Wordcount: 10386 words

Exploring Copula-based Bayesian Model Averaging with multiple ANNs for $PM_{2.5}$ ensemble forecasts

Yanlai Zhou[1,2], Fi-John Chang[1], Hua Chen[3], Hong Li[2]

[1] Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei, 10617, Taiwan.

[2] Department of Geosciences, University of Oslo, P.O. Box 1047 Blindern, N-0316 Oslo, Norway.

[3] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, 430072, China.

**Correspondence author*: Fi-John Chang (changfj@ntu.edu.tw); Yanlai Zhou (yanlai.zhou@whu.edu.cn).

Abstract

Quantifying predictive uncertainty of ensemble air quality forecast is very crucial and challenging. This study integrated a Copula-based Bayesian Model Averaging (CBMA) and multiple deterministic artificial neural networks (ANNs) to make accurate ensemble probabilistic $PM_{2.5}$ forecasts. The new approach (CBMA), has a flexible structure that grants the posterior distribution to have any shape owing to the Copula function. The CBMA approach could remove the data transformation and bias correction procedures as it is done in the original BMA, which was taken as the benchmark. The air quality in Taipei City of Taiwan was selected as a study case to evaluate the applicability and reliability of the proposed approach. Three kinds of air quality monitoring stations denoted heavy traffic loads, intensive commercial trading and human intervention, and a natural circumstance with fewer human activities respectively. The forecasts of $PM_{2.5}$ concentrations were regarded as a math function involving meteorological and air quality variables, using long-term (2010-2018) hourly observational datasets. Firstly, four deterministic ANN models were established and evaluated to provide inputs for ensemble forecasting. Then, the two post-processing techniques (i.e. CBMA and BMA) were employed to produce ensemble probabilistic forecasts based on the forecasts obtained from multiple ANN models. The results demonstrated that the CBMA not only could outperform the BMA but also could provide a practical and reliable approach as a complement to multiple deterministic ANN models to create ensemble probabilistic forecasts. From horizons t+1 up to t+4, the CBMA approach could drive up the Containing Ratio (CR) values by 3.12% - 9.58% as well as reduce the average Relative Band-width (RB) values by 8.63% - 34.48% and the Continuous Ranked Probability Score (CRPS)

values by 7.62% - 32.89%, in comparison with the BMA one. Consequently, the predictive uncertainty could be alleviated while model reliability and $PM_{2.5}$ forecast accuracy could be considerably increased.

Keywords: Air quality; Ensemble forecast; Uncertainty; Bayesian Model Averaging (BMA); Copula function

1. Introduction

Suspended atmospheric particulate matter (e.g. $PM_{2.5}$, aerodynamic diameter less than 2.5 μm) is one of main air pollutants (Huang et al., 2014; Zhang et al., 2018). Natural sources and anthropogenic sources transformation of precursor emissions in the atmosphere such as SO_2 to Sulphates and NO_X to Nitrates may also cause $PM_{2.5}$ (Berardis and Eleonora, 2017; Van Fan et al., 2018). The electricity generation process using overmuch fossil fuels would produce plenty of precursor emissions and trigger air pollution. The accurate and reliable air quality predictions can provide technical guidelines for the trade-off between fossil fuels energy and renewable energy outputs toward cleaner production. Air quality predictions and environmental impacts of the electricity generation process not only increase efficiencies in the uses of energy but also are in the interest of cleaner production in power industries (Han et al., 2019). It is essential to make accurate and reliable air quality forecasts in advance to mitigate environmental impacts and health risks. There is a noticeably growing trend to move away from purely deterministic air quality forecasting to probabilistic air quality forecasting (Krapu and Borsuk, 2019; Zhang, 2017; Zhai and Chen, 2018).

Some promising techniques (Table 1) have been used to quantify uncertainties in air quality forecasts, for instance, (1) pre-processing techniques: Fuzzy Clustering (FC) method, Wavelet Transform (WT) and bias-correction method (Dunea et al., 2015; Feng et al., 2015; Gong and Ordieres, 2016; Lohani et al., 2014; Lyu et al., 2017; Monteiro et al., 2013) and (2) post-processing techniques: Multiple Linear Regression (MLR), Kalman filtering, Generalized Likelihood Uncertainty Estimation (GLUE), Bayesian Uncertainty Processor (BUP) and Bayesian Model Averaging (BMA) (Aznarte, 2017; Djalalova et al., 2015; Garner and Thompson, 2013; Kaminska, 2018; Pucer et al., 2018; Zhai and Chen, 2018). Ensemble forecasting techniques are commonly used to characterize diverse uncertainties in air quality forecasts (Bai et al., 2018; Thielen-del and Bruen, 2019). According to the comparative analysis for various probabilistic forecasting techniques (Table 1), the BMA, as one of the smart post-processing methods, employed for weather firstly forecasting, is being broadened to air quality modeling applications, which exhibits ensemble forecasts' advantage (Herr and Krzysztofowicz, 2015; Pucer et al., 2018; Krapu and Borsuk, 2019). Ensemble forecasts with post-processing techniques are commonly used to supplement the information provided by point-value deterministic predictions. Modular design, ensemble modeling and hybridization with deterministic models are yielding new tools for probabilistic air quality forecasting (Liu et al., 2019).

Table 1 Comparison analysis of probabilistic forecasting methods

Methods	Categories	Pros	Cons
Pre-processing	Fuzzy Clustering (FC)	Quantifying the input uncertainty possessing *fuzzy characteristics*	Only for the input uncertainty possessing a specific characteristic, not for
	Wavelet Transform (WT)	Quantifying the input uncertainty possessing *periodic or seasonal*	

		characteristics	model structure and
	Bias-correction	Quantifying the input uncertainty possessing *systematic bias error*	parameters uncertainty
Post-processing	Multiple Linear Regression (MLR)	Quantifying the overall predictive uncertainty of model structure and parameters possessing *linear features*	Only for the uncertainty possessing linear features
	Kalman filtering	Quantifying the overall predictive uncertainty of model structure and parameters possessing *systematic bias error*	Only for the single-model independently
	Generalized Likelihood Uncertainty Estimation (GLUE)	Quantifying the overall predictive uncertainty of model structure and parameters possessing *nonlinear features*	Only for the single-model independently
	Bayesian Uncertainty Processor (BUP)	Quantifying the overall predictive uncertainty of model structure and parameters possessing *Gaussian features*	Only for the single-model independently and meeting Gaussian assumption
	Bayesian Model Averaging (BMA)	Quantifying the overall predictive uncertainty of *multi-model structure and parameters*	Only for the specific form of posterior distributions

Any ensemble forecast approach relies upon model diversity that different models produce, with specific emphasis and different aspects of the features they want to model (Li et al., 2013; Raftery et al., 2005). Artificial Neural Networks (ANNs) used as data-driven methods to model air quality and meteorological systems have evolved rapidly over the last few decades (Ryan, 2016; Shen et al., 2018). For instance, the Back Propagation Neural Networks (BPNN), the Adaptive Neural Fuzzy Inference System (ANFIS), the Random Forest (RF), the Quantile Regression Neural Networks (QRNN), the Radial Basis Function (RBF), the Extreme Learning Machine (ELM), the Non-linear AutoRegressive with eXogenous inputs neural network (NARX), the Support Vector Machine (SVM) and the Long-Short Term Memory (LSTM) have been widely used to model air quality and meteorological forecasts (e.g. Akbari et al.,

2018; Ausati and Amanollahi, 2016; Cannon, 2011; Chang et al., 2016; Gao et al., 2018; Nieto et al., 2018; Prasad et al., 2016; Taghavifar et al., 2016; Voukantsis et al., 2011; Yeganeh et al., 2018; Yu et al., 2016; Zhu et al., 2018; Zhou et al., 2019 a,b). The factors of geographical location, meteorological conditions, population, traffic density and industrial activities have an impact on the physical-chemical composition and the concentration of airborne particles (e.g. Fanizza et al., 2018; Li et al., 2018; Sun et al., 2016; Yu and Stuart, 2017). The mass concentration of atmospheric particulate matter (e.g. $PM_{2.5}$) relies on a series of natural and anthropogenic processes, furthermore, the main contribution stems from secondary particles (Lin and Zhu, 2018; Lyu et al., 2016; Wu et al., 2018). Secondary particles' formation is attributed to a lot of factors: ozone, carbon monoxide, carbon dioxide, organic carbon, sulphur dioxide, nitrogen oxides and meteorological environments like temperature, precipitation, wind speed and direction as well as relative ambient humidity (Berardis and Eleonora, 2017; Coelho et al., 2014). Moreover, ensemble forecasting provides a practical and reliable approach that serves as a complement to ANN models for simulating and understanding of particle formation, transport, transformation and deposition mechanisms in the primary, secondary and natural sources and processes (Chen et al., 2018). Hence, it is interesting to make an in-depth study on ANN models for improving forecast reliability and accuracy and on the conversion of the deterministic forecasts into probabilistic forecasts using post-processing ensemble techniques.

Predictive uncertainties are closely associated with the spatial discretization of

physical processes, model structure and parameterization. Because any air quality model is considered the brief conceptualization of complicated chemical-physical processes in the atmospheric system, the hypotheses in the conceptual model induce air quality forecasts to get inaccurate. To decrease model uncertainty, the model averaging method is commonly adopted to integrate an ensemble of multiple models by using a linear sum of diverse models. Such model-averaging approaches bring deterministic outputs' linear average and make a combined single-value, for instance, equal weights averaging, MLR, Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC)-based model averaging (Breiman and Friedman, 1997; Buckland et al., 1997; Granger and Ramanathan, 1984; Leslie and Holland, 1991). Even if these model averaging methods have achieved good practicality and applicability, some researchers (e.g. Hoeting et al., 1999; Raftery et al., 2005) contended that the weights cannot thoroughly characterize single models' contribution and advocated BMA's application. The BMA approach can integrate the Probability Density Function (PDF) of different model predictions by means of making a weighted one. The applications of BMA approach in meteorological forecasts motivated its several usages in air quality forecasts (e.g. Mok et al., 2018; Pucer et al., 2018; Pannullo et al., 2016; Weber et al., 2016). However, the conditional PDF in standard BMA is supposed to conform to a Gaussian distribution (Raftery et al., 2005), which is appropriate for some specific predicted variables (e.g. atmospheric pressure and temperature). For other variables (e.g. $PM_{2.5}$, PM_{10}, Ozone, precipitation and wind speed), the Gaussian distribution would be a bad choice, whereas other PDF

distributions (e.g. Gamma, Gumbel, Pearson type III, Generalized Extreme Value, etc) would be good choices for fitting predicted variables (Mok et al., 2018; Pucer et al., 2018). Additionally, the standard BMA application need transform the model outputs (or predicted variables) from original space to the Gaussian space. To prevent such information loss during data space transformation, multivariate Copula functions have been used in meteorological and hydrological fields (e.g. Chen and Guo, 2019; Khajehei and Moradkhani, 2017; Nelsen, 2006; Zhang and Singh, 2019), owing to their outstanding capability of modeling the nonlinear dependence of multiple variables and their allowance of some flexibility in choosing an arbitrary marginal distribution. In a study by Madadgar and Moradkhani (2014), the combination of multivariate Copula function and BMA (CBMA) can relax the Gaussian assumption of PDFs. The CBMA approach showed superior practicality and reliability in hydrologic forecasts (e.g. rainfall-runoff processes). Whereas the review of the available literature indicates the CBMA has not been applied in air quality forecasts. Consequently, it is imminent to implement an in-depth study on the exploration of CBMA for quantifying and reducing the uncertainty encountered in ensemble air quality forecasts.

The research gaps and how did this work fulfill research gaps were described as follows. First, the current ensemble models for air quality forecasting mainly involved single-output ANNs and/or shallow learning ANNs whereas multi-output ANNs and deep learning ANNs were rarely applied in the ensemble forecast of air quality. Accordingly, the integration of the single-output, the multi-output, the shallow

learning, and the deep learning ANN models were proposed to configure the four members of the ensemble scheme for air quality forecasting (i.e. point forecasts). Second, the contribution of this study was attributed to exploring and extending our previous works (i.e. deterministic ANN models) (Zhou et al., 2019a,b) for making probabilistic ensemble $PM_{2.5}$ forecasts. Third, the family of Bayesian ensemble forecast methods consists of BMA and CBMA. The BMA method has been widely adopted for air quality forecasting. Despite the CBMA is an existing method, it has been rarely employed in the air quality forecast field. Accordingly, the CBMA method was introduced to create a probabilistic ensemble scheme for air quality forecasting based on point forecasts driven by multiple ANNs.

The novelties of this study relied on: multiple ANNs with various characteristics were for the first time integrated into a novel ensemble scheme for air quality forecasting while the combination of Copula function and BMA (CBMA) was taken as an existing method but rarely used in the air quality forecast field.

In this study, a CBMA-based approach was proposed for integrating CBMA and multiple ANNs to reduce the prediction intervals of ensemble $PM_{2.5}$ forecasts. Firstly, multiple ANN models were constructed for creating deterministic $PM_{2.5}$ forecasts independently. Then for comparison purpose, the CBMA approach and the BMA approach were implemented to transform the deterministic $PM_{2.5}$ forecasts of multiple ANN models into the ensemble probabilistic $PM_{2.5}$ forecasts respectively. The regional $PM_{2.5}$ forecasts in Taipei City of Taiwan were taken as a study case to assess the applicability as well as reliability of the proposed ensemble forecast method.

2. Methods

Figure 1 illustrated the ensemble forecast architecture that integrated the four deterministic ANN models (Figure 1 (a)) with the BMA (Figure 1 (b)) or the CBMA ensemble forecast approach (Figure 1 (c)). The deterministic point forecasts were created by multiple ANN models independently. The ensemble forecast could be improved by CBMA, as compared with the benchmark method (i.e. BMA). The used methods were briefly described as below.

2.1 Deterministic ANN models

In this study, the selected ensemble members included single-output ANFIS (S-ANFIS) (Jang, 1993), multi-output SVM (M-SVM) (Xu et al., 2013; Zhou et al., 2019a), single-output NARX (S-NARX) (Leontaritis and Billings, 1985) and multi-output deep learning LSTM (M-LSTM) (Zhou et al., 2019b) models. All models were the artificial neural network models and constructed for deterministic $PM_{2.5}$ forecasting. The models have the same network structures (i.e. input layer, hidden layer & output layer) whereas the models have different machine learning mechanisms. The S-ANFIS can extract the static and fuzzy feature between air quality and other factors, the M- SVM can extract the non-linear relationship between them, the S-NARX can extract dynamic feature between them, while the M-LSTM can extract the long and short-term relationship between them.

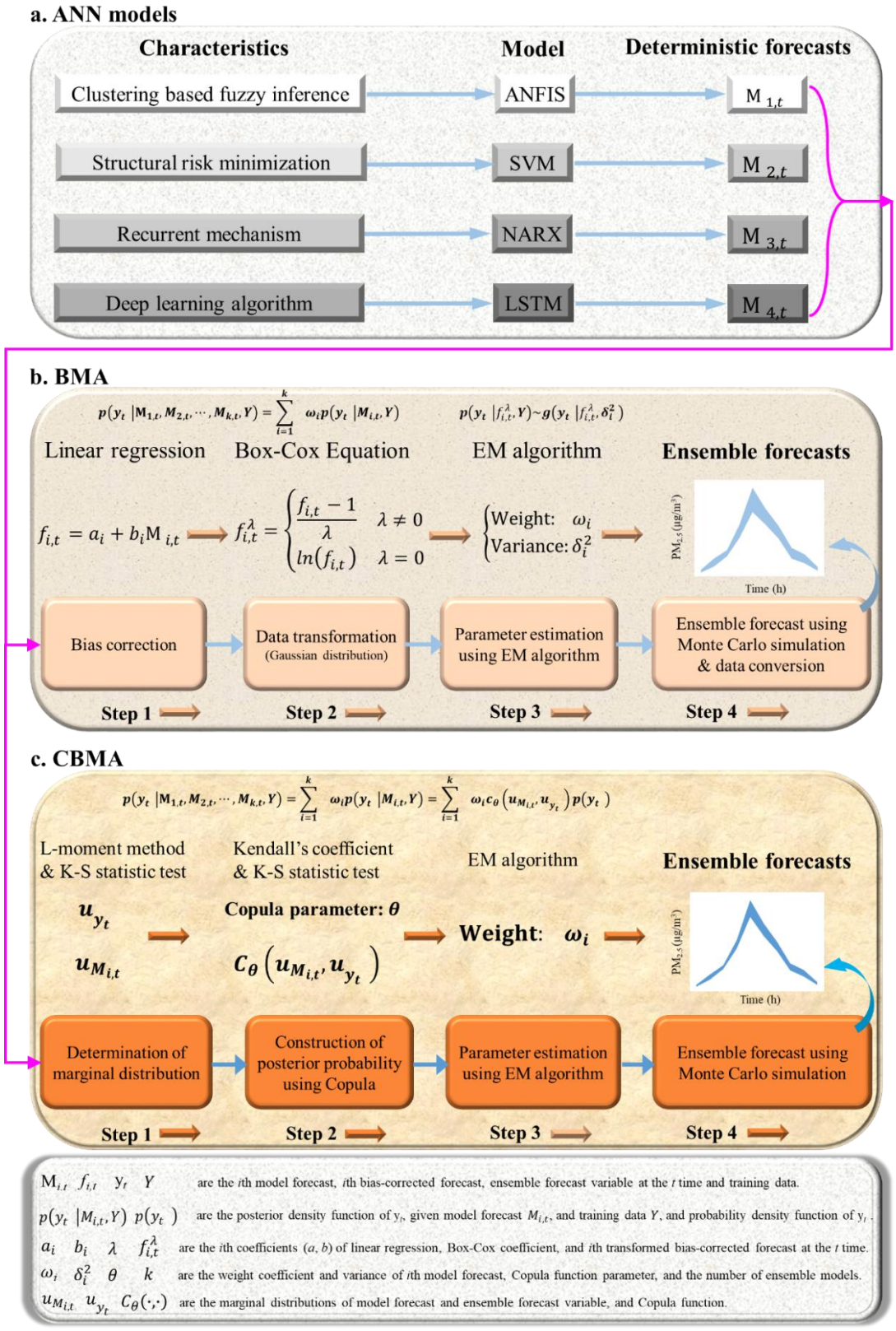


Figure. 1 Ensemble probabilistic forecast architecture. (a) ANN models. (b) Bayesian Model Averaging (BMA). (c) Copula-based Bayesian Model Averaging (CBMA) for ensemble forecasting.

In the case of static ANNs (i.e. S-ANFIS and M-SVM), a typical three-layered static feedforward neural network, which is comprised of multiple elements including nodes and weight connections that link nodes. In the case of recurrent ANNs (i.e. S-NARX and M-LSTM), the recurrent neural network involves three layers and constitutes recurrent connections from the outputs, which can delay several unit times to produce new inputs. More detailed descriptions of the four models and their parameters setting for air quality forecasting can be found in the references (Ausati and Amanollahi, 2016; Ghazi and Khadir, 2009; Prasad et al., 2016; Zhou et al., 2019 a,b).

Air quality data with specific time-lags (e.g. $PM_{2.5}$, PM_{10}, ozone, oxynitride, nitrogen dioxide, nitric oxide, sulfur dioxide, carbon monoxide, etc) and meteorological data with specific time-lags (e.g. precipitation, temperature, wind speed and direction as well as relative humidity) constituted the input variables while multi-step-ahead air quality forecasts (e.g. $PM_{2.5}$ concentration from t+1 up to t+4, horizon = 4) constituted the output variables.

The differences of four ANN models were summarized as (1) the two models (S-ANFIS & S-NARX) possessed single-output model structures where the two models (M-SVM & M-LSTM) possessed multi-output model structures; (2) the former three models (i.e. S-ANFIS, M-SVM & S-NARX) were classified as shallow neural networks (i.e. number of hidden layers = 1) whereas the fourth model (i.e. M-LSTM) was classified as deep learning neural networks (i.e. number of hidden layers $\geq$ 2); and (3) the former two models (i.e. S-ANFIS & M-SVM) were classified

as static (i.e. non-recurrent) neural networks whereas the latter two models (i.e. S-NARX & M-LSTM) were classified as dynamic (i.e. recurrent) neural networks. Moreover, the S-ANFIS and S-NARX models need to construct multiple independent models to output air quality forecast at diverse monitoring stations whereas the M-SVM and M-LSTM models require only one forecast model to make air quality multi-outputs. That is to say, the selected four ensemble members can provide model diversity for the applications of the following ensemble forecast approaches.

2.2 Bayesian Model Averaging (BMA)

BMA is a post-processing technique used to integrate the forecast results that are created by different models in virtue of making an ensemble PDF. The predicted distribution of a realization of the observation y_t, considering the multiple forecasts of *k* models $\{\mathrm{M}_{1,t}, \mathrm{M}_{2,t}, \cdots, \mathrm{M}_{k,t}\}$, and the observed data *Y* within the training stage can be formulated as follows.

$$p(y_t|\mathrm{M}_{1,t}, \mathrm{M}_{2,t}, \cdots, \mathrm{M}_{k,t}, Y) = \sum_{i=1}^{k} \omega_i p(y_t|\mathrm{M}_{i,t}, Y) \quad (1)$$

where $p(y_t|\mathrm{M}_{1,t}, \mathrm{M}_{2,t}, \cdots, \mathrm{M}_{k,t}, Y)$ is the predictive distribution of the realization of the observation y_t, given the independent forecasts of *k* models $\{\mathrm{M}_{1,t}, \mathrm{M}_{2,t}, \cdots, \mathrm{M}_{k,t}\}$, and the observed data *Y*. $p(y_t|\mathrm{M}_{i,t}, Y)$ is the posterior distribution of function of y_t, given model forecast $M_{i,t}$, and training data Y. ω_i is the weight coefficient of *i*th model. The general implementation procedure of the BMA approach consisting of four basic steps can be found in Appendix A.

2.3 Copula-based Bayesian Model Averaging (CBMA)

From what has been discussed above, the predictive distribution in the BMA approach

is generally confined to a specific parameter distribution (e.g. Gaussian distribution) and is computed by a weighted sum of forecast PDFs. Therefore, we clarified a general procedure that fused multivariate Copula function into the original BMA approach (CBMA) to relax the limitations of unbiased forecasts and Gaussian distribution.

Let $u_{\mathrm{M}_{i,t}}$ and u_{y_t} be the sampling values in CDFs of U_{M} and U_y respectively. Let $p_{\mathrm{M}}(u_{\mathrm{M}_{i,t}})$ and $p_y(u_{y_t})$ be the Probability Density Functions (PDFs) of the forecast variables of multiple models ($\mathrm{M}_{i,t}$) and realization of observation (y_t) respectively. Using the PDF of Copula function, a joint PDF of ($u_{\mathrm{M}_{i,t}}$, u_{y_t}) and a conditional probability can be constructed as follows.

$$p\left(u_{\mathrm{M}_{i,t}}, u_{y_t}\right) = c_{\theta_i}\left(u_{\mathrm{M}_{i,t}}, u_{y_t}\right) \cdot p_{\mathrm{M}}(u_{\mathrm{M}_{i,t}}) \cdot p_y(u_{y_t}) \tag{2a}$$

$$p\left(u_{y_t}\middle|u_{\mathrm{M}_{i,t}}\right) = \frac{p\left(u_{\mathrm{M}_{i,t}}, u_{y_t}\right)}{p_{\mathrm{M}}(u_{\mathrm{M}_{i,t}})} = c_{\theta_i}\left(u_{\mathrm{M}_{i,t}}, u_{y_t}\right) \cdot p_y(u_{y_t}) \tag{2b}$$

where $p\left(u_{\mathrm{M}_{i,t}}, u_{y_t}\right)$ is the joint PDF of ($u_{\mathrm{M}_{i,t}}$, u_{y_t}). $c_{\theta_i}\left(u_{\mathrm{M}_{i,t}}, u_{y_t}\right)$ is the Copula joint PDF of ($u_{\mathrm{M}_{i,t}}$, u_{y_t}) and θ_i is the parameter of the Copula function. $p\left(u_{y_t}\middle|u_{\mathrm{M}_{i,t}}\right)$ is the conditional probability of u_{y_t}, given the value of $u_{\mathrm{M}_{i,t}}$. Then, the conditional probability (Eq. (2b)) is used to replace the posterior probability (Eq. (1)) and the predicted distribution of the realization of observation y_t is updated as follows.

$$\begin{aligned} p\left(y_t\middle|\mathrm{M}_{1,t}, \mathrm{M}_{2,t}, \cdots, \mathrm{M}_{k,t}, Y\right) &= \sum_{i=1}^{k} \omega_i p\left(y_t\middle|\mathrm{M}_{i,t}, Y\right) \\ &= \sum_{i=1}^{k} \omega_i c_{\theta_i}\left(u_{\mathrm{M}_{i,t}}, u_{y_t}\right) \cdot p_y(u_{y_t}) \end{aligned} \tag{3}$$

As seen in Eq. (3), the posterior distribution $p\left(y_t\middle|\mathrm{M}_{i,t}, Y\right)$ is directly calculated

without needs to use both bias-correction methods (Eq. (1) in Appendix A) and Gaussian data transformation (Eq. (2) in Appendix A). The general implementation procedure of the CBMA approach consisting of four basic steps can be found in Appendix B.

It is noted that the differences between BMA approach and CBMA approach include: (1) the former demands a particular conditional PDFs (e.g. Gaussian), or data transformation (Non-Gaussian PDFs) and bias-correction for model forecasts whereas the latter has a flexible structure and relaxes the type of conditional PDFs and (2) the former needs to estimate the parameters of weight (ω_i) and variance (δ_i^2) whereas the latter needs to estimate the Copula parameter (θ_i) and the weight (ω_i).

The general implementation programming of used machine learning models (ANFIS, SVM, NARX, LSTM) and Copula function can be obtained from the Statistics and Machine Learning Toolbox of the Matlab software (website: https://ww2.mathworks.cn/products/statistics.html#machine-learning).

2.4 Evaluation criteria

The Root-Mean-Square Error (RMSE) and the goodness-of-fit with respect to the benchmark (G_{bench}) were introduced to assess the accuracy of the deterministic forecast model. The two indicators were defined as follows.

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

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

where $\widehat{\mathrm{Y}}(t)$ and $\mathrm{Y}(t)$ are the model forecast and observation at the *t*-th time, respectively. $\mathrm{Y_{bench}}(t)$ is the observation moved backwards by *n*th time lags, for

instance, for the horizon t + n, $\mathrm{Y_{bench}}(t)$=$\mathrm{Y}(t-n)$.

To evaluate the performance of probabilistic forecast models, the Containing Ratio (CR), the average Relative Band-width (RB) and the Continuous Ranked Probability Score (CRPS), were adopted for assessing the goodness of the prediction bounds (Gneiting and Raftery, 2007; Gneiting, 2008; Xiong and O'Connor, 2008). Their mathematical formulas were described below.

$$\mathrm{N}(t)=\begin{cases}1, & if\left(q_l(t)\le\hat{Z}(t)\le q_u(t)\right)\\ 0, & else\end{cases} \tag{6a}$$

$$\mathrm{CR}=\frac{\sum_{t=1}^{\mathrm{N}}\mathrm{N}(t)}{\mathrm{N}}\times 100\% \tag{6b}$$

$$\mathrm{RB}=\frac{1}{\mathrm{N}}\sum_{t=1}^{\mathrm{N}}\left(\frac{q_u(t)-q_l(t)}{\mathrm{Z}(t)}\right) \tag{7}$$

$$\mathrm{CRPS}=\int_{-\infty}^{+\infty}[F^f(x)-F^o(x)]^2dx \tag{8}$$

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 the t time respectively. $\mathrm{F}^f(x)$ and $\mathrm{F}^o(x)$ are the cumulative distribution functions of the forecast and observation distributions, respectively. x is the variable of the cumulative distribution function. The value of $\mathrm{N}(t)$ is either 0 or 1, in which 0 indicates the observed data falls outside of its prediction bounds while 1 indicates the observed data falls within its prediction bounds. These evaluation criteria indicate that models with higher $\mathrm{G_{bench}}$ and CR values but lower RMSE, RB and CRPS values would produce better performances.

3. Study area and materials

The study area (Figure 2) was briefly introduced as follows. With the economy and

population fast boosting, one of the hot topics in Taiwan focused on air quality deterioration. People in Taipei City were compelled to handle a high-level intervention of $PM_{2.5}$. Air pollution not just induced respiratory diseases but also caused a matter of life or death. Hence, it is imperative to make accurate and reliable $PM_{2.5}$ forecasts so as to adequately process the health risk caused by regional air pollution.

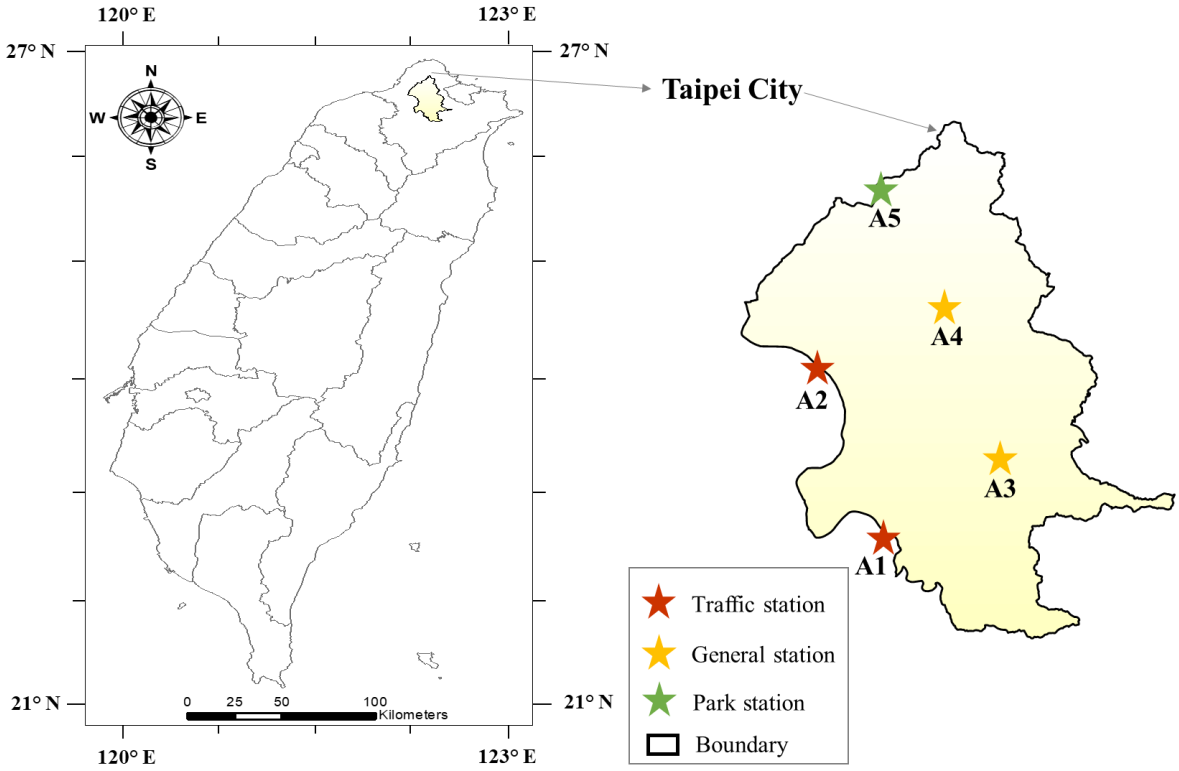


Figure. 2 Distribution of air quality monitoring stations (A1-A5) in Taipei City.

The positions of Taipei City and 5 air quality monitoring stations were presented in Figure 2. Stations A1 (Yonghe) and A2 (Sanchong) where the stations located in areas of heavy traffic are traffic stations, Stations A3 (Songshan) and A4 (Shilin) where the stations located in areas of intensive human activities and commercial trading are general stations, and Station A5 (Yangming) where the station located in the Yang-Ming Park is a park station. The Environmental Protection Administration

(EPA) in Taiwan (https://taqm.epa.gov.tw/taqm/en/b0101.aspx) provided a convenient open data platform where researchers could get access to many kinds of Taiwan-related data such as air quality and meteorological datasets. Hourly data of air quality factors (eight variables: $PM_{2.5}$, PM_{10}, ozone, oxynitride, nitrogen dioxide, nitric oxide, sulfur dioxide, carbon monoxide) and meteorological factors (five variables: precipitation, temperature, wind speed and direction as well as relative humidity) over a span of 9 years (2010-2018) were available. A total of 78,888 (=[(2×366)+(7×365)]×24) hourly datasets were used in this study, where 35,064 data (4 years) were used for model training while the remaining 26,304 data (3 years) and 17,520 data (2 years) were used for model validating and testing respectively. The data standardization that centered the mean to 0 and the standard deviation to 1, was conducted to decrease the negative effect of the different scales of input data on the model's learning ability.

Figure 3 presented the statistic indexes of seasonal and annual $PM_{2.5}$ concentration at five air quality monitoring stations. We noticed that the statistic indexes of the maximum, average and standard derivation at traffic stations (A1 and A2) were the highest while those in the park station (A5) were the lowest, which could be due to the primary source of particulate matter of a station. Based on the highest values ($\geqslant$ 0.5) of the Kendall tau coefficients (Maidment, 1993), 1h-4h time lags were identified for air quality factors at traffic stations (A1 and A2) and 1h-2h time lags were identified for air quality factors at general and park stations (A3, A4 and A5) while 1h-4h time lags were identified for meteorological factors at all

stations (Zhou et al., 2019 a,b).

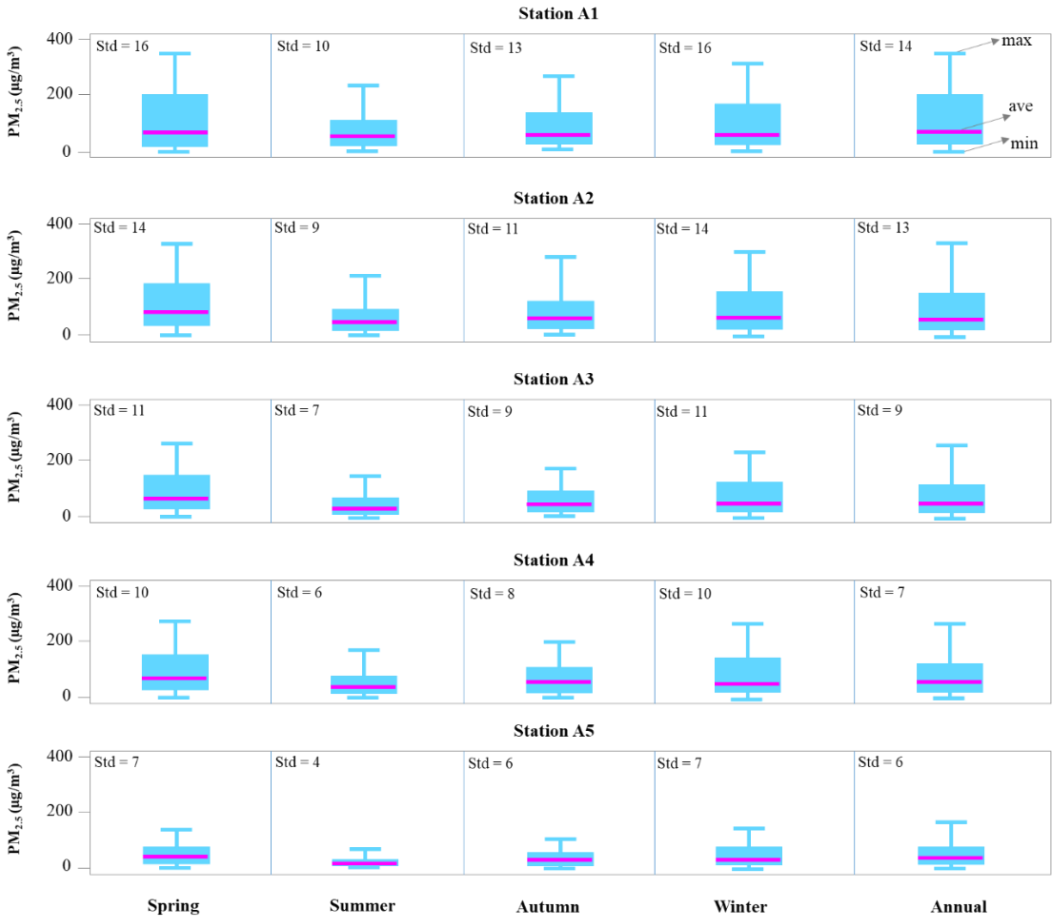


Figure. 3 Statistic indexes of seasonal $PM_{2.5}$ concentrations from 2010 to 2018 (9 years) at five air quality monitoring stations in Taipei City. The abbreviations (max, ave, min, std) denote the maximum, average, minimum and standard deviation respectively.

4. Results and discussion

Both the CBMA approach and the BMA approach were employed to integrate the $PM_{2.5}$ forecasts of four deterministic ANN models and the BMA approach served as a benchmark. The results and findings were presented in the order of the deterministic $PM_{2.5}$ forecasts of four ANN models (Section 4.1), the determination of marginal distributions and the Copula function (Section 4.2), the ensemble $PM_{2.5}$ forecasts and

the summarization (Section 4.3), shown as follows.

4.1 Deterministic $PM_{2.5}$ forecasts of four ANN models

The four models (S-ANFIS, M-SVM, S-NARX and M-LSTM) were applied for deterministic forecasting $PM_{2.5}$ concentrations of five monitoring stations (A1-A5) from horizons t+1 up to t+4 respectively. The RMSE and G_{bench} scores over the testing stages were calculated for each ANN model (Table 2). The RMSE and G_{bench} scores indicated that the performance of static models (S-ANFIS & M-SVM) was not as good as the recurrent models (S-NARX & M-LSTM) at the traffic stations (A1 & A2) and the general stations (A3 & A4). While the single-output modes (S-ANFIS & S-NARX) performed better than the multi-output models (M-SVM & M-LSTM) at the park station (A5). This was mainly due to the different learning mechanisms (or model structures) used for the configuration of each model and the simulation of the different air pollutant generating processes. The secondary processes (e.g. stations A1 & A2) and primary processes (e.g. stations A3 & A4) represented complex and indirect air pollutant generating mechanisms and then required complex ANN models (e.g. S-NARX or M-LSTM) with recurrent or deep learning algorithms (complex model structure and a large number of parameters) to characterize such processes. The natural processes (e.g. station A5) represented simplex and direct air pollutant generating mechanisms and then only required simplex ANN models (e.g. S-ANFIS or S-NARX) with single-output structure (simplex mode structure and fewer parameters) to characterize such processes. As a reminder, the multi-model ensemble strategy was a means to exploit the diversity of skillful predictions from different

models. Hence, from a perspective of regional air quality forecast accuracy, it needs an ensemble technique (BMAs) to combine $PM_{2.5}$ forecasts of different deterministic models and improves their effectiveness for regional air quality forecasts.

Table 2 Comparison performance of four ANN models for deterministic $PM_{2.5}$ forecasts from horizons t+1 up to t+4 in the testing stage at different stations.

Station	Horizon	ANFIS		SVM		NARX		LSTM	
		RMSE (μg/m^3)	G_{bench}	RMSE (μg/m^3)	G_{bench}	RMSE (μg/m^3)	G_{bench}	RMSE (μg/m^3)	G_{bench}
A1	t+1	4.86	0.92	4.62	0.93	4.51	0.95	4.41	0.95
	t+2	5.77	0.87	5.68	0.90	5.43	0.91	5.11	0.92
	t+3	8.35	0.82	8.09	0.85	7.32	0.86	6.14	0.87
	t+4	11.33	0.78	11.26	0.80	10.78	0.82	9.55	0.84
A2	t+1	4.66	0.91	4.58	0.92	4.47	0.94	4.31	0.95
	t+2	5.33	0.87	5.26	0.88	5.06	0.90	4.83	0.92
	t+3	7.49	0.76	7.34	0.78	7.12	0.82	7.04	0.85
	t+4	10.38	0.71	10.16	0.73	9.94	0.75	9.45	0.77
A3	t+1	4.13	0.92	4.05	0.93	3.88	0.93	3.65	0.94
	t+2	5.23	0.88	5.11	0.90	4.92	0.91	4.71	0.92
	t+3	6.20	0.78	6.13	0.81	6.05	0.83	5.94	0.85
	t+4	9.29	0.72	9.15	0.74	8.67	0.76	8.21	0.78
A4	t+1	3.86	0.90	3.72	0.92	3.61	0.93	3.55	0.94
	t+2	5.04	0.88	4.97	0.90	4.89	0.91	4.72	0.92
	t+3	6.30	0.80	6.21	0.83	6.13	0.86	6.06	0.88
	t+4	8.92	0.73	8.85	0.75	8.66	0.81	8.57	0.83
A5	t+1	2.51	0.91	2.58	0.89	2.12	0.93	2.65	0.88
	t+2	3.61	0.88	3.78	0.86	3.36	0.90	3.90	0.84
	t+3	5.18	0.84	5.57	0.81	5.03	0.87	5.86	0.78
	t+4	7.26	0.78	7.59	0.75	7.03	0.81	7.92	0.74

4.2 Determination of marginal distribution and Copula function

In the CBMA application for different horizons (t+1 ~ t+4), it needs to identify the best CDFs for fitting the observations ($y_t, \dots, y_{t+4}$) and ith model forecasts ($M_{i,t+1}, \dots, M_{i,t+4}$). It seems reasonable to consider that the observations ($y_t, \dots, y_{t+4}$) follow the same marginal CDF of the variable (y_t) and therefore only the cumulative distribution functions of the observation (y_t) and model forecasts ($M_{i,t+1}, \dots, M_{i,t+4}$)

need to be fitted (Koutsoyiannis and Montanari, 2015; Liu et al., 2018).

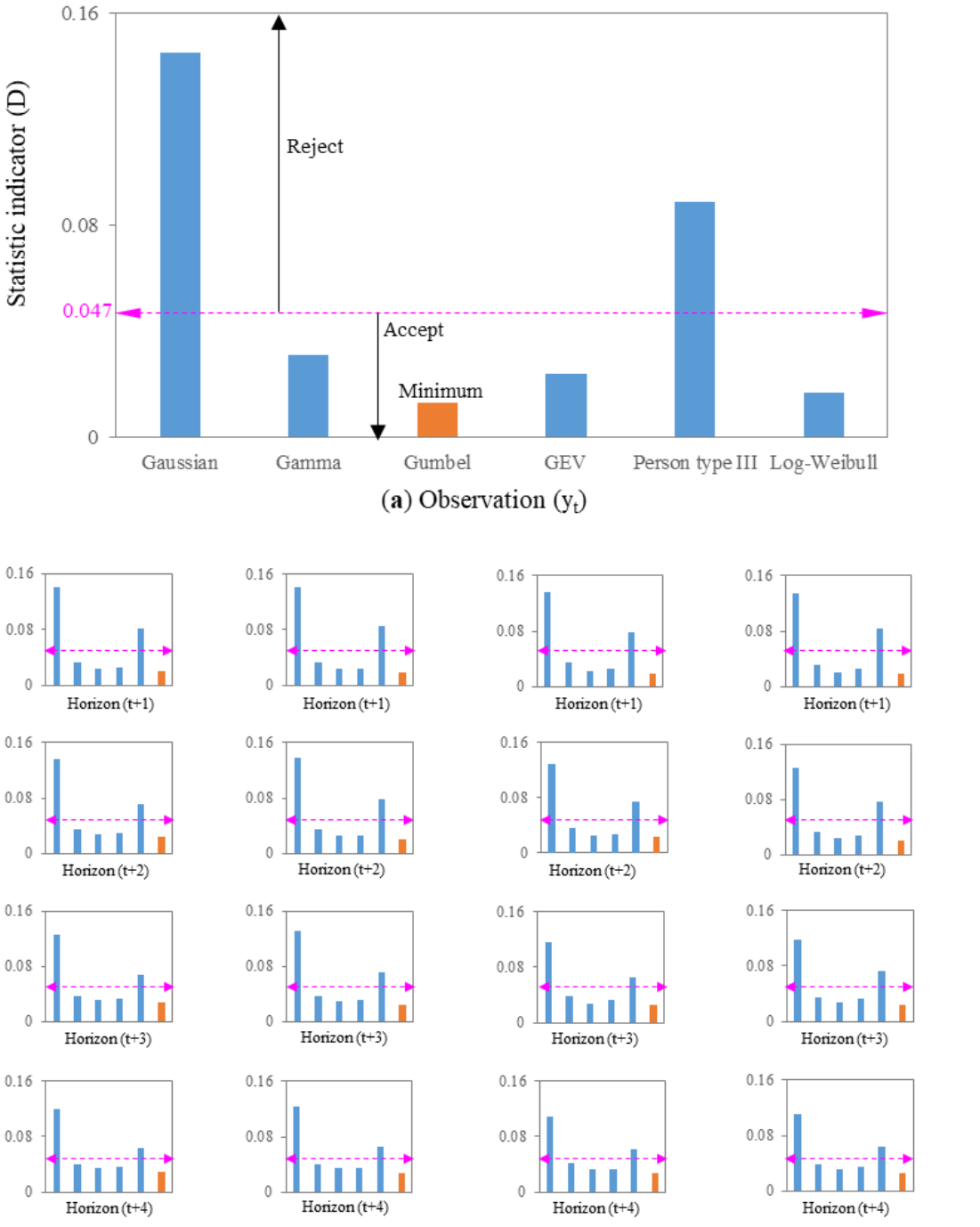


Figure. 4 Statistic indicator (D) values for verifying the null hypothesis at the 5% significance level in the training stages at the Traffic Station A1. The critical value of statistic indicator (D) = 0.047. The large values (≥ 0.047) of statistic indicator (D) indicate that the null hypothesis for candidate distribution would be rejected at the 5% significance level and the small values (< 0.047) of statistic indicator (D) indicate that the null hypothesis for candidate distribution cannot be rejected at the 5% significance level.

Take traffic Station A1 for example, Figure 4 summarized the K-S statistic indicator values (D), which was used to verify the null hypothesis at the 5% significance level in the training stages. The null hypothesis was defined as the marginal distribution follows one candidate distribution, against the alternative that it did not follow such a candidate distribution. The results suggested that the null hypothesis for all four candidate distributions could not be rejected at the 5% significance level (critical value = 0.047), other than Gaussian and Pearson type III distributions. That is to say, both observations and model forecasts have non-Gaussian and Pearson type III properties. The Gumbel distribution provided minimal D values for all observations while the Log-Weibull distribution provided minimal D values for all model forecasts. In other words, the Gumbel distribution and the Log-Weibull distribution would be considered as the best fitted distributions for observations (y_t) and model forecasts ($\mathrm{M}_{i,t+1}, \ldots, \mathrm{M}_{i,t+4}$) respectively.

When the best marginal distribution was determined, a Copula function should be selected to model the joint distribution between model forecasts ($\mathrm{M}_{i,t}$) and observations (y_t). Take the Traffic Station A1 for example, Table 3 presented the estimated parameters of the three candidate copula functions and the values of statistic indicator (D) in the training stages. The null hypothesis was defined as the joint distribution follows one candidate Copula function, against the alternative that it did not follow such a candidate Copula function. The results revealed that the null hypothesis for two candidate distributions could not be rejected at the 5% significance level (critical value = 0.047), other than Clayton Copula function. The smallest K-S

statistic indicator (D) was produced by the Gumbel-Hougaard Copula function. Consequently, the Gumbel-Hougaard Copula function could be considered as the best fitted joint distribution between observations (y_t) and model forecasts ($M_{i,t}$).

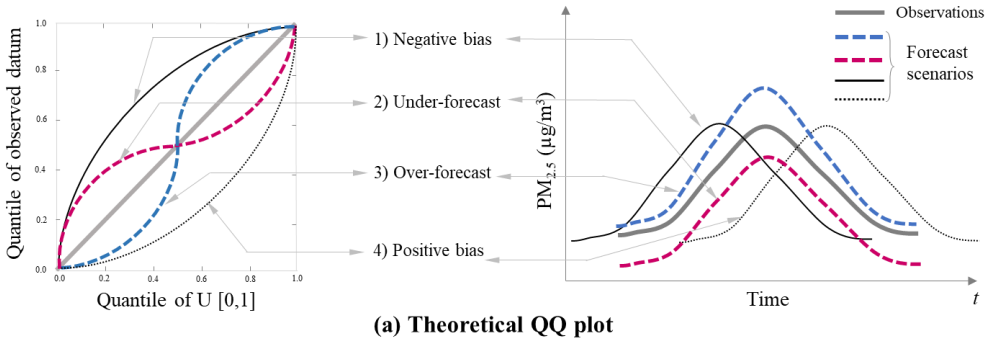
Table 3 Estimated parameters of the candidate copula functions and the values of statistic indicator (D) in the training stages at the Traffic Station A1

Model	Variables	Gumbel-Hougaard		Frank		Clayton	
		θ	D	θ	D	θ	D
S-ANFIS	(y_t, M_{t+1})	9.3	***0.017**	21.4	0.023	16.6	0.092
	(y_t, M_{t+2})	9.0	**0.022**	20.7	0.027	16.0	0.105
	(y_t, M_{t+3})	8.6	**0.029**	19.8	0.034	15.2	0.111
	(y_t, M_{t+4})	8.1	**0.036**	18.6	0.040	14.2	0.124
M-SVM	(y_t, M_{t+1})	9.6	**0.015**	22.0	0.020	17.2	0.094
	(y_t, M_{t+2})	9.1	**0.021**	20.9	0.025	16.2	0.102
	(y_t, M_{t+3})	8.8	**0.027**	20.2	0.031	15.6	0.114
	(y_t, M_{t+4})	8.3	**0.034**	19.0	0.038	14.6	0.128
S-NARX	(y_t, M_{t+1})	10.3	**0.012**	22.3	0.016	18.6	0.085
	(y_t, M_{t+2})	9.8	**0.016**	21.2	0.020	17.6	0.094
	(y_t, M_{t+3})	9.3	**0.021**	20.1	0.025	16.6	0.109
	(y_t, M_{t+4})	8.9	**0.030**	19.3	0.032	15.8	0.120
M-LSTM	(y_t, M_{t+1})	10.2	**0.014**	22.5	0.017	18.4	0.089
	(y_t, M_{t+2})	9.8	**0.017**	21.6	0.021	17.6	0.097
	(y_t, M_{t+3})	9.4	**0.020**	20.7	0.027	16.8	0.105
	(y_t, M_{t+4})	9.0	**0.028**	19.9	0.034	16.0	0.117

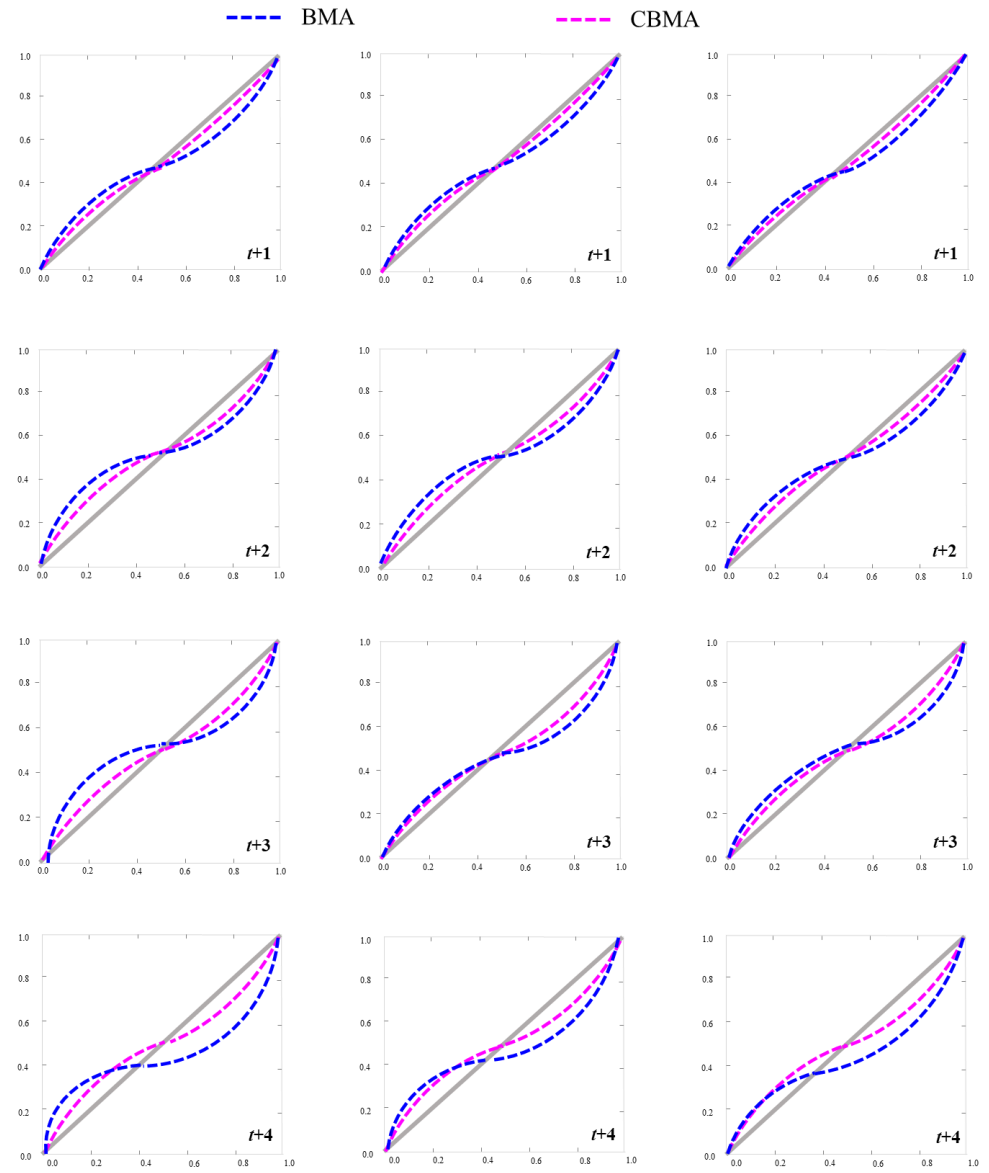
*A number in bold denotes the smallest value of statistic indicator (D) in its category. The values of y_t are the observed $PM_{2.5}$ concentrations of Traffic Station A1 at the current time *t*. The values of M_{t+1}, M_{t+2}, M_{t+3}, M_{t+4} are the ANN models forecasts of $PM_{2.5}$ concentrations of Traffic Station A1 at the horizons from *t*+1 to *t*+4. The critical value of statistic indicator (D) = 0.047. The large values ($\geqslant$ 0.047) of statistic indicator (D) indicate that the null hypothesis for candidate distribution would be rejected at the 5% significance level and the small values (< 0.047) of statistic indicator (D) indicate that the null hypothesis for candidate distribution cannot be rejected at the 5% significance level.

4.3 Ensemble $PM_{2.5}$ forecasts

Moreover, QQ plots were employed to assess the reliability of ensemble $PM_{2.5}$ forecasts. Figure 5 presented the predictive QQ plots used for ensemble $PM_{2.5}$ forecasting (e.g. traffic Station A1, general Station A3 & park Station A5) from horizons *t*+1 up to *t*+4 in the testing stages, respectively.



(a) Theoretical QQ plot



(b) Traffic Station A1 **(c) General Station A3** **(d) Park Station A5**

Figure. 5 Predictive Quantile-Quantile (QQ) plots for ensemble $PM_{2.5}$ forecasts from horizons t+1 up to t+4 in the testing stages at the traffic Station A1, general Station A3 and park Station A5 respectively. 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.

According to Figures 5(b) and 5(c), it was easy to find that the QQ plot generated by the CBMA approach was closer to the 1:1 line, in comparison to that of the BMA one. That is to say, the CBMA approach produced higher reliability and smaller bias than the BMA one. The results pointed out that the CBMA approach could effectively quantify predictive uncertainty owing to its better agreement between the predictive distribution and the observations. This finding demonstrated that the CBMA approach performed significantly better from the perspective of reliability.

For the ensemble $PM_{2.5}$ forecasts (e.g. Traffic Station A1, General Station A3 & Park Station A5) at horizons from t+1 up to t+4, the values of CR, RB and CRPS scores were listed in Figure 6. For the Traffic Station A1 and the General Station A3, the CBMA approach produced better performance in all horizons whereas the BMA performed well only at horizons up to t+2 (e.g. CR was higher than 90%, RB was lower than 0.15 and CRPS was lower than 12 $\mu g/m^3$ at the Station A1). For the Park Station A5, the BMA approach performed as well as the CBMA approach in all horizons. Take the Traffic Station A1 for example, the BMA approach produced small CR values, whereas the CBMA approach produced small RB and CRPS values. For horizon t+4, the CBMA approach could improve the CR value by 9.58 % as well as reduce the RB value by 34.48 % and the CRPS value by 32.89 %, as compared to the BMA one. That is to say, the CBMA approach not only could largely increase ensemble forecast accuracy at the goodness of the prediction bounds (in terms of CR and CRPS values) but also could decrease the impact of $PM_{2.5}$ concentration magnitude on the band-width of the prediction bounds (in terms of RB values)

simultaneously.

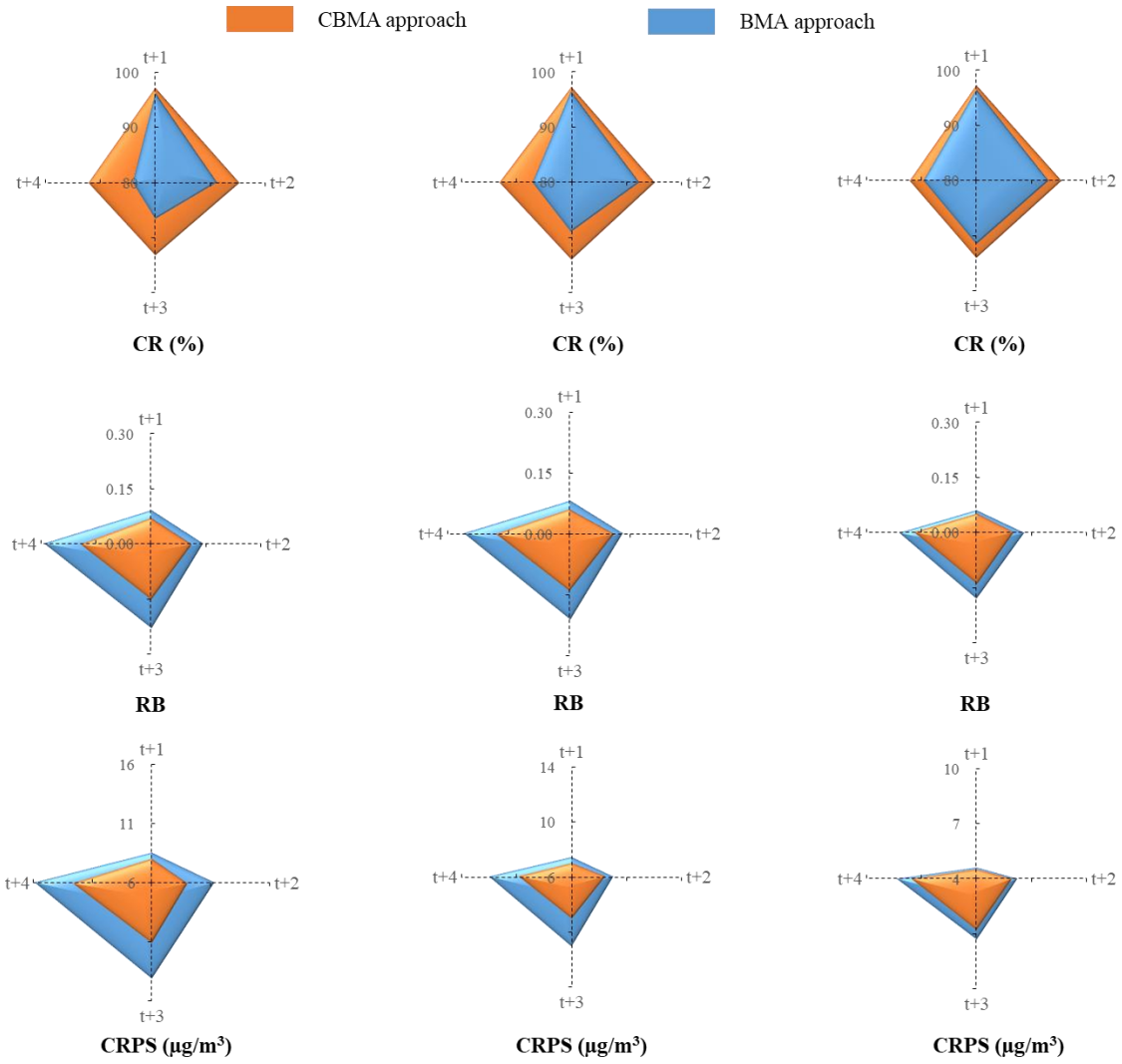


Figure. 6 CBMA and BMA performance of ensemble $PM_{2.5}$ forecasts in the testing stage at the traffic Station A1, general Station A3 and park Station A5. All of indicator values are computed for the 90% prediction intervals.

The results demonstrated that the CBMA approach had higher reliability and generalizability for ensemble $PM_{2.5}$ forecasting, in comparison to the BMA one. The reason for causing the forecast accuracy of the CBMA approach superior to the BMA one consisted of: the Copula functions in CBMA approach were able to eliminate forecast bias and characterize the correlations between observed values and forecast

values so that the forecast errors and biases could be reduced significantly. In consequence, the simple bias correction with linear regression method would not need in the application of Copula functions in the CBMA approach, as compared to the BMA one.

The weights created by CBMA and BMA approaches in the training stage were presented in Figure 7. Each data point suggested the weights (CBMA & BMA) for various models and different horizons (t+1 ~ t+4); there were 16 data points (4 (models) × 4 (horizons) = 16) in each subplot. Except in the Park Station A5 (correlation coefficient R = 0.65), the correlation of weights (R = -0.11 ~ 0.37) was very small, which could clearly demonstrate the different performance of CBMA and BMA. It was noted from Eqs. (5) and (8) that the weight of each model in the BMA (or CBMA) approach was expressed as a function of the latent variable $v_i^t(j)$ (or $z_i^t(j)$) and the posterior probability of training data was employed to compute it. Hence, the CBMA approach was not only restricted to the model diversity and the shape of posterior distributions, but it also had an effect on the weights assigned to each forecast model and the performance of EM algorithm.

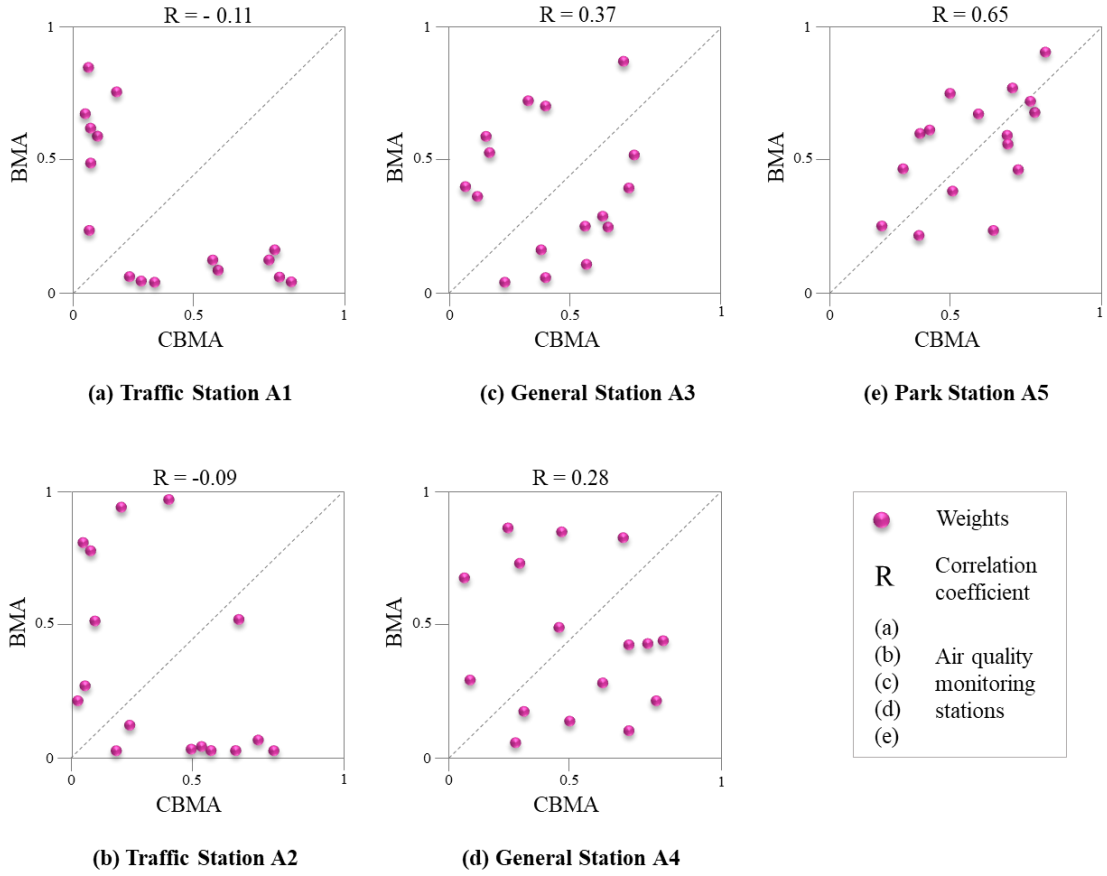


Figure. 7 Comparing the weights of four ANN models for the horizons (t+1 to t+4) in the training stage after the application of CBMA and BMA for each station.

To clearly differentiate the capabilities of the BMA and the CBMA approaches, three $PM_{2.5}$ events with maximal $PM_{2.5}$ concentrations reaching 80 μg/m^3 (low), 160 μg/m^3 (medium) and 250 μg/m^3 (high), respectively, were applied for testing both approaches by evaluating whether the observed $PM_{2.5}$ concentrations fell within the 90% prediction interval at horizon *t*+4 in the testing stages, as shown in Figure 8. It revealed that: (1) most of the observed $PM_{2.5}$ concentrations fell within the 90% prediction intervals generated by both approaches, (2) the CBMA approach provided better results in terms of predictive distribution, and (3) the CBMA approach was magically superior to that of the BMA one. From air pollutant mechanisms' perspective, primary emission's impact related to meteorological circumstances (i.e.

park Station A5) on the BMA and CBMA approaches was not significant, while secondary emission's impact related to meteorological circumstances (i.e. traffic Station A1 and general Station A3) on the BMA and CBMA approaches made a significant difference. For Taipei City, a fast urban growth city, regional air quality exchanges with traffic burdens, commercial trading and intensive human activities frequently. A high $PM_{2.5}$ event driven by secondary processes was closely related with regional transportation of aged secondary aerosol or secondary transformation of gaseous pollutants, whereas a medium-low $PM_{2.5}$ event driven by the primary or natural process was closely related with local weather conditions and primary emissions. Both CBMA and BMA approaches produced a better performance at the traffic station (A1) and general station (A3) than at the park station (A5). In other words, the CBMA approach not only greatly improved the ensemble forecast accuracy of $PM_{2.5}$ concentration at traffic station and general station, but also performed as well as the BMA approach at the park station.

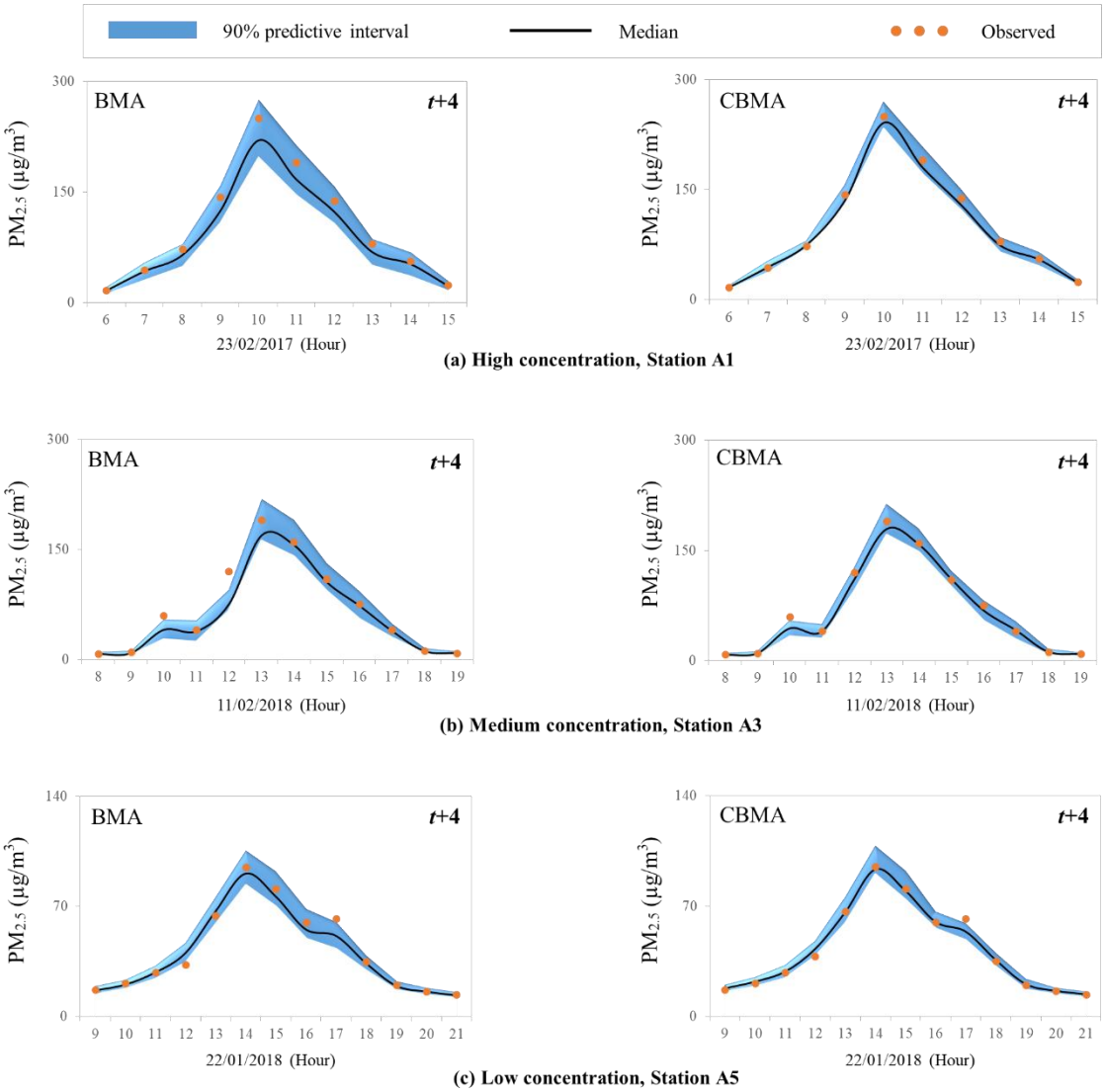


Figure. 8 BMA and CBMA ensemble $PM_{2.5}$ forecasts for air quality monitoring Stations A1, A3 and A5 at horizon t+4 respectively. Three $PM_{2.5}$ events with maximal $PM_{2.5}$ concentrations exceeding (a) 250 µg/m^3 (high concentration, Station A1), (b) 160 µg/m^3 (medium concentration, Station A3) and (c) 80 µg/m^3 (low concentration, Station A5) were selected for testing the constructed models, respectively.

In brief, from the standpoint of model performance, RMSE and G_{bench} were employed to evaluate the accuracy of deterministic $PM_{2.5}$ forecasts while QQ plot, CR, RB and CRPS indicators were employed to evaluate the reliability (QQ plot) and sharpness (CR, RB and CRPS) of ensemble $PM_{2.5}$ forecasts. The CBMA approach not only could produce more stable and accurate ensemble forecasts but also could reduce the predictive distributions encountered in multi-step-ahead $PM_{2.5}$ forecasts to small

ranges, by means of removing the requirements of data transformation and bias correction procedures, in comparison to the BMA approach. In light of methodological transferability, future research would extend the CBMA methodology on ensemble forecasting or comparison analysis studies between data-driven models and physically-based models (e.g. Weather Research and Forecasting Models).

5. Conclusions

This study explored a CBMA approach for modeling ensemble $PM_{2.5}$ forecasts of 5 air quality monitoring stations located in the Taipei City of Taiwan and the standard BMA one was selected as a benchmark. First, four ANN models with different complexities were used for $PM_{2.5}$ forecasts of each air quality monitoring station. And then, the CBMA approach and the BMA approach were compared in ensemble $PM_{2.5}$ forecasting.

The results demonstrated that the CBMA approach displayed better ensemble forecast skill in comparison to the BMA one. First, in terms of CR, RB and CRPS indicator values, forecast accuracy and reliability increased significantly after applying the CBMA approach in all air quality monitoring stations. For horizons t+1 up to t+4, the CBMA approach would increase the values of CR indicator by 3.12% - 9.58% as well as decrease the values of RB indicator by 8.63% - 34.48% and the values of CRPS indicator by 7.62% - 32.89%, as compared to the BMA one. Second, results of QQ plots indicated that less bias, and more reliable forecast results when the CBMA approach was employed as a post-processing technique for multiple

deterministic models. In comparison to BMA, the CBMA approach could create a more precise predictive distribution with small uncertainty. The CBMA approach produced much better forecasts on the air quality concentrations at longer forecast horizons and significantly alleviated underpredicting phenomena. The reason that the CBMA approach succeeded in attaining favorable ensemble forecasts would be owing to the core strategy: the use of the Copula function could capture the dependence structure between variables, which avoided their transformation in the Gaussian space as it was done in the BMA approach.

Therefore, the CBMA approach in place of the BMA one would be in the interest of reducing the predictive uncertainty of real-time $PM_{2.5}$ forecasting. In the application of the CBMA approach, the key point was to detect and select a suitable marginal PDF for each observation and model forecast, and then a Copula function was constructed for modelling a joint PDF between observation and model forecast. Finally, it was worth noting that the computational time (less than 2 minutes) of the proposed approach was extremely short and therefore it could be applied with success to real-time air quality forecasting.

From the perspective of regional $PM_{2.5}$ characteristics, Taipei City acts as Taiwan's political, economic and cultural center, while its air quality concentrations are attributed to high traffic influences, high human activities and commercial trading influences in comparison to these in other cities of Taiwan. The proposed methodology could be effectively employed not only to model the heterogeneities in different air pollutant-generating mechanisms (e.g., primary and secondary

mechanisms, and natural situations) and different seasons, but also to provide reliable and accurate probabilistic regional $PM_{2.5}$ forecasts in the interest of Taiwan's social and industrial development.

Appendix A

General implementation procedure of BMA approach

Step 1: Implement bias-correction. One requirement of BMA application is that the model forecasts $M_{i,t}$ should be bias-corrected due to the non-bias assumption. A bias-correction with linear regression method suggested by Raftery et al. (2005) was adopted prior to BMA execution and the original model forecast results ($\mathrm{M}_{i,t}$) ought to be substituted by the bias-corrected forecast variables ($f_{i,t}$).

$$f_{i,t} = a_i + b_i \mathrm{M}_{i,t} \quad (1)$$

where $f_{i,t}$ and $\mathrm{M}_{i,t}$ are the *i*th bias-corrected value and original model forecast respectively. a_i and b_i are the linear regression coefficients of *i*th model forecast.

Step 2: Transform data space. Another requirement of BMA application is that the bias-corrected values ($f_{i,t}$) should be converted to special datasets with a Gaussian space. Box-Cox transformation proposed by Box and Cox (1964) was used to conduct data transformation and was described as below.

$$f_{i,t}^{\lambda} = \begin{cases} \frac{f_{i,t}-1}{\lambda} & \lambda \neq 0 \\ ln\left(f_{i,t}\right) & \lambda = 0 \end{cases} \quad (2)$$

where $f_{i,t}^{\lambda}$ and λ are the bias-corrected value of *i*th model at the *t* time and Box-Cox coefficient respectively. In this study, the artificial covariate method (Dag et al., 2013) was employed to determine the optimal value of Box-Cox coefficient while the K-S

test statistic (Lilliefors, 1967) was employed to prove the Gaussianity of the transformed data. After implementation of the data transformation, the posterior distribution $p(y_t|f_{i,t}^{\lambda}, Y)$ would follow a Gaussian distribution $p(y_t|f_{i,t}^{\lambda}, Y) \sim g(y_t|f_{i,t}^{\lambda}, \delta_i^2)$.

Step 3: Estimate parameters. A log-likelihood function was adopted to estimate the parameters of weight (ω_i) and variance (δ_i^2) and was formulated as follows.

$$l_{\varphi} = log\left(\sum_{i=1}^{k} \omega_i p(y_t|f_{i,t}^{\lambda}, Y)\right) \quad (3)$$

where $\boldsymbol{\varphi}$ is the vector of parameters $\{\omega_i, \delta_i^2, i = 1,2,\cdots,k\}$.

The Expectation-Maximization (EM) suggested by Raftery et al. (2005) was utilized to search the optimal parameters of weight (ω_i) and variance (δ_i^2) when a termination criterion (early stopping or the maximal iteration) was achieved. As the EM algorithm proceeds, the parameters of weight (ω_i) and variance (δ_i^2) were updated as follows.

$$\omega_i(j) = \frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} v_i^t(j) \quad (4a)$$

$$\delta_i^2(j) = \frac{\sum_{t=1}^{\mathrm{T}} v_i^t(j)\left(y_t - f_{i,t}^{\lambda}\right)^2}{\sum_{t=1}^{\mathrm{T}} v_i^t(j)} \quad (4b)$$

$$v_i^t(j) = \frac{\omega_i(j-1) \cdot g\left(y_t|f_{i,t}^{\lambda}, \delta_i^2(j-1)\right).}{\sum_{i=1}^{k} \omega_i(j-1) \cdot g\left(y_t|f_{i,t}^{\lambda}, \delta_i^2(j-1)\right)} \quad (4c)$$

$$l_{\varphi}(j) = log\left(\sum_{i=1}^{k} \omega_i(j) \sum_{t=1}^{\mathrm{T}} g\left(y_t|f_{i,t}^{\lambda}, \delta_i^2(j)\right)\right) \quad (4d)$$

where T is the number of the training datasets and $v_i^t(j)$ is the latent variable for the *i*th model at the *t* time in the *j*th iteration.

Step 4: Create BMA ensemble forecasts. After the parameters of weight (ω_i) and variance (δ_i^2) were estimated, we used the Monte Carlo simulation method to generate

BMA ensemble forecasts (Raftery, 2005; Zhou et al., 2016). The procedure was described as follows.

a) Generate an integer value of i in [1, 2, …, k] by using the corresponding probabilities [ω_1, ω_2, …, ω_k]. Set the initial cumulative weight $\omega_0^* = 0$ and calculate cumulative weight $\omega_i^* = \omega_{i-1}^* + \omega_i$ for i = 1, 2, …, k. Create a random variable u between 0 and 1. If $\omega_{i-1}^* \leq u \leq \omega_i^*$, it indicates that the ith model forecast would be selected and used in the next step.

b) Generate a realization of the observation y_t using the PDF $g\left(y_t \middle| f_{i,t}^{\lambda}, \delta_i^2\right)$.

c) Repeat the above two steps (a) & b)) for K times. K is the number of Monte Carlo simulation and set as 1000 in this study. At last, data conversion is needed to convert the ensemble forecasts from a Gaussian space to their original space. Furthermore, 90 % confidence intervals between the 5 % and 95 % quantities were employed to reveal the uncertainty of BMA ensemble forecasts.

Appendix B

General implementation procedure of CBMA approach

Step 1: Configure the marginal distributions of the forecast variable of each model ($M_{i,t}$) and realization of observation (y_t) respectively. Let $U_M = P_M(M_{i,t})$ and $U_y = P_y(y_t)$ are the Cumulative Distribution Functions (CDFs) of the forecast variable of each model ($M_{i,t}$) and realization of observation (y_t) respectively. It would specify and determine the marginal distribution $P_M(M_{i,t})$ of the forecast variable of each model ($M_{i,t}$) and the marginal distribution $P_y(y_t)$ of the realization of

observation (y_t) to construct posterior probability in the next step. Seven different probability distributions, including Gaussian, Gamma, Gumbel, Pearson type III, Generalized Extreme Value (GEV) and Log-Weibull were tried in this study and summarized in Table 1.

Table 1 Candidate univariate distributions adopted for fitting the marginal distributions

Distribution	Probability distribution function (pdf)	Range	Parameters
Gaussian	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	$-\infty<x<+\infty$	μ σ
Gamma	$f(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x)$	$x>0$	α β
Gumbel	$f(x)=\alpha\exp\left[-\alpha(x-\mu)-e^{-\alpha(x-\mu)}\right]$	$-\infty<x<+\infty$	α μ
GEV	$f(x)=\frac{1}{\sigma}\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{(-1/\gamma)-1}\cdot\exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\}$	$-\infty<x<+\infty$	μ σ γ
Person type III	$f(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}(x-\mu)^{\alpha-1}\exp[-\beta(x-\mu)]$	$x>\mu$	α β μ
Log-Weibull	$f(x)=\frac{1}{\alpha(x-\mu+1)}\left[\frac{\ln(x-\mu+1)}{\alpha}\right]^{\beta-1}\cdot\exp\left\{-\left[\frac{\ln(x-\mu+1)}{\alpha}\right]^{\beta}\right\}$	$x>\mu$	α β μ

Owing to the wide practicality of the L-moments method (Hosking, 1990; Zhou et al., 2014) and the K-S statistic test method (Lilliefors, 1967; Razali and Wah, 2011), these two methods were used to estimate the distribution parameters and find the best marginal distribution respectively. The 5% significance level was applied to deciding whether a fitted distribution was acceptable or not, and then the probability distribution that possessed the minimum K-S test statistic indicator D (i.e. the maximum difference between the values of the empirical and the expected cumulative distributions) value was recommended as the best fitted distribution.

Step 2: Apply Copula function to constructing the posterior probability of forecast variable of each model $p(y_t|\mathrm{M}_{i,t},Y)$. Various family members of Copula functions

have been introduced by Nelsen (2006). The Archimedean Copula functions have impressive practicality in hydrologic and meteorological research domains because it is easy to construct (e.g. Chen and Guo, 2019; Zhang and Singh, 2019).

Table 2 Candidate bivariate Archimedean copula functions

Copula function	Joint distribution function	Parameter
Gumbel-Hougaard	$C(u_1,u_2\|\theta)=\exp\left\{-\left[(-\ln u_1)^\theta+(-\ln u_2)^\theta\right]^{1/\theta}\right\}$	$*\tau=1-\frac{1}{\theta}$ $\theta\geq 1$
Frank	$C(u_1,u_2\|\theta)=-\frac{1}{\theta}\ln\left[1+\frac{[\exp(-\theta u_1)-1][\exp(-\theta u_2)-1]}{\exp(-\theta)-1}\right]$	$\tau=1+\frac{4}{\theta}\left[\frac{1}{\theta}\int_0^\theta \frac{t}{\exp(t)}d_t-1\right]$ $-\infty<\theta<+\infty$
Clayton	$C(u_1,u_2\|\theta)=\left({u_1}^{-\theta}+{u_2}^{-\theta}-1\right)^{-1/\theta}$	$\tau=\frac{\theta}{2+\theta}$ $\theta>0$

* τ is the Kendall's coefficient.

In this study, three Copula functions were tested, including Gumbel-Hougaard, Clayton and Frank from Archimedean Copula functions (Table 2). Then, the Kendall's coefficient and the K-S statistic test method were employed to estimate the parameter of Copula functions and choose the best Copula function. The copula function possessing the smallest K-S statistic indicator (D) at the 5% significance level would be selected as the most suitable one.

Step 3: Apply EM algorithm to estimating weight parameter (ω_i) of each model. After the posterior distribution is constructed, its weight parameter was estimated by using the EM algorithm by means of a few adjustments in the Eq. (4) of Appendix A.

$$\omega_i(j)=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} z_i^t(j) \tag{5a}$$

$$z_i^t(j)=\frac{\omega_i(j-1)\cdot p\left(u_{y_t}\middle|u_{\mathrm{M}_{i,t}}\right)}{\sum_{i=1}^{k}\omega_i(j-1)\cdot p\left(u_{y_t}\middle|u_{\mathrm{M}_{i,t}}\right)}=\frac{\omega_i(j-1)\cdot c_{\theta_i}\left(u_{\mathrm{M}_{i,t}},u_{y_t}\right)\cdot p_y(u_{y_t})}{\sum_{i=1}^{k}\omega_i(j-1)\cdot c_{\theta_i}\left(u_{\mathrm{M}_{i,t}},u_{y_t}\right)\cdot p_y(u_{y_t})} \tag{5b}$$

$$l_\varphi(j)=log\left(\sum_{i=1}^{k}\omega_i(j)\sum_{t=1}^{\mathrm{T}} c_{\theta_i}\left(u_{\mathrm{M}_{i,t}},u_{y_t}\right)\cdot p_y(u_{y_t})\right) \tag{5c}$$

where $z_i^t(j)$ is the latent variable for the *i*th model at the *t* time in the *j*th iteration based on the Copula conditional probability. As seen, the estimation of variance parameter (δ_i^2) has not occurred in Eq. (5). Furthermore, the posterior probability $p\left(u_{y_t}\middle|u_{\mathrm{M}_{i,t}}\right)$ is calculated only one time in Eq. (5) and that remains the same for all the iterations. While the posterior probability $g\left(y_t\middle|f_{i,t}^{\lambda},\delta_i^2(j-1)\right)$ in the BMA (Eq. (4) in Appendix A) should be computed and updated when the variance parameter (δ_i^2) changes.

Step 4: Apply the Monte Carlo simulation method for producing the realization of observation (y_t). a) Generate an integer value of *i* in [1, 2, …, k] by using the corresponding probabilities [ω_1, ω_2, …, ω_{k}]. Set the initial cumulative weight $\omega_0^*=0$ and calculate cumulative weight $\omega_i^*=\omega_{i-1}^*+\omega_i$ for *i* = 1, 2, …, k. Create a random variable *u* between 0 and 1. If $\omega_{i-1}^*\le u\le\omega_i^*$, it indicates that the *i*th model forecast would be selected and used in the next step. b) Generate a realization of observation y_t using the conditional PDF $c_{\theta_i}\left(u_{\mathrm{M}_{i,t}},u_{y_t}\right)\cdot p_y(u_{y_t})$. c) Repeat the above two steps (a) & b)) for K times. K is the number of Monte Carlo simulation and set as 1000 in this study. Similarly, 90 % confidence intervals were employed to reveal the uncertainty of CBMA ensemble forecasts.

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-), the Research Council of Norway (FRINATEK Project 274310) and the National Key

Research and Development Program of China (2018YFC0407904). The datasets provided by the Environmental Protection Administration, Taiwan, are acknowledged. The authors would like to thank the Editors and anonymous Reviewers for their constructive comments that greatly contributed 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.

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.

Bai, L., Wang, J., Ma, X., & Lu, H., 2018. Air pollution forecasts: An overview. International Int. J. Environ. Res. Public Health, 15(4), 780.

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.

Box, G. E. P., and Cox D. R., 1964, An analysis of transformations, J. R. Stat. Soc. Ser. B, 26(2), 211-252.

Breiman, L., & Friedman, J. H., 1997. Predicting multivariate responses in multiple linear regression. J. R. Stat. Soc., 59(1), 3-54.

Buckland, S. T., Burnham, K. P., & Augustin, N. H., 1997. Model selection: an integral part of inference. Biometrics, 603-618.

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., & Tsai, M. J., 2016. A nonlinear spatio-temporal lumping of radar rainfall for modeling multi-step-ahead inflow forecasts by data-driven techniques. J. Hydrol., 535, 256-269.

Chen L, & Guo S., 2019. Copulas and its application in hydrology and water resources. Springer Verlag, Singapore.

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.

Dag, O., Asar O., & Ilk O., 2013. A methodology to implement Box-Cox transformation when no covariate is available, Commun. Stat. Simul. Comput., 43(7), 1740–1759.

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. Environmental monitoring and assessment, 187(7), 477.

Fanizza, C., De Berardis, B., Ietto, F., Soggiu, M. E., Schirò, R., Inglessis, M., & Incoronato, F., 2018. Analysis of major pollutants and physico-chemical characteristics of PM2.5 at an urban site in Rome. Sci. Total Environ., 616, 1457-1468.

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.

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.

Gneiting T, & Raftery AE., 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. J. Am. Stat. Assoc., 102(477):359–378.

Gneiting, T., 2008. Probabilistic forecasting. Journal of the Royal Statistical Society: Series A (Statistics in Society), 171(2), 319-321.

Granger, C. W., & Ramanathan, R., 1984. Improved methods of combining forecasts. J. Forecasting, 3(2), 197-204.

Ghazi, S., & Khadir, M. T., 2009. Recurrent Neural Network for Multi-steps ahead prediction of PM_{10} concentration. Autom. Syst. Eng., 3(2), 13-21.

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.

Han, X., Chen, N., Yan, J., Liu, J., Liu, M., & Karellas, S., 2019. Thermodynamic analysis and life cycle assessment of supercritical pulverized coal-fired power plant integrated with No.0 feedwater pre-heater under partial loads. J. Clean. Prod., 233, 1106-1122.

Herr, H. D., & Krzysztofowicz, R., 2015. Ensemble Bayesian forecasting system Part I: Theory and algorithms. J. Hydrol., 524, 789-802.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T., 1999. Bayesian model averaging: a tutorial. Stat. Sci., 382-401.

Hosking, J.R.M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. J. R. Stat. Soc., 52 (1), 105-124.

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.

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.

Kaminska, J., 2018. Probabilistic forecasting of nitrogen dioxide concentrations at an urban road intersection. Sustainability, 10(11), 4213.

Khajehei, S., & Moradkhani, H., 2017. Towards an improved ensemble precipitation forecast: A probabilistic post-processing approach. J. Hydrol., 546, 476-489.

Koutsoyiannis, D., & Montanari, A., 2015. Negligent killing of scientific concepts: the stationarity case. Hydrol. Sci. J., 60 (7-8), 1174-1183.

Krapu, C., & Borsuk, M., 2019. Probabilistic programming: A review for environmental modellers. Environ. Model. Softw., 114, 40-48.

Leontaritis, I. J., & Billings, S. A., 1985. Input-output parametric models for non-linear systems. Part I: Deterministic non-linear systems, Part II: stochastic non-linear systems. Int. J. Control, 41(2): 323-344.

Leslie, L. M., & Holland, G. J., 1991. Predicting regional forecast skill using single and ensemble forecast techniques. Mon. Weather Rev., 119(2), 425-435.

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, L., Xu, C. Y., & Engeland, K., 2013. Development and comparison in uncertainty assessment based Bayesian modularization method in hydrological modeling. J. Hydrol., 486, 384-394.

Lilliefors, H. W., 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown, J. Am. Stat. Assoc., 62, 399-402.

Lin, B., & Zhu, J., 2018. Changes in urban air quality during urbanization in China. J. Clean. Prod., 188, 312-321.

Liu, H., Duan, Z., & Chen, C., 2019. A hybrid framework for forecasting PM2.5 concentrations using multi-step deterministic and probabilistic strategy. Air Qual., Atmos. Health, 1-11.

Liu, Z., Guo, S., Xiong, L., & Xu, C. Y., 2018. Hydrological uncertainty processor based on a copula function. Hydrol. Sci. J., 63(1), 74-86.

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.

Lyu, W., Li, Y., Guan, D., Zhao, H., Zhang, Q., & Liu, Z., 2016. Driving forces of Chinese primary air pollution emissions: an index decomposition analysis. J. Clean. Prod., 133, 136-144.

Maidment, D. R., 1993. Handbook of hydrology. New York: McGraw-Hill.

Madadgar, S., and Moradkhani, H., 2014, Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging, Water Resour. Res., 50, 9586-9603.

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.

Mok, K. M., Yuen, K. V., Hoi, K. I., Chao, K. M., & Lopes, D., 2018. Predicting ground-level ozone concentrations by adaptive Bayesian model averaging of statistical seasonal models. Stoc. Environ. Res. Risk. Assess., 32(5), 1283-1297.

Nelsen, R.B., 2006. An Introduction to Copulas. Second. New York: Springer.

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.

Pannullo, F., Lee, D., Waclawski, E., & Leyland, A. H., 2016. How robust are the estimated effects of air pollution on health? Accounting for model uncertainty using Bayesian model averaging. Spat. Spatio-Temporal. Epidemiol., 18, 53-62.

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.

Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon. Weather Rev., 133(5), 1155-1174.

Razali, N. M., & Wah, Y. B., 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. J. Statis. Model. Anal., 2(1), 21-33.

Ryan, W. F., 2016. The air quality forecast rote: Recent changes and future challenges. J. Air Waste Manage. Assoc., 66(6), 576-596.

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F. J., et al., 2018. HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community. Hydrol. Earth Syst. Sci., 22(11), 5639-5656.

Sun, Y., Du, W., Fu, P., Wang, Q., Li, J., Ge, X., & Zhao, J., 2016. Primary and secondary aerosols in Beijing in winter: sources, variations and processes. Atmos. Chem. Phys., 16(13), 8309-8329.

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.

Thielen-del Pozo, J., & Bruen, M., 2019. Overview of Forecast Communication and Use of Ensemble Hydrometeorological Forecasts. Handbook of Hydrometeorological Ensemble Forecasting, 1037-1045.

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 PM10 and PM2.5 concentrations using artificial neural networks, in Thessaloniki and Helsinki. Sci. Total Environ., 409(7), 1266-1276.

Weber, S. A., Insaf, T. Z., Hall, E. S., Talbot, T. O., & Huff, A. K., 2016. Assessing the impact of fine particulate matter (PM2.5) on respiratory-cardiovascular chronic diseases in the New York City Metropolitan area using Hierarchical Bayesian Model estimates. Environ. Res., 151, 399-409.

Wu, J., Zheng, H., Zhe, F., Xie, W., & Song, J., 2018. 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.

Xu, S., An, X., Qiao, X., Zhu, L., & Li, L., 2013. Multi-output least-squares support vector regression machines. Pattern. Recognit. Lett., 34(9), 1078-1084.

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., & 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 L, & Singh V P., 2019. Copulas and their applications in water resources engineering. Cambridge University Press.

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., & Guo, S., 2014. Risk analysis for flood control operation of seasonal flood-limited water level incorporating inflow forecasting error. Hydrol. Sci. J., 59(5), 1006-1019.

Zhou, Y., Guo, S., Xu, C. Y., Chen, H., Guo, J., & Lin, K., 2016. Probabilistic prediction in ungauged basins (PUB) based on regional parameter estimation and Bayesian model averaging. Hydrol. Res., 47(6), 1087-1103.

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.