IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 March 2023, accepted 4 April 2023, date of publication 10 April 2023, date of current version 13 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3265725

APPLIED RESEARCH

Long Short-Term Memory Bayesian Neural Network for Air Pollution Forecast

HUYNH A. D. NGUYEN[1], (Student Member, IEEE), QUANG P. HA[1], (Senior Member, IEEE), HIEP DUC[2], MERCHED AZZI[2], NINGBO JIANG[2], XAVIER BARTHELEMY[2], AND MATTHEW RILEY[2]

[1]Faculty of Engineering and Information Technology (IT), University of Technology Sydney, Ultimo, NSW 2007, Australia
[2]Department of Planning and Environment, NSW, Parramatta, NSW 2141, Australia

Corresponding author: Huynh A. D. Nguyen (huynhanhduy.nguyen@uts.edu.au)

This work was supported in part by the University of Technology Sydney under Project PRO20-11875 and Project PRO21-13285, and in part by the Department of Planning and Environment, Parramatta, NSW, Australia.

ABSTRACT This paper presents a data fusion framework to enhance the accuracy of air-pollutant forecast in the state of New South Wales (NSW), Australia using deep learning (DL) as a core model. Here, we propose a long short-term memory Bayesian neural network (LSTM-BNN) to improve performance of the predictive profiles via quantifying uncertainties and adjusting model parameters. For this, we develop a new inferring technique for kernel density estimation with subdivision tuning to ensure both forecast accuracy and computational efficiency with a limited number of samples from the prediction distributions. Moreover, a novel algorithm called spatially-adjusted multivariate imputation by chained equation is also developed to take into account spatial correlations between nearby air-quality stations for correctly imputing the incoming data, and hence, to enable forecasting at a local scale. The LSTM-BNN framework is evaluated with observed datasets collected from stations and modeling outputs generated by the Conformal Cubic Atmospheric Model - Chemical Transport Model (CCAM-CTM) currently used in NSW. The airborne pollutants under investigation are $PM_{2.5}$ and ozone, which frequently exceed the standards. The results obtained from data fusion with our framework demonstrated high performance of the proposed LSTM-BNN model in air-pollutant prediction with reductions of over 30% in root mean square error compared to CCAM-CTM and over 50 % in inferring time compared to a DL model with Gaussian-based inference. Accuracy and reliability of the proposed model were also achieved with air pollution forecast in various seasons and suburbs.

INDEX TERMS Air pollution, deep learning, long short term memory, spatial-correlated imputation, time-series forecast, uncertainty.

I. INTRODUCTION

Air pollution, regardless of emissions sources, has become an important topic in environmental research and management [1], [2], particularly in urban areas [3]. To effectively control air quality, it requires an accurate and reliable solution for urban air pollution forecasting. In the state of New South Wales (NSW), Australia, the key airborne pollutants such as particulate matters ($PM_{2.5}$, PM_{10}) and ozone (O_3) are monitored in real-time as well as regularly predicted by numerical modeling, using the dispersion model (Chemical Transport Model - CTM [4] or Community Multiscale Air Quality Modeling System - CMAQ) integrated with meteorological models (the Conformal Cubic Atmospheric Model - CCAM, or weather research and forecast (WRF) model [5]). Although these models are frequently upgraded to predict air-pollutant concentrations at a large scale, the accuracy of the forecasts is limited due to the dependency on initial assumptions and emission inventory defined during the simulation of the complex process of emissions, dispersion and transformation of air pollutants in physical-chemical reactions [4]. In particular, concentrations of fine particles

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

and ozone frequently exceed the healthy level and are quite difficult to predict. The essential requirement is to improve accuracy of the forecast and maintain stable performance by exploiting available sources of environmental data.

Air pollution prediction is a challenging task due to the dynamic and non-stationary nature of the time series of pollutants. For this, statistical methods have been widely applied, integrating spatial correlations with trends and seasonal patterns to estimate the chronological dependency of historical and future values [6]. Commonly-used techniques based on autoregressive integrated moving-average (ARIMA) with different variants (e.g., VARIMA, SARIMA) [7], [8] have contributed to a large number of studies, involving data that are often stationary, detrended or deseasonalized. These conditions may however be unfulfilled for real-world applications, wherein temporal distributions are prone to not only locations but also environmental perturbations and data drifts [9].

Improving the accuracy and reliability of the forecast would require an approach that can learn from the big data to be less dependent on the assumptions of model-based or statistics-based methods. In this regard, machine learning (ML) techniques such as artificial neural networks [10], support vector machine [11], random forest [1], K-nearest neighbors or naïve Bayes [12] have been applied to the time-series prediction of airborne pollutants. However, these shallow-learning methods require intensive data processing procedures involving high computational latency before training and during prediction [9]. As an advanced version of ML, deep learning (DL) models, with the ability to learn ''deeper'' from multiple layers, can produce superior performance comparable to human experts [13]. Among the state-of-the-art DL networks, the convolutional neural network (CNN) and recurrent neural network (RNN) demonstrate their high performance for comprehensive learning of spatial and temporal features respectively in many applications [14], [15], including air quality forecasting [16]. The long short-term memory (LSTM) [17] as a variant of RNN, has proved its robustness in capturing long-term dependency of time and causal features of the inputs, especially pollutant concentrations and meteorological values. Therefore, the LSTM is particularly promising in air-quality estimation [18], [19].

For air-quality forecast, most of the learning-based models result in point-wise estimation at each time step, which can be considered as deterministic. As such, good predictions can be achieved only when training data and remaining data share the same statistical properties (e.g., the same distribution). This condition is impractical for deep learning with air pollution data in the presence of spatial-temporal correlations and influences of external changes in emissions, weather patterns and the multifaceted factors of environmental volatility [20]. Indeed, as mentioned in a recent survey [21], uncertainties associated with those conditions and incoming data imperfectness, such as missing and out-of-distribution values, pose a challenge in deep learning, for which the integration of probabilistic methods such as Bayesian reasoning into deep neural networks [22] is worth exploring to deal with uncertainties.

In this paper, we propose a new deep-learning model that can incorporate spatially dispersed features when learning the time series of airborne pollutant concentrations via the fusion of historical observations and predicted values from the CCAM-CTM output. The proposed model can handle data uncertainties and imperfectness by using a recursive neural network with Bayesian inference forming the long short-term memory-Bayesian neural network (LSTM-BNN). Besides, we aim to optimally approximate the probability density function (PDF) to produce the distribution of the pollution forecast at each time step. Therefore, Bayesian modeling of uncertainties with variational inference is applied to both training and forecasting tasks. The quantification of uncertainties from the proposed technique presents an effective treatment of bias inference associated with the conventional Gaussian assumption for distributions of the forecast values and significantly reduces the sampling numbers.

The contributions of this paper include:

- An effective LSTM-BNN framework using an LSTM single-step recursive forecast model in combination with Bayesian reasoning for data fusion of air pollution observations and existing numerical estimations.
- A new algorithm for kernel density estimation with subdivision tuning (KDEST), developed for the air pollutant's probability density function with a reduced amount of samples of forecast distributions for uncertainty quantification.
- A new imputation algorithm for spatially-adjusted multivariate imputation by chained equations (SAMICE), developed to adaptively impute missing observation data of the target location based on correlation with neighbor stations.
- Possible application of the proposed model with spatial inferences to be integrated with the system managing all stations in a region to achieve the required accuracy and reliability of the forecast, as verified through extensive experiments.

The paper is organized as follows. After the Introduction, Section II presents the proposed LSTM-BNN framework. Section III is devoted to the handling of uncertainties with the approximation for Bayesian inference and the proposed KDEST technique. Section IV presents the imputation of spatio-temporal distributions of air pollutants and the SAMICE algorithm development. The results and discussion from comparison and ablation analysis for different model configurations in various seasons are included in Section V, indicating the potential of applying the proposed approach for NSW suburbs. Finally, a conclusion is drawn in the last section, Section VI.

II. DEEP LEARNING FRAMEWORK FOR AIR-POLLUTANT FORECAST

In this section, we introduce the datasets used for training, validating and forecasting along with their collection and preprocessing before modeling with our proposed framework with LSTM-BNN model.

A. AIR-QUALITY DATA FROM OBSERVATIONS AND NUMERICAL MODEL

The air-quality data are collected from two sources: (i) observations (OBS), measured from air quality stations, and (ii) numerical model's predicted values from CCAM-CTM.

1) OBSERVATIONS - OBS

The real observations are open databases managed and published by NSW government through the application programming interface (API), which provides air pollution information from over 50 state-run stations over the whole NSW [23]. This includes the main pollutants such as $PM_{2.5}$, PM_{10}, O_3, NO, NO_2, CO, SO_2, and NH_3 along with visibility and meteorological variables (i.e., wind speed, wind direction, air temperature, relative humidity and rainfall).

2) CCAM-CTM

From the numerical models, pollutant concentrations are available for up to 72-hour forecasts obtained from the combination of two numerical models currently used in NSW state of Australia:

- The Conformal Cubic Atmospheric Model (CCAM) is a 3D cubic atmospheric model which uses a non-hydrostatic, semi-implicit, semi-Lagrangian dynamical core to simulate climate and weather at fine resolutions. It accounts sufficiently for the local topography, atmospheric processes and associated climate impacts or extreme weather features (e.g., tropical cyclones or bushfires) [24].
- The Chemical Transport Model (CTM) is currently deployed for predictions of particles ($PM_{2.5}$ and PM_{10}), NO, NO_2 and O_3. This model employs data of emissions and anthropogenic sources from the air quality inventory of NSW-Sydney Greater Metropolitan Region (GMR), calculated emissions for marine aerosol, wind-blown dust, volatile organic compounds (VOC), as an integration of the sources and distribution sizes of air pollutants [25].

The combined CCAM-CTM has been implemented in NSW since 2017 to flexibly scale the predictions at different resolutions (80 km × 80 km, 27 km × 27 km, 9 km × 9 km, and 3 km × 3 km) respectively in accordance with four grid domains, namely Australia, NSW, GMR and Sydney basin for modeling accurately the transportation of air pollutants across a wide region [4]. In our study, we use the GMR domain (60 × 60 grid cells at 9 km × 9 km) for CCAM-CTM values based on the average distance between the air-quality monitoring stations [3].

3) ACCURACY AND RELIABILITY

An essential requirement for air pollution forecasting is to maintain its accuracy. Here, the CCAM-CTM model requires highly accurate capture of variable emissions sources as the main inputs in order to infer estimation outputs via multiple chemical reactions and physical equations [25]. However, as a result of inaccurate predictions of organic compounds and other chemical species, the model displays unreliable results at different seasons such as overestimation and underestimation of $PM_{2.5}$ in winter and summer, respectively [4].

Another problem is data leakage in OBS data due to missing information or imperfect conditions. This may occur at stations and low-cost sensors from unexpected failures of instruments or various impacts of volatile environment [26]. The missing information problem degrades the capacity of learning the dynamic features and other extreme events from the time series. Moreover, data-driven models are ineffective because of incomplete inputs or absent variables.

The drifting effect remains also a problem for air quality prediction. According to a recent report from 2012-2018 in the NSW GMR [27], the pollutant concentrations vary significantly from year to year, especially for the particle levels. Therefore, the forecast performance is inevitably affected by the concept drift problem in air-quality data when using any learning technique with a pre-trained model [9]. As such, historical data may appear to be insufficient to handle the prediction in the coming periods given chaotic changes of air pollution.

To overcome these issues, we develop an effective technique for kernel density estimation with subdivision tuning (KDEST) to improve prediction accuracy and smoothen the distribution shape of limited samples obtained from our LSTN-BNN model. Besides, an algorithm for spatially-adjusted multivariate imputation by chained equations (SAMICE) is proposed to handle any missing information or abrupt changes in concentration levels to update the spatio-temporal distributions of new incoming data from nearby stations.

4) DATA PARTITION AND MODEL CONFIGURATION

Each variable in this work constitutes approximately 30,000 hourly-averaged values from March 2018 to August 2021 used for training, validating and testing the model. The raw data will be scanned to remove the outliers (negative or extreme values 3 times higher than averages), resample missing time steps, and impute the missing values. After preprocessing, the inputs are transformed into matrices of three dimensions, i.e. number of samples, number of time steps, and number of features, to create a set of spatio-temporal data for fitting to the forecast model. Then, the transformed dataset is divided into training and validating sets respectively with splitting ratios of 80% and 10%, while the testing set with a splitting ratio of 10% of the total samples. This selection with the sliding windows method in training

TABLE 1. Model configuration.

Hyperparameters	Values and types	Notes
Input layer	12-24-36-48-60-72	No. historical values
Hidden layers	3 layers	128 nodes per layer
Dropout layers	3 layers	Dropout ratio: 0.1 - 0.5
Output layer	6-12-24-36-48-60-72	Forecast horizons
Batch size	128	No. samples per epoch
Learning rate	$1 - 3 \times 10^{-4}$	Depend on pollutant
Optimizer	Adam	Adaptive learning rate
Activation function	ReLU	Rectified Linear Unit
Early stopping	$\nu = 20$	Stop after ν epochs

and testing processes accounts for the temporal nature of the time series used in the LSTM forecast model. Indeed, assigning 80% of the total for the training set can cover all seasonal patterns, extreme events (e.g., bushfires in black summer 2019-2020 in NSW, Australia) and other episodes of air pollutants. Hence, the distributions of various features can be considered as fully learned by the proposed model during the training process. The 10% of hourly-averaged data (approximate 3000 values) assigned for each validating and testing datasets, equivalent to 125 days, can sufficiently evaluate the generic capacity of our model for a particular season of the year. Before training, the data were normalized in the interval [0, 1] to increase the speed of convergence and reduce the prediction bias.

From empirical experiments with our real data, the hyperparameters selected in our model are summarized in Table 1. Here, the activation function is the Rectified Linear Unit (ReLU), and Adam optimizer is chosen for training because it applies an adaptively stochastic optimization method that is suitable for time series. Finally, the early stopping method is also applied with $\nu = 20$ epochs to reduce overfitting.

The proposed model architecture and related functions are developed with a high-level neural network API in Python, namely Keras running on Tensorflow, an open-source library for machine learning tasks. We train our program on the Interactive High Performance Computing (iHPC) server with NVIDIA Quadro RTX 6000 GPU.

B. FUSION OF OBSERVATION AND MODEL DATA

Methods for multistep-ahead predictions can be categorized as (i) direct or one-shot forecast and (ii) recursive forecast. The former produces a sequence of multiple time steps at one prediction, which is suitable for stationary data with seasonal patterns (e.g., temperature). This method may however face the uncertainty problem in air pollutants, causing instability in predictive performance, especially for the long-term forecast (e.g., 48h-, 60h- or 72h-future values). On the other hand, the latter method, using a one-step model recursively with its new input updated by the latest prediction values, can flexibly produce multiple forecast time steps in an iterative manner to reach the standard period of 72-hour forecast.

To mitigate the accumulated errors, we apply a recursive model for single-step forecast with the fusion of real-world observations and predicted CCAM-CTM data. Here, as only historical observations are available, a recent forecast value from the model output is fed back to join the input sequence for the next time-step forecast. Taking advantage of multistep-ahead forecast in physical modeling, the predicted data from CCAM-CTM model are combined with observations and previous forecast values to enhance the knowledge of future trends which contribute to reducing the predictive error at each forecast time step. During operations, OBS data obtained from the monitoring stations and low-cost sensor networks are updated hourly to the input sequence for replacing the previous forecast to suppress the model uncertainty of previous forecast. Apart from improving prediction performance, this method also allows for the prevention of data leakage in DL with neural networks.

C. PROPOSED ARCHITECTURE

In our approach, a recurrent neural network (RNN) model is utilized as the main core to formulate our proposed framework owing to its robustness of modeling sequences and flexibility with respect to different scenarios of prediction such as one-to-one, one-to-many, many-to-one or many-to-many time-step predictions [13]. To control the flow of information from the input sequences with the long-term patterns of time series of airborne pollutants, overcome the vanishing and exploding issues in RNNs and, more importantly, improve the forecast accuracy via quantification of uncertainties, we propose a hybrid deep learning model using LSTM-BNN, based on the LSTM network in integration with Bayesian inference. Here, to implement the LSTM-BNN predictive model, recurrent layers are stacked intermittently with Monte-Carlo (MC) dropout layers for regularization, prevention of overfitting and quantification of uncertainties during prediction [28].

The sequential structure of an LSTM layer includes multiple memory cells with inputs x_t of air-pollutant concentrations and two other states: the previous cell state C_{t-1} and the hidden state h_{t-1}. The LSTM cell state is described by the following equation [17]:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \tag{1}$$

where f_t and i_t are respectively the forget and input gates:

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f), \tag{2}$$
$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i), \tag{3}$$
$$\tilde{C}_t = tanh(W_C.[h_{t-1}, x_t] + b_C). \tag{4}$$

The hidden state h_t is determined as a function of the cell state C_t:

$$h_t = o_t * tanh(C_t), \tag{5}$$

where the output gate o_t is determined as:

$$o_t = \sigma(W_f.[h_{t-1}, x_t] + b_o), \tag{6}$$

σ and *tanh* represent respectively sigmoid and hyperbolic activation functions:

$$\sigma(x) = \frac{1}{1+e^{-x}}, \tag{7}$$

$$tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}. \tag{8}$$

The DL parameters W_f, W_i, W_C, and b_f, b_i and b_C are respectively the weights and biases of the three gates, to be iteratively updated following the pattern of air-quality features during training. Their values depend on the level of correlation between patterns. The cell state C_t is updated by an element-wise product ($*$) of the forget gate with the previous state ($f_t * C_{t-1}$) to skip the unimportant features and add up with the new feature from the input gate ($i_t * \tilde{C}_t$).

The proposed LSTM-BNN framework is depicted in Fig. 1, where its inputs combine CCAM-CTM and the real-world OBS collected from monitoring stations and low-cost sensor networks with data being normalized in [0, 1] and formulated as a matrix of m rows of historical time-step values and n columns featuring air pollutants of interest as well as meteorological, spatial and temporal variables. Since the raw data are subject to intermittent missing and/or erroneous values due to sensors' noises, external disturbances as well as stochastic dynamics of air pollutants, it is essential to develop effective methods not only for the treatment of uncertainties as mentioned above but also for data imputation, to be addressed in the next sections.

III. BAYESIAN INFERENCE FOR MODEL UNCERTAINTY HANDLING

As uncertainties in DL models are inevitable, we integrate the Bayesian inference method in the LSTM model for uncertainty quantification to improve the accuracy and reliability of the forecast.

A. BAYESIAN INFERENCE IN NEURAL NETWORK

The degree of belief on the neural network model specifications based on new information (observations) can be inferred from the posterior $p(\omega|x)$ determined by the Bayesian theorem:

$$p(\omega|x) = \frac{p(x,\omega)}{P(x)} \Longrightarrow p(\omega|x) = \frac{p(x|\omega)p(\omega)}{P(x)}, \tag{9}$$

where ω represents the learnable parameters of the network, x denotes the input data of air-quality values as well as auxiliary variables (e.g., temporal, meteorological or topographical features) given to the model for training and predicting, $p(x|\omega)$ is the likelihood of input values given ω with the prior $p(\omega)$, and $p(x)$ is the marginal likelihood for the input distribution. The prior is initialized model's weights updated from the previous batches of data in each training epoch, sampled from parameters ω, assumed to follow the Gaussian distribution ($\omega \sim \mathcal{N}(\mu_\omega, \sigma_\omega)$).

For every new parameter ω_i sampled in the distribution $p(\omega|X_{new})$ of model's posterior, a new predictive value ($y_{forecast}$ - future air-pollutant value) is generated from the new inputs X_{new} of OBS and CCAM-CTM values. The distribution of these predictions then formulates the posterior of forecast or predictive distribution $P(y_{forecast}|X_{new})$:

$$p(y_{forecast}|X_{new}) = \int p(y_{forecast}|\omega)p(\omega|X_{new})d\omega. \tag{10}$$

Since the posterior of weights $p(\omega|X_{new})$ is intractable, its approximation can be sought via (i) sampling the model parameters, or (ii) variational inference to find an equivalent distribution $q(\omega)$. The latter method, faster and less computationally expensive, is applied in this study. For variational inference, the aim is to minimize the distance between the posterior distribution $p(\omega|x)$ and its equivalence. A measure for this distance is the Kullback-Leibler (KL) divergence defined as [29]:

$$D_{KL}(q(\omega)||p(\omega|x)) = \int q(\omega) log \frac{q(\omega)}{p(\omega|x)} d\omega. \tag{11}$$

Thus, to achieve the closest approximation of $p(\omega|x)$, the KL divergence should be minimized:

$$\min_{\theta} D_{KL}(q(\omega,\theta)||p(\omega|x)) = \min_{\theta} \mathbb{E}_{q(\omega,\theta)}\left[logq(\omega,\theta) - logp(\omega|x)\right], \tag{12}$$

where $\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta)$ is an intrinsic parameter to be obtained from the minimization, $\mathbb{E}_{q(\omega,\theta)}$ is the expectation of the distribution $q(\omega,\theta)$. To proceed with the minimization of the KL divergence for airborne pollutant distributions, we used the dropout technique with layer weights (ω) initialized by an L_2-regularization, which has been proved to be an approximate solution to (12) [28]. In consistence with the mathematical establishment that all-layered MC dropout best approximates a Bayesian neural network [30], the MC dropout has been verified as superior to other state-of-the-art uncertainty estimation techniques, particularly with strong robustness to noise [31]. Here, dropout is implemented by skipping some hidden nodes in a layer of the neural network to form a varying configuration for the network at an inference time.

First, through multiple samplings of forecast values $y_{forecast}$ given input X_{new}, we form the equivalent distribution $p(y_{forecast}|X_{new})$ as obtained from Bayesian inference. With a given dropout probability p_{drop}, by repeated random sampling using the Monte Carlo (MC) algorithm to omit neurons in our LSTM-BNN, the obtained distribution can be considered as the closest to posterior $p(\omega|x)$, considered as equivalent to the result obtained from the minimization of the KL divergence (12). Furthermore, since the effective management of uncertainties by Bayesian inference using the MC-dropout technique may incur a tradeoff in computing expenses, we present the following a new algorithm to estimate the smooth distribution and enhance inference accuracy by reducing the number of samples.

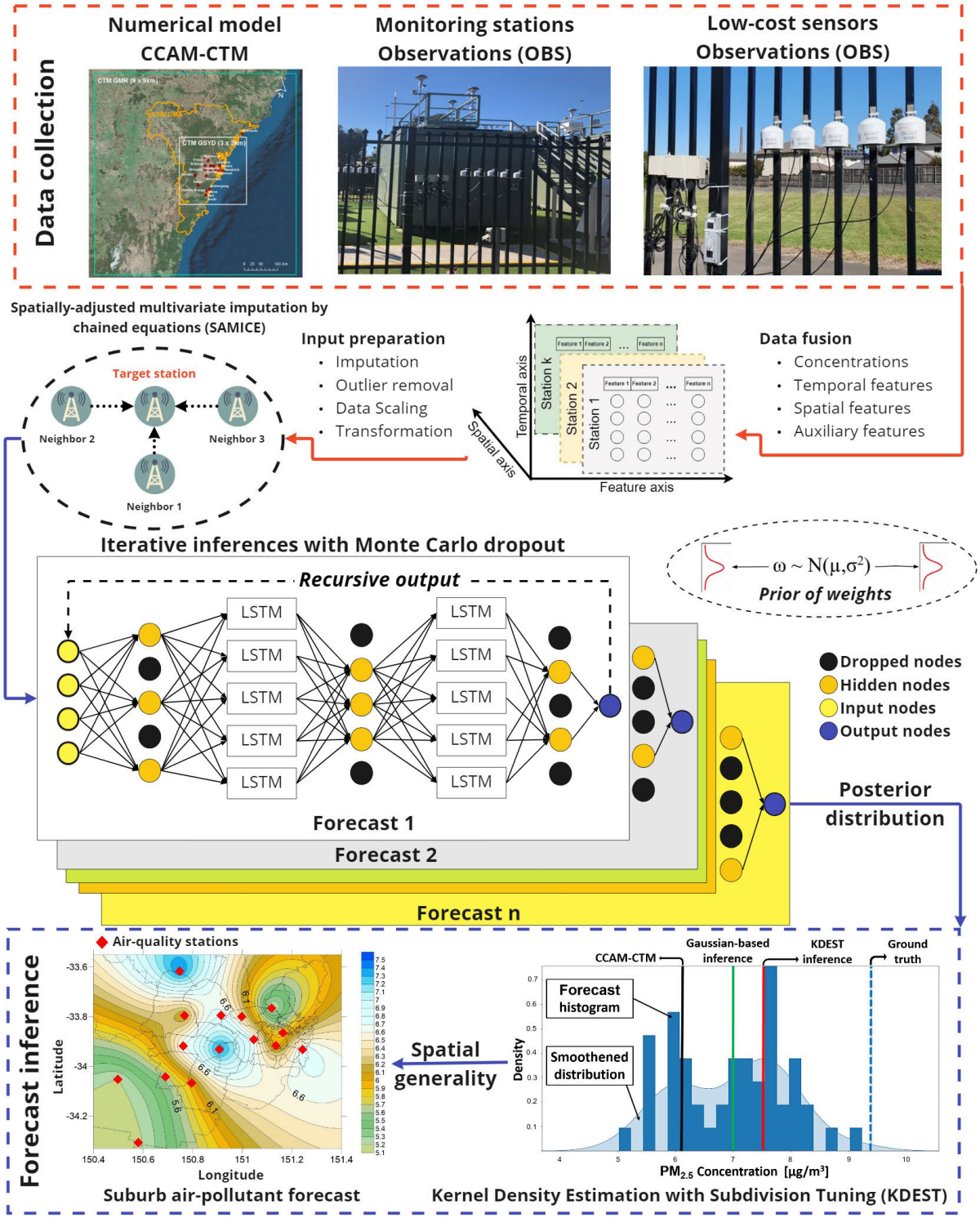
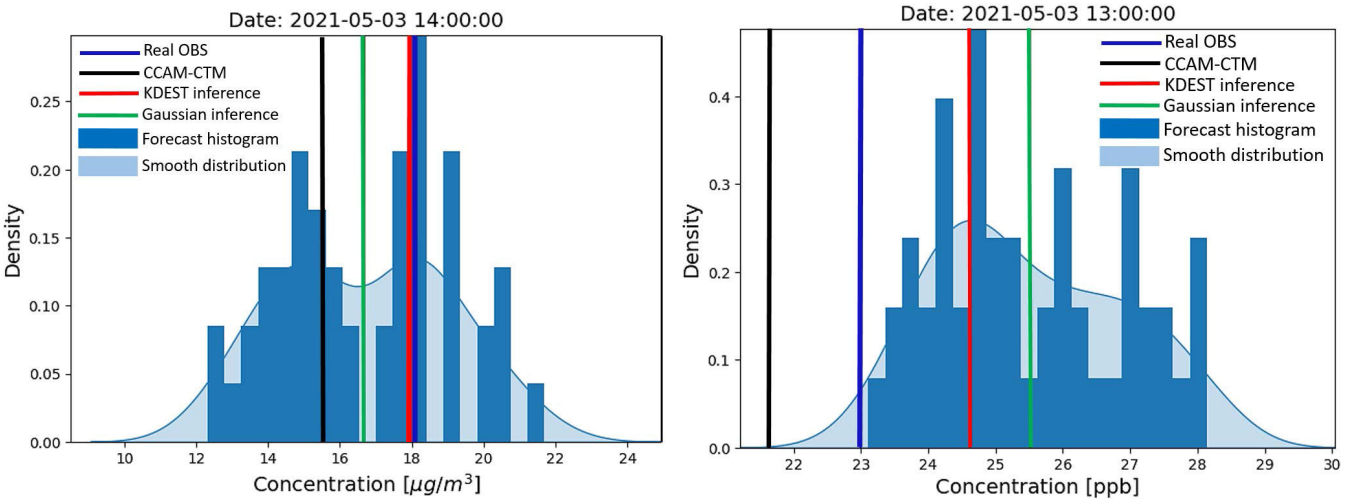


FIGURE 1. Structure of the proposed LSTM-BNN framework.

B. KERNEL DENSITY ESTIMATION WITH SUBDIVISION TUNING

The Gaussian distribution assumption is widely used in probabilistic models for applications with a large number of samples for each time step according to the central limit theorem (CLT) [32]. As such, in air-quality forecasting, the cost of computation is quite expensive for multi-step ahead estimation over a large region. Moreover, in addition



(a) $PM_{2.5}$ (b) Ozone

FIGURE 2. Histograms and inferences of CCAM-CTM (black line), Gaussian inference by the mean of distribution (green line) and the maximum likelihood estimation (red line) from KDEST, compared to the ground truth OBS (blue line).

to inevitable uncertainties, direct Gaussian-based inferences from observations containing non-normal distributions could incur some bias issues. As a remedy for that and also taking into account the recursive fusion of CCAM-CTM and OBS for large-area prediction, we propose to subdivide the data for adjusting the parameters of the kernel density estimation (KDE), a non-parametric method commonly used to infer the smooth shape of distributions from the observed data [33]. The idea is to obtain an optimal approximation of a probability density function (PDF) of each forecast time step.

To estimate the distribution density at point x, we consider the weighted distances with its neighbor points x_i, $1 \leq i \leq n$, where n is the number of the distribution's samples. Now, we firstly define an estimate of the distribution density via a kernel function [34]:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big), \tag{13}$$

where the kernel is a Gaussian function $K(x, h) \propto \exp(-\frac{x^2}{2h^2})$ having its positive bandwidth h. This KDE parameter controls the tradeoff between the bias (underfit) and variance (overfit) of the estimation [35]. The large bandwidth may cause a high bias with a very smooth density distribution and vice versa. Here, optimized values for the bandwidth are sought in accordance with various sets of samples from the predictive distribution by using the proposed technique for kernel density estimation with subdivision tuning (KDEST). Our development is based on the unbiased cross-validation method for kernel nonparametric density estimation. The integrated square error (ISE) of the density estimate is then:

$$\begin{aligned} ISE &= \int [(f(x) - \hat{f}(x)]^2 dx \\ &= \int f^2(x)dx - 2\int f(x)\hat{f}(x)dx + \int \hat{f}^2(x)dx), \end{aligned} \tag{14}$$

where $f(x)$ is the real density function of the forecast posterior distribution. As $f(x)$ does not involve the bandwidth h, it can be ignored in the minimization of ISE for the optimal bandwidth. Hence, when minimizing ISE in h, the first term, $\int f^2(x)dx$, is thus omitted while the second term containing the statistic mean of the estimate $\hat{f}(x)$ in (14) becomes approximately $-2(\frac{1}{n}\sum_{i=1}^{n}\hat{f}(x_i))$. Therefore, the minimization of ISE can be rendered to the minimization of an unbiased cross variation J_{ISE}:

$$J_{ISE} = \int \hat{f}^2(x)dx - \frac{2}{n}\sum_{i=1}^{n}\hat{f}(x_i). \tag{15}$$

To proceed, for a given bandwidth range $\mathcal{H}$, we randomly divide the posterior distribution into κ partitions ($\kappa \geq 2$), each of k samples, i.e. $n = \kappa k$. To tune for an optimal bandwidth h_{opt} in the range, we consider first $(\kappa - 1)$ partitions to compute the estimated density function (13) for each $h_j \in \mathcal{H}$ with an iteration step Δh as,

$$\hat{f}_j(x) = \frac{1}{(n-k)h_j} \sum_{i=1}^{n-k} K_j\Big(\frac{x - x_i}{h_j}\Big), \tag{16}$$

and use the last partition of k data points to obtain the average index (15):

$$J_{ISE}(h_j) = \frac{1}{k}\Big(\sum_{i=1}^{k}\int \hat{f}_j^2(x)dx\Big) - \frac{2}{k}\sum_{i=1}^{k}\hat{f}_j(x_i). \tag{17}$$

The procedure of finding optimized bandwidth (h_{opt}) is summarized in the pseudo-code of Algorithm 1. After subdivision tuning and cross-variation optimization, the kernels obtained are used to infer the forecast values of the posterior distribution.

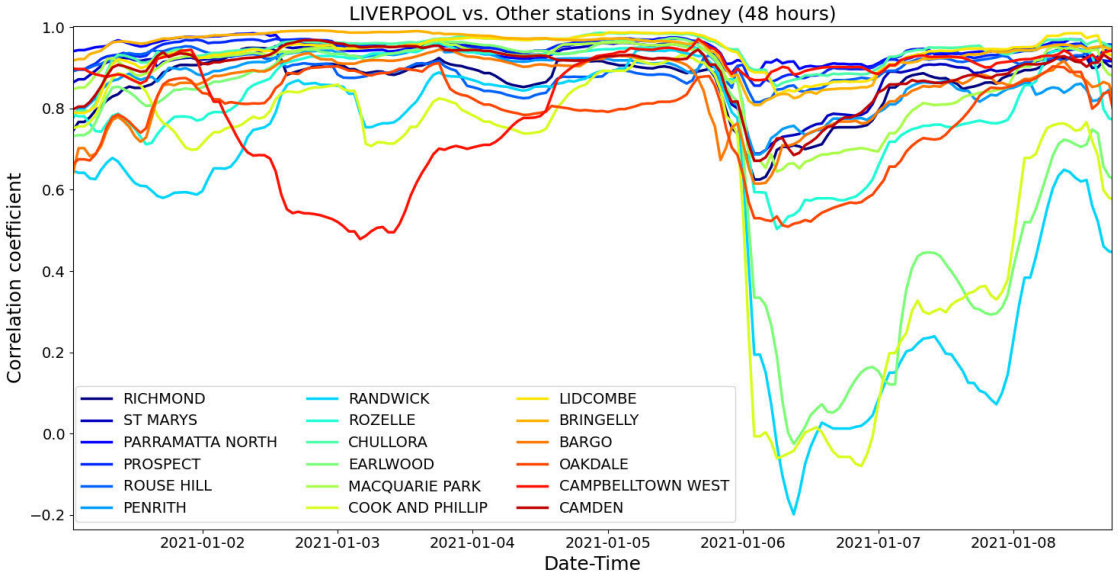


FIGURE 3. Correlations of ozone at Liverpool with respect to other stations in the 1st week of January 2021.

Algorithm 1 KDEST

Input: Forecast posterior distribution at a time step
Output: Optimal bandwidth (h_{opt})

1: Define bandwidth range $\mathcal{H}$.
2: Divide the posterior distribution into κ partitions of k data points each.
3: **for each** $h_j \in \mathcal{H}$ **do** (with $h_j = h_{j-1} + \Delta h$)
4: Compute the estimate (16) and index (17).
5: **end for**
6: $h_{opt} = argmin\ J_{ISE}(h_j)$

To illustrate our inference for forecasting three air pollutants $PM_{2.5}$ and ozone on the 3rd of May 2021 with $n = 20$ samples, we selected the partition number $\kappa = 4$ and the bandwidth h in the range $\mathcal{H}$=[0.1, 1.5] with Δh=0.1. The histogram, in dark blue, of sampled distributions typically for $PM_{2.5}$ and ozone concentrations are depicted in of Fig. 2(a) and (b), respectively, with the inference values obtained respectively from KDEST (red line), Gaussian-based mean of samples (green line), CCAM-CTM (black line), and ground truth of observations (blue line). It is discernible of the skewed and abnormal distribution in the discrete samples of the posteriors. Accordingly, when the forecast values are inferred with posterior distributions assumed to be Gaussian [28], there will be biased predictions. In this regard, KDEST can reduce the gap to the ground truth. Unlike a recent probabilistic study [36], here KDEST results in a more accurate and smoother probability density function of the posterior, and hence contributes to improving the forecast accuracy.

IV. SPATIO-TEMPORAL DISTRIBUTION IMPUTATION

Missing information and imperfectness in data recording are important issues in air quality prediction. This section is devoted to the imputation techniques we developed in this work for air-pollutant forecasts.

A. CORRELATION OF SPATIO-TEMPORAL PROFILES OF AIR POLLUTANTS

Correlations of measurements at regional air-quality stations are widely used to select appropriate features of air pollutants or meteorologies for imputing, training and forecast in association with the spatial relation analysis [37]. For levels of a pollutant collected at 19 air-quality stations over the Greater Metropolitan Region (GMR) of Sydney in 2021 [23], the variation of correlations with respect to locations is quite excessive and obviously represents a concern for data imputation. Temporal short-term correlations also vary episodically due to complex dispersion, meteorological impacts, emissions conditions and chemical reactions of the pollutants, resulting in intermediate correlations of an air pollutant at a target station with other stations. For example, at Liverpool station in early 2021, distinguished changes in the correlation for ozone occurred at Cook & Phillip, Randwick and Earlwood with respect to Liverpool, in a window of 48-hour observations during the 1st week of January 2021 as shown in Fig. 3. Similarly for particle concentrations, the correlation changes were observed owing to the impact of local meteorologies (e.g., wind, rainfall, air humidity and others) [3]. The above rationale has motivated us to develop a technique to remove or scale down the influence of low-correlated stations during the imputation for incoming model inputs. In this paper, to enhance the forecast performance, we propose a correlation-based adjustment algorithm for imputing the missing information between stations in the cluster or region of interest under the context of real-world operation for our DL model.

B. SPATIALLY-ADJUSTED MULTIVARIATE IMPUTATION BY CHAINED EQUATIONS

In environmental monitoring, the observations measured are occasionally absent due to volatile impacts of outdoor conditions, system failures or communication problems [26]. Information loss may cause model corruptions with any incomplete length of inputs. Besides, the forecast becomes unreliable by using only statistical properties of historical data to infer the absent information due to the concept drift problem [9]. As a remedy, to deal with the drift of incoming data by referring to nearby stations through correlations between stations, we develop here a novel online imputation technique, called spatially-adjusted multivariate imputation by chained equations (SAMICE). The proposed technique, modified from the multiple imputation method by chained equations (MICE) [38], also treats a feature with missing values as a dependent variable and other remaining variables as predictors in a multiple regression model. Here, not all features in the dataset but only the most correlated variables are involved in the regressive-based imputation from the predictive distributions of the fitted model.

Let us consider the whole OBS dataset $Y \in \mathbb{R}^{m\times n}$ (m samples and n stations). We denote Y_i the set of target observations at the i^{th} station, $Y_j, j \neq i$ the set of observations collected from the neighbor stations, and Y_{-i} the set of missing measurements at the i^{th} station ($Y_{-i} \subset Y_i$). The missing values y_{-i} are the responses of the i^{th} regression model for imputing missing values in Y_i by using information from $y_j \in Y_j$. This regression model is often defined as,

$$y_{-i} = y_j \cdot \alpha_{-i} + \beta_{-i}, \tag{18}$$

where α_{-i} and β_{-i} are respectively the regression coefficients and intercepts.

The idea behind our SAMICE algorithm is to utilize only the most correlated variables to improve the regressive-based imputation. Here, a faster convergence with higher reliability is expected to result by reducing uncertainties from predictions after multiple cycles of imputation. For that, the correlation between the target station i and neighbor station j is first obtained from the Pearson's correlation coefficient:

$$r_{ij} = \frac{\sum_{i,j}(y_i - \bar{y}_i)(y_j - \bar{y}_j)}{\sqrt{\sum_i (y_i - \bar{y}_i)^2 \sum_j (y_j - \bar{y}_j)^2}}, \tag{19}$$

where $\bar{y}_i$ and $\bar{y}_j$ are the sample mean respectively at stations i and j. We select a threshold r_{thr} for intermediate coefficients of correlation during the forecast. Those values with a lower correlation than the threshold are to be removed, otherwise, they are accounted for the set $Y_{j-remain}$ for valid values $y_{j-remain}$ remaining. If all coefficients are below the threshold, the mean of the target feature ($\bar{y}_i$) calculated from the available observations will be filled in for the missing values.

The correlation-based SAMICE regression model is now formulated to impute missing values at station i as follows,

$$y_{-i} = \begin{cases} y_{j-remain} \cdot \alpha_{-i} + \beta_{-i} & (\text{if } r_{ij} \geq r_{thr}) \\ \bar{y}_i & (\text{if } r_{ij} < r_{thr}, \forall j), \end{cases} \tag{20}$$

and its pseudo-code is presented in Algorithm 2.

Algorithm 2 SAMICE

Input:

1: Sequences of incoming data $y_i \in Y_i$ from target station i^{th}.
2: Sequences of incoming data $y_j \in Y_j$ from neighbor station j^{th}.

Output: Imputed data for $Y_{-i} \subset Y_i$ in target station i^{th}

3: Set the spatial correlation threshold r_{thr} ($0 < r_{thr} < 1$).
4: **for** j in $(n-1)$ stations **do**
5: Compute Pearson's correlation coefficients r_{ij} as per (19).
6: **if** $r_{ij} \geq r_{thr}$ **then**
7: $Y_{j-remain} \leftarrow Y_j$
8: **end if**
9: **end for**
10: **if** $Y_{j-remain} = \emptyset$ **then**
11: $y_{-i} \leftarrow \bar{y}_i$
12: **end if**
13: Obtain imputed values to form the set Y_{-i} of imputed values.

V. RESULTS AND DISCUSSION

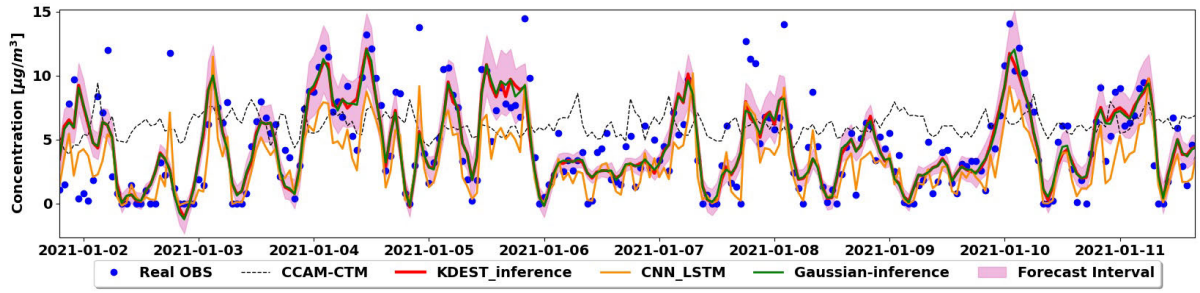
In this work, we considered the air-pollutant forecast in two main periods: summer (January 2021) and winter (late May and early June 2021) to evaluate the performance and reliability of our LSTM-BNN model in comparison with the current CCAM-CTM for respectively two key pollutants, the ozone and $PM_{2.5}$. These evaluation periods are selected from the fact that photochemistry plays a major role of the high level of ozone concentrations over the sunny and hot months during summer while during winter the smoke from fire heaters significantly contributes to $PM_{2.5}$ concentrations in Australia. A comprehensive ablation study was conducted on various choices of the proposed KDEST and SAMICE algorithms to reveal their advantages. The results obtained are also compared with a hybrid CNN-LSTM model to show the LSTM-BNN superior performance.

A. EVALUATION METRICS

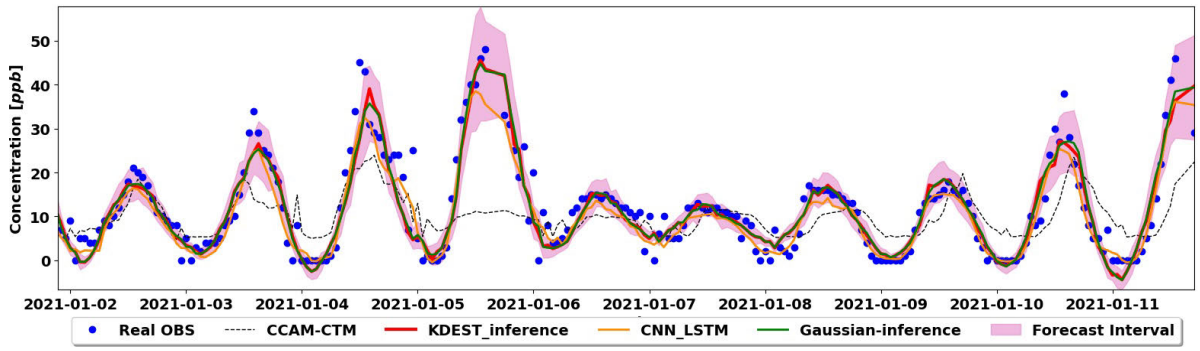
For performance evaluation on the forecast of the time-series data for the concerned air pollutants, collected at a number of monitoring stations in NSW, widely-adopted metrics are used here:

- *The mean absolute error (MAE)*:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{21}$$



(a) $PM_{2.5}$



(b) Ozone

FIGURE 4. Comparison of recursive forecast profiles with LSTM-BNN: KDEST-50 (red), GAUSSIAN-300 (green), CCAM-CTM (dashed black), CNN-LSTM (orange) and ground truth OBS (blue dot) from 02nd January 2021 to 11th January 2021 in Liverpool.

- *The root mean square error (RMSE)*:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{22}$$

- *The Pearson's correlation (r)*:

$$r = \frac{\sum(x_i - \hat{x}_i)(y_i - \hat{y}_i)}{\sqrt{\sum(x_i - \hat{x}_i)^2 \sum(y_i - \hat{y}_i)^2}}, \tag{23}$$

and - *The coefficient of determination (R^2)*:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}, \tag{24}$$

where y_i and $\hat{y}_i$ here are respectively the measured observations and forecast values of variable y at the i^{th} instant, (similarly to variable x), and n is the number of inspected samples. The lower values of *RMSE* and *MAE* or higher values of r and R^2 indicate better performances.

B. LSTM-BNN PERFORMANCE

In the following, we illustrate the high performance of the proposed LSTM-BNN model in forecasting the two air pollutants of interest ($PM_{2.5}$ and ozone) in NSW via comparison and ablation analyses using the above metrics.

1) RECURSIVE FORECAST

Here, the LSTM-BNN data are sampled with 50 and 300 values per distribution respectively for KDEST inference (KDEST-50) and Gaussian-based inference (GAUSSIAN-300) to benchmark with the measurements of observations (OBS) as the ground truth, and the predicted values of CCAM-CTM as a physical model currently used for air quality estimation in NSW state. We also compared the results with those obtained from a hybrid deep learning model, the CNN-LSTM constructed by 1D-CNN layers concatenated with LSTM layers [16]. From the predicted profiles for $PM_{2.5}$ and ozone shown respectively in Figs. 4(a) and (b), it can be seen that profiles from the proposed LSTM-BNN model share similar patterns with the ground truth OBS for both airborne pollutants over the studied period from 2nd January 2021 to 11th January 2021.

While ozone prediction with CCAM-CTM often has large errors, the LSTM-BNN approach can provide its forecast rather accurately, even at higher concentrations of the pollutant, contributed by its diurnal characteristics. Forecast

TABLE 2. Ozone prediction with different recursive models in May 2021 in Liverpool.

Models	MAE (ppb)	RMSE (ppb)	Pearson r	R-squared R^2	Time (second)
CCAM-CTM	8.120	13.450	0.469	0.182	N/A
LSTM	7.882	9.932	0.809	0.633	21.5
CNN-LSTM	7.751	9.686	0.801	0.631	32.6
KDEST-05	7.393	9.450	0.811	0.649	73.5
KDEST-10	7.654	9.436	0.831	0.687	100
KDEST-20	7.269	9.127	0.834	0.692	172
KDEST-30	7.248	9.148	0.856	0.739	250
KDEST-50	7.147	8.910	0.875	0.768	398
GAUSSIAN-300	7.436	9.430	0.811	0.651	547

TABLE 3. Direct forecast performance for $PM_{2.5}$ [$\mu g/m^3$] with different combinations of input lengths and output horizons in June 2021 in Liverpool.

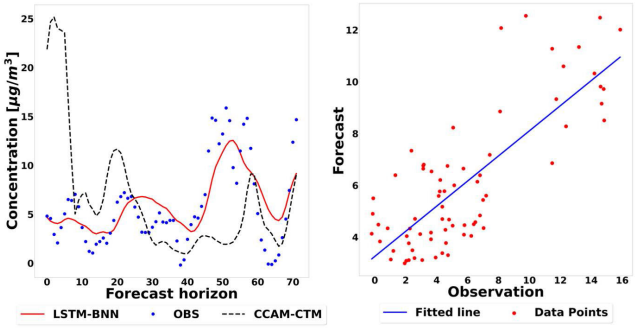
Input Timesteps	Output Timesteps	MAE	RMSE	R^2	Input Timesteps	Output Timesteps	MAE	RMSE	R^2	Input Timesteps	Output Timesteps	MAE	RMSE	R^2
	6	0.339	0.364	0.92		6	0.508	0.708	0.947		6	0.448	0.537	0.967
	12	0.414	0.503	0.855		12	0.465	0.593	0.863		12	0.569	0.698	0.785
	24	0.737	0.952	0.781		24	0.702	0.844	0.71		24	0.864	1.011	0.724
12	36	0.792	1.019	0.755	24	36	1.131	1.285	0.601	36	36	0.964	1.185	0.687
	48	1.398	1.654	0.612		48	1.102	1.383	0.621		48	1.185	1.504	0.543
	60	1.72	2.036	0.578		60	1.445	1.88	0.521		60	1.273	1.651	0.511
	72	2.116	2.415	0.313		72	1.752	2.09	0.412		72	1.452	1.842	0.442
	6	0.406	0.439	0.964		6	0.27	0.302	0.926		6	0.455	0.486	0.901
	12	0.477	0.603	0.871		12	0.459	0.548	0.763		12	0.464	0.587	0.906
	24	0.73	0.855	0.739		24	0.673	0.774	0.711		24	1.194	1.502	0.712
48	36	1.04	1.326	0.671	60	36	1.328	1.621	0.614	72	36	1.193	1.38	0.714
	48	1.124	1.349	0.655		48	1.428	1.766	0.578		48	1.237	1.485	0.698
	60	1.403	1.831	0.612		60	1.862	2.103	0.412		60	1.576	1.875	0.662
	72	1.324	1.705	0.623		72	1.281	1.559	0.589		72	1.746	2.103	0.512

values of the CNN-LSTM model in general present a good fit to OBS as LSTM-BNN but display underpredictions at some peaks of concentrations such as the forecasts in the midday of the 5^{th} of January 2021. For fine particle level, the LSTM-BNN model can accurately forecast with small deviations from the observations except for some minor underestimation at some extreme peaks, as shown in Fig. 4(a). The CNN-LSTM model performs well but only at low concentrations of $PM_{2.5}$, while there are large gaps with respect to the high level of the air pollutant due to uncertainties involved. The band covering ± 5 % of the predicted distributions is shown in the figures (shaded in pink) to represent a level of robustness of the prediction. For the LSTM-BNN model in comparison with other techniques, this coverage in percentage is the highest, as depicted in Fig. 4 (b), with more than 90 % for ozone, the airborne pollutant that varies diurnally in a large range.

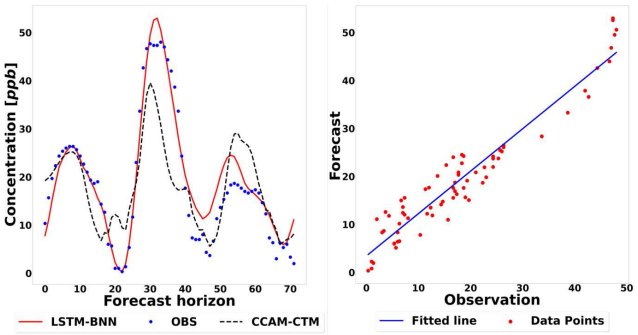
Notably, with the proposed KDEST, the number of samples can be reduced from 300 down to 50 without performance loss. This can result in some improvement in computational efficiency and enable possibilities for prediction with missing data. To further illustrate the advantage of the proposed KDEST algorithm, we conducted an ablation study with smaller numbers of sampled data in addition to KDEST-50, i.e., 5 (KDEST-5), 10 (KDEST-10), 20 (KDEST-20) and 30 (KDEST-30).

An extensive comparison was conducted for predictions with the CCAM-CTM model, two deterministic DL networks including an LSTM model and the hybrid CNN-LSTM model, the proposed LSTM-BNN with KDEST at various numbers of sampled data, and a Gaussian-inference LSTM-BNN (GAUSSIAN-300) model without KDEST. Table 2 summarizes the comparison based on the metrics MAE, $RMSE$, Pearson's coefficient r, R^2 and time of simulation typically for prediction of ozone in May 2021.

It can be seen from the ablation study that the proposed LSTM-BNN with KDEST sampled at 30 data points (KDEST-30) is about the best for forecasting ozone with 10.73% improvement in MAE, 31.9% improvement in RMSE as compared to CCAM-CTM, and 54.3% reduction in the processing time in comparison to the LSTM-BNN with Gaussian inference at 300 samples. Moreover, predictions



(a) $PM_{2.5}$



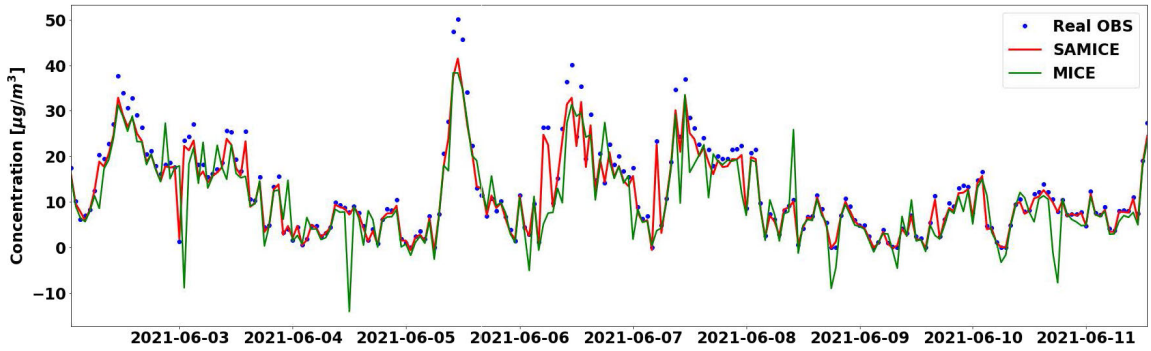
(b) Ozone

FIGURE 5. Forecast performance (left) and scatter plot (right) of direct-forecast models for $PM_{2.5}$ and ozone at 72-h horizon in Liverpool comparing to real observations (OBS) and predicted values of CCAM-CTM model.

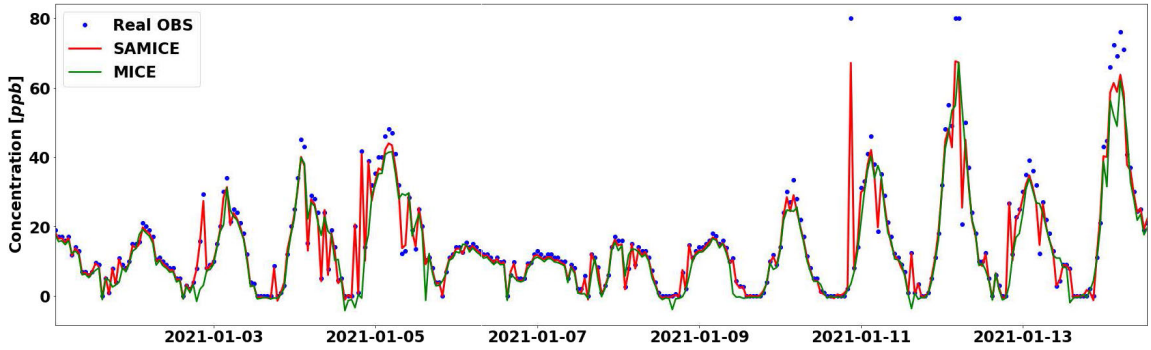
using the proposed model with KDEST-30 have a higher coefficient of determination (R^2) by over 14% compared to those from LSTM and CNN-LSTM. Similar results can be obtained for $PM_{2.5}$ in different seasons of the year. The findings of this ablation study and comparison analysis demonstrate the effectiveness of our Bayesian inference integrated with the LSTM deep learning network in dealing with uncertainties.

2) DIRECT FORECAST

Numerical models like CCAM-CTM often require a large number of values from air emissions inventory and meteorologies as their inputs. Also, unavailable inputs from



(a) $PM_{2.5}$



(b) Ozone

FIGURE 6. Predicted air-pollutant profiles from incoming data with random missing values at a ratio of 0.2, imputed by MICE and SAMICE for the targeted station at Liverpool.

TABLE 4. Performance comparison between MICE and SAMICE of $PM_{2.5}$ and ozone with the threshold of Pearson's coefficient at 0.9 for the targeted station at Liverpool.

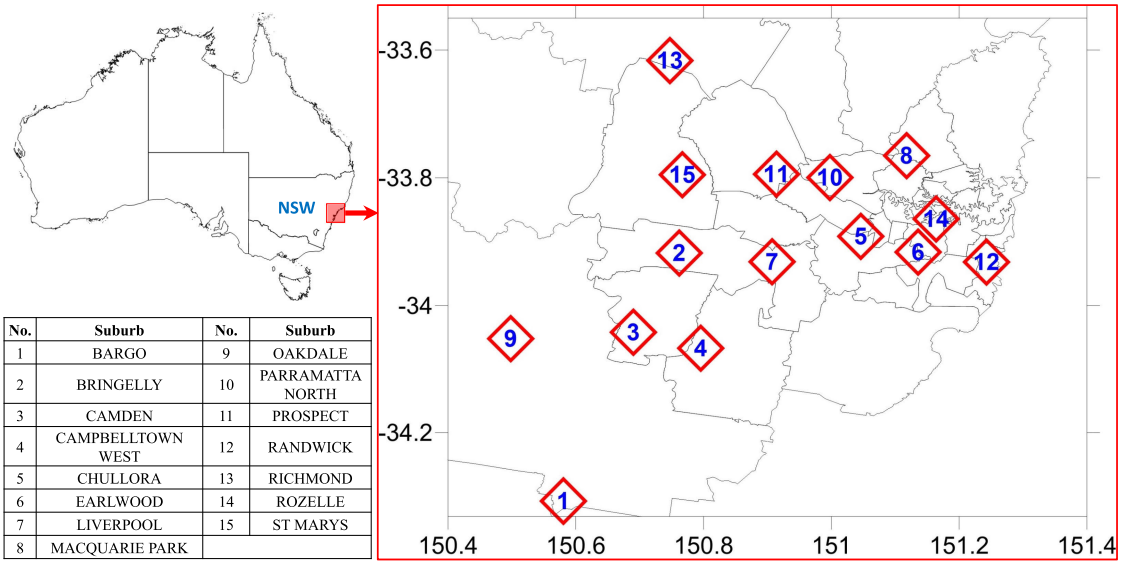
Dropping ratio	$PM_{2.5}$						Ozone					
	MAE			RMSE			MAE			RMSE		
	MICE	SAMICE	Improved Percentage	MICE	SAMICE	Improved Percentage	MICE	SAMICE	Improved Percentage	MICE	SAMICE	Improved Percentage
0.1	1.158	0.912	21 %	1.974	1.587	20 %	0.133	0.024	82 %	0.414	0.031	93 %
0.2	2.630	1.729	34 %	4.118	2.386	42 %	0.152	0.031	80 %	0.392	0.041	90 %
0.3	3.004	1.529	49 %	4.369	2.146	51 %	0.254	0.019	93 %	0.475	0.033	93 %
0.4	4.330	2.867	34 %	11.167	3.528	68 %	0.245	0.095	61 %	0.614	0.108	82 %
0.5	5.202	3.361	35 %	22.455	5.432	76 %	0.247	0.121	51 %	0.716	0.114	84 %

these models may cause an accuracy problem in prediction. As such, we consider here the use of the proposed LSTM-BNN model to forecast multiple values with only historical data of OBS and CCAM-CTM. For this, we conducted an extensive ablation study for direct predictions with short-term forecast horizons of 6, 12, 24, 36, 48, 60 and 72 hours ahead. The profiles (left) and scatter plots (right) for $PM_{2.5}$ and ozone concentrations in June 2021 are depicted respectively in Fig. 5(a) and (b). It can be seen therein that our proposed model (red line) is much better in terms of forecast accuracy in comparison with the ground truth observations (blue dot) for both airborne pollutants. The predicted values of CCAM-CTM (dashed black line) are in poor correlation for $PM_{2.5}$ and display underprediction for ozone as compared OBS. Indeed, the scatter plots for the LSTM-BNN forecast also show outperformance over the existing CCAM-CTM model wherein the coefficient of correlation with real observations is greater than 0.9 for ozone even for a large output horizon of 72 hours (3-day ahead forecast).

A comprehensive experiment was also conducted on different combinations of input lengths and output horizons. Table 3 summarizes typically the performance evaluation for the direct forecast of fine particles in the wintertime. It shows that the forecast accuracy is acceptable with *MAE* ranging between 0.339 and 2.116 $\mu g/m^3$, and *RMSE* between 0.364 and 2.415 $\mu g/m^3$. Moreover, the proposed model with the direct forecast is quite reliable and stable with an input length of 36-48 hours, as can be seen from the table. These results can be also obtained for ozone and in the summertime.

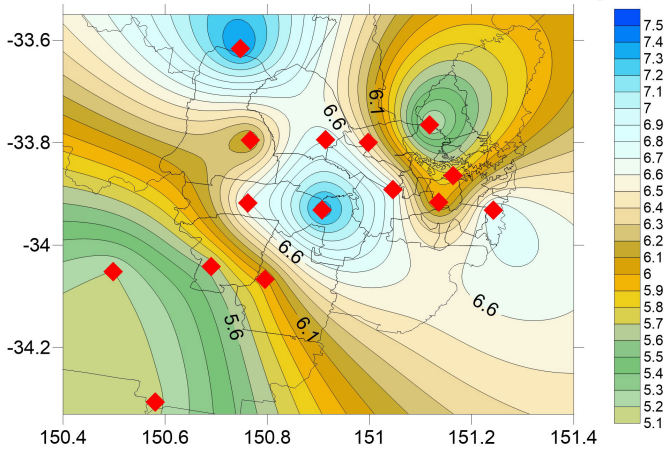
C. SUBURBAN SCALE AIR POLLUTION FORECAST

With the availability of data recorded at air-quality stations only, the missing information or observation at a location required for LSTM-BNN can be imputed by using the

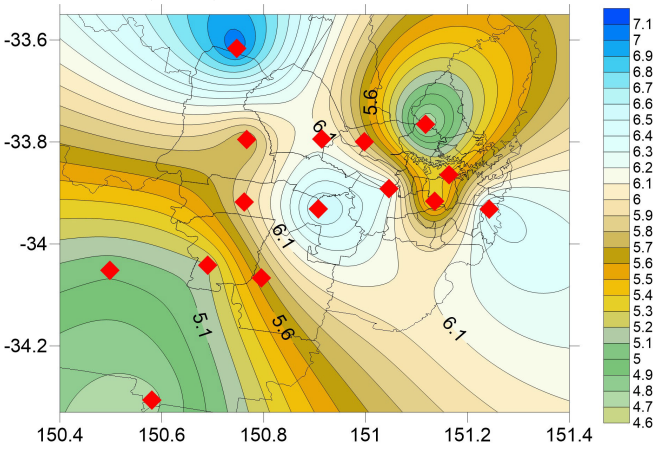


No.	Suburb	No.	Suburb
1	BARGO	9	OAKDALE
2	BRINGELLY	10	PARRAMATTA NORTH
3	CAMDEN	11	PROSPECT
4	CAMPBELLTOWN WEST	12	RANDWICK
5	CHULLORA	13	RICHMOND
6	EARLWOOD	14	ROZELLE
7	LIVERPOOL	15	ST MARYS
8	MACQUARIE PARK		

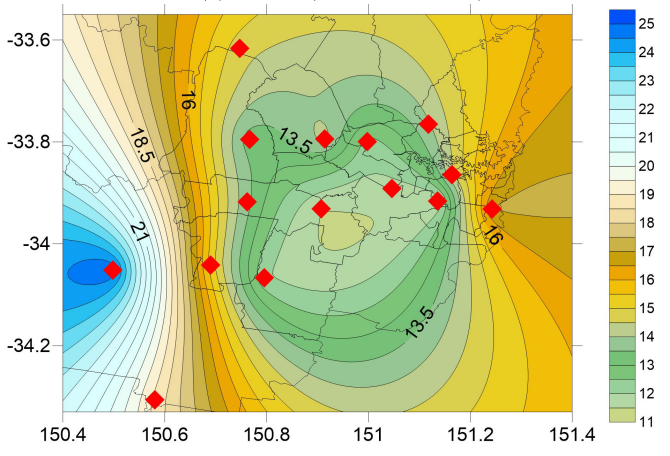
(a) Locations of 15 air-quality stations in GMR, NSW, Australia



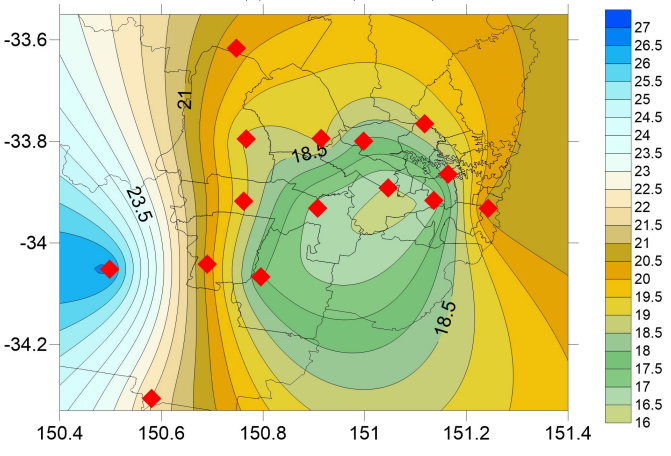
(b) $PM_{2.5}$ (real observations)



(c) $PM_{2.5}$ (forecast)



(d) Ozone (real observations)



(e) Ozone (forecast)

FIGURE 7. Spatial distributions of real observations versus 72h-forecast of $PM_{2.5}$ [$\mu g/m^3$] and ozone [*ppb*] on 04^{th} June 2021 (Note: the red marks are locations of air-quality stations, x-axis and y-axis are latitude and longitude).

proposed SAMICE algorithm, based on correlations with nearby monitoring stations. In this work, 15 stations located in [−150°4, −151°4]-longitude and [−34°33, −33°55]-latitude are selected with the number of missing values less than 30% of the total recorded observations over the three-year period (2018-2021). As in [2], the model outputs are interpolated according to the gridding values over the whole region via kriging to obtain the distribution map along with the stations.

1) SPATIAL DATA IMPUTATION WITH SAMICE

As mentioned previously, our LSTM-BNN framework with SAMICE imputation can provide forecasts of air pollutants in terms of spatial distributions at an interested location. This merit can be verified by comparing profiles of forecast values with imputation by the conventional MICE and proposed SAMICE algorithm.

Benchmarked to the ground-truth observations, the accuracy enhancement of SAMICE over MICE can be seen in Figs. 6(a) and (b) for the predicted profiles of $PM_{2.5}$ and ozone concentrations on the 4^{th} of June 2021. It is clearly seen that the forecast profile with SAMICE (red line) has better fit to the real OBS (blue dots) where the gaps at peaks of pollutants are smaller than forecast profile applying MICE (green line) for imputing the inputs of model. It indicates that the uncertainty of input values is reduced by omitting low-correlated stations with our proposed SAMICE.

Table 4 summarizes the outperformance of SAMICE for prediction of the three air pollutants in consideration with randomly dropping data at ratios varying from 0.1 to 0.5 and the threshold 0.9 for Pearson's coefficient r. As can be seen, the accuracy improvement can be achieved from 20% up to 93% in terms of MAE and RMSE. This improvement is attributed to our model's capacity of reducing the gaps between measurements and predictions, especially at extreme values, by correlation-based adjustment of the estimated disperses of the air pollutants at a location. This spatial feature can be implemented easily for state-run air quality stations or low-cost wireless sensor networks for monitoring systems.

2) SUBURBAN AIR-POLLUTANT DISTRIBUTIONS

Applying the proposed framework for the Sydney GMR gridding, spatial distributions of the air pollutant forecast can be obtained at any suburb or location of interest shown in the map of Fig.7(a). Figures 7(b) - (e) present the comparisons between the distributions of real observations (left) and 72-hour forecast (right) respectively for $PM_{2.5}$ and ozone on the 04^{th} of June 2021.

The spatial distribution maps present quite accurately the forecast dispersion of three air pollutants as per evaluation given in Table 4 for Liverpool station. For example, particles tend to move South East while ozone displays a high concentration on the west during winter 2021. More importantly, this allows for possibly predicting potential risk of air pollution, particularly in any suburb or local area along with the meteorology forecast and, given the availability of low-cost wireless sensor networks, which is promising for microclimate analysis.

VI. CONCLUSION

This paper has presented a long short-term memory Bayesian neural network (LSTM-BNN) as a new deep learning model to improve accuracy and reliability of the air pollution forecast, particularly for two main air pollutants $PM_{2.5}$ and ozone, in the state of New South Wales, Australia. The proposed network utilizes both single-step recursive forecast and multistep ahead direct forecast approaches for fusing observations and data from the currently-used CCAM-CTM. The resulting model provides the predictive distributions as posteriors at each time step instead of point-wise estimations as in deterministic models. Here, the Monte-Carlo dropouts approximate Bayesian inferences to quantify uncertainties in real-world data and designed model. For achieving higher forecast accuracy over the Gaussian-based inference while mitigating the computational latency, we developed the kernel density estimation algorithm with subdivision tuning (KDEST) to substantially reduce the number of distribution samples required. We also developed a new algorithm for spatially-adjusted imputation by chained equations (SAMICE) for considering spatial distributions of air pollutants based on the correlation to monitoring data from air-quality stations or low-cost wireless sensor networks. Extensive experiments with real-world data collected from state-run stations and predicted values by CCAM-CTM have demonstrated the effectiveness of our model in terms of accuracy and reliability of the forecast as compared to observations. Moreover, the proposed model has provided promising results in forecasting air pollution at a local scale for suburbs. This contributes to not only the enhancement of prediction performance but also the possibility of management of urban air quality in moving towards smart livelihoods. Work is in progress for integrating the framework into a dashboard for the state authority. Furthermore, the proposed SAMICE algorithm may incur a bias in data imputation at a target station using the mean of observations at the neighbor stations. Besides, developing a stand-alone model operating on a private-sector application independent of the CCAM-CTM predictions remains a challenge for the recursive forecast method proposed in this paper. These limitations will be rooms for our future developments.

REFERENCES

[1] K. Huang, Q. Xiao, X. Meng, G. Geng, Y. Wang, A. Lyapustin, D. Gu, and Y. Liu, "Predicting monthly high-resolution $PM_{2.5}$ concentrations with random forest model in the North China plain," *Environ. Pollut.*, vol. 242, pp. 675–683, Nov. 2018.

[2] S. Araki, H. Hasunuma, K. Yamamoto, M. Shima, T. Michikawa, H. Nitta, S. F. Nakayama, and S. Yamazaki, "Estimating monthly concentrations of ambient key air pollutants in Japan during 2010–2015 for a national-scale birth cohort," *Environ. Pollut.*, vol. 284, Sep. 2021, Art. no. 117483.

[3] H. A. D. Nguyen and Q. P. Ha, "Wireless sensor network dependable monitoring for urban air quality," *IEEE Access*, vol. 10, pp. 40051–40062, 2022.

[4] L. Chang, H. Duc, Y. Scorgie, T. Trieu, K. Monk, and N. Jiang, "Performance evaluation of CCAM-CTM regional airshed modelling for the New South Wales greater metropolitan region," *Atmosphere*, vol. 9, no. 12, p. 486, Dec. 2018.
[5] J. Michalakes, J. Dudhia, D. Gill, T. Henderson, J. Klemp, W. Skamarock, and W. Wang, "The weather research and forecast model: Software architecture and performance," in *Use of High Performance Computing in Meteorology*. Singapore: World Scientific, 2005, pp. 156–168.
[6] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Wiley, 2015.
[7] P. J. G. Nieto, F. S. Lasheras, E. García-Gonzalo, and F. J. de Cos Juez, "PM_{10} concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, Varma and ARIMA: A case study," *Sci. Total Environ.*, vol. 621, pp. 753–761, Apr. 2018.
[8] A. I. Middya and S. Roy, "Pollutant specific optimal deep learning and statistical model building for air quality forecasting," *Environ. Pollut.*, vol. 301, May 2022, Art. no. 118972.
[9] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 12, pp. 2346–2363, Dec. 2019.
[10] Q. P. Ha, H. Wahid, H. Duc, and M. Azzi, "Enhanced radial basis function neural networks for ozone level estimation," *Neurocomputing*, vol. 155, pp. 62–70, Sep. 2015.
[11] Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao, Y.-S. Wang, and C.-C. Kang, "Multi-output support vector machine for regional multi-step-ahead $PM_{2.5}$ forecasting," *Sci. Total Environ.*, vol. 651, pp. 230–240, Feb. 2019.
[12] A. Tella, A.-L. Balogun, N. Adebisi, and S. Abdullah, "Spatial assessment of PM_{10} hotspots using random forest, K-nearest neighbour and Naïve Bayes," *Atmos. Pollut. Res.*, vol. 12, no. 10, Oct. 2021, Art. no. 101202.
[13] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso, "Deep learning for time series forecasting: A survey," *Big Data*, vol. 9, no. 1, pp. 3–21, Feb. 2021.
[14] M. Aladem and S. A. Rawashdeh, "A single-stream segmentation and depth prediction CNN for autonomous driving," *IEEE Intell. Syst.*, vol. 36, no. 4, pp. 79–85, Jul. 2021.
[15] Q. Zhu, T. H. Dinh, M. D. Phung, and Q. P. Ha, "Hierarchical convolutional neural network with feature preservation and autotuned thresholding for crack detection," *IEEE Access*, vol. 9, pp. 60201–60214, 2021.
[16] 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.
[17] K. Greff, R. K. Srivastava, J. Koutnìk, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
[18] N. Jin, Y. Zeng, K. Yan, and Z. Ji, "Multivariate air quality forecasting with nested long short term memory neural network," *IEEE Trans. Ind. Informat.*, vol. 17, no. 12, pp. 8514–8522, Dec. 2021.
[19] C.-Y. Lo, W.-H. Huang, M.-F. Ho, M.-T. Sun, L.-J. Chen, K. Sakai, and W.-S. Ku, "Recurrent learning on $PM_{2.5}$ prediction based on clustered airbox dataset," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 10, pp. 4994–5008, Oct. 2022.
[20] Y. Han, J. C. K. Lam, V. O. K. Li, and Q. Zhang, "A domain-specific Bayesian deep-learning approach for air pollution forecast," *IEEE Trans. Big Data*, vol. 8, no. 4, pp. 1034–1046, Aug. 2022.
[21] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243–297, Dec. 2021.
[22] Q. Zhu and Q. P. Ha, "A bidirectional self-rectifying network with Bayesian modeling for vision-based crack detection," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 3017–3028, Mar. 2023.
[23] M. Riley, J. Kirkwood, N. Jiang, G. Ross, and Y. Scorgie, "Air quality monitoring in NSW: From long term trend monitoring to integrated urban services," *Air Qual. Climate Change*, vol. 54, no. 1, pp. 44–51, 2020.
[24] Australian Government and Commonwealth Scientific and Industrial Research Organisation (CSIRO). (Jan. 19, 2021). *High-Resolution Regional Climate and Weather Modelling*. Accessed: Sep. 15, 2021. [Online]. Available: https://www.csiro.au/en/research/natural-environment/atmosphere/ccam
[25] H. N. Duc, T. Trieu, Y. Scorgie, M. Cope, and M. Thatcher, "Air quality modelling of the Sydney region using CCAM-CTM," *Air Qual. Climate Change*, vol. 51, no. 1, pp. 29–33, 2017.
[26] S. Metia, H. A. D. Nguyen, and Q. P. Ha, "IoT-enabled wireless sensor networks for air pollution monitoring with extended fractional-order Kalman filtering," *Sensors*, vol. 21, no. 16, p. 5313, Aug. 2021.
[27] DPIE. (2020). *Air Quality Study for the NSW Greater Metropolitan Region*. Accessed: May 24, 2022. [Online]. Available: https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Air/air-quality-study-nsw-greater-metropolitan-region-200488.pdf
[28] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. conf. Mach. Learn.*, 2016, pp. 1050–1059.
[29] 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, 2017.
[30] A. Mobiny, P. Yuan, S. K. Moulik, N. Garg, C. C. Wu, and H. Van Nguyen, "DropConnect is effective in modeling uncertainty of Bayesian deep networks," *Sci. Rep.*, vol. 11, no. 1, p. 5458, Mar. 2021.
[31] P. Goel and L. Chen, "On the robustness of Monte Carlo dropout trained with noisy labels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2021, pp. 2219–2228. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPRW53098.2021.00251
[32] S. M. Ross, *Introductory Statistics*. New York, NY, USA: Academic, 2017.
[33] L. Jiang, S. He, and H. Zhou, "Spatio-temporal characteristics and convergence trends of $PM_{2.5}$ pollution: A case study of cities of air pollution transmission channel in Beijing-Tianjin-Hebei region, China," *J. Cleaner Prod.*, vol. 256, May 2020, Art. no. 120631.
[34] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: Experiences from the scikit-learn project," in *Proc. ECML PKDD Workshop, Lang. Data Mining Mach. Learn.*, 2013, pp. 108–122.
[35] B. Liu, Y. Yang, G. I. Webb, and J. Boughton, "A comparative study of bandwidth choice in kernel density estimation for Näive Bayesian classification," in *Advances in Knowledge Discovery and Data Mining*. Berlin, Germany: Springer, 2009, pp. 302–313.
[36] A. Murad, F. A. Kraemer, K. Bach, and G. Taylor, "Probabilistic deep learning to quantify uncertainty in air quality forecasting," *Sensors*, vol. 21, no. 23, p. 8009, Nov. 2021.
[37] A. Alimissis, K. Philippopoulos, C. G. Tzanis, and D. Deligiorgi, "Spatial estimation of urban air pollution with the use of artificial neural network models," *Atmos. Environ.*, vol. 191, pp. 205–213, Oct. 2018.
[38] S. Van Buuren, J. P. L. Brand, C. G. M. Groothuis-Oudshoorn, and D. B. Rubin, "Fully conditional specification in multivariate imputation," *J. Statist. Comput. Simul.*, vol. 76, no. 12, pp. 1049–1064, 2006.

HUYNH A. D. NGUYEN (Student Member, IEEE) received the B.E. degree in mechatronics engineering from Can Tho University, Vietnam, in 2009, and the M.Sc. degree in mechatronics engineering from the University of Siegen, Germany, in 2015. He is currently pursuing the Ph.D. degree with the Faculty of Engineering and Information Technology, University of Technology Sydney (UTS), Australia. His research interests include the IoT technology, data processing and analysis, machine learning, and deep learning.

QUANG P. HA (Senior Member, IEEE) received the B.E. degree in electrical engineering from the Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 1983, the first Ph.D. degree in complex systems and control from Moscow Power Engineering Institute, Moscow, Russia, in 1993, and the second Ph.D. degree in intelligent systems from the University of Tasmania, Australia, in 1997. He is currently an Associate Professor with the School of Electrical and Data Engineering, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. His research interests include automation, robotics, and control systems. He was on the editorial board of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, from 2009 to 2013, *Mathematical Problems in Engineering*, and *Electronics*. He is currently an Associate Editor of *Automation in Construction*, *Robotica*, and *Frontiers in Robotics and AI*.

HIEP DUC received the B.Eng. degree from The University of Western Australia, in 1978, and the Ph.D. degree in biomedical engineering from The University of Sydney, in 1987. He is currently with the Department of Planning and Environment, NSW, as a Senior Scientist. He has collaborated with scientists and researchers from academic institutions, and research institutes within Australia and overseas. He published many scientific papers in peer-reviewed journals. His current research interests include the studies of health effects due to air pollution, Bayesian statistics, climate change impact, and air quality modeling.

MERCHED AZZI received the B.Sc. degree in physics from the University of Picardie, Amiens, France, in 1982, and the M.Sc. and Ph.D. degrees in chemical engineering from the University of Technology of Compiègne, Compiègne, France, in 1983 and 1986, respectively. He is currently an Atmospheric Scientist with the Department of Planning Industry and Environment, Climate Change and Atmospheric Science, NSW, Australia.

NINGBO JIANG received the bachelor's and master's degrees in atmospheric sciences, the master's degree in statistics, and the Ph.D. degree in environmental science. He is currently a Principal Climate and Atmospheric Scientist with the Department of Planning and Environment, NSW. He has experience working in several organizations across different countries, including Sun Yat-sen University, The University of Auckland, the National Institute of Water and Atmospheric Research Ltd., Macquarie University, the City University of Hong Kong, and some New South Wales Government agencies. His current research interests include air quality forecast, model evaluation, machine learning, data summarization and visualization, and processes influencing regional climate, air quality, and occurrence of bushfires.

XAVIER BARTHELEMY received the Ph.D. degree from Institut de Mècanique des Fluides de Toulouse, Toulouse, France, in 2004. He spent seven years as a Researcher with the Water Research Laboratory, University of New South Wales, Manly Vale, NSW, Australia. He is currently a Climate and Atmospheric Scientist with the Department of Planning and Environment, Climate Change and Atmospheric Science, NSW.

MATTHEW RILEY is currently the Science Director with the Department of Planning, Industry and Environment, NSW, Australia. He has a proven track record of delivering policy-ready environmental research and services that deliver significant public goods. He has delivered multimillion-dollar programs in climate change impacts and adaptation, greenhouse gas emissions, air pollution, and energy efficiency programs for the NSW Government that have led to significant benefits for the people and businesses of NSW. He leads a team of climate and atmospheric researchers, technicians, programmers, and analysts that study and observe the climate, urban air quality and meteorology, greenhouse gas emissions, and energy systems of NSW.