Science of the Total Environment 711 (2020) 134792

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Explore a Multivariate Bayesian Uncertainty Processor driven by artificial neural networks for probabilistic $PM_{2.5}$ forecasting

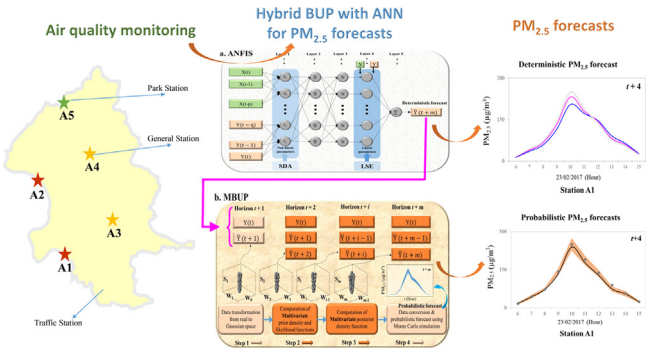
Yanlai Zhou [a,b], Li-Chiu Chang [c], Fi-John Chang [a,*]

[a] Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan
[b] Department of Geosciences, University of Oslo, P.O. Box 1047, Blindern, N-0316 Oslo, Norway
[c] Department of Water Resources and Environmental Engineering, Tamkang University, New Taipei City 25137, Taiwan

HIGHLIGHTS

- ANN models produce multi-step-ahead deterministic $PM_{2.5}$ forecasts accurately.
- MBUP extracts nonlinear multivariate ($\geq$3) dependence between forecasts & observations.
- MBUP reduces the predictive uncertainty of probabilistic $PM_{2.5}$ forecasts.
- MBUP mitigates the underpredictions commonly encountered in ANN models.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 4 July 2019
Received in revised form 10 September 2019
Accepted 1 October 2019
Available online 31 October 2019

Editor: Pavlos Kassomenos

Keywords:
Air quality
Probabilistic forecast
Bayesian Uncertainty Processor
Artificial intelligence
Taipei City

ABSTRACT

Quantifying predictive uncertainty inherent in the nonlinear multivariate dependence structure of multi-step-ahead $PM_{2.5}$ forecasts is challenging. This study integrates a Multivariate Bayesian Uncertainty Processor (MBUP) and an artificial neural network (ANN) to make accurate probabilistic $PM_{2.5}$ forecasts. The contributions of the proposed approach are two-fold. First, the MBUP can capture the nonlinear multivariate dependence structure between observed and forecasted data. Second, the MBUP can alleviate predictive uncertainty encountered in $PM_{2.5}$ forecast models that are configured by ANNs. The reliability of the proposed approach was assessed by a case study on air quality in Taipei City of Taiwan. We consider forecasts of $PM_{2.5}$ concentrations as a function of meteorological and air quality factors based on long-term (2010–2018) hourly observational datasets. Firstly, the Back Propagation Neural Network (BPNN) and the Adaptive Neural Fuzzy Inference System (ANFIS) were investigated to produce deterministic forecasts. Results revealed that the ANFIS model could learn different air pollutant emission mechanisms (i.e. primary, secondary and natural processes) from the clustering-based fuzzy inference system and produce more accurate deterministic forecasts than the BPNN. The ANFIS model then provided inputs (i.e. point estimates) to probabilistic forecast models. Next, two post-processing techniques (MBUP and the Univariate Bayesian Uncertainty Processor (UBUP)) were separately employed to produce probabilistic forecasts. The Bayesian Uncertainty Processors (BUPs) can model the dependence structure (i.e. posterior density function) between observed and forecasted data using a prior density function and a likelihood density function. Here in BUPs, the Monte Carlo simulation was introduced to create a probabilistic predictive interval of $PM_{2.5}$ concentrations. The results demonstrated that the MBUP not only outperformed the UBUP but also suitably characterized the complex nonlinear multivariate dependence structure between observations and forecasts. Consequently, the proposed approach could reduce

* Corresponding author.
E-mail address: changfj@ntu.edu.tw (F.-J. Chang).

https://doi.org/10.1016/j.scitotenv.2019.134792

predictive uncertainty while significantly improving model reliability and $PM_{2.5}$ forecast accuracy for future horizons.

1. Introduction

In the recent era, the impacts of air pollution on human health has become a major global concern (Huang et al., 2014; Van Fan et al., 2018). Natural sources (e.g. sea salt from oceans, volcanic eruptions, and windblown dust) and anthropogenic sources (e.g. motor vehicle exhaust, heat and power generation, industrial processes, and open burning activities) transformation of precursor emissions in the atmosphere such as SO_2 to Sulphates and NO_X to Nitrates may also cause $PM_{2.5}$ (Zhang et al., 2018). In recent years, particulate matter air pollution has also raised a serious concern in Taiwan (Li et al., 2017). Plenty of studies were devoted to establishing various models to predict $PM_{2.5}$ concentrations (e.g. Al-Saadi, et al., 2005; Biancofiore, et al., 2017; Feng, et al., 2015; Niu et al., 2016). The quantification and interpretation of model uncertainty becomes a crosscutting issue that is common to air quality communities and many other disciplines. There is a noticeably growing trend to move away from purely deterministic forecasts towards probabilistic forecasts where modelers seek to describe the likelihood of events and their associated probabilities (Dabberdt and Miller, 2000; Krapu and Borsuk, 2019). Real-time $PM_{2.5}$ forecasting increasingly moves towards probabilistic forecasting in place of traditional deterministic forecasting to meet the demands of the rising public awareness of human health (Berardis and Eleonora, 2017).

Accuracy and reliability of air quality forecasts is affected by several sources, including meteorological and initial input uncertainties (i.e. erroneous sensor readings), as well as the inherent uncertainty (i.e. model structure and parameters) of the forecast model. One of the primary techniques explored to reflect different uncertainties in air quality forecasts is probabilistic forecasting (Coccia and Todini, 2011; Dabberdt and Miller, 2000). Probabilistic forecasts can be produced by three commonly used approaches. The first approach integrates a deterministic forecast model with a probabilistic pre-processing technique such as Fuzzy Clustering (FC), Wavelet Transform (WT) and the bias-correction method (e.g. Dunea et al., 2015; Gong and Ordieres, 2016; Lyu et al., 2017). The second approach is a stochastic (or probabilistic) forecast model, for instance, Quantile Regression Neural Network (QRNN), Random Forest (RF) and Gradient Boosting Decision Tree (GBDT) (e.g. Cannon, 2011; Yu et al., 2016; Liu et al., 2019). The third approach integrates a deterministic forecast model with a probabilistic post-processing technique such as Multiple Linear Regression (MLR), Kalman Filtering, Generalized Likelihood Uncertainty Estimation (GLUE), Bayesian Uncertainty Processor (BUP) and Bayesian Model Averaging (BMA) (e.g. Aznarte, 2017; Djalalova et al., 2015; Pucer et al., 2018; Zhai and Chen, 2018). The first approach focuses on quantifying the impacts of input uncertainty on air quality forecasts because pre-processing techniques are widely used either to simulate the stochasticity or to correct the bias/error of input variables. The second approach intends to quantify the impacts of model structure uncertainty on air quality forecasts because a stochastic forecast model is inherently random, with uncertain factors built into the model. The third approach aims at quantifying the impacts of the overall predictive uncertainty associated with model structure and parameters on air quality forecasts because techniques involved can transform deterministic forecasts into a conditional probability distribution based on the statistical theorem. In this study, we pay special attention to exploring an approach that combines a deterministic forecast model and a probabilistic post-processing technique for improving $PM_{2.5}$ forecasts, considering "perfect" data as model inputs. When referring to "perfect" input data in this study, it implies that there is no uncertainty in sensor readings.

Probabilistic forecasts coupled with post-processing techniques are commonly used to supplement the information provided by point-value predictions (Herr and Krzysztofowicz, 2015). The Bayesian Forecasting System (BFS) proposed by Krzysztofowicz (1999) offers an ideal-theoretic post-processing framework of uncertainty quantification for probabilistic air quality forecasting, where the theoretical structure can be integrated with the results obtained from empirically validated models and numerical computation methods (Herr and Krzysztofowicz, 2015; Zhang, 2017). The Bayesian Uncertainty Processor (BUP) is a vital component of the BFS, and it can be adopted to quantify forecast uncertainty and produce probabilistic forecasts under the hypothesis that there is no input uncertainty, considering "perfect" meteorological data (e.g. rainfalls) as model inputs (Krapu and Borsuk, 2019; Pucer et al., 2018; Ryan, 2016). Probabilistic forecasting exists at the intersection of Bayesian statistics, machine learning and post-processing techniques, and therefore it is contributive to the air quality modeling community. Based on the BUP approach proposed by Krzysztofowicz (1999), various probabilistic forecast methods were developed and widely applied to making forecasts on meteorological and air quality time series (e.g. Garner and Thompson, 2013; Gong and Ordieres, 2016; Zhai and Chen, 2018). The BUP family contains the univariate BUP (UBUP, Krzysztofowicz, 2002) and the multivariate BUP (MBUP, Krzysztofowicz and Maranzano, 2004). The UBUP approach can only quantify the nonlinear bivariate (=2) dependence between observed and forecasted data at each horizon, whereas the MBUP approach can characterize the nonlinear multivariate ($\geq$3) dependence between observed and forecasted data for two horizons at a time (Krapu and Borsuk, 2019; Zhang, 2017). That is to say, the MBUP approach generalizes the UBUP approach and quantifies the stochastic dependence of the consecutive observed and forecasted data between two horizons, in addition to the predictive uncertainty at each horizon. A review of literature indicates that the MBUP approach has not yet been applied to probabilistic air quality forecasting. Both Bayesian and frequentist approaches are feasible and valid, and their predictive performances need to be assessed. Probabilistic forecasting not only poses challenges but also creates outreach opportunities. Consequently, it is imperative to conduct in-depth research on the exploration of the MBUP for quantifying and reducing the uncertainty encountered in multi-step-ahead $PM_{2.5}$ forecasting by characterizing the nonlinear multivariate dependence structure between observed and forecasted data.

To the best of our knowledge, the process of transformation from a deterministic forecast into a probabilistic forecast involves the quantification of the uncertainty inherent in the deterministic forecast. There are two main types of models, physically-based (chemical transport) and data-driven (or artificial intelligence) ones, used to predict $PM_{2.5}$ concentrations (Zhang, 2017). The main merit of physically-based models is their talent to smartly imitate the $PM_{2.5}$ emission mechanisms, whereas the demerits of these models are that they fail to dynamically or adaptively learn the stochastic and/or fuzzy $PM_{2.5}$ emission process induced by the changing environment when encountering scale-related

(transferability) problems (Coelho et al., 2014; Wu et al., 2018a). The main advantage of data-driven models is their capability to tackle highly stochastic and nonlinear prediction problems by means of dynamically adjusting model structures, algorithms and parameters. Besides data-driven models can achieve great performance within dozens of seconds during model construction even though such case study has a complex context of air quality characteristics and plenty of meteorological datasets (Chen et al., 2018; Voukantsis et al., 2011). Artificial Neural Networks (ANNs) are data-driven methods. ANNs have evolved rapidly over the last few decades and have been widely applied to predicting air quality time series (e.g. Akbari Asanjan et al., 2018; Ausati and Amanollahi, 2016; Gao et al., 2018; Nieto et al., 2018; Prasad et al., 2016; Taghavifar et al., 2016; Yeganeh et al., 2018; Zhu et al., 2018; Zhou et al., 2019a,b). If accurate deterministic $PM_{2.5}$ prediction could be made in advance, negative impacts of data-driven models on multi-step-ahead probabilistic forecasts could be minimized. Due to the lack of prior knowledge of $PM_{2.5}$ emission mechanisms, data-driven models may suffer from instability and are prone to systematically underpredicting air quality concentrations for extreme events, i.e. the events that impose the most adverse effects on human health. To cope with these challenges, probabilistic forecasting provides a practical and reliable approach that serves as a complement to data-driven models (Aznarte, 2017; Djalalova et al., 2015; Lyu et al., 2017; Kaminska, 2018; Mok et al., 2017; Monteiro et al., 2013). Fuzzy logic can easily provide heuristic reasoning but has difficulty in generating exact solutions. In comparison to the other data-driven models mentioned above, the ANFIS model is adopted to make deterministic $PM_{2.5}$ forecasts in this study on the grounds that: (1) the ANFIS merges the neural network and the fuzzy logic technique to supply adequate solutions while delivering qualitative heuristic knowledge about the solutions, and (2) its fuzzy if-then rules provide insights into the non-linear, stochastic and fuzzy relationship between air quality and meteorology (Ausati and Amanollahi, 2016; Chang and Chang, 2006; Prasad et al., 2016). Therefore, in-depth research on data-driven models is needed for improving model reliability and forecast accuracy and on the conversion of the deterministic forecasts into probabilistic forecasts using post-processing techniques.

The novelties of this study are two-fold. Firstly, this study focuses on probabilistic $PM_{2.5}$ forecasting, whereas our previous works (Zhou et al., 2019a,b) concentrate on deterministic $PM_{2.5}$ forecasting. Secondly, this study explores a Multivariate Bayesian Uncertainty Processor (i.e. MBUP) approach while other studies (e.g. Huang et al., 2018; Liu et al., 2008) explore Univariate Bayesian Uncertainty Processor (i.e. UBUP) approaches for probabilistic forecasting. The MBUP is introduced for the first time to quantify the nonlinear multivariate ($\geq$3) dependence between observed and forecasted $PM_{2.5}$ data. We explore a MBUP to effectively quantify and reduce the predictive uncertainty encountered in multi-step-ahead $PM_{2.5}$ forecasting. At first, BPNN and ANFIS models are utilized to produce deterministic $PM_{2.5}$ forecasts and their performances are evaluated to identify the model that offers more accurate and reliable deterministic forecasts for use in probabilistic forecasting. For comparison purpose, both MBUP and UBUP approaches are implemented separately to transform deterministic $PM_{2.5}$ forecasts into probabilistic $PM_{2.5}$ forecasts. The reliability and applicability of probabilistic forecasting approaches is assessed by a case study of regional multi-ahead-step $PM_{2.5}$ forecasts for Taipei City of Taiwan. The remainder of this study is organized as follows. Section 2 presents the framework of the proposed methods, including deterministic forecasts as well as the UBUP and MBUP probabilistic forecasts. Section 3 introduces the study area and materials. Section 4 presents the results and discussion on the methods applied to deterministic and probabilistic $PM_{2.5}$ forecasting. Conclusions are then drawn in Section 5.

2. Methods

The goal of this study is to create probabilistic forecasts and reduce the predictive interval to a small range. We integrate BUP and ANN to improve probabilistic $PM_{2.5}$ forecasts, where ANN models (ANFIS and BPNN) are adopted to produce deterministic $PM_{2.5}$ forecasts while the post-processing techniques (MBUP and UBUP) are adopted to create probabilistic $PM_{2.5}$ forecasts. Fig. 1 illustrates the probabilistic forecasting architecture, where the ANFIS (Fig. 1(a)) is incorporated separately into the UBUP (Fig. 1(b)) and the MBUP (Fig. 1(c)) probabilistic forecast approaches. The BPNN serves as the benchmark of deterministic forecasts while the UBUP serves as the benchmark for probabilistic forecasts. The methods used in this study are briefly introduced as follows.

2.1. Deterministic $PM_{2.5}$ forecast models

The learning algorithm of the ANFIS (Jang, 1993) is a hybrid algorithm composed of the steepest descent algorithm (SDA) and the least squares estimation (LSE). Owing to its capability in extracting highly nonlinear, stochastic and fuzzy features, the ANFIS is suitable for predicting multi-step-ahead time series (e.g. Barzegar, et al., 2018; Chang et al., 2014, 2016; Dehghani, et al., 2019). The ANFIS is configured by five layers consisting of: (1) Input layer, (2) Fuzzy-AND operation, (3) Normalization, (4) Consequent layer, and (5) Output layer (Fig. 1(a)). A detailed description of the ANFIS model can be found in Jang (1993).

To demonstrate the reliability and accuracy of the ANFIS model, the BPNN model is implemented for comparison purpose. The main difference between BPNN and ANFIS models is that the former has a simplex neural network architecture while the latter hybrids neural network and fuzzy logic techniques. The distinguishing characteristic of the ANFIS is its ability to update parameters using a hybrid learning algorithm. In this study, the SDA is employed to tune the nonlinear parameters $\{a_i, b_i, c_i\}$ while the LSE is introduced to identify the linear parameters $\{p_i, q_i, r_i\}$. In addition, the Levenberg-Marquardt back propagation algorithm (Yu and Wilamowski, 2011) is used to train the BPNN model. Then the trained ANN producing the best performance in the validation stage is selected as the final model for evaluating model reliability with test datasets.

2.2. Probabilistic forecasting using the UBUP

The UBUP proposed by Krzysztofowicz (2002) can identify the nonlinear bivariate dependence structure between observed and forecasted data at each lead time (or forecast horizon) m (m starts from 1 up to M) step by step. M is the number of time steps. The core theories of the UBUP are briefly described as follows.

Let predictor H be the observed data whose realization h is forecasted. Let estimator S be the output data generated by the corresponding deterministic forecast model constructed by ANNs whose realization s constitutes a point estimate of H. Let random variable H_0 represent the observed data at the current time ($m = 0$). H_m and S_m ($m = 1, 2, \ldots, M$) are the observed data and the corresponding deterministic forecasted data at lead time m, respectively. The groundwork for the UBUP concentrates on identifying the empirical dependence structure of the joint process $\{(H_m, S_m), m = 1, \ldots, M\}$ between observed and forecasted datasets. Uncertainty of probabilistic forecasts at each lead time m can be quantified by a univariate posterior density function, as shown below.

$$\phi_m(h_m|h_0, s_m) = \frac{f_m(s_m|h_0, h_m)\hat{A} \cdot g_m(h_m|h_0)}{\int_{-\infty}^{+\infty} f_m(s_m|h_0, h_m)\hat{A} \cdot g_m(h_m|h_0)\mathrm{d}h_m} \quad (1)$$

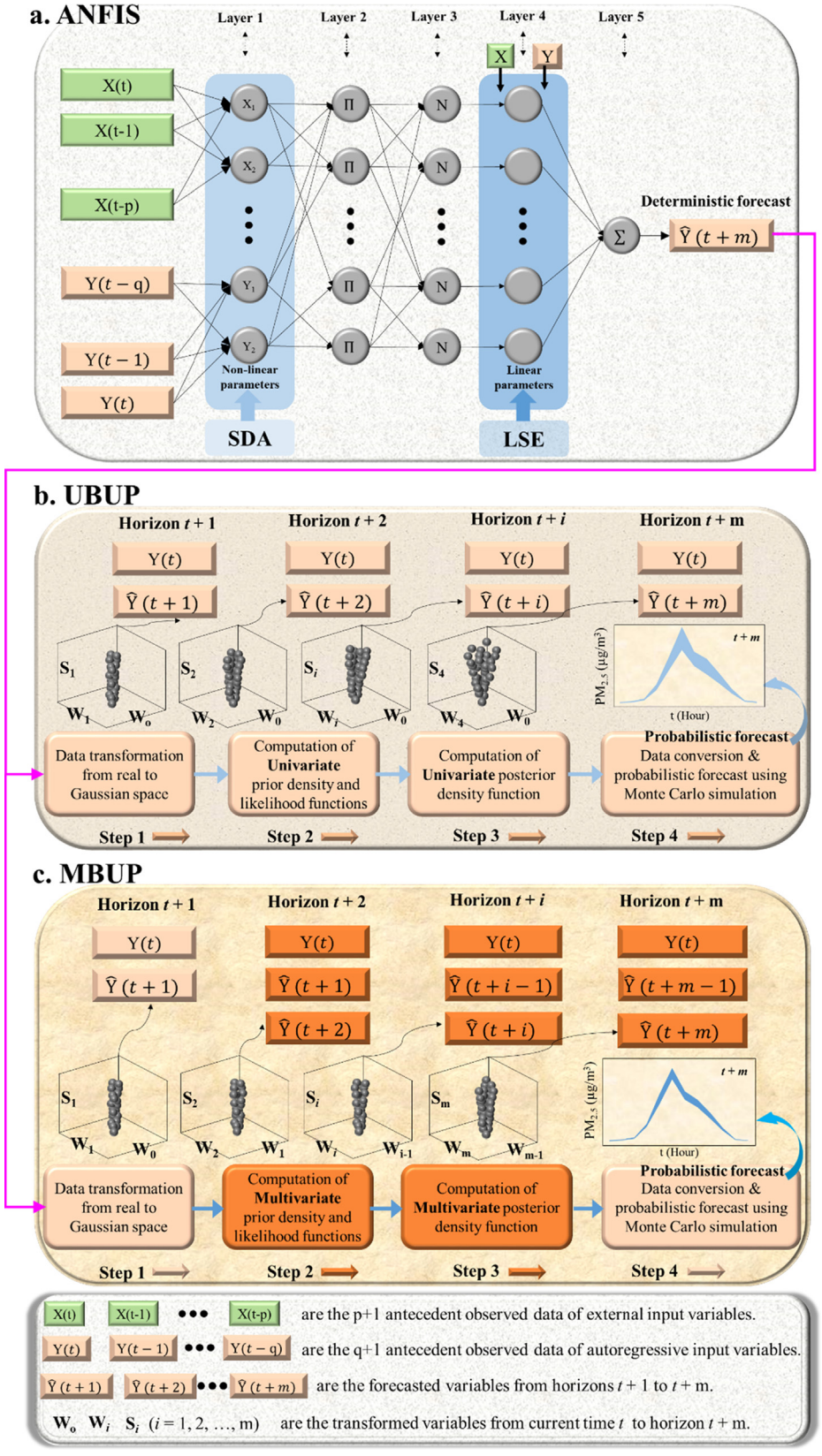


Fig. 1. Architecture of probabilistic forecasting. (a) ANFIS neural network model. (b) Univariate Bayesian Uncertainty Processor (UBUP) approach. (c) Multivariate Bayesian Uncertainty Processor (MBUP) approach. SDA denotes the steepest descent algorithm. LSE denotes the least squares estimation.

where $f_m(s_m|h_0, h_m)$ is the density function of the forecasted data S_m at lead time m. h_0 and h_m are the realizations of the observed datasets at the current time and lead time m, respectively. The conditional hypotheses for $f_m(s_m|h_0, h_m)$ are that the observed data $H_m = h_m$ at lead time m and the observed data $H_0 = h_0$ at current time ($m = 0$). $g_m(h_m|h_0)$ is the transition prior density of the data

H_m at lead time m, where the conditional hypothesis is that the observed data $H_0 = h_0$ at current time ($m = 0$). $\phi_m(h_m|h_0, s_m)$ is the posterior density function used to quantify the uncertainty of H_m, which remains after the deterministic forecast (i.e. ANNs) model is constructed and produces the forecasted $S_m = s_m$. s_m is the realization of the deterministic forecasted data at lead time m. Besides, $\phi_m(h_m|h_0, s_m)$ is the posterior density function for the joint process $\{(H_m, S_m), m = 1, 2, \ldots, M\}$ between observed and forecasted datasets based on the hypothesis that the observed data is $H_0 = h_0$ at current time ($m = 0$). It is understood that when $m = 1$, the conditional densities shown in Eq. (8) can be formulated as $\phi_1(h_1|h_0, s_1)$, $f_1(s_1|h_0, h_1)$ and $g_1(h_1|h_0)$

2.3. Probabilistic forecasting using the MBUP

From what has been discussed above, the UBUP can characterize the nonlinear bivariate dependence structure between observed and forecasted data. But it fails to capture the nonlinear multivariate ($\geq$3) dependence between observed and forecasted data. In consequence, Krzysztofowicz and Maranzano (2004) proposed the MBUP for extracting the nonlinear multivariate dependence features among multi-step-ahead forecasts. The MBUP can provide a family of posterior density functions of the actual datasets $\{H_m, m = 1, 2, \ldots, M\}$, conditional on a realization of the model data $\{S_m, m = 1, 2, \ldots, M\}$ output from a deterministic forecast model, and an observation of the current data H_0. The groundwork for the MBUP concentrates on identifying the empirical dependence structure of the joint process $\{(H_m, H_{m-1}, S_m), m = 1, \ldots, M\}$ between observed and forecasted data. The uncertainty of probabilistic forecasts from the 1st step to the Mth step can be quantified by multivariate posterior density functions, as shown below.

$$\phi_M(h_M|h_0, s_M) = \prod_{m=1}^{M} \phi_m(h_m|h_0, h_{m-1}, s_m) \tag{2a}$$

$$\phi_m(h_m|h_0, h_{m-1}, s_m) = \frac{f_m(s_m|h_0, h_{m-1}, h_m) \cdot g_m(h_m|h_0, h_{m-1})}{\int_{-\infty}^{+\infty} f_m(s_m|h_0, h_{m-1}, h_m) \cdot g_m(h_m|h_0, h_{m-1}) \mathrm{d}h_m} \tag{2b}$$

where $f_m(s_m|h_0, h_{m-1}, h_m)$ is the density function of the forecasted data S_m, conditional on the hypotheses that the observed data $H_m = h_m$ and $H_{m-1} = h_{m-1}$, and the observed datum $H_0 = h_0$ at the current time $m = 0$. $g_m(h_m|h_{m-1}, h_0)$ is the transition prior density from the datum $H_{m-1} = h_{m-1}$ to the datum H_m, conditional on the hypothesis that the observed datum is $H_0 = h_0$ at the current time $m = 0$. $\phi_m(h_m|h_0, h_{m-1}, s_m)$ is the multivariate posterior density function used to quantify the uncertainty about H_m and H_{m-1}, which remains after the deterministic forecast model (e.g. ANFIS) produces the forecasted $S_m = s_m$. $\phi_M(h_M|h_0, s_M)$ is a family of multivariate posterior density functions for the joint process $\{(H_m, S_m), m = 1, \ldots, M\}$ between observed and forecasted data based on the hypothesis that the observed datum $H_0 = h_0$ at the current time $m = 0$.

It is noted that the differences between the UBUP and the MBUP involve two main aspects. Firstly, from the standpoint of model structure the former is a univariate posterior density function whereas the latter is a family of multivariate posterior density functions. Secondly, from the standpoint of model function the former is employed to extract the nonlinear bivariate dependence among data pair $\{(H_m, S_m), m = 1, 2, \ldots, M\}$ whereas the latter is employed to extract the nonlinear multivariate ($\geq$3) dependence among data pair $\{(H_m, H_{m-1}, S_m), m = 1, 2, \ldots, M\}$. The parameters of the BUPs consist of the parameters of the marginal prior distributions, the parameters of the marginal initial distributions, and the dependence parameters of the posterior distributions.

According to the structures shown in Eqs. (1) and (2), the posterior density function in the BUPs (UBUP & MBUP) is closely associated with the prior density function and the likelihood function. The Normal Quantile Transform (NQT) strategy (Krzysztofowicz, 2002) is one of the most common techniques used to compute the posterior density function and the likelihood function. The implementation procedures for the UBUP and the MBUP (Fig. 1 (b) and (c)) contain the following four steps (Krzysztofowicz and Maranzano, 2004).

Step 1 Perform data transformation from the real space to the Gaussian space for both actual and forecasted datasets $\{(H_m, S_m), m = 1, \ldots, M\}$.
Step 2 Compute the univariate (multivariate) prior density function and the likelihood function of the UBUP (MBUP).
Step 3 Compute the univariate (multivariate) posterior density function of the UBUP (MBUP).
Step 4 Implement data transformation from the Gaussian space to the real space and carry out the Monte Carlo simulation for probabilistic forecasting.

3. Study area and materials

The study area is illustrated in Fig. 2. Taiwan has undergone fast-growing economy and population for decades. Air quality deterioration has become a hot topic in Taiwan in recent years. Taipei City is the center of politics, commerce and culture in Taiwan. This city covers an area of 272 km^2 and has a population of 2.68 million in 2018. People across Taipei City nowadays are forced to cope with a high-level invasion of $PM_{2.5}$. Air pollution is not just about sore throats and respiratory diseases but a matter of life or death. Therefore, healthy and green urban development demands for accurate multi-step-ahead $PM_{2.5}$ forecasts that adequately deal with the high variability of regional air quality.

Fig. 2 illustrates the locations of Taipei City and five air quality monitoring stations (also responsible for monitoring meteorological factors) in the study area. Stations A1 (Yonghe) and A2 (Sanchong) are traffic stations, located in urban areas with heavy traffic. Stations A3 (Songshan) and A4 (Shilin) are general stations, located in urban areas with little traffic. Station A5 (Yangming) is a park station, located in the Yangmingshang Park whose elevation ranging between 200 m and 1120 m. Traffic stations are obligated to monitor the emissions from primary air pollutants. General stations are obligated to monitor the emissions from secondary air pollutants. A park station represents the natural situation that has less human intervention, and air quality is monitored here for conservation purpose. More description about the functions of the five stations can be found in the official statement released by the Environmental Protection Administration in Taiwan (TW EPA, https://taqm.epa.gov.tw/taqm/en/b0101.aspx).

Hourly data of eight air quality factors ($PM_{2.5}$, PM_{10}, O_3, NO_x, NO_2, NO, SO_2, CO) and five meteorological factors (rainfall, temperature, wind speed, wind direction, and relative humidity) over a span of 9 years (2010–2018) are available in the study area. Data shuffling operation would make sure that the datasets allocated into training, validation and testing stages are representative of the overall distribution of the data. Moreover, data shuffling can serve the purposes of reducing variance, ensuring model generalizability, and avoiding overfitting. A total of 78,888 (=[(2 × 366) + (7 × 365)] × 24) hourly datasets are used and randomly allocated into three independent datasets, where 35,064 datasets (4 years) are used for model training while the remaining 26,304 datasets (3 years) and 17,520 datasets (2 years) are used for model validating and testing, respectively. To reduce the negative effect of the different scales of input data on model learning, all thirteen input variables are transformed into the same scale. For obtaining a

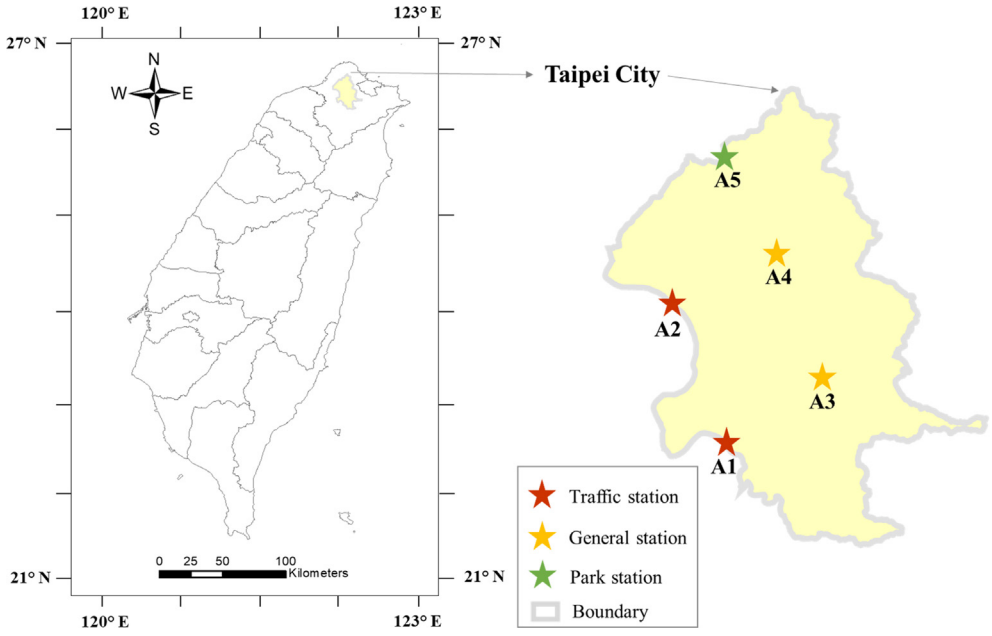


Fig. 2. Distribution of air quality monitoring stations in Taipei City. Stations A1 (Yonghe) and A2 (Sanchong) are traffic stations, Stations A3 (Songshan) and A4 (Shilin) are general stations, and Station A5 is a park station.

stable convergence from the developed model, a transformation that centers the mean to 0 and the standard deviation to 1 is carried out during data pre-processing. The transformation formula is defined as follows.

$$X^*(t) = \frac{X(t) - \overline{X}}{\sigma} \tag{3}$$

where $X^*(t)$ is the transformation of input data in the t^{th} time. $\overline{X}$ and σ are the average and standard deviation of input data, respectively.

The performance of deterministic forecast models is evaluated by three criteria, i.e. the Root-Mean-Square-Error (RMSE), the goodness-of-fit with respect to the benchmark (G_{bench}) and the Critical Success Index (CSI), defined as follows.

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{Y}(t) - Y(t)\right)^2},\ RMSE \geq 0 \tag{4}$$

$$G_{bench} = \left(1 - \frac{\sum_{i=1}^{T}\left(\hat{Y}(t) - Y(t)\right)^2}{\sum_{i=1}^{T}\left(Y(t) - Y_{bench}(t)\right)^2}\right) \times 100\%,\ G_{bench} \leq 100\% \tag{5}$$

$$CSI = \left(\frac{Hits}{Hits + Misses + False\ alarms}\right) \times 100\%,\ CSI \leq 100\% \tag{6}$$

where $\hat{Y}(t)$ and $Y(t)$ is the forecasted and observed values of the output variable at the t^{th} time, respectively. $Y_{bench}(t)$ is the observed data shifted backwards by one or more time lags, e.g. for the nth-step-ahead forecast, and $Y_{bench}(t) = Y(t-n)$. Hits denotes the intersection area between the observed and forecasted values. Misses denotes the range of the observed value that is missed by the range of the forecasted value. False alarms denotes the part of the forecasted value that does not overlap a range of the observed value. It is clear from these definitions that models with higher G_{bench} and CSI values but lower RMSE values would produce better performances.

The performance of probabilistic forecast models is evaluated by two criteria, i.e. the Containing Ratio (CR) and the average Relative Band-width (RB), which assess the goodness of the prediction bounds (Biondi and De Luca, 2013; Xiong and O'Connor, 2008). The CR is defined as the ratio of the number of observed data enveloped by its prediction bounds to the total number of observed data. The RB is defined as the ratio of the band-width of the prediction bounds to the corresponding observed data, which can be used to eliminate the impact of data magnitude on the band-width of the prediction bounds. Their mathematical formulas are described below.

$$N(t) = \begin{cases} 1, & if\ \left(q_l(t) \leq \hat{Z}(t) \leq q_u(t)\right) \\ 0, & else \end{cases} \tag{7a}$$

$$CR = \frac{\sum_{t=1}^{N} N(t)}{N} \times 100\% \tag{7b}$$

$$RB = \frac{1}{N}\sum_{t=1}^{N}\left(\frac{q_u(t) - q_l(t)}{Z(t)}\right) \tag{8}$$

where $q_l(t)$ and $q_u(t)$ are the lower and upper boundaries of the forecasted data corresponding to a given confidence level at time t, respectively. The value of $N(t)$ is either 0 or 1, in which 0 indicates the observed datum falls outside of its prediction bounds while 1 indicates the observed datum falls within its prediction bounds. It is clear from these definitions that models with higher CR values but lower RB values would produce better performances.

4. Results and discussion

The goal of this study is to produce probabilistic forecasts and reduce the predictive distribution to a small range. The results show that point forecasts can be improved by the ANFIS (compared to those of the benchmark BPNN) and probabilistic forecasts can be improved by the MBUP (compared to those of the benchmark UBUP). We would like to remark that it only consumes about 40-second computation time to produce the deterministic (within 30 s) and the probabilistic (within 10 s) $PM_{2.5}$ forecasts for Taipei City. The computation was conducted by a HP computer (Intel®

CoreTM i5). More results and findings are presented: preliminary analysis (Section 4.1), deterministic $PM_{2.5}$ forecasts (Section 4.2), and probabilistic $PM_{2.5}$ forecasts and summarization (Section 4.3).

4.1. Preliminary analysis of $PM_{2.5}$

Table 1 presents the statistic indexes of seasonal $PM_{2.5}$ concentration at Stations A1–A5. We notice that the values of the maximum and average concentrations as well as standard derivation at traffic stations (A1 and A2) are the highest while those in the park station (A5) are the lowest, which could be corresponding to the primary sources of particulate matters at a station. For instance, air pollutant emission from vehicle exhaust is the primary source of particulate matters at traffic stations, air pollutant emission from residential and commercial activities is the primary source of particulate matters at general stations, and atmospheric transport is the primary trigger of particulate matters at the park station. In other words, transportation is a stronger driving force of air pollutants than human activities in Taipei City. In comparison to other input variable selection techniques (e.g. Principal Component Analysis, and Filter Algorithms), the Partial Mutual Information (PMI) method (Sharma, 2000) has the merit that it can estimate both linear and nonlinear dependence between two variables while possessing wider applicability in the environmental domain (Bowden et al., 2005; Fernando et al., 2009; Galelli et al., 2014, code sources: http://ivs4em.deib.polimi.it/). Other than the Pearson and Spearman correlation coefficients, the Kendall tau coefficient (Maidment, 1993) has the merit that input variables do not need to meet the hypothesis of following a Gaussian distribution owing to its gift of non-parametric statistical analysis, and therefore this method is commonly used to extract the nonlinear correlation between two variables. As a result, the PMI and Kendall tau coefficient methods are introduced to implement the input variable selection in this study. According to the highest values of the PMI ($\geq$0.1) (Bowden et al., 2005) and the Kendall tau coefficient ($\geq$0.5) (Zhou et al., 2019a,b), time lags identified by both methods are the consistent. In brief, the time lags of air quality factors are set as 1 h–4 h for traffic stations (A1 and A2) and 1 h–2 h for general and park stations (A3, A4, –A5) while the time lags of meteorological factors are set as 1 h–4 h.

Table 2
Improvement rates of two indicators (G_{bench}, RMSE and CSI) in the testing stages of the multi-step-ahead deterministic $PM_{2.5}$ forecast models (the ANFIS model in comparison with the BPNN model).

Station name	Horizon	Improvement rate* (%)		
		G_{bench}	RMSE	CSI
A1	t + 1	2.15	8.08	2.45
	t + 2	13.01	16.31	11.27
	t + 3	15.20	29.89	13.52
	t + 4	21.27	36.51	19.46
A2	t + 1	3.16	8.08	2.78
	t + 2	11.05	17.08	12.38
	t + 3	11.45	29.24	14.07
	t + 4	14.52	37.09	16.10
A3	t + 1	2.18	5.98	2.20
	t + 2	3.28	8.17	4.84
	t + 3	5.57	11.50	6.63
	t + 4	8.19	16.59	10.37
A4	t + 1	2.87	6.65	2.15
	t + 2	3.88	9.48	4.39
	t + 3	5.57	11.56	7.24
	t + 4	7.29	15.56	9.86
A5	t + 1	1.97	3.83	2.09
	t + 2	2.39	5.68	3.38
	t + 3	3.76	6.58	4.22
	t + 4	4.39	7.72	5.66
Regional	t + 1	2.27	6.33	2.11
	t + 2	6.12	13.95	8.14
	t + 3	6.91	19.15	9.36
	t + 4	9.73	22.30	11.13

*Improvement rate of $G_{bench} = \frac{(G_{bench}\{ANFIS\} - G_{bench}\{BPNN\})}{G_{bench}\{BPNN\}} \times 100\%$.

Improvement rate of RMSE $= \frac{(RMSE\{BPNN\} - RMSE\{ANFIS\})}{RMSE\{BPNN\}} \times 100\%$.

Improvement rate of CSI $= \frac{(CSI\{ANFIS\} - CSI\{BPNN\})}{CSI\{BPNN\}} \times 100\%$.

4.2. Performance of deterministic forecasts

As known, a longer lead time implies lower forecast accuracy. To be consistent with the time periods (AM: 8:00–11:00 and PM: 14:00–17:00) of intensive human activities, lead times up to 4 h (t + 1–t + 4) are selected as the forecasting horizon steps to provide

Table 1
Statistic indexes of seasonal $PM_{2.5}$ concentration at five air quality monitoring stations in Taipei City.

Season	Statistic index ($\mu g/m^3$)	Air Quality Monitoring Stations				
		A1	A2	A3	A4	A5
Spring	max	377	358	259	278	147
	average	22	25	18	15	11
	min	0	0	0	0	0
	std*	16	14	11	10	7
Summer	max	226	215	155	167	88
	average	13	14	11	9	6
	min	0	0	0	0	0
	std	10	9	7	6	4
Autumn	max	264	251	181	195	103
	average	17	18	13	11	7
	min	0	0	0	0	0
	std	13	11	9	8	6
Winter	max	358	340	246	264	140
	average	22	24	18	15	10
	min	0	0	0	0	0
	std	16	14	11	10	7
Annual	max	377	358	259	278	147
	average	20	22	15	14	9
	min	0	0	0	0	0
	std	14	13	9	9	6

*Standard deviation.

high forecast accuracy while reducing human health risks. Therefore, forecast horizons (lead time = 4 h) from $t+1$ up to $t+4$ at a time step of 1 h are specified to validate the two deterministic $PM_{2.5}$ forecast models constructed by the ANFIS and the BPNN. After implementing trial-and-error procedures during model configuration based on the training and validation datasets, the parameters of the ANFIS model constructed for each station (e.g. Station A1) are summarized here. There are 52 ($=N_1$) input variables, 3 ($=N_2$) membership functions (high, medium and low), 468 ($N_3 = 3 \times N_1 \times N_2$) nonlinear parameters in Layer 1, 156 ($N_4 = N_1 \times N_2$) fuzzy rules using the subtractive clustering algorithm (Lohani et al., 2014; Chang et al., 2015), 1 ($=N_5$) output variable under the single output pattern, and 159 ($N_6 = N_2 \times (N_1 + N_5)$) linear parameters in Layer 4. In the SDA, the parameters of learning rate (η), decreasing factor (α), increasing factor (β) and maximal generation (G_{max}) are set as 0.01, 0.9, 1.1 and 1000, respectively. The parameters of the BPNN model consist of the maximal generation (G_{max}), the initial learning rate (*mu*), the increasing factor of *mu*, the decreasing factor of *mu* and the maximal value of *mu* of both models, which are set as 1000, 0.001, 10, 0.1 and 1000, respectively.

Taking horizon t + 4 shown in Table 2 for example, the improvement rates in terms of G_{bench}, RMSE and CSI values reach 21.27%, 36.51% and 19.46% at Station A1, respectively, but only remain

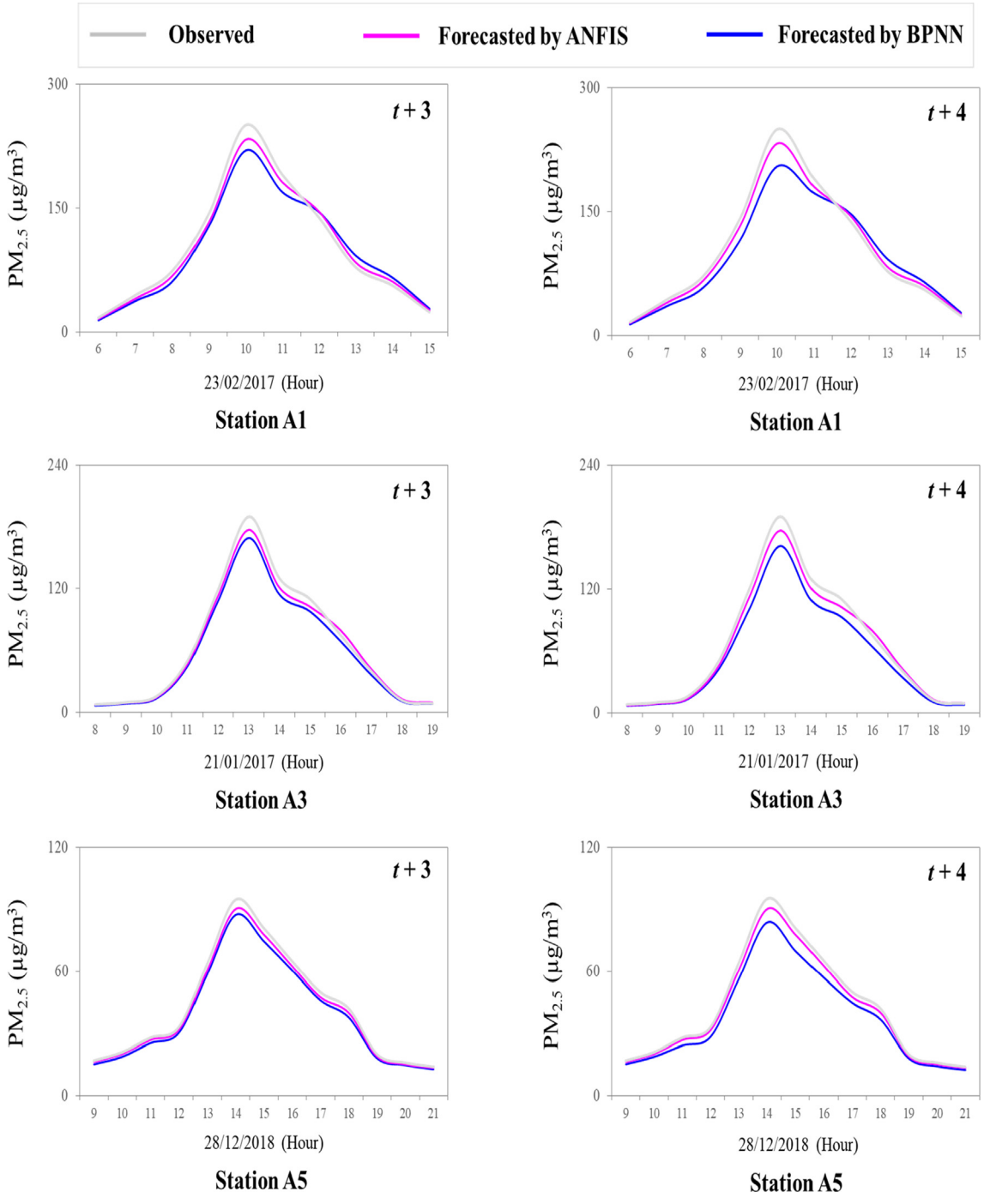


Fig. 3. Multi-step-ahead $PM_{2.5}$ forecast results of BPNN and ANFIS models in the testing stages at horizon t + 4 at the traffic Station A1 (Yonghe), the general Station A3 (Songshan) and the park Station A5 (Yangming). The test event with maximal $PM_{2.5}$ concentration reaching 250 μg/m^3 occurred at Station A1. The test event with maximal $PM_{2.5}$ concentration reaching 180 μg/m^3 occurred at Station A3. The test event with maximal $PM_{2.5}$ concentration reaching 90 μg/m^3 occurred at Station A5.

4.39%, 7.72% and 5.66% at Station A5, respectively. There are three main findings. 1) The BPNN model produces unstable and inferior performance for $PM_{2.5}$ forecasting at each air quality monitoring station and in the whole study region (Taipei City). 2) The ANFIS model performs the best not only at individual air quality monitoring stations but also in the whole region (Taipei City). 3) The ANFIS model has the best performance in the testing stages at all stations. It appears that the ANFIS model produces much higher G_{bench} and CSI values but much smaller RMSE values than the BPNN model in both training and testing stages.

The results demonstrate that forecasts at horizons higher than $t+1$ are more accurate using the subtractive clustering mechanism and the fuzzy inference system of the ANFIS model. In other words, the subtractive clustering-based fuzzy inference system can alleviate time shift phenomena. This reveals that the uncertainty of model input and model reliability as well as forecast accuracy for future horizons can be improved significantly.

To clearly differentiate the abilities of BPNN and ANFIS models, both models are tested by three selected $PM_{2.5}$ events (Park Station A5 with its maximal $PM_{2.5}$ concentration reaching 90 $\mu g/m^3$, General Station A3 with its maximal $PM_{2.5}$ concentration reaching 180 $\mu g/m^3$ and Traffic Station A1 with its maximal $PM_{2.5}$ concentration reaching 250 $\mu g/m^3$). Model performance is assessed by the goodness-of-fit between observations and forecasts at horizons $t+3$ and $t+4$ in the testing stages, as shown in Fig. 3. The results indicate that the ANFIS model is able to forecast well at horizons $t+3$ and $t+4$ whereas the BPNN model has obvious time-lag phenomena and produces comparatively large gaps between observations and forecasts. That is to say, the BPNN model fails to forecast $PM_{2.5}$ concentration adequately at horizons more than $t+3$. It demonstrates that the ANFIS model can effectively trace the trails of $PM_{2.5}$ events, significantly mitigate time-lag effects, and produce accurate and reliable multi-step-ahead $PM_{2.5}$ forecasts. From the perspective of air pollutant emission mechanisms (e.g. Yu and Stuart, 2017; Li et al., 2018; Lin and Zhu, 2018; Wu et al., 2018b), the primary emission associated with meteorological conditions (e.g. Park Station A5) makes insignificant influence on BPNN and ANFIS models but the secondary emission associated with meteorological conditions (e.g. Traffic Station A1) makes a significant difference between BPNN and ANFIS models. Taipei City has undergone fast development, and the regional air quality of the city frequently interacts with intensive human activities, traffic loads and commercial trading. A high $PM_{2.5}$ event usually corresponds to the secondary processes either from regional transportation of the aged secondary aerosol or the secondary transformation of gaseous pollutants. A $PM_{2.5}$ event driven by the primary or natural process would be expected to correlate with local weather conditions and the primary emissions. The BPNN model produces better performance at the general stations (A3 and A4) and the park station (A5) than at traffic stations (A1 and A2). Nevertheless, the ANFIS model gains better improvement rates of G_{bench} and RMSE (see Table 2) at traffic stations (A1 and A2) and general stations (A3 and A4) than at the park station (A5). The ANFIS model not only can greatly improve the forecast accuracy at traffic stations (secondary processes) and general stations (primary processes) by clustering the emission mechanisms of the aged secondary aerosol but also can perform as well as the BPNN model at the park station (natural processes).

Although the forecasts made by the ANFIS model provide substantial evidence of good model performance and gain much more confidence in air quality forecasting, the forecasted values, unfortunately, are prone to systematically under-predicting $PM_{2.5}$ series for extreme $PM_{2.5}$ events (Fig. 3). As known, uncertainties reside in inputs (e.g. meteorological and air quality factors) such that the structure and parameters of the ANFIS model could be the sources of time-lag effects encountered in forecasting. Bearing this in mind as motivation, two processing approaches (UBUP & MBUP) are used to quantify the predictive uncertainty under the hypothesis that there is no input uncertainty. The following subsection will concentrate on the comparison between UBUP and MBUP approaches for probabilistic $PM_{2.5}$ forecasting based on the results obtained from the ANFIS model.

4.3. Performance of probabilistic $PM_{2.5}$ forecasts

For making the median forecasts (e.g. at Traffic Station A1) at horizons from $t+1$ up to $t+4$, the values of the G_{bench}, RMSE and CSI closely associated with UBUP and MBUP approaches in the testing stages are listed in Table 3. It reveals that both approaches produce better forecast results than the deterministic forecast model configured by the ANFIS. Additionally, the MBUP approach is supe-

Table 4
Model performance of probabilistic $PM_{2.5}$ forecasts in the testing stage at the traffic Station A1.

Model	Indicator	Horizon			
		$t+1$	$t+2$	$t+3$	$t+4$
MBUP[1]	CR(%)	97.20	95.33	93.12	92.05
	RB	0.09	0.11	0.15	0.19
UBUP[1]	CR(%)	97.20	92.20	87.57	83.55
	RB	0.09	0.13	0.21	0.26

All of the CR and RB values are computed for the 90% prediction intervals.
[1] MBUP is the Multivariate Bayesian Uncertainty Processor.
[2] UBUP is the Univariate Bayesian Uncertainty Processor.

Table 3
Comparison of the median forecasts on $PM_{2.5}$ concentration in the testing stages at the traffic Station A1.

Horizon		t + 1	t + 2	t + 3	t + 4
MBUP[1] (Probabilistic forecasts)	G_{bench}(%)	99.1	98.2	96.1	94.4
	RMSE($\mu g/m^3$)	3	5	7	10
	CSI (%)	92.7	91.3	89.6	87.4
UBUP[2] (Probabilistic forecasts)	G_{bench}(%)	99.1	97.8	94.7	92.7
	RMSE($\mu g/m^3$)	3	6	9	12
	CSI (%)	92.7	89.2	85.3	82.8
ANFIS[3] (Deterministic forecast)	G_{bench}(%)	99	96.3	93.3	91.6
	RMSE($\mu g/m^3$)	5	7	10	13
	CSI (%)	91.0	88.5	83.3	80.7

[1] MBUP is the Multivariate Bayesian Uncertainty Processor.
[2] UBUP is the Univariate Bayesian Uncertainty Processor.
[3] ANFIS is the Adaptive Neural Fuzzy Inference System.

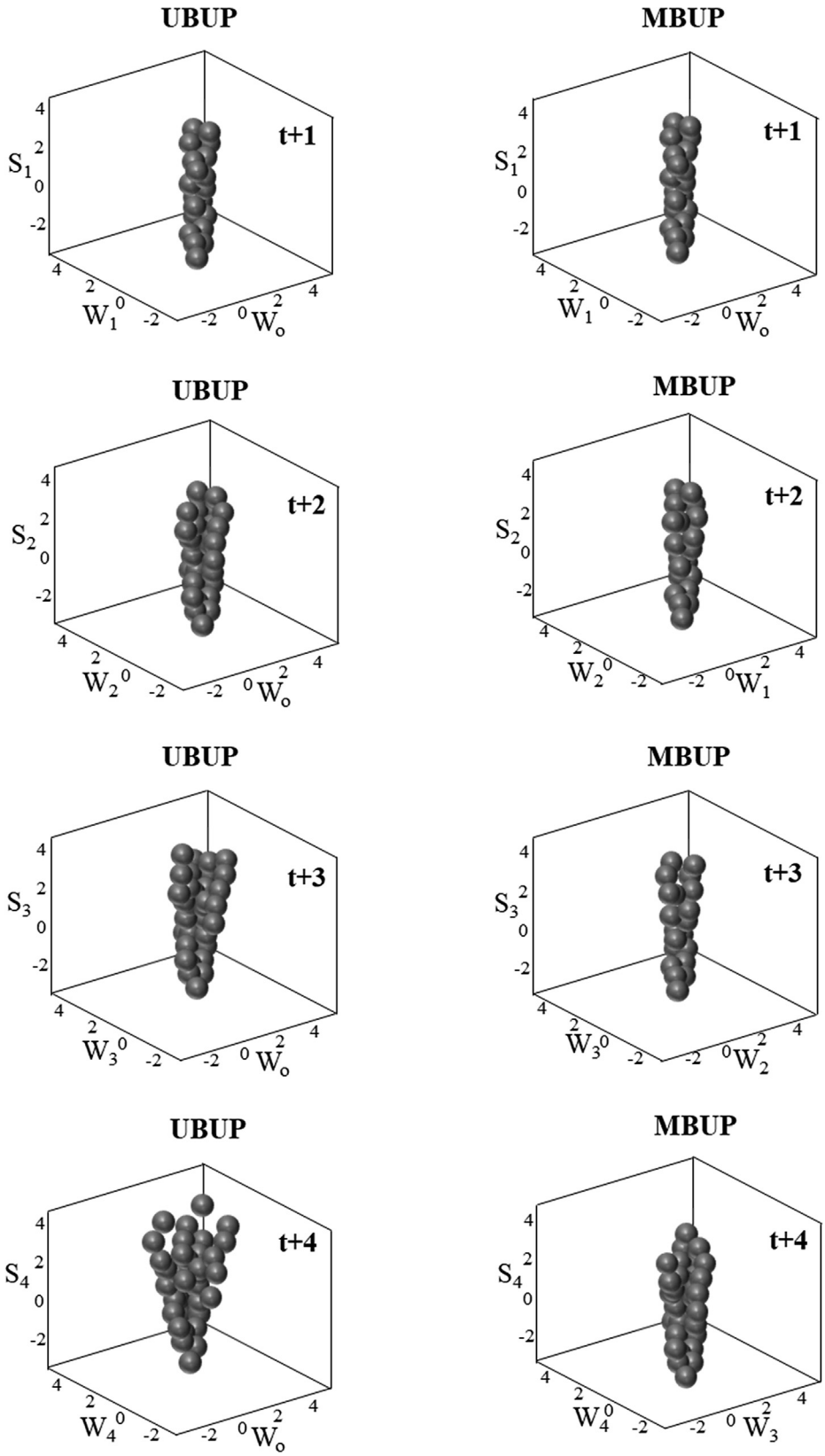


Fig. 4. Comparison of transformed data using the UBUP and MBUP approaches for probabilistic $PM_{2.5}$ forecasts at the traffic Station A1 from horizons $t + 1$ up to $t + 4$ in the testing stages. W_0 and W_m ($m = 1, 2, 3, 4$) are the transformed Gaussian datum of the observed $PM_{2.5}$ concentration at the current time and at lead time m, respectively. S_m ($m = 1, 2, 3, 4$) is the transformed Gaussian datum of the deterministic forecast of $PM_{2.5}$ concentration at lead time m.

rior to the UBUP one in the testing stages. Take horizon $t+4$ for example, the MBUP approach can improve the G_{bench} and the CSI values by 3.09% and 8.30%, respectively, and reduce the RMSE value by 23.08% in the testing stage, in comparison to the deterministic forecast model. The predictability of the MBUP approach for future horizons is significantly better than that of the deterministic forecast model. From the perspective of indicator characteristics, the RMSE and CSI are sensitive to medium-high $PM_{2.5}$ concentrations while the G_{bench} is sensitive to the total $PM_{2.5}$ concentrations. That means the MBUP approach not only can largely increase forecast accuracy at medium-high magnitudes but also can improve the goodness-of-fit to the total $PM_{2.5}$ concentrations at the same time.

For the probabilistic forecasts (e.g. Traffic Station A1) at horizons from $t+1$ up to $t+4$ in the testing stage, the values of CR and RB corresponding to UBUP and MBUP approaches are listed in Table 4. It reveals that the MBUP approach produces better performance at all horizons whereas the UBUP only performs well at horizons up to t + 2 (CR is higher than 90%, and RB is lower than 0.15). For horizon $t+4$, the MBUP approach can improve the CR value by 10.17% as well as reduce the RB value by 26.92% in the testing stage, as compared to the UBUP one. That is to say, the MBUP approach not only can significantly increase probabilistic forecast accuracy indicated by a narrow prediction bound (in terms of CR values) but also can eliminate the impact of $PM_{2.5}$ concentration magnitude on the band-width of the prediction bounds (in terms of RB values) simultaneously.

The reasons that the UBUP approach has inferior forecast accuracy than the MBUP one from horizons t + 2 to t + 4 consist of: the former employs a univariate strategy to model the predictive uncertainty of $PM_{2.5}$ forecasts at each time step independently, whereas the latter employs a multivariate strategy to model the predictive uncertainty of $PM_{2.5}$ forecasts between two time steps. In other words, the former only can consider the nonlinear bivariate dependence between observed and forecasted data while the latter can extract the nonlinear multivariate ($\geq$3) dependence structure between observed and forecasted data owing to the substantial difference between Eqs. (1) and (2) (m = 2, 3, 4). In addition, the forecast accuracy of the UBUP approach is equivalent to that of the MBUP one at horizon t + 1 (Table 4) because there is no difference between Eqs. (1) and (2) at this horizon (m = 1). Fig. 4 further demonstrates that the transformed Gaussian data used in the MBUP approach substantially differ from those of the UBUP one at horizons t + 2 up to t + 4.

Furthermore, QQ plots are adopted to assess the reliability of probabilistic forecasts. Fig. 5 presents the predictive QQ plots of probabilistic $PM_{2.5}$ forecasting (e.g. at Traffic Station A1) from horizons $t+3$ up to $t+4$ in the testing stages. The predictive QQ plots shown in Fig. 5(a) can be elaborated point by point as follows. (1) If all the points fall on the 1:1 line, the predictive distribution agrees perfectly with the observations. (2) If the *probability* value corresponding to the observed datum is 1.0 or 0.0, the corresponding observed datum lies outside the prediction range, which implies predictive uncertainty is significantly underestimated. (3) If the *probability* values corresponding to the observed data cluster around the midrange, predictive uncertainty is overestimated. (4) If the *probability* values corresponding to the observed data cluster around the tails, the predictive uncertainty is underestimated. (5) If the *probability* values corresponding to the observed data at the theoretical median are higher/lower than theoretical quantiles, the predictions systematically underpredict/overpredict the observed data (DeChant and Moradkhani, 2015; Laio and Tamea, 2007). According to Fig. 5(b) and (c), it indicates that the points of the QQ plot generated by the MBUP approach are closer to the

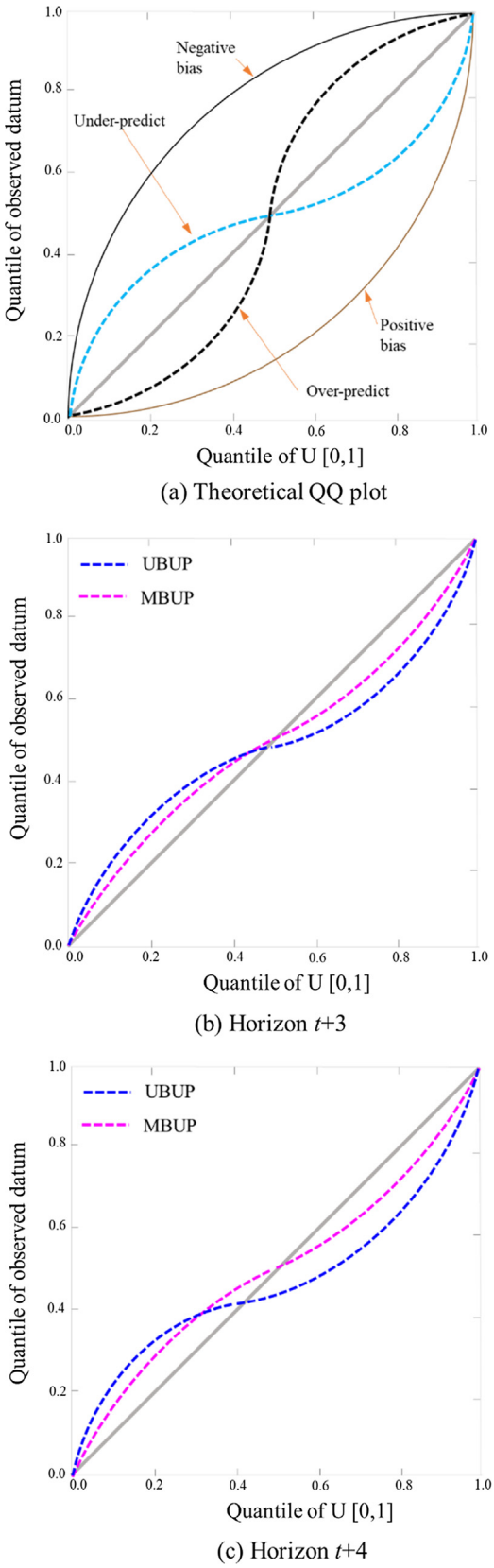


Fig. 5. Predictive Quantile-Quantile (QQ) plots for probabilistic $PM_{2.5}$ forecasts at the traffic Station A1 from horizons $t+3$ up to $t+4$ in the testing stages. The quantile of observed datum is the probability value corresponding to the observed datum while the quantile of U[0, 1] is the probability value corresponding to the forecasted datum.

1:1 line, in comparison to that of the UBUP one. In other word, the MBUP approach produces higher reliability but smaller bias than the UBUP one.

The results demonstrate that the MBUP approach can effectively quantify predictive uncertainty owing to the better agreement between the predictive distribution and the observations. This finding indicates that the MBUP approach performs significantly better from the perspective of reliability.

To clearly differentiate the abilities of the UBUP and the MBUP approaches, three $PM_{2.5}$ events that test the deterministic forecast models are selected to test both approaches through assessing if the observed $PM_{2.5}$ concentrations lie within the 90% prediction interval at horizon $t+4$ in the testing stages, as shown in Fig. 6. The results show that: (1) most of the observed $PM_{2.5}$ concentrations fall within the 90% prediction intervals generated by both approaches, and (2) the MBUP approach provides a predictive distribution narrower than that of the UBUP one. The goal of probabilistic forecasting is to maximize the sharpness of the predictive distributions, where sharpness refers to the concentration of the predictive distributions. Thus, the MBUP approach is superior to the UBUP one.

In brief, the MBUP approach not only can produce more stable and accurate probabilistic forecasts but also can alleviate the phenomenon of systematically underpredicting $PM_{2.5}$ series for extreme $PM_{2.5}$ events by means of extracting the nonlinear multivariate dependence structure among observed and forecasted vari-

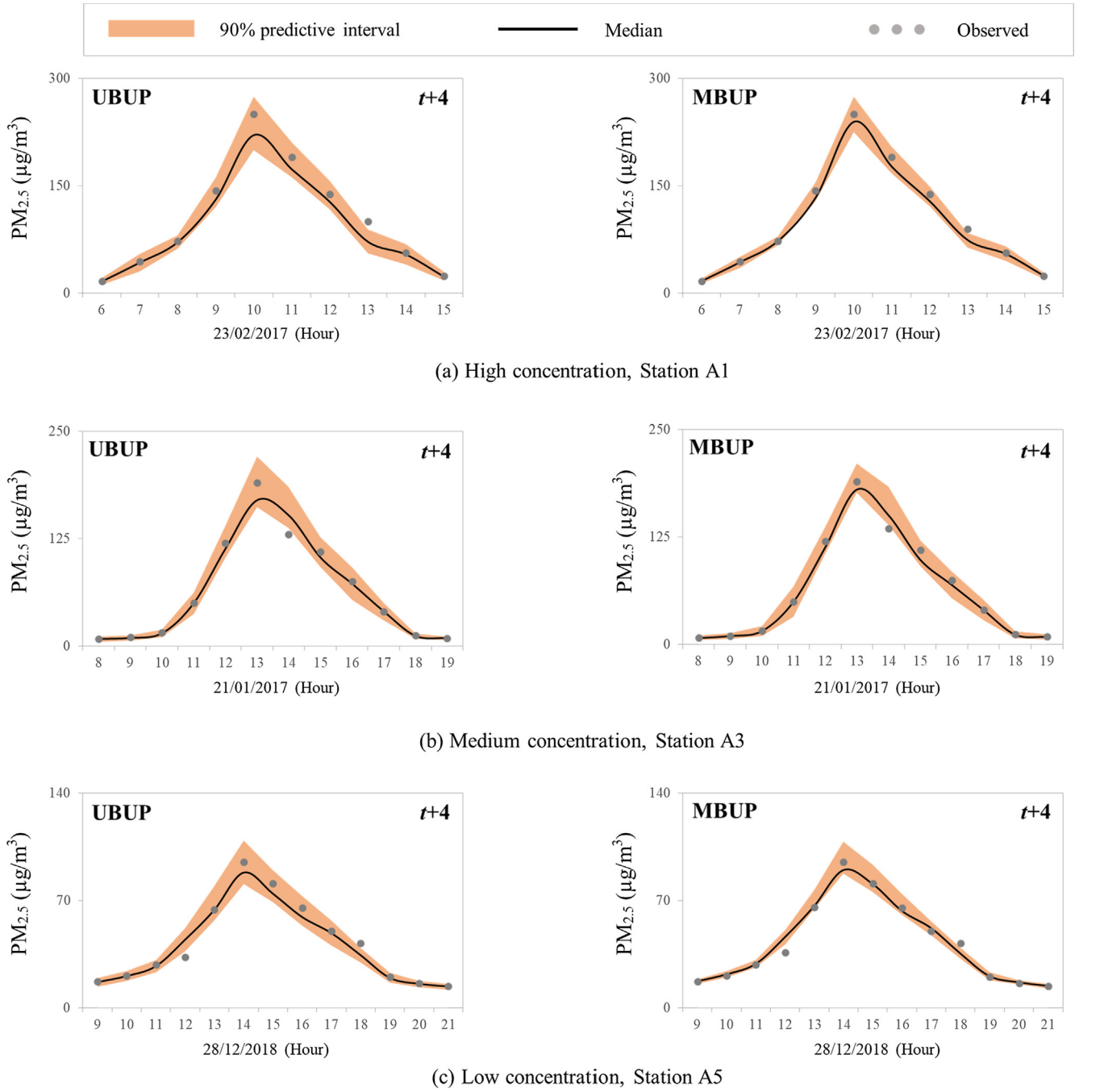


Fig. 6. Probabilistic $PM_{2.5}$ forecasts for air quality monitoring Stations A1, A3 and A5 at horizon $t+4$. Three $PM_{2.5}$ events with maximal $PM_{2.5}$ concentrations exceeding (a) 250 μg/m^3 (high concentration, Station A1), (b) 180 μg/m^3 (medium concentration, Station A3) and (c) 90 μg/m^3 (low concentration, Station A5) were selected for testing the constructed models.

ables H_m, H_{m-1} and S_m. In addition, it is worth noting that even though the MBUP approach can mitigate the phenomenon of underprediction, it cannot fully eliminate the under-estimated predictive uncertainty.

5. Conclusion

In this study, we explored a MBUP approach for modeling probabilistic $PM_{2.5}$ forecasts. The demand for the MBUP approach in place of the UBUP one is driven by real-world applications in the interest of reducing the predictive uncertainty of real-time $PM_{2.5}$ forecasting. At first, two ANNs (BPNN and ANFIS) were configured to establish deterministic forecast models for the regional $PM_{2.5}$ concentrations of Taipei City in Taiwan. The comparison analysis between BPNN and ANFIS models was to identify a model that provides accurate deterministic forecasts for making probabilistic forecasts. Then, two BUP post-processing techniques (MBUP and UBUP) were explored to convert the deterministic forecasts of the ANFIS model into probabilistic forecasts. As compared with the UBUP approach, the main merit of the MBUP approach lies in capturing the nonlinear multivariate ($\geq$3) dependence structure between observed and forecasted data as well as in alleviating the uncertainty encountered in multi-step-ahead $PM_{2.5}$ forecasting.

The results of the two deterministic forecast models applied to the regional $PM_{2.5}$ series of Taipei City demonstrated that the ANFIS model prominently outperformed the comparative BPNN model for all the training, validation and testing cases at different horizons. It meant that the ANFIS model could provide much more accurate forecasts on the $PM_{2.5}$ series at long forecast horizons and significantly alleviate time shift phenomena than the BPNN model. The reason that the ANFIS model succeeded in achieving satisfactory multi-step-ahead forecasts could be owing to the key strategy: the incorporation of the clustering-based fuzzy inference system into ANNs for learning different air pollutant emission mechanisms (i.e. primary, secondary and natural processes). However, the ANFIS model also encountered the technical bottleneck of underpredicting $PM_{2.5}$ peaks.

The MBUP can explicit configure the nonlinear multivariate dependence between observed and forecasted data and quantify the predictive uncertainty of probabilistic forecasts. The results of the two probabilistic post-processing techniques applied to processing the deterministic forecasts of the ANFIS model demonstrated that the MBUP approach was distinguishably superior to the UBUP one for all the training, validation and testing cases at different horizons, in terms of CR and RB values as well as the 90% prediction intervals. The results indicated that the MBUP approach could provide much accurate forecasts on the $PM_{2.5}$ series at long forecast horizons and significantly alleviate underprediction phenomena than the UBUP one. The reason that the MBUP approach succeeded in attaining favorable probabilistic forecast results could be owing to the core strategy: the effective extraction of the nonlinear multivariate dependence structure between observed and forecasted data for lessening predictive uncertainty by virtue of the multivariate Bayesian conditional probability distribution. The limitation of the MBUP approach is the need to re-calibrate the parameters of the MBUP when considering the uncertainty of meteorological forecasting. In light of methodology reliability, future works can concentrate on comparing BUPs with other post-processing probabilistic forecasting techniques (e.g. GLUE and Kalman Filtering) for $PM_{2.5}$ forecasting. In light of methodology transferability, future research could extend the MBUP methodology from small- and medium-scale (a local city) datasets to large-scale ones (examining a large number of time series at a regional or national scale).

Declaration of Competing Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

This work was supported by the Ministry of Science and Technology, Taiwan (MOST 106-3114-M-002-001-A, 108-2119-M-002-017-A, 106-2811-B-002-087), and the Research Council of Norway (FRINATEK Project 274310). The datasets provided by the Environmental Protection Administration in Taiwan are acknowledged. The authors would like to thank the Editors and anonymous Reviewers for their constructive comments that are greatly contributive to improving the manuscript.

References

Akbari Asanjan, A., Yang, T., Hsu, K., Sorooshian, S., Lin, J., Peng, Q., 2018. Short-term precipitation forecast based on the PERSIANN system and LSTM recurrent neural networks. J. Geophys. Res. [Atmos.] 123 (22), 12–543.

Al-Saadi, J., Szykman, J., Pierce, R.B., Kittaka, C., Neil, D., Chu, D.A., MacDonald, C., 2005. Improving national air quality forecasts with satellite aerosol observations. Bull. Am. Meteorol. Soc. 86 (9), 1249–1262.

Ausati, S., Amanollahi, J., 2016. Assessing the accuracy of ANFIS, EEMD-GRNN, PCR, and MLR models in predicting $PM_{2.5}$. Atmos. Environ. 142, 465–474.

Aznarte, J.L., 2017. Probabilistic forecasting for extreme NO_2 pollution episodes. Environ. Pollut. 229, 321–328.

Barzegar, R., Moghaddam, A.A., Adamowski, J., Ozga-Zielinski, B., 2018. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. Stoch. Environ. Res. Risk Assess. 32 (3), 799–813.

Berardis, D., Eleonora, M., 2017. Analysis of major pollutants and physico-chemical characteristics of $PM_{2.5}$ at an urban site in Rome. Sci. Total Environ. 617, 1457–1468.

Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., Di Carlo, P., 2017. Recursive neural network model for analysis and forecast of PM10 and PM2.5. Atmos. Pollut. Res. 8 (4), 652–659.

Bowden, G.J., Maier, H.R., Dandy, G.C., 2005. Input determination for neural network models in water resources applications. Part 1. background and methodology. J. Hydrol. 301, 75–92.

Cannon, A.J., 2011. Quantile regression neural networks: Implementation in R and application to precipitation downscaling. Comput. Geosci. 37 (9), 1277–1284.

Chang, F.J., Chang, Y.T., 2006. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. Adv. Water Resour. 29 (1), 1–10.

Chang, F.J., Chung, C.H., Chen, P.A., Liu, C.W., Coynel, A., Vachaud, G., 2014. Assessment of arsenic concentration in stream water using neuro fuzzy networks with factor analysis. Sci. Total Environ. 494, 202–210.

Chang, F.J., Chiang, Y.M., Ho, Y.H., 2015. Multistep-ahead flood forecasts by neuro-fuzzy networks with effective rainfall–run-off patterns. J. Flood Risk Manage. 8 (3), 224–236.

Chang, F.J., Wang, Y.C., Tsai, W.P., 2016. Modelling intelligent water resources allocation for multi-users. Water Resour. Manage. 30 (4), 1395–1413.

Chen, S., Kan, G., Li, J., Liang, K., Hong, Y., 2018. Investigating China's urban air quality using big data, information theory, and machine learning. Pol. J. Environ. Stud. 27 (2), 1–8.

Coelho, M.C., Fontes, T., Bandeira, J.M., Pereira, S.R., Tchepel, O., Dias, D., et al., 2014. Assessment of potential improvements on regional air quality modelling related with implementation of a detailed methodology for traffic emission estimation. Sci. Total Environ. 470 (2), 127–137.

Coccia, G., Todini, E., 2011. Recent developments in predictive uncertainty assessment based on the model conditional processor approach. Hydrol. Earth Syst. Sci. 15 (10), 3253–3274.

Dabberdt, W.F., Miller, E., 2000. Uncertainty, ensembles and air quality dispersion modeling: applications and challenges. Atmos. Environ. 34 (27), 4667–4673.

DeChant, C.M., Moradkhani, H., 2015. On the assessment of reliability in probabilistic hydrometeorological event forecasting. Water Resour. Res. 51 (6), 3867–3883.

Dehghani, M., Seifi, A., Riahi-Madvar, H., 2019. Novel forecasting models for immediate-short-term to long-term influent flow prediction by combining ANFIS and Grey Wolf optimization. J. Hydrol. (In press)

Djalalova, I., Delle Monache, L., Wilczak, J., 2015. $PM_{2.5}$ analog forecast and Kalman filter post-processing for the community multiscale air quality (CMAQ) model. Atmos. Environ. 108, 76–87.

Dunea, D., Pohoata, A., Iordache, S., 2015. Using wavelet-feedforward neural networks to improve air pollution forecasting in urban environments. Environ. Monit. Assess. 187 (7), 477.

Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J., 2015. Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. Atmos. Environ. 107, 118–128.

Fernando, T.M.K.G., Maier, H.R., Dandy, G.C., 2009. Selection of input variables for data driven models: an average shifted histogram partial mutual information estimator approach. J. Hydrol. 367, 165–176.
Galelli, S., Humphrey, G.B., Maier, H.R., Castelletti, A., Dandy, G.C., Gibbs, M.S., 2014. An evaluation framework for input variable selection algorithms for environmental data-driven models. Environ. Model. Soft. 62, 33–51.
Gao, S., Zhao, P., Pan, B., Li, Y., Zhou, M., Xu, J., Shi, Z., 2018. A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network. Acta Oceanol. Sin. 37 (5), 8–12.
Garner, G.G., Thompson, A.M., 2013. Ensemble statistical post-processing of the national air quality forecast capability: enhancing ozone forecasts in Baltimore, Maryland. Atmos. Environ. 81, 517–522.
Gong, B., Ordieres, M.J., 2016. Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: case study of Hong Kong. Environ. Model. Softw. 84, 290–303.
Herr, H.D., Krzysztofowicz, R., 2015. Ensemble Bayesian forecasting system Part I: Theory and algorithms. J. Hydrol. 524, 789–802.
Huang, G., Lee, D., Scott, E.M., 2018. Multivariate space-time modelling of multiple air pollutants and their health effects accounting for exposure uncertainty. Stat. Med. 37 (7), 1134–1148.
Huang, Y., Shen, H., Chen, H., Wang, R., Zhang, Y., Su, S., et al., 2014. Quantification of global primary emissions of $PM_{2.5}$, PM_{10}, and TSP from combustion and industrial process sources. Environ. Sci. Technol. 48 (23), 13834–13843.
Jang, J.S., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man. Cybern. 23 (3), 665–685.
Kaminska, J., 2018. Probabilistic forecasting of nitrogen dioxide concentrations at an urban road intersection. Sustainability 10 (11), 4213.
Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resour. Res. 35 (9), 2739–2750.
Krzysztofowicz, R., 2002. Bayesian system for probabilistic river stage forecasting. J. Hydrol. 268 (1–4), 16–40.
Krzysztofowicz, R., Maranzano, C.J., 2004. Bayesian system for probabilistic stage transition forecasting. J. Hydrol. 299 (1–2), 15–44.
Krapu, C., Borsuk, M., 2019. Probabilistic programming: a review for environmental modellers. Environ. Model. Softw. 114, 40–48.
Laio, F., Tamea, S., 2007. Verification tools for probabilistic forecasts of continuous hydrological variables. Hydrol. Earth Syst. Sci. 11 (4), 1267–1277.
Li, H., You, S., Zhang, H., Zheng, W., Lee, W.L., Ye, T., Zou, L., 2018. Analyzing the impact of heating emissions on air quality index based on principal component regression. J. Clean. Prod. 171, 1577–1592.
Li, N., Chen, J.P., Tsai, C.I., et al., 2017. Potential impacts of electric vehicles on air quality in Taiwan. Sci. Total Environ. 566, 919–928.
Lin, B., Zhu, J., 2018. Changes in urban air quality during urbanization in China. J. Clean. Prod. 188, 312–321.
Liu, X., Tan, W., Tang, S., 2019. Bagging-GBDT ensemble learning model for city air pollutant concentration prediction. IOP Conference Series: Earth and Environmental Science.
Liu, Y., Guo, H., Mao, G., Yang, P., 2008. A Bayesian hierarchical model for urban air quality prediction under uncertainty. Atmos. Environ. 42 (36), 8464–8469.
Lohani, A.K., Goel, N.K., Bhatia, K.K.S., 2014. Improving real time flood forecasting using fuzzy inference system. J. Hydrol. 509, 25–41.
Lyu, B., Zhang, Y., Hu, Y., 2017. Improving $PM_{2.5}$ air quality model forecasts in China using a bias-correction framework. Atmosphere 8 (8), 147.
Maidment, D.R., 1993. Handbook of Hydrology. McGraw-Hill, New York.
Monteiro, A., Ribeiro, I., Tchepel, O., Sá, E., Ferreira, J., Carvalho, A., Schaap, M., 2013. Bias correction techniques to improve air quality ensemble predictions: focus on O_3 and PM over Portugal. Environ. Model. Assess. 18 (5), 533–546.
Mok, K.M., Miranda, A.I., Yuen, K.V., Hoi, K.I., Monteiro, A., Ribeiro, I., 2017. Selection of bias correction models for improving the daily PM_{10} forecasts of WRF-EURAD in Porto, Portugal. Atmos. Pollut. Res. 8 (4), 628–639.
Nieto, P.J.G., Lasheras, F.S., García-Gonzalo, E., Juez, F.J.D.C., 2018. 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. 621, 753–761.
Niu, M., Wang, Y., Sun, S., Li, Y., 2016. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM2.5 concentration forecasting. Atmos. Environ. 134, 168–180.
Prasad, K., Gorai, A.K., Goyal, P., 2016. Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. Atmos. Environ. 128, 246–262.
Pucer, J.F., Pirš, G., Štrumbelj, E., 2018. A Bayesian approach to forecasting daily air-pollutant levels. Know. Inf. Syst. 57 (3), 635–654.
Ryan, W.F., 2016. The air quality forecast rote: Recent changes and future challenges. J. Air Waste Manage. Assoc. 66 (6), 576–596.
Sharma, A., 2000. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 – a strategy for system predictor identification. J. Hydrol. 239 (1–4), 232–239.
Taghavifar, H., Taghavifar, H., Mardani, A., Mohebbi, A., Khalilarya, S., Jafarmadar, S., 2016. Appraisal of artificial neural networks to the emission analysis and prediction of CO_2, soot, and NO_x of n-heptane fueled engine. J. Clean. Prod. 112, 1729–1739.
Van Fan, Y., Perry, S., Klemeš, J.J., Lee, C.T., 2018. A review on air emissions assessment: transportation. J. Clean. Prod. 194, 673–684.
Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., Kolehmainen, M., 2011. Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki. Sci. Total Environ. 409 (7), 1266–1276.
Wu, L., Li, N., Yang, Y., 2018a. Prediction of air quality indicators for the Beijing-Tianjin-Hebei region. J. Clean. Prod. 196, 682–687.
Wu, J., Zheng, H., Zhe, F., Xie, W., Song, J., 2018b. Study on the relationship between urbanization and fine particulate matter ($PM_{2.5}$) concentration and its implication in China. J. Clean. Prod. 182, 872–882.
Yeganeh, B., Hewson, M.G., Clifford, S., Tavassoli, A., Knibbs, L.D., Morawska, L., 2018. Estimating the spatiotemporal variation of NO_2 concentration using an adaptive neuro-fuzzy inference system. Environ. Model. Softw. 100, 222–235.
Yu, H., Wilamowski, B.M., 2011. Levenberg-marquardt training. Ind. Electron. Handb. 5 (12), 1–16.
Yu, H., Stuart, A.L., 2017. Impacts of compact growth and electric vehicles on future air quality and urban exposures may be mixed. Sci. Total Environ. 576, 148–158.
Yu, R., Yang, Y., Yang, L., Han, G., Move, O., 2016. RAQ-a random forest approach for predicting air quality in urban sensing systems. Sensors 16 (1), 86.
Zhai, B., Chen, J., 2018. Development of a stacked ensemble model for forecasting and analyzing daily average $PM_{2.5}$ concentrations in Beijing, China. Sci. Total Environ. 635, 644–658.
Zhang, Y., 2017. Air quality modelling: Current status, major challenges and future prospects. Air Qual. Clim. Change 51 (3), 41.
Zhang, Y., Lang, J., Cheng, S., Li, S., Zhou, Y., Chen, D., Wang, H., 2018. Chemical composition and sources of PM_1 and $PM_{2.5}$ in Beijing in autumn. Sci. Total Environ. 630, 72–82.
Zhu, S., Lian, X., Wei, L., Che, J., Shen, X., Yang, L., Li, J., 2018. $PM_{2.5}$ forecasting using SVR with PSOGSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors. Atmos. Environ. 183, 20–32.
Zhou, Y., Chang, F.J., Chang, L.C., Kao, I.F., Wang, Y.S., 2019a. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. J. Clean. Prod. 209, 134–145.
Zhou, Y., Chang, F.J., Chang, L.C., Kao, I.F., Wang, Y.S., Kang, C.C., 2019b. Multi-output support vector machine for regional multi-step-ahead $PM_{2.5}$ forecasting. Sci. Total Environ. 651, 230–240.