020, the absolute measurement, and the predictive results of the model for NO_2 , at the UB-5 station.

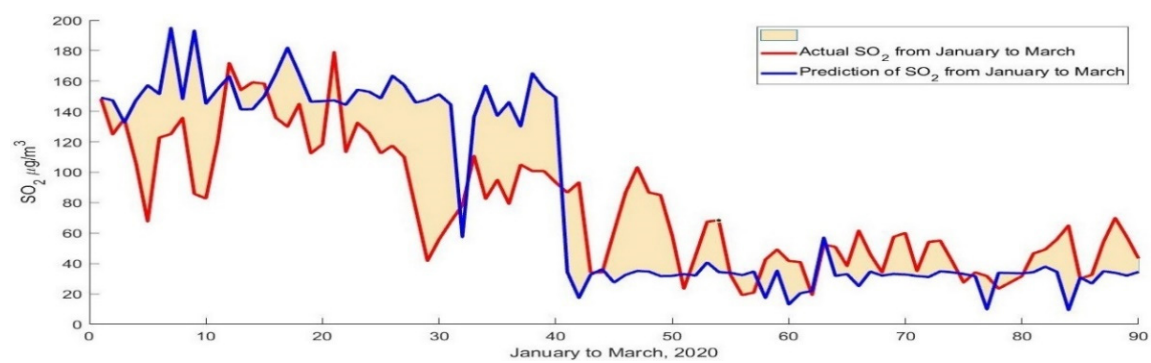


Figure A5. The first three months of 2020, the absolute measurement, and the predictive results of the model for SO_2 , at the UB-7 station.

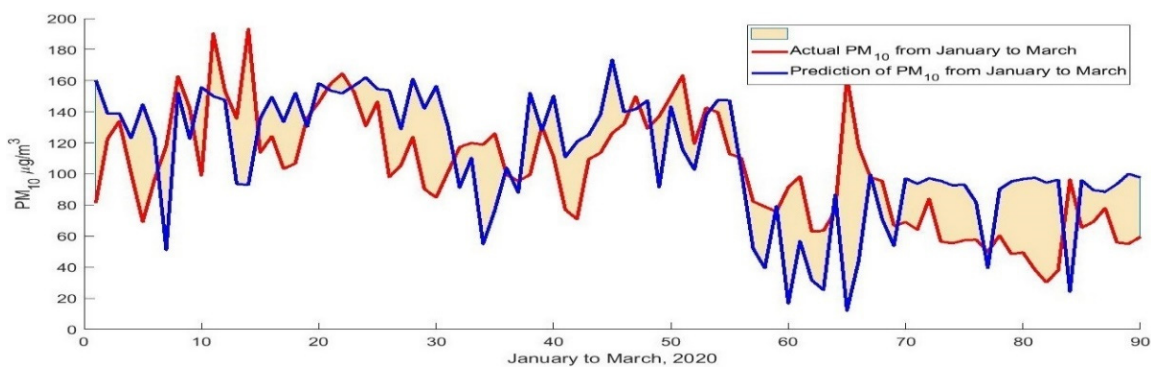


Figure A6. The first three months of 2020, the absolute measurement, and the predictive results of the model for PM_{10} , at the APRD-2 station.

The two graphics in Figures A7 and A8 show a comparison of the real and predicted $PM_{2.5}$ values of the method for the UB-1 station, as well as the prediction of CO for the APRD-4 station, respectively.

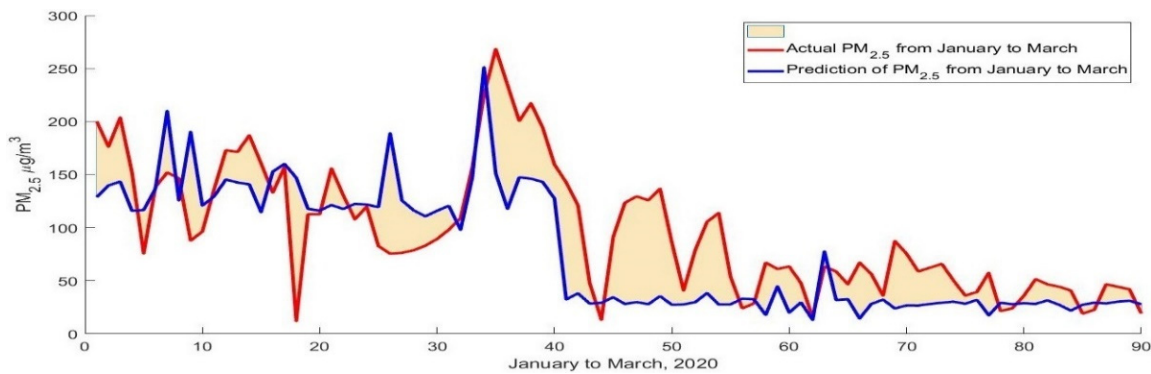


Figure A7. The first three months of 2020, the absolute measurement, and the predictive results of the model for $PM_{2.5}$, at the UB-1 station.

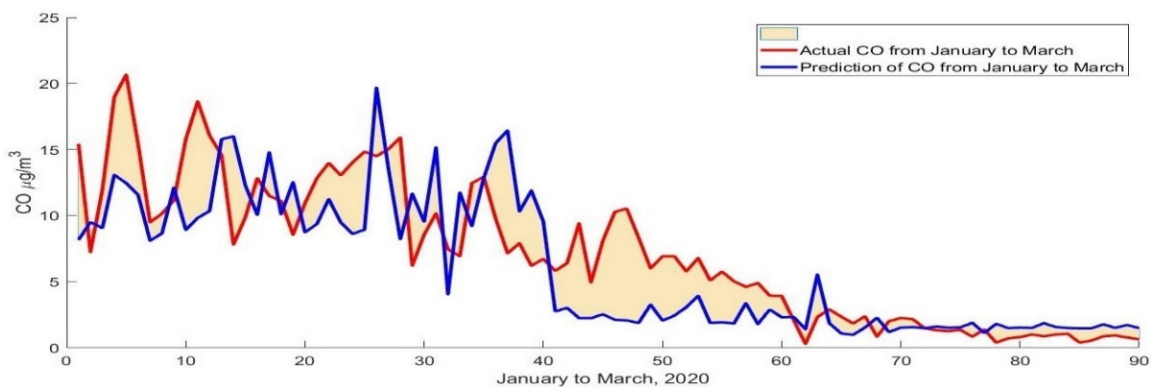


Figure A8. The first three months of 2020, the absolute measurement, and the predictive results of the model for CO, at the APRD-4 station.

Two of the corresponding contingency tables for NO_2 and $PM_{2.5}$ are shown in the following figure.

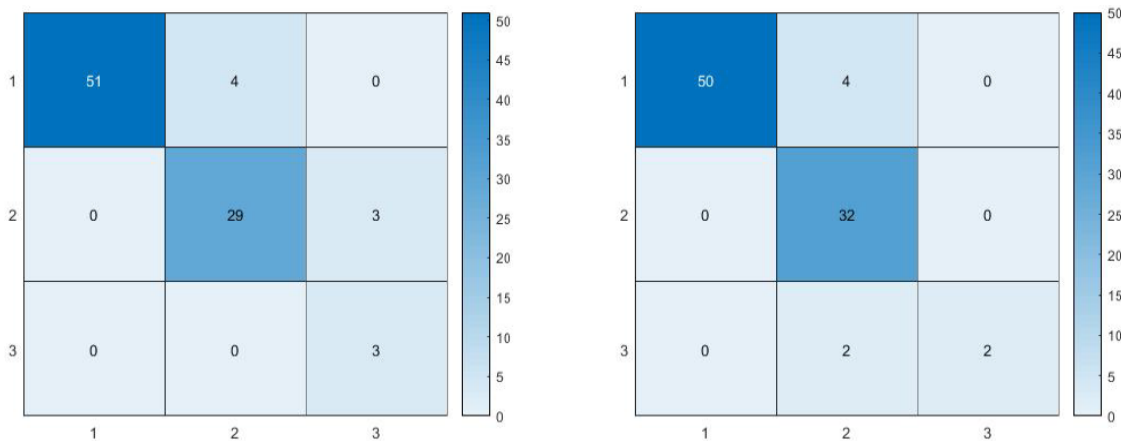


Figure A9. Contingency tables for NO_2 and $PM_{2.5}$.

References

1. Ritchie, H. Urbanization. 2018. Available online: <https://ourworldindata.org/urbanization> (accessed on 25 December 2021).
2. Ganbat, G.; Baik, J.J. Wintertime winds in and around the Ulaanbaatar metropolitan area in the presence of a temperature inversion. *Asia-Pac. J. Atmos. Sci.* **2016**, *52*, 309–325. [CrossRef]
3. Byamba, B.; Ishikawa, M. Municipal Solid Waste Management in Ulaanbaatar, Mongolia: Systems Analysis. *Sustainability* **2017**, *9*, 896. [CrossRef]

4. World Bank. Air Quality Analysis of Ulaanbaatar: Improving Air Quality to Reduce Health Impacts. 2011. Available online: <https://openknowledge.worldbank.org/handle/10986/26802> (accessed on 25 December 2021).
5. Guttikunda, S.K.; Lodoysamba, S.; Bulgansaikhan, B.; Dashdondog, B. Particulate pollution in Ulaanbaatar, Mongolia. *Air Qual. Atmos. Health* **2013**, *6*, 589–601. [[CrossRef](#)]
6. Luvsan, M.-E.; Shie, R.-H.; Purevdorj, T.; Badarch, L.; Baldorj, B.; Chan, C.-C. The influence of emission sources and meteorological conditions on SO₂ pollution in Mongolia. *Atmos. Environ.* **2012**, *61*, 542–549. [[CrossRef](#)]
7. Curbing Air Pollution in Mongolia's Capital, World Bank Report. 2012. Available online: <https://www.worldbank.org/en/news/feature/2012/04/25/curbing-air-pollution-in-mongolia-capital> (accessed on 25 December 2021).
8. Reich, S.; Gomez, D.; Dawidowski, L. Artificial neural network for the identification of unknown air pollution sources. *Atmos. Environ.* **1999**, *33*, 3045–3052. [[CrossRef](#)]
9. Moustiris, K.P.; Ziomas, I.C.; Paliatsos, A.G. 3-Day-Ahead Forecasting of Regional Pollution Index for the Pollutants NO₂, CO, SO₂, and O₃ Using Artificial Neural Networks in Athens, Greece. *Water Air Soil Pollut.* **2010**, *209*, 29–43. [[CrossRef](#)]
10. Rahman, N.L.; Muhammad, H.; Latif, M.T.; Suhartono, S. Forecasting of Air Pollution Index with Artificial Neural Network. *J. Teknol.* **2013**, *63*, 59–64. [[CrossRef](#)]
11. Azid, A.; Juahir, H.; Toriman, M.E.; Kamarudin, M.K.A.; Saudi, A.S.M.; Hasnam, C.N.C.; Aziz, N.A.A.; Azaman, F.; Latif, M.T.; Zainuddin, S.F.M.; et al. Prediction of the Level of Air Pollution Using Principal Component Analysis and Artificial Neural Network Techniques: A Case Study in Malaysia. *Water Air Soil Pollut.* **2014**, *225*, 2063. [[CrossRef](#)]
12. Maleki, H.; Sorooshian, A.; Goudarzi, G.; Baboli, Z.; Birgani, Y.T.; Rahmati, M. Air pollution prediction by using an artificial neural network model. *Clean Technol. Environ. Policy* **2019**, *21*, 1341–1352. [[CrossRef](#)]
13. Elangasinghe, M.A.; Singhal, N.; Dirks, K.N.; Salmond, J.A. Development of an ANN-based air pollution forecasting system with explicit knowledge through sensitivity analysis. *Atmos. Pollut. Res.* **2014**, *5*, 696–708. [[CrossRef](#)]
14. Cabaneros, S.M.; Calautit, J.K.; Hughes, B.R. A review of artificial neural network models for ambient air pollution prediction. *Environ. Model. Softw.* **2019**, *119*, 285–304. [[CrossRef](#)]
15. Liu, Z.; Peng, C.; Xiang, W.; Tian, D.; Deng, X.; Zhao, M. Application of artificial neural networks in global climate change and ecological research: An overview. *Chin. Sci. Bull.* **2010**, *55*, 3853–3863. [[CrossRef](#)]
16. Salami, E.S.; Ehteshami, M. Application of neural networks modeling to environmentally global climate change at San Joaquin Old River Station. *Model. Earth Syst. Environ.* **2016**, *2*, 38. [[CrossRef](#)]
17. Zhang, J.; Ding, W. Prediction of Air Pollutants Concentration Based on an Extreme Learning Machine: The Case of Hong Kong. *Int. J. Environ. Res. Public Health* **2017**, *14*, 114. [[CrossRef](#)] [[PubMed](#)]
18. Zhan, Y.; Luo, Y.; Deng, X.; Chen, H.; Grieneisen, M.L.; Shen, X.; Zhu, L.; Zhang, M. Spatiotemporal prediction of continuous daily PM_{2.5} concentrations across China using a spatially explicit machine learning algorithm. *Atmos. Environ.* **2017**, *155*, 129–139. [[CrossRef](#)]
19. Delavar, M.R.; Gholami, A.; Shiran, G.R.; Rashidi, Y.; Nakhaeizadeh, G.R.; Fedra, K.; Hatefi Afshar, S. A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 99. [[CrossRef](#)]
20. Ly, H.-B.; Le, L.M.; Phi, L.V.; Phan, V.-H.; Tran, V.Q.; Pham, B.T.; Le, T.-T.; Derrible, S. Development of an AI Model to Measure Traffic Air Pollution from Multisensor and Weather Data. *Sensors* **2019**, *19*, 4941. [[CrossRef](#)]
21. Arnaudo, E.; Farasin, A.; Rossi, C. A Comparative Analysis for Air Quality Estimation from Traffic and Meteorological Data. *Appl. Sci.* **2020**, *10*, 4587. [[CrossRef](#)]
22. Lee, M.; Lin, L.; Chen, C.Y. Forecasting Air Quality in Taiwan by Using Machine Learning. *Sci. Rep.* **2020**, *10*, 4153. [[CrossRef](#)]
23. Enebish, T.; Chau, K.; Jadamba, B. Meredith Franklin, Predicting ambient PM_{2.5} concentrations in Ulaanbaatar, Mongolia with machine learning approaches. *J. Expo. Sci. Environ. Epidemiol.* **2020**, *31*, 699–708. [[CrossRef](#)]
24. Franklin, M.; Chau, K.; Kalashnikova, O.V.; Garay, M.J.; Enebish, T.; Sorek-Hamer, M. Using Multi-Angle Imaging SpectroRadiometer Aerosol Mixture Properties for Air Quality Assessment in Mongolia. *Remote Sens.* **2018**, *10*, 1317. [[CrossRef](#)]
25. Bellinger, C.; Mohamed Jabbar, M.; Zaiane, O. A systematic review of data mining and machine learning for air pollution epidemiology. *BMC Public Health* **2017**, *17*, 907. [[CrossRef](#)]
26. Zhu, D.; Cai, C.; Yang, T.; Zhou, X. A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization. *Big Data Cogn. Comput.* **2018**, *2*, 5. [[CrossRef](#)]
27. Woschank, M.; Rauch, E.; Zsifkovits, H. A Review of Further Directions for Artificial Intelligence, Machine Learning, and Deep Learning in Smart Logistics. *Sustainability* **2020**, *12*, 3760. [[CrossRef](#)]
28. Olga, F.; Qin, W.; Markus, S.; Pierre, D.; Wan-Jui, L.; Melanie, D. Potential, challenges and future directions for deep learning in prognostics and health management applications. *Eng. Appl. Artif. Intell.* **2020**, *92*, 103678. [[CrossRef](#)]
29. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2015**, arXiv:1412.6980. Available online: <https://arxiv.org/abs/1412.6980> (accessed on 25 December 2021).
30. Tien, H.D.; Duc, M.N.; Evaggelia, T.; Angel, L.A.; Valerio, P.L.; Frank, J.P.; Wilfried, P.; Nikos, D. Matrix Completion with Variational Graph Autoencoders: Application in Hyperlocal Air Quality Inference. *arXiv* **2018**, arXiv:1811.01662. Available online: <https://arxiv.org/abs/1811.01662> (accessed on 25 December 2021).
31. Do, T.H.; Tsiligiani, E.; Qin, X.; Hofman, J.; La Manna, V.P.; Philips, W.; Deligiannis, N. Graph-Deep-Learning-Based Inference of Fine-Grained Air Quality From Mobile IoT Sensors. *IEEE Internet Things J.* **2020**, *7*, 8943–8955. [[CrossRef](#)]

32. Li, X.; Peng, L.; Hu, Y. Deep learning architecture for air quality predictions. *Environ. Sci. Pollut. Res.* **2016**, *23*, 22408–22417. [[CrossRef](#)] [[PubMed](#)]
33. Li, L. Geographically Weighted Machine Learning and Downscaling for High-Resolution Spatiotemporal Estimations of Wind Speed. *Remote Sens.* **2019**, *11*, 1378. [[CrossRef](#)]
34. Li, L.; Franklin, M.; Girguis, M.; Lurmann, F.; Wu, J.; Pavlovic, N.; Breton, C.; Gilliland, F.; Habre, R. Spatiotemporal imputation of MAIAC AOD using deep learning with downscaling. *Remote Sens. Environ.* **2020**, *237*, 111584. [[CrossRef](#)] [[PubMed](#)]
35. Li, L.; Fang, Y.; Wu, J.; Wang, J.; Ge, Y. Encoder-Decoder Full Residual Deep Networks for Robust Regression and Spatiotemporal Estimation. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4217–4230. [[CrossRef](#)] [[PubMed](#)]
36. Li, L.; Girguis, M.; Lurmann, F.; Pavlovic, N.; McClure, C.; Franklin, M.; Wu, J.; Oman, L.D.; Breton, C.; Gilliland, F.; et al. Ensemble-based deep learning for estimating PM_{2.5} over California with multisource big data including wildfire smoke. *Environ. Int.* **2020**, *145*, 106143. [[CrossRef](#)] [[PubMed](#)]
37. Li, L. A Robust Deep Learning Approach for Spatiotemporal Estimation of Satellite AOD and PM_{2.5}. *Remote Sens.* **2020**, *12*, 264. [[CrossRef](#)]
38. Fan, J.; Li, Q.; Hou, J.; Feng, X.; Karimian, H.; Lin, S. A Spatiotemporal Prediction Framework for Air Pollution Based on Deep RNN ISPRS Annals of Photogrammetry. *Remote Sens. Spat. Inf. Sci.* **2017**, *IV-4/W2*, 15–22. [[CrossRef](#)]
39. Qi, Z.; Wang, T.; Song, G.; Hu, W.; Li, X.; Zhang, Z. Deep Air Learning: Interpolation, Prediction, and Feature Analysis of Fine-Grained Air Quality. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 2285–2297. [[CrossRef](#)]
40. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2013**, arXiv:1312.6114. Available online: <https://arxiv.org/abs/1312.6114> (accessed on 25 December 2021).
41. Seo, J.K.; Kim, K.C.; Jargal, A.; Lee, K.; Harrach, B. A Learning-Based Method for Solving Ill-Posed Nonlinear Inverse Problems: A Simulation Study of Lung EIT. *SIAM J. Imaging Sci.* **2019**, *12*, 1275–1295. [[CrossRef](#)]
42. Hyun, C.M.; Baek, S.H.; Lee, M.; Lee, S.M.; Seo, J.K. Deep Learning-Based Solvability of Underdetermined Inverse Problems in Medical Imaging. *arXiv* **2020**, arXiv:2001.01432. Available online: <https://arxiv.org/abs/2001.01432> (accessed on 25 December 2021). [[CrossRef](#)] [[PubMed](#)]
43. Tsagaan, M.; Ganbat, B.; Renchin, S.; Khurlee, U.; Ichin, O. A Deep Variational Autoencoder Based Inverse Method for Active. Energy Consumption of Mining Plants and Ball Grinding Circuit Investigation. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2021**. [[CrossRef](#)]