Predicting Multivariate Air Pollution: A Gaussian-Mixture Nested Factorial Variational Autoencoder Approach

Prasanjit Dey, Soumyabrata Dev, *Member, IEEE*, and Bianca Schoen Phelan

***Abstract*— In recent years, global concern for human health has escalated due to the persistent threat of air pollution, resulting in a surge of chronic diseases and premature mortality. Poor air quality not only has adverse effects on human health but also poses negative impacts on vegetation, society, and the economy. Hence, it is imperative to invest more effort in accurately predicting multivariate air pollutants to offer practical and relevant solutions. However, many machine learning (ML) and deep learning (DL) models face significant challenges when dealing with the complexities of multivariate air pollution dynamics and the ill-posed nature of the data. In this letter, we propose a Gaussian-mixture nested factorial variational autoencoder (NF-VAE), specifically designed for multivariate air pollution prediction. To assess the performance of the proposed framework, we conducted experimental validation using air pollution data from six monitoring sites in Chinese cities. Three statistical indicators have been used to evaluate forecasting accuracy. The experimental results demonstrate the satisfactory performance of the NF-VAE model in predicting six pollutants for six different sites. Furthermore, the results indicate that the proposed NF-VAE model can effectively enhance efficiency gains, demonstrating improvements of at least 31% for RMSE, 22% for MAE, and 13% for R^2 compared with popular DL models, namely, long short-term memory (LSTM), gated recurrent unit (GRU), bidirectional LSTM (BiLSTM), and bidirectional GRU (BiGRU).**

***Index Terms*— Air pollutant, deep learning (DL), factorial variational autoencoder, latent space, machine learning (ML).**

I. Introduction

AIR quality is a critical concern globally, particularly in urban areas experiencing rapid population growth and increased industrialization. The degradation of air quality poses significant threats to public health, with documented adverse effects, including diseases, allergies, and, tragically, human fatalities [1], [2]. The air quality index (AQI) serves as a pivotal indicator of air quality, encompassing six major air pollutant components: fine particulate matter ($PM_{2.5}$), respirable PM_{10}, nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3) [3]. Monitoring and forecasting air quality have become essential endeavors for mitigating these health risks and improving overall well-being.

Machine learning (ML) and deep learning (DL) models for predicting air pollutants have garnered significant attention in the research community in recent years [4], [5]. These models leverage historical air quality data to make informed predictions, offering valuable insights into air pollution dynamics. For instance, Wang et al. [6] introduced a hybrid ML model integrating a biphasic decomposition method with an extreme learning machine (ELM) for AQI prediction, demonstrating promising results. Zhang and Li [7] proposed a DL model, fusing convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, to predict AQI with high accuracy. Meanwhile, Leong et al. [8] leveraged support vector machines (SVMs) to forecast the air pollution index (API) in specific regions. Zhang et al. [9] combined semi-supervised models, namely, empirical mode decomposition (EMD) and bidirectional LSTM (BiLSTM) neural networks for predicting $PM_{2.5}$ concentration. Chen et al. [10] proposed an integrated dual LSTM model to forecast air quality in a specific zone. Song et al. [11] introduced an LSTM-Kalman time prediction model, which uses LSTM to capture information from historical data and fine-tunes the time data sequence using Kalman filtering.

While these approaches have made significant strides in air quality prediction, there remains a need for more advanced techniques capable of handling the multivariate nature of air pollution, where various pollutants intercorrelate, exhibit complex temporal patterns, and present characteristics of an ill-posed problem [12]. The existing methods often struggle to capture these intricate relationships, leading to a lack of generalization capability. In dealing with multivariate and dynamic pollution data, the challenges arise from ill-posed problems lacking essential characteristics such as well-posedness, stability, and existence. Unlike well-posed problems, ill-posed ones, prevalent in pollution data's complexity, struggle to offer

Manuscript received 29 February 2024; revised 30 May 2024; accepted 6 June 2024. Date of publication 18 June 2024; date of current version 26 June 2024. This work was supported in part by the Science Foundation Ireland Center for Research Training in Digitally-Enhanced Reality (d-real) under Grant 18/CRT/6224; in part by the Science Foundation Ireland at the ADAPT SFI Research Center, University College Dublin, under Grant 13/RC/2106_P2; and in part by ADAPT, SFI Research Center for AI-Driven Digital Content Technology funded by the Science Foundation Ireland through the SFI Research Centers Program. *(Corresponding author: Prasanjit Dey.)*

Prasanjit Dey is with the ADAPT SFI Research Centre, School of Computer Science, Technological University Dublin, Dublin, D07 EWV4 Ireland (e-mail: d22124678@mytudublin.ie).

Soumyabrata Dev is with the ADAPT SFI Research Centre, School of Computer Science, University College Dublin, Dublin, D04 V1W8 Ireland (e-mail: soumyabrata.dev@ucd.ie).

Bianca Schoen Phelan is with the School of Computer Science, Technological University Dublin, Dublin, D07 EWV4 Ireland (e-mail: bianca.schoenphelan@tudublin.ie).

This article has supplementary downloadable material available at https://doi.org/10.1109/LGRS.2024.3416343, provided by the authors.

Digital Object Identifier 10.1109/LGRS.2024.3416343

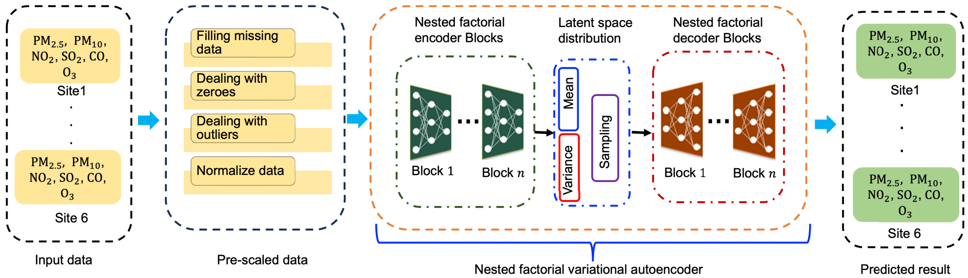

Fig. 1. Proposed forecasting framework for predicting air pollutants using Gaussian-mixture NF-VAE.

unique and stable generalization capabilities. Small variations or errors in initial data can result in significant uncertainties, sometimes making it impossible to determine generalization capability.

To address this problem, we have proposed an innovative Gaussian-mixture nested factorial variational autoencoder (NF-VAE). These variational autoencoders (VAEs), renowned for their ability to capture structured and compact latent encodings of data, stand out as promising candidates for achieving these objectives. Our method surpasses the limitations of the existing models as it addresses the intricate challenge of multivariate air pollution prediction and tackles the ill-posed nature of air pollution data. The NF-VAE framework empowers us to explore and untangle the complex relationships among various air quality components, ultimately leading to more precise and insightful multivariate predictions.

The main contributions of the letter are as follows.

1) The first contribution is primarily related to the development of an innovative Gaussian-mixture NF-VAE framework. The proposed framework is capable of handling multivariate pollutants prediction and addressing the ill-posed problem of the dynamic nature of air pollution data. It has the potential to improve the forecasting of multivariate ambient air pollution.
2) The second contribution lies in the validation of the proposed framework for multivariate data with multiple outputs.
3) The third and final contribution of this study involves a comprehensive comparative analysis between the proposed forecasting model and several powerful DL models. Through meticulous evaluation using air pollution data from six sites in China, the study demonstrates that the NF-VAE framework exhibits superior forecasting performance for various air pollutants. Notably, it outperforms well-established DL models such as LSTM, gated recurrent unit (GRU), BiLSTM, and bidirectional GRU (BiGRU), affirming its effectiveness in forecasting multivariate pollutants.

The following sections II of this letter delve into our NF-VAE approach. Section III presents a comparative analysis using real-world air quality data. Finally, Section IV describes the conclusions. To access the code for reproducing this research, one can find it in this repository: https://github.com/Prasanjit-Dey/NF-VAE.

II. Proposed Methodology

In this section, we describe the non-Gaussian characteristics of air pollution data. Following this, we introduce the proposed Gaussian-mixture NF-VAE framework for predicting $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3 levels from six monitoring sites. The overall framework of the proposed NF-VAE is illustrated in Fig. 1. The middle part of Fig. 1 represents the NF-VAE, which primarily consists of nested factorial encoder and decoder blocks. Each block represents an encoder/decoder network.

A. Variational Autoencoder

To gain a clearer understanding of the VAE, let us begin with a brief overview of the autoencoder. An autoencoder is a neural network designed to learn efficient codings of input data through unsupervised learning. It comprises three core components: the encoder, the latent space, and the decoder. The encoder compresses input data, such as air pollution measurements, into a latent-space representation. This latent space stores the essential information in a compressed form, optimizing the data distribution within this space. The decoder reconstructs the input data from the latent representation, aiming to produce outputs as close as possible to the original inputs.

B. Non-Gaussian Characteristics of Air Pollution Data

The behavior of pollutants (a), including $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3, is characterized by a significant diversity in scales. This wide-ranging variability leads to their characterization as having a strong non-Gaussian nature. For the sake of simplicity, let us assume that the parameter "a" follows a Gaussian (normal distribution). In this hypothetical scenario, different components or features extracted at various scales should not exhibit correlations or dependencies. However, this expected independence is not observed in practice, particularly when dealing with real-world data such as air pollution data. In reality, the dependencies between different scales are shown to be crucial for accurately capturing the non-Gaussian stochastic (random) structure of the data [13]. These dependencies can be captured by considering the following correlation matrix:

$$\mathbb{E}\{X(a)X(a)^T\} = \mathbb{E}\begin{bmatrix} W_a(W_a)^T & W_a(W|W_a|)^T \\ W|W_a|(W_a)^T & W|W_a|(W|W_a|)^T \end{bmatrix}. \quad (1)$$

The matrix includes three types of coefficients. Correlation coefficients $\mathbb{E}\{W_a(W_a)^T\}$ describe the roughness of the data. $\mathbb{E}\{W_a(W|W_a|)^T\}$ captures signed interactions between the coefficients of the air pollution data. They have the capability to detect sing-asymmetry and time-asymmetry within "a." Finally, coefficients $\mathbb{E}\{W|W_a|(W|W_a|)^T\}$ capture correlation between sign envelopes $|W_a|$ at different time intervals.

C. Gaussian-Mixture NF-VAE

To efficiently predict concentrations of $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3, a generative model is essential for extracting valuable features from non-Gaussian air pollution data. Gaussian-mixture VAEs [14], [15] are known for their ability to learn highly structured, low-dimensional latent representations of data, making them promising candidates for achieving our predictive goal. As our objective of predicting multivariate air pollutants from various input sources with different timestamps, we propose a Gaussian-mixture NF-VAE. This proposed model extracts essential features for predicting multivariate air pollutants with varying timestamps through a three-step process: 1) using a nested joint encoder to extract feature maps; 2) learning low-dimensional Gaussian-mixture latent variables for each timestamp; and 3) within the nested decoder, decoding the latent features of each timestamp to predict multivariate air pollutants. The subsequent sections provide a detailed description of our proposed generative model.

1) Generative Model: Let $a = a_0, \ldots, a_{t-1}$ denote the observed features of each pollutant, where t represents the timestamp. Our objective is to approximate the target joint distribution $p(a)$ through variational inference [16] using samples from this distribution as training data. To achieve this, we define the following generative model:

$$p_\theta(a, b, c) = p_\theta(a|c)p_\theta(c|b)p_\theta(b) = \prod_{i=0}^{t-1} p_\theta(a_i|c_i)p_\theta(c_i|b_i)p_\theta(b_i) \tag{2}$$

where c_i and b_i, for $i = 0, \ldots, t-1$, represent the Gaussian-mixture and categorical latent variables corresponding to the ith time stamp, respectively. Besides, $b = b_0, \ldots, b_{t-1}$, and $c = c_0, \ldots, c_{t-1}$ denote the collections of latent variables for all the time stamps. We choose the following parametric distributions for these random variables:

$$\begin{aligned} p_\theta(b_i) &= \mathrm{Cat}(e_i) \\ p_\theta(c_i|b_i) &= \mathcal{N}\left(c_i|\mu_{c,i}(b;\theta), \mathrm{diag}\left(\sigma^2_{c,i}(b_i;\theta)\right)\right) \\ p_\theta(a_i|c_i) &= \mathcal{N}\left(a_i|\mu_{a,i}(c;\theta), \mathrm{diag}\left(\sigma^2_{a,i}(c_i;\theta)\right)\right) \end{aligned} \tag{3}$$

where $p_\theta(c_i|b_i)$ denotes a Gaussian distribution having a mean defined as $\mu_{c,i}(b;\theta)$ and diagonal covariance of $\sigma^2_{c,i}(b_i;\theta)$. These parameters are simply learnable vectors for each b_i. This configuration results in a Gaussian mixture model for c_i. Given c_i, we define a_i as a Gaussian distribution having a mean of $\mu_{a,i}(c;\theta)$ and diagonal standard deviation $\sigma^2_{a,i}(c_i;\theta)$, both of which are parameterized using deep neural networks. The generative model mentioned above involves independent decoders, meaning there is a mapping from the latent variable to a covariance representation for each timestamp in the prediction task. However, the training of this generative model requires using a latent posterior distribution inference model.

TABLE I

HYPERPARAMETER SETTINGS FOR NE-VAE, LSTM, GRU, BI-LSTM, AND BI-GRU ACCORDING TO THE BEST MODELS' PERFORMANCE

Models	Batch Size	Learning Rate	Latent size	Epochs	Optimizer	Activation	Loss
NE-VAE	256	10^{-3}	32	100	Adam	ReLU	MSE
LSTM	256	10^{-3}	–	100	Adam	ReLU	MSE
GRU	256	10^{-3}	–	100	Adam	ReLU	MSE
Bi-LSTM	256	10^{-3}	–	100	Adam	ReLU	MSE
Bi-GRU	256	10^{-3}	–	100	Adam	ReLU	MSE

2) Inference Model: The earlier described model requires the process of marginalizing the Gaussian-mixture and categorical latent distributions to evaluate the likelihood of the parametric distribution $p_\theta(a)$. However, due to the high dimensionality of this distribution, the computational complexity of this process is very high. To address this challenge, we approximate the latent posterior distribution $q(b, c|a)$. It explores potential features across multivariate pollutants at various time stamps. To predict the multivariate pollutants, we used an approximation to the latent posterior distribution using the following factorization:

$$q_\phi(c, b|a) = q_\phi(c|b, a)q_\phi(b|a) = \prod_{i=0}^{t-1} q_\phi(c_i|b_i, a)q_\phi(b_i|a). \tag{4}$$

We use the subsequent parameterizations to train an amortized latent posterior model

$$\begin{aligned} q_\phi(b_i|a) &= \mathrm{Cat}(\pi_i(a;\phi)) \\ q_\phi(c_i|b_i, a) &= \mathcal{N}\left(c_i|\mu_{c,i}(a, b_i;\phi), \mathrm{diag}\left(\sigma^2_{c,i}(a, b_i;\phi)\right)\right) \end{aligned} \tag{5}$$

where $i \in \{0, \ldots, t-1\}$, $\pi_i(a;\phi)$ denotes the forecast membership probabilities for the covariance input "a" at the ith timestamp. As the inferred latent variable encodes information regarding the forecast membership of "a," we explicitly incorporated both "a" and b_i into the deep neural network to parameterize its mean $\mu_{c,i}(a, b_i;\phi)$ and diagonal covariance $\sigma^2_{c,i}(a, b_i;\phi)$.

3) Objective Function: To train the NF-VAE, the objective is to minimize the reverse Kullback–Leibler (KL) divergence between the parameterized distribution and the true joint distribution

$$\mathcal{KL}(p(a)\,||\,p_\theta(a)) = \mathbb{E}_{a\sim p(a)}[\log p(a) - \log p_\theta(a)] = \mathbb{E}_{a\sim p(a)}[-\log p_\theta(a)] + \mathrm{const.} \tag{6}$$

III. EXPERIMENTAL RESULT

A. Dataset and Parameter Description

The dataset comprises six major air pollutants ($PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3), gathered from six monitoring stations in Beijing, China, available via OpenAQ [17]. It spans from December 27, 2017, to August 9, 2021, recorded at hourly intervals, totaling 25 941 h. We used 90% of the data (the first three years) for training, reserving the last year's data (February 17, 2021, to August 8, 2021) for testing the models

TABLE II

QUANTITATIVE COMPARISON WITH NE-VAE AND FOUR OTHER METHODS FOR SIX POLLUTANTS ACROSS SIX MONITORING SITES. THE METRICS INCLUDE RMSE, MAE, AND R^2. WE USE TWO COLORS TO LABEL THE TOP TWO METHODS FOR EACH POLLUTANT AT EACH SITE: GREEN INDICATES THE BEST MODEL RESULT, AND BLUE INDICATES THE SECOND-BEST MODEL RESULT

Site	Pollutant	NE-VAE			LSTM			GRU			BiLSTM			BiGRU		
		RMSE↓	MAE↓	R^2 ↑	RMSE↓	MAE↓	R^2 ↑	RMSE↓	MAE↓	R^2 ↑	RMSE↓	MAE↓	R^2 ↑	RMSE↓	MAE↓	R^2 ↑
Site1 (6167)	$PM_{2.5}$	7.13	5.06	0.951	8.02	5.37	0.935	8.67	5.93	0.924	7.78	5.29	0.938	8.10	5.52	0.933
	PM_{10}	13.1	8.86	0.930	19.3	12.5	0.861	19.2	12.6	0.862	20.2	13.2	0.848	20.3	13.1	0.846
	NO_2	6.90	4.89	0.938	10.3	6.69	0.865	9.90	6.46	0.876	10.3	6.68	0.864	10.3	6.91	0.864
	SO_2	1.26	0.84	0.942	3.86	1.11	0.512	4.00	1.28	0.475	3.81	1.32	0.524	3.62	1.16	0.572
	CO	97.5	59.4	0.921	126.0	82.5	0.857	123.2	77.9	0.863	130.4	80.7	0.847	130.3	86.0	0.847
	O_3	10.1	7.11	0.959	10.3	6.41	0.958	9.89	6.21	0.961	10.3	6.38	0.957	10.3	6.58	0.958
Site2 (6168)	$PM_{2.5}$	7.16	5.05	0.955	8.28	5.44	0.937	8.42	5.77	0.935	7.82	5.36	0.944	8.23	5.66	0.938
	PM_{10}	11.3	7.90	0.936	16.1	10.5	0.881	15.9	10.6	0.884	17.9	12.3	0.852	16.7	11.0	0.871
	NO_2	6.26	4.68	0.945	9.45	6.27	0.878	9.30	6.33	0.882	9.65	6.56	0.873	9.89	6.76	0.866
	SO_2	1.73	1.00	0.888	1.96	1.15	0.771	1.84	1.09	0.799	1.90	1.08	0.787	1.91	1.11	0.782
	CO	94.2	58.5	0.920	128.6	79.2	0.848	129.4	81.7	0.846	129.2	78.0	0.847	131.8	83.2	0.841
	O_3	9.77	6.95	0.962	10.7	6.81	0.954	9.70	5.78	0.963	10.3	6.30	0.958	10.0	6.07	0.960
Site3 (6169)	$PM_{2.5}$	10.2	5.11	0.890	22.6	6.00	0.437	13.9	5.81	0.786	22.5	5.95	0.446	14.9	5.67	0.756
	PM_{10}	11.6	7.94	0.924	16.5	10.8	0.859	16.3	10.9	0.862	17.7	11.5	0.837	16.5	10.8	0.859
	NO_2	6.94	4.67	0.936	10.0	6.54	0.850	10.3	6.85	0.841	10.1	6.69	0.848	10.4	6.82	0.839
	SO_2	1.02	0.70	0.951	3.03	1.02	0.555	3.02	1.01	0.558	3.05	1.17	0.549	3.08	1.10	0.542
	CO	119.6	59.2	0.882	203.5	73.9	0.655	206.1	75.3	0.646	202.7	70.6	0.658	204.5	74.2	0.652
	O_3	9.90	7.18	0.960	11.0	6.91	0.949	10.4	6.24	0.955	10.8	6.50	0.952	10.7	6.65	0.952
Site4 (6218)	$PM_{2.5}$	7.79	5.71	0.944	9.35	6.49	0.918	10.2	7.29	0.901	9.33	6.34	0.918	9.28	6.55	0.919
	PM_{10}	13.1	9.10	0.931	15.1	9.92	0.922	16.4	10.6	0.908	15.1	9.99	0.923	16.1	9.86	0.912
	NO_2	6.45	4.57	0.926	10.7	6.66	0.798	10.7	6.92	0.800	10.7	6.75	0.798	10.9	7.02	0.791
	SO_2	1.39	1.00	0.933	3.49	2.08	0.571	3.54	2.34	0.559	3.81	2.18	0.489	3.70	2.27	0.517
	CO	87.9	61.9	0.919	139.1	89.9	0.805	133.4	87.4	0.821	121.9	81.3	0.850	124.7	84.2	0.843
	O_3	10.0	7.10	0.958	11.5	6.77	0.944	10.7	6.19	0.951	11.6	6.69	0.943	11.3	6.82	0.946
Site5 (6273)	$PM_{2.5}$	9.40	6.34	0.946	11.0	6.61	0.927	10.5	6.04	0.933	10.6	6.04	0.932	10.6	6.24	0.933
	PM_{10}	18.1	11.4	0.935	17.6	10.9	0.952	17.7	10.6	0.951	17.7	10.9	0.951	19.3	11.4	0.942
	NO_2	4.72	3.38	0.942	6.47	4.37	0.899	6.62	4.48	0.895	6.54	4.45	0.897	6.94	4.86	0.884
	SO_2	3.81	2.44	0.938	12.8	3.96	0.331	13.5	3.64	0.264	12.3	3.72	0.390	12.3	3.60	0.381
	CO	72.6	52.7	0.944	107.0	70.4	0.879	108.3	70.5	0.874	109.1	71.0	0.875	110.6	74.6	0.871
	O_3	9.76	6.81	0.969	10.5	6.55	0.965	9.96	6.01	0.969	10.3	6.33	0.966	10.4	6.47	0.965
Site6 (6274)	$PM_{2.5}$	8.68	6.09	0.957	10.6	6.38	0.937	9.60	6.02	0.948	9.61	6.02	0.948	10.8	7.32	0.934
	PM_{10}	15.8	10.3	0.942	17.7	11.2	0.942	16.8	10.3	0.948	16.8	10.7	0.948	17.6	11.2	0.943
	NO_2	5.95	4.24	0.926	9.27	6.12	0.837	9.22	5.94	0.838	9.06	5.89	0.844	9.37	6.12	0.833
	SO_2	1.97	1.43	0.926	4.58	2.52	0.613	4.63	2.51	0.604	4.85	2.63	0.566	4.95	2.93	0.548
	CO	83.3	59.2	0.948	139.4	90.1	0.860	141.9	92.3	0.855	138.2	87.7	0.862	139.9	89.6	0.859
	O_3	9.67	6.80	0.965	10.0	5.92	0.964	9.70	6.01	0.966	10.0	6.36	0.963	10.6	7.20	0.958

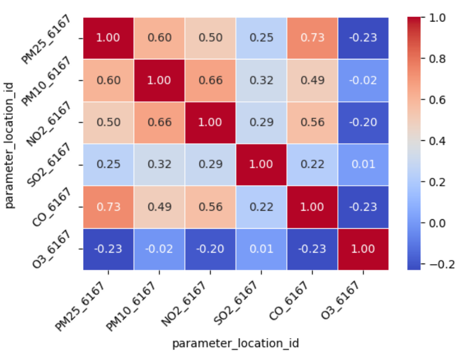

Fig. 2. Correlation matrix among ambient air pollutants ($PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3) in site 1 (6167).

(see supplementary material for dataset location). We have included a detailed summary of the hyperparameters used for each model in Table I, which enables reproducing the results. In addition, Fig. 2 shows the correlation coefficients among $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3 over Site1 (6167) (see supplementary for other sites). This figure highlights a moderate positive correlation among $PM_{2.5}$, PM_{10}, NO_2, SO_2, and CO. The efficient correlation within the dataset supports the robust training of our proposed NE-VAE model, as well as other comparison models.

B. Result Analysis

The experiments conducted in this study are designed with the primary objective of analyzing the performance of the NE-VAE model alongside four other DL models in predicting six different pollutants across six distinct sites. In the context of multivariate forecasting, each model undergoes a supervised training process, aiming to capture and understand the temporal dependencies present in the time-series data of each pollutant. The primary aim is to acquire a robust understanding of long-term predictions and enhance the models' predictive capabilities.

To evaluate the performance of the NE-VAE model, we conducted a quantitative comparison with four other methods: LSTM, GRU, BiLSTM, and BiGRU, across six pollutants at six monitoring sites. Table II shows that the proposed NE-VAE emerges as the most effective approach for addressing multivariate air pollution forecasting problems, showcasing high efficiency and satisfactory accuracy. Notably, the NE-VAE framework outperforms popular DL models such as LSTM, GRU, BiLSTM, and BiGRU in forecasting the levels of all the investigated pollutants ($PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3) measured at six Chinese location—Site1 (6167), Site2 (6168), Site3 (6169), Site4 (6218), Site5 (6273), and Site6 (6274). As anticipated, the NE-VAE framework achieves the lowest forecasting error (RMSE, MAE) and the highest R^2 score for most of the six sites, reinforcing its effectiveness in providing accurate predictions for diverse environmental variables.

For a visual illustration, the observed ground-truth concentrations (averaged over 24 h for each data point) alongside the forecast concentrations using NE-VAE, LSTM, GRU, Bi-LSTM, and Bi-GRU models are depicted in Fig. 3 for CO from Site1 (6167). The results from other stations show relatively similar outcomes. From Fig. 3, it is evident that the proposed NE-VAE model demonstrates a strong ability to accurately forecast future trends in CO concentration dynamics.

Similarly, Fig. 4 provides insight into the average validation metrics for $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3 pollutants across six sites, using the proposed NE-VAE framework in comparison to LSTM, GRU, BiLSTM, and BiGRU models. The results indicate a substantial performance advantage for the NE-VAE framework, showcasing efficiency gains of at least 31% for RMSE, 22% for MAE, and 13% for R^2 when contrasted with the LSTM, GRU, BiLSTM, and BiGRU

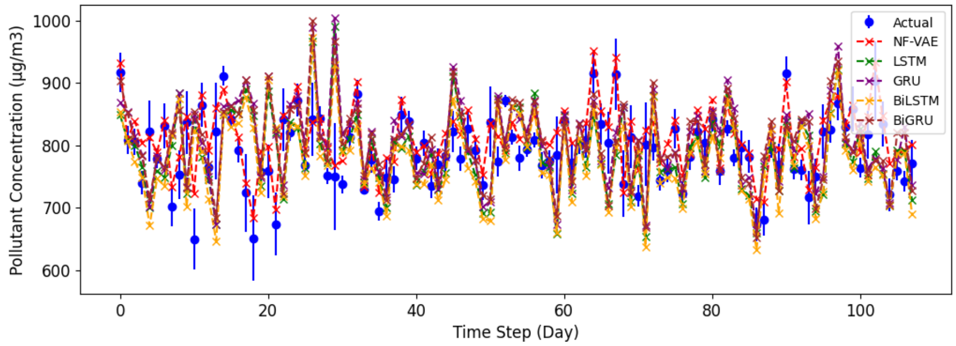

Fig. 3. Comparison of ground-truth CO concentrations versus predicted CO concentrations using NE-VAE, LSTM, GRU, Bi-LSTM, and Bi-GRU for Site1 (6167). The blue dot with a line represents the average standard deviation (error) of all the models for each data point (averaged over 24 h).

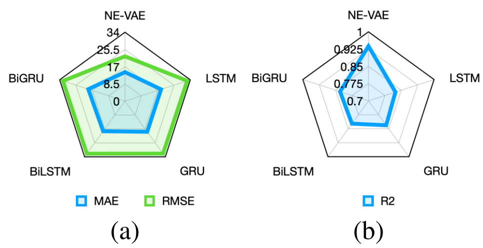

Fig. 4. Average validation metrics [(a) RMSE, MAE, and (b) (R^2)] for six pollutants across six sites in multivariate forecasting using the proposed NE-VAE, LSTM, GRU, BiLSTM, and BiGRU models.

models. These findings underscore the efficacy of the proposed NE-VAE approach in achieving superior predictive accuracy for a diverse set of pollutants, emphasizing its potential as a robust model for multivariate time-series forecasting in environmental studies.

IV. Conclusion

Air pollution is a global issue with detrimental health effects, exacerbated by industrial advancements. Monitoring ambient air quality is crucial. This letter introduces the Gaussian-mixture NF-VAE as an effective framework for improving air pollution forecasting. Our study shows that the NF-VAE model outperforms LSTM, GRU, BiLSTM, and BiGRU in forecasting six key pollutants ($PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3), using RMSE, MAE, and R^2 as evaluation metrics. One limitation of our proposed NF-VAE model is the assumption of a Gaussian-mixture distribution. While the model assumes that the data follow a certain pattern (a mixture of Gaussian distributions), the complex and diverse nature of air pollution dynamics at some sites may not align well with this assumption. To address this, we plan to use data preprocessing techniques and introduce a multiscale version of the model that combines NF-VAE with a wavelet-based multiresolution representation in the future. These steps will enhance the model's performance and applicability to diverse air pollution scenarios.

Acknowledgment

The authors would like to thank Loghlen Rickard from University College Dublin for introducing them to the dataset that has been used in this letter.

References

[1] M. Abbasi-Kangevari et al., "Effect of air pollution on disease burden, mortality, and life expectancy in North Africa and the Middle East: A systematic analysis for the Global Burden of Disease study 2019," *Lancet Planet. Health*, vol. 7, no. 5, pp. 358–369, 2023.

[2] F. de la Barrera, F. Barraza, P. Favier, V. Ruiz, and J. Quense, "Megafires in Chile 2017: Monitoring multiscale environmental impacts of burned ecosystems," *Sci. Total Environ.*, vols. 637–638, pp. 1526–1536, Oct. 2018.

[3] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the Internet of Things using big data analytics," *Comput. Netw.*, vol. 101, no. 4, pp. 63–80, Jun. 2016.

[4] S. Du, T. Li, Y. Yang, and S.-J. Horng, "Deep air quality forecasting using hybrid deep learning framework," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2412–2424, Jun. 2021.

[5] W. Huang, T. Li, J. Liu, P. Xie, S. Du, and F. Teng, "An overview of air quality analysis by big data techniques: Monitoring, forecasting, and traceability," *Inf. Fusion*, vol. 75, pp. 28–40, Nov. 2021.

[6] D. Wang, S. Wei, H. Luo, C. Yue, and O. Grunder, "A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine," *Sci. Total Environ.*, vol. 580, pp. 719–733, Feb. 2017.

[7] J. Zhang and S. Li, "Air quality index forecast in Beijing based on CNN-LSTM multi-model," *Chemosphere*, vol. 308, Dec. 2022, Art. no. 136180.

[8] W. C. Leong, R. O. Kelani, and Z. Ahmad, "Prediction of air pollution index (API) using support vector machine (SVM)," *J. Environ. Chem. Eng.*, vol. 8, no. 3, Jun. 2020, Art. no. 103208.

[9] L. Zhang, P. Liu, L. Zhao, G. Wang, W. Zhang, and J. Liu, "Air quality predictions with a semi-supervised bidirectional LSTM neural network," *Atmos. Pollut. Res.*, vol. 12, no. 1, pp. 328–339, 2021.

[10] H. Chen, M. Guan, and H. Li, "Air quality prediction based on integrated dual LSTM model," *IEEE Access*, vol. 9, pp. 93285–93297, 2021.

[11] X. Song, J. Huang, and D. Song, "Air quality prediction based on LSTM-Kalman model," in *Proc. IEEE 8th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, May 2019, pp. 695–699.

[12] M. Han and M. Xu, "Laplacian echo state network for multivariate time series prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 238–244, Jan. 2018.

[13] R. Morel, G. Rochette, R. Leonarduzzi, J.-P. Bouchaud, and S. Mallat, "Scale dependencies and self-similar models with wavelet scattering spectra," 2022, *arXiv:2204.10177*.

[14] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[15] B. Paige et al., "Learning disentangled representations with semi-supervised deep generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[16] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Amer. Stat. Assoc.*, vol. 112, no. 518, pp. 859–877, Apr. 2017.

[17] *Openaq*. Accessed: Jan. 13, 2023. [Online]. Available: https://openaq.org/