Wordcount: 6579 words

Seamless integration of convolutional and back-propagation neural networks for regional multi-step-ahead $PM_{2.5}$ forecasting

Pu-Yun Kow[a], Yi-Shin Wang[a], Yanlai Zhou[a,b], I-Feng Kao[a], Maikel Issermann[a], 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

Correspondence to: Fi-John Chang (changfj@ntu.edu.tw); Li-Chiu Chang (changlc@mail.tku.edu.tw)

Abstract

The fine particulate matter (e.g. $PM_{2.5}$) gains an increasing concern of human health deterioration. Modelling $PM_{2.5}$ concentrations remains a substantial challenge due to the limited understanding of the dynamic processes as well as uncertainties residing in the emission data and their projections. This study proposed a hybrid model (CNN-BP) engaging a Convolutional Neural Network (CNN) and a Back Propagation Neural Network (BPNN) to make accurate $PM_{2.5}$ forecasts for multiple stations at multiple horizons at the same time. The hourly datasets of six air quality and two meteorological factors collected from 73 air quality monitoring stations in Taiwan during 2017 formed the case study. A total of 639,480 hourly datasets were collected and allocated into training (409,238, 64%), validation (102,346, 16%), and testing

(127,896, 20%) stages. The forecasts of $PM_{2.5}$ concentrations were first characterized as a function of air quality and meteorological variables. Then the proposed CNN-BP approach effectively learned the dominant features of input data and simultaneously produced accurate regional multi-step-ahead $PM_{2.5}$ forecasts (73 stations; t+1−t+10). The results demonstrate that the proposed CNN-BP model is remarkably superior to the BPNN, the random forest and the long short term memory neural network models owing to its higher forecast accuracy and excellence in creating reliable regional multi-step-ahead $PM_{2.5}$ forecasts. Besides, the CNN-BP model not only has the power to cope with the curse of dimensionality by adequately handling heterogeneous inputs with relatively large time-lags but also has the capability to explore different $PM_{2.5}$ mechanisms (local emission and transboundary transmission) for the five regions (R1-R5) and the whole Taiwan. This study shows that multi-site (regional) and multi-horizon forecasting can be achieved by exactly one model (i.e. the proposed CNN-BP model), hitting a new milestone. Therefore, the CNN-BP model can facilitate real-time $PM_{2.5}$ forecast service and the forecasts can be made publicly available online.

Keywords: $PM_{2.5}$ forecast; Deep learning; Convolutional neural network; Back Propagation neural network; Multi-step-ahead forecasts; Taiwan

1. Introduction

Air quality deteriorations have attracted intensive public attention for decades, and fine aerosols (e.g. $PM_{2.5}$) in suspended particulates are one of the critical indicators of health hazards and air pollution. Air pollutants with particle sizes smaller than 2.5 microns are difficult to control. Besides, the composition of fine particles is too complex to be blocked by the cilia in the respiratory tract, and therefore they are labeled as "pulmonary particulate matter" (Kong et al., 2017; Nurkiewicz et al., 2011; Tang et al., 2017; Yang et al., 2018; Zhou et al., 2018). Once being inhaled, it will reach the lungs, invade the alveoli and enter into the blood vessels, causing serious harms to human health (Lai et al., 2019; Li et al., 2017; Qiu et al., 2013; Tsai and Kuo, 2005). In recent years, air pollution caused by industrial development and transportation intensity upon rapid urbanization has become a severe issue in Taiwan. Besides, in winter a large number of aerosols are entrained in the northeast monsoon over the West Pacific Ocean (Hsu et al., 2006), coupled with a gradual expansion of long-range transboundary air pollution (Chan et al., 2006; Du et al., 2010; Hsiao et al., 2017; Hsu et al., 2016; Widiana et al., 2019). It is observed more and more people in Taiwan are substantially affected by air pollution. According to the statistics released by the Environmental Protection Administration in Taiwan (TW EPA), the primary sources of air pollution in Taiwan are building construction (37%), traffic pollution

(23%), industrial emissions (23%) and others (17%) (EPM, 2015). This projects that $PM_{2.5}$ is a multi-sources pollutant in relation mainly to industrial and automobile emissions from physical and chemical processes (Li et al., 2019). Therefore, many efforts have been made to forecast $PM_{2.5}$ concentrations (Cheng et al., 2019; Fernando et al., 2012; Loy-Benitez et al., 2019); nevertheless, challenges have arisen in the course of regional multi-step-ahead forecasting when facing high spatio-temporal variability in $PM_{2.5}$ concentrations. This creates a thirst for in-depth research on modelling approaches needed for regional multi-step-ahead $PM_{2.5}$ forecasting.

Modelling is an important tool for understanding the linkages between emissions and observations as well as for predicting ambient concentrations under a self-consistent framework. For instance, air quality forecasting is considered critical to early warning and control management of air pollution. Air quality forecast models can be broadly classified into physically-based models and machine learning models. Physically-based models have received extensive attention over the last decades, while notorious complexity and high uncertainty raised in modelling $PM_{2.5}$ has made their development full of thorns and challenges (Karambelas et al., 2018). Machine learning models such as the most commonly used Artificial Neural Networks (ANNs) have served to effectively characterize $PM_{2.5}$ as a function of its affecting factors for rapidly depicting the interdependence between air quality and meteorological systems, and thereby have been considered as a better choice for air quality forecasting (Cheng et al.,2019; Feng et al., 2015; Feng et al., 2019; Fernando et al., 2012; Gao et al., 2018; Loy-Benitez et al., 2019; Ma et al.,2019; Ma et al., 2020; Mihăiţă et al., 2019; Wang et al., 2020). A variety of machine learning techniques have been used to predict $PM_{2.5}$ concentrations, such as the backpropagation neural network (BPNN) (Elbayoumi et al., 2015), the neuro-fuzzy neural network (Ausati and Amanollahi, 2016; Mishra et al., 2015), the long short term memory neural network (LSTM) (Bai et al., 2019; Zhao et al., 2019; Zhou et al., 2019a), the random forest (RF) (Liu et al., 2018; Stafoggia et al., 2019), and the support vector machine (SVM) (Zhou et al., 2019b). Hybridization approaches integrating different machine learning techniques have also been explored in recent years to improve $PM_{2.5}$ prediction reliability and accuracy, with satisfactory forecast results (e.g. Du et al., 2018; Huang et al., 2018; Jiang, et al., 2017; Mahajan, et al., 2018; Mishra, et al., 2015; Niu, et al., 2016).

It is noted that the methods mentioned above have been usually adopted to construct site-specific data-driven models for individual air quality monitoring station. High spatio-temporal variability in $PM_{2.5}$ concentrations also occurs at plenty of monitoring stations spreading over a large region. These issues inevitably create great challenges in regional multi-step-ahead $PM_{2.5}$ forecasting. Bearing this in mind as a motivation, this study intends to develop a novel hybrid deep learning model for multiple site/horizon $PM_{2.5}$ forecasting, with missions to extract the spatio-temporal correlation features and interdependence of multivariate air quality-related and meteorological time series data, explore $PM_{2.5}$ mechanisms (local emission & transboundary transmission), and make $PM_{2.5}$ forecasts for multiple sites at multiple horizons simultaneously.

To achieve these goals, we propose a hybridization approach (CNN-BP) that seamlessly integrates a Convolutional Neural Network (CNN) and a BPNN in the interest of improving the reliability and accuracy of regional multi-step-ahead $PM_{2.5}$ forecasts, where. One of the study goals is to extend the prediction interval is exended from one hour up to ten hours. Four machine learning models (i.e. CNN-BP, BPNN, RF, and LSTM) are independently constructed for creating regional multi-step-ahead $PM_{2.5}$ forecasts based on hourly observed data collected at 73 air quality monitoring stations spreading over the whole Taiwan, where the two static (BPNN and RF) and one dynamic (LSTM) models are taken as benchmarks for the purpose of comparison. The proposed CNN-BP forecast model (Figure 1) is a meta model enabled to predict multiple site/horizon attributes at once (i.e. 730 forecasts (73 stations x 10 horizons) each time), and the real-time regional multi-step-ahead $PM_{2.5}$ forecasts can be visualized in a 2D map using the Kriging method. Following the Intodruction Section, this study is organized to outline the study area and materials in Section 2, introduce the methods in Section 3, show and discuss the multi-step-ahead $PM_{2.5}$ forecast results in Section 4, and make concluding remarks in Section 5.

2. Study area and materials

2.1. Study area

The fast booming economy and high population density of Taiwan has made air quality deterioration rank high on the hot topic list in recent years. Air pollution not only induces respiratory diseases but is also a matter of life and death. Therefore, it is imperative to make accurate and reliable $PM_{2.5}$ forecasts for assisting in the reduction of the health risk associated with air pollution. Air quality monitoring stations in Taiwan constitute the case study, and the study area is partitioned into five regions according to geographic locations, i.e. R1−northern region, R2−central region, R3−southern region, R4−eastern region, and R5−surrounding islands (Figure 2). Four machine learning models are separately constructed to produce regional multi-step-ahead $PM_{2.5}$ forecasts.

2.2. Data collection and statistical analysis

The TW EPA provides an open data platform accessible to the public, where environmental monitoring datasets such as local air quality and meteorological conditions are on demand (EPA, 2019). This highly facilitates the collection of reliable data for research use. It is noted that there are a total of 76 ground-based air quality monitoring stations in Taiwan but this study adopts data only from 73 stations because 3 stations encountered the problem of massive missing data. The attribiutes of the five regions defined in this study are give as follows. R1 (26 stations) centers on economic activities with heavy traffic loads and intensive commercial trading. R2 (17 stations) and R3 (23 stations) encompass industrial areas locating chemical and thermal power plants. R4 (4 stations) embraces high-elevation mountains with natural scenery famous for tourism. R5 (3 stations) contains three groups of small islands surrounding Taiwan. This study utilizes hourly data of six air quality factors ($PM_{2.5}$, PM_{10}, SO_2, CO, NO_2, and O_3) and two meteorological factors (ambient temperature and relative humidity) collected between 1/1/2017 and 31/12/2017. A total of 639,480 hourly datasets (=24 hours x 365 days x 73 stations) were collected, where 409,238 datasets (64%) were for model training, and the remaining 102,346 datasets (16%) and 127,896 datasets (20%) were for model validating and testing, respectively. The prediction interval is set as one hour, in accordance with the data collection interval of the 73 stations.

Table 1 presents the results of the statistical analyses on air quality and meteorological data for use in this study. The results indicate that high NO_2, SO_2 and CO concentrations occur in R1 (northern region), providing evidence that R1 suffers mainly from vehicle exhaust emissions. R2 (central region) and R3 (southern region) have the highest mean values of $PM_{2.5}$ and PM_{10} concentrations due to their thermal power plants. R2, R3, and R5 (surrounding islands) have relatively large standard deviations of $PM_{2.5}$ concentrations. It is conceivable that long-term prediction of $PM_{2.5}$ concentrations at the three regions should be very challenging because their high variations in concentrations would create a barrier to capturing the future trends of $PM_{2.5}$ concentrations. As for R4 (eastern region), its mean and standard deviation of $PM_{2.5}$ concentrations are the lowest on account of few industrial and commercial activities here. It is worth noting that R5 (surrounding islands) is less industrialized but has relatively high concentrations in $PM_{2.5}$, PM_{10} and O_3, which suggest that air quality in R5 could be largely affected by transboundary transmissions (Yuan et al., 2004). As known, the deposition process of $PM_{2.5}$ is highly correlated with relative humidity because moisture adheres to fine particles and accumulates to a larger size (Hernandez et al., 2017; Hien et al., 2002; Lou et al., 2017; Tai et al., 2010). The high relative humidity over the whole Taiwan, with an annual average exceeding 70%, implies that $PM_{2.5}$ in Taiwan is profoundly affected by relative humidity (Hien et al., 2002; Lou et al., 2017).

3. Methodology

3.1. Problems and motivations

Accurate air quality forecasting at longer lead times is the key to early warning and management of air pollution. Our goal is to anticipate changes in $PM_{2.5}$ concentrations at monitoring stations over time. Air quality forecasting becomes highly challenging under the rapidly changing conditions of weather and pollutant emission, in addition to the influence imposed by plenty of nonlinear and dynamic factors. Therefore, it is difficult to precisely predict air quality for a region at a specific time. Artificial Intelligence (AI) has been significantly empowered to bridge the gap between the capabilities of humans and machines. The advancements in Computer Vision with Deep Learning has been constructed, primarily by the CNN. This study explores a hybridization approach (CNN-BP) driven by CNN and BPNN for producing reliable and accurate regional $PM_{2.5}$ forecasts at longer horizons. Three ANN models (static-BPNN and RF; and dynamic-LSTM) form the benchmarks for comparison purpose in this study.

Figure 3 shows the graphical illustration of the proposed air quality forecasting framework, which consists of three main components: the CNN for learning the spatial pattern of each sample in time series (Figure 3(a)); the BPNN for extracting the interdependency and temporal features from the corresponding time series data (Figure 3(b)); and the hybrid CNN-BP model for producing multi-site and multi-horizon $PM_{2.5}$ forecasts (Figure 3(c)). To be more precise, the proposed CNN-BP model implements a two-phase procedure engaging feature extraction from air quality and meteorological samples (CNN, configured with three one-dimensional convolutional layers) and forecasting (BPNN, configured with two fully connected hidden layers) for multi-site multi-horizon $PM_{2.5}$ forecasting. The methods involved are briefly introduced as follows.

3.2. Convolutional Neural Network (CNN)

The CNN configured with a deep learning algorithm is a type of ANNs. It has the merit to effectively differentiate one sample from the others owing to feature extraction, where each sample is assigned importance to gain various objects in the sample and to extract its high-level features/characteristics for differentiating itself from the other samples. The CNN typically has three layers: the convolutional layer that extracts features from the inputs to form a feature map matrix, the pooling layer that reduces the spatial size of the convolved feature, and the fully connected layer that flattens the output into one column vector and feed it into a feed-forward neural network. Therefore, the model can successfully capture the spatio-temporal dependencies in each sample and distinguish between dominating and certain low-level features in samples. The CNN has been widely used in natural language processing and image processing, and it has also been applied to time series forecasting recently (Borovykh et al., 2017; Li et a., 2017). The implementation of the CNN is briefly introduced below.

In this study, there are 73 stations and each station has 8,760 samples (24 hours x 365 days). Each sample allows 4 types of time lags (6-h, 12-h, 24-h and 36-h), and there are 8 input variables (6 air quality and 2 meteorological variables) at each time lag. The number of filters is set as 100, and the filtering process is conducted on each sample. It is noted that the CNN has a concept of "weight sharing", that is, a filter does not change its weight values when screening each sample during training and validation stages. This leads to lesser parameters required for the CNN during model construction than for other feed-forward ANNs. As a result, the CNN is easier to train, with an avoidance of overfitting, which makes the CNN an attractive deep learning algorithm. More details of the CNN can be found in Chen et al. (2019).

3.3. Back Propagation Neural Network (BPNN)

The BPNN is a fully connected neural network with three layers, i.e. one input layer, one hidden layer, and one output layer (Figure 3(b)). This static neural network enables two main actions: propagation (forward and backward) and weight adjustment. In the forward propagation, an input signal is assigned a weight by the activation function (i.e. the Rectified Linear Unit (ReLU) function in this study) in the hidden layer, and then the weighted signal is passed to the output layer for calculating the output value. After the forward propagation finishes, the backward propagation will be activated if the difference (error) between the output value and the target output value falls outside the tolerable error range. More detailes of the BPNN can be found in Hecht-Nielsen (1992).

3.4. Hybrid of CNN and BPNN (CNN-BP)

The proposed CNN-BP approach aims at simultaneously producing $PM_{2.5}$ forecasts of 73 monitoring stations at horizons t+1 up to t+10, and its two-phase implementation procedure engaging feature extraction by CNN and $PM_{2.5}$ forecasting by BPNN (Figure 3(c)). The CNN-BP model seamlessly connects a CNN (configured with three one-dimensional convolutional layers, determined by trial and error procedures) to a BPNN (configured with two fully connected hidden layers, determined by trial and error procedures).

The CNN is a powerful tool for feature extraction because each output of the CNN contains multiple time-attributes. The similarity in patterns among samples can be considered as an auxiliary to improve forecast accuracy; that is to say, learning of similar patterns greatly assists in forecasting, especially for multiple stations and multiple horizons. In brief, the CNN can effectively capture the spatial dependencies and distinguish between dominating and certain low-level features and classify them.

Following feature extraction, the flatten layer that links the feature map of the

CNN with the fully connected hidden layer of the BPNN has a mission to reshape each multi-dimensional input into a one-dimensional input (Jin et al., 2014). Then, the two fully connected hidden layers and the output layer in the BPNN constitute the forecasting phase of the CNN-BP model. It is worth noting that we set up ten neurons in the output layer for producing ten-dimensional outputs (i.e. horizons t+1 up to t+10 of $PM_{2.5}$ concentrations in this study).

The ReLU function is employeed as the activation function for the three ANN models in this study, and its formula is shown below.

$$f(x) = \begin{cases} x \text{ if } x > 0 \\ 0 \text{ if } x \leq 0 \end{cases} \tag{9}$$

where $f(x)$ is a linear function when x is greater than 0, otherwise 0. The ReLU function is very powerful and has several advantages: a) able to solve the problem of gradient disappearance or explosion; b) able to mimic the computational structure of the human brain; c) has a fast calculation speed; and d) easier to converge than the sigmoid activation function ((Glorot et al., 2011; Romero et al., 2015).

3.5. Random Forest (RF)

The RF evolved from the decision tree is an ensemble machine learning method and has two cores: "random" for random feature selection and "forest" for bagging (Ho, 1995). It is implemented by establishing a multitude of decision trees at the training stage and then outputting the class that is the status of the classes (classification) or mean prediction (regression) of the individual trees (Karginova et al., 2012; Liaw and Wiener, 2002). The RF has been widely used in the field of data mining and time series forecasting (Feng et al., 2019; Kamińska, 2018; Kumar, 2018). In this study, the random forest regressor that averaging the outputs of the individual trees is used.

The training procedure of the RF regressor is described as follows.

1) Draw a bootstrap sample from the original data.

2) For each bootstrap sample, grow a regression tree with the following modification at each node: choosing the best split-point predictors among the m predictors and picking the best variable associated with the best split.

3) Predict new data by aggregating the predictions of the *n* trees, where the final output is produced by aggregating and averaging the prediction results of all decision trees

More details of the RF can be found in Liaw and Wiener (2002).

3.6. Long Short Term Memory Neural Network (LSTM)

The LSTM is a well-known recurrent architecture in the deep learning field. It has two capabilities, namely long-term memory and short-term memory, owing to its internal self-looped cell that can pass the previous state to the next (Hochreiter, 1998). The LSTM unit is composed of six parts, including the input block, three gates (input, forget and output gates), the self-looped cell, and the output block. The training procedure of the LSTM is briefly described as follows.

(1) The input gate determines what information to produce and what information to add to the current cell state based on the output of the previous state and the input of the current state.

(2) The forget gate determines what information to remove from the current cell state. In other words, information considered unimportant can be forgotten. Otherwise, it will be “memorized” by the LSTM cell.

(3) The output gate determines the output state and the output of the LSTM cell.

More details of the LSTM can be found in Zhou et al. (2019).

3.7. Techniques to prevent overfitting

To avoid overfitting during model training, this study employees the nonlinear L2 regularization technique and the early stopping criterion. The nonlinear L2 regularization can discourages the learning of a model to avoid overfitting by adding penalty terms (Chang et al., 2010). Model training will terminate with early stopping when the model reaches the minimum validation loss (Prechelt, 1998), which means the network stops training if the validation error is no longer reducible after n iterations (n=15 in this study).

3.8. Evaluation indicators

We use three indicators to evaluate model performance, which are the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The formula of the three indicators are given below.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{L} |\mathrm{o_i} - p_i|}{L} \tag{10}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{L} (\mathrm{o_i} - p_i)^2}{L}} \tag{11}$$

$$\mathrm{R}^2 = \frac{L\sum_{i=1}^{L} o_i p_i - \sum_{i=1}^{L} o_i \sum_{i=1}^{L} p_i}{\sqrt{L\sum_{i=1}^{L} o_i^2 - \left(\sum_{i=1}^{L} o_i\right)^2}\sqrt{L\sum_{i=1}^{L} p_i^2 - \left(\sum_{i=1}^{L} p_i\right)^2}} \tag{12}$$

where $\mathrm{o_i}$ denotes observed data, p_i denotes forecasted values, and L denotes the data length.

The MAE is the average of the absolute difference between forecasted values and observed data, which measures the average magnitude of forecast errors. The RMSE is an error index favorable for the assessment on forecast accuracy of peak values due to significant magnification of forecast errors. A model with higher R^2 values but lower RMSE and MAE values performs better. In general, higher R^2 values may coincide with smaller RMSE and MAE values.

4. Results and discussion

In this study, the CNN-BP model is proposed to produce regional multi-step-ahead (t+1− t+10) $PM_{2.5}$ forecasts based on hourly data of six air quality and two metrological factors from 73 air quality monitoring stations in Taiwan. Two static ANN models (i.e. BPNN and RF) and one dynamic ANN model (i.e. LSTM) are employed for comparison purpose. The results and findings are presented in the order of the data preprocessing, model construction, and model comparison, shown as follows.

4.1. Data preprocessing

To overcome the different scales of heterogeneous data and/or over-fitting encountered in model training, data normalization is the first step of data pre-processing. The goals of normalization are to adjust the values of variables in datasets to a common scale and to make sure different features take on similar ranges of values so that gradient descents can converge effectively. The second step of data pre-processing is to randomly allocate the normalized samples into training (64%), validation (16%) and testing (20%) datasets.

4.2. Model construction

4.2.1. Parameter setting

A large number of trial and error procedures are executed to identify parameters the most suitable for each ANN model based on the training and validation datasets, and the obtained parameter settings of the ANN models are presented in Table 2.

4.2.2. Input factor selection

For the determination of the best input combination, two input scenarios

(Scenario 1 & Scenario 2) are designed to assess the impacts of air quality and/or meteorological factors on $PM_{2.5}$ forecasts. Scenario 1 considers six air quality factors ($PM_{2.5}$, PM_{10}, SO_2, CO, NO_2 and O_3) as model inputs because these six factors are the components of the Air Quality Index (AQI) defined by the TW EPA (Hao and Liu, 2016; Moisan et al., 2018; Zhang et al., 2018). Scenario 2 considers $PM_{2.5}$ forecasting as a function of the same air quality factors and two meteorological factors (i.e. ambient temperature and relative humidity).

Figure 4 shows the forecast performance of CNN-BP, BPNN and RF models at horizons t+6 and t+10 for the whole of Taiwan under Scenarios 1 and 2. The results reveal that the three ANN models perform better (higher R^2 and lower RMSE values) under Scenario 2 than under Scenario 1, especially obvious for the CNN-BP model. This gives a useful hint that high correlation is implicitly expressed between $PM_{2.5}$ concentrations and the two meteorological factors. Previous studies provided the following findings. When there is high relative humidity in the air, the deposition process of $PM_{2.5}$ will occurs (Li et al., 2017). In summer (high temperature), temperature inversion may occur so that a layer of cool air at the surface would be overlaid by a layer of warmer air due to difference in air density; while in winter (low temperature), a slower inversion of temperature may occur due to similar and lower air densities in upper atmospheres (Wallace and Kanaroglou, 2009). This points out that temperature is another important factor affecting $PM_{2.5}$ concentrations. Considering our results shown in Figure 4 and the findings from the previous studies, the air quality and meteorological factors of Scenario 2 are determined to form the input combination for carrying out regional multi-step-ahead $PM_{2.5}$ forecasting in this study. The next step is to identify the time-lag of the input variables needed for multi-step-ahead forecasting.

To fully investigate the time-lag effect, several historical temporal patterns (i.e. 6-h, 12-h, 24-h, and 36-h) of all the eight input variables are incorporated into model training and testing stages of the three ANN models. Taking the CNN-BP model as an example, Table 3 shows the model performance in training/validation and testing stages at horizons t+6 and t+10 based on the inputs with different time-lags for the whole of Taiwan. The results clearly explain that the model based on inputs with a 24-h time-lag patterns would produce the best performance (the highest R^2 values and the smallest RMSE values) in both training/validation and testing stages. When the time-lag increases from 6-h to 24-h, there is a significant improvement in model performance over time. When the time-lag extends from 24-h to 36-h, the model performance, however, deteriorates in both training and testing stages. Such phenomenon might be because uncertainty keeps increasing and more noise information involves as the time-lag exceeds 24-h (a day), which prohibits the model from being well trained and well validated. Consequently, the eight inputs with 24-h time-lag are determined as the final input combination of the three ANN models for carrying out the following analyses.

4.3. Comparison of ANN models for $PM_{2.5}$ forecasts

4.3.1. Regional multi-step-ahead $PM_{2.5}$ forecasts

Figure 5 gives the comparison of regional multi-step-ahead $PM_{2.5}$ forecasts (R1−R5, the whole Taiwan) obtained from three ANN models in the training and testing stages at horizons t+6 and t+10. The comparative results demonstrate that the CNN-BP model can produce the most accurate $PM_{2.5}$ forecasts in terms of the smallest RMSE values in both training and testing stages at horizons t+6 and t+10. The reason is that the CNN-BP model can adequately handle inputs with relatively large time-lags to cope with the curse of dimensionality. In other words, this model can more effectively and deeply learn and extract useful information (knowledge) from high-dimensional datasets (input-output patterns), as compared with BPNN and RF models.

We next take the regional CNN-BP model as an example for further evaluation. In the training stages, it is noticed from Figure 5(a) that the southern region (R3) has the largest RMSE values (5.38 μg/m^3 at t+6, and 5.98 μg/m^3 at t+10), followed by the central region (R2, 4.99 μg/m^3 at t+6, and 5.63 μg/m^3 at t+10), while the lowest

RMSE values (3.74 μg/m^3 at t+6, and 4.22 μg/m^3 at t+10) occur in the eastern region (R4). In the testing stages, Figure 5(b) indicates that the southern region (R3) gives the largest RMSE values (6.77 μg/m^3 at t+6, and 7.67 μg/m^3 at t+10), followed by the central region (R2, 6.32 μg/m^3 at t+6, and 7.29 μg/m^3 at t+10), while the lowest RMSE values (4.49 μg/m^3 at t+6, and 4.95 μg/m^3 at t+10) occur in the eastern region (R4). It appears that the constructed CNN-BP model (one model) could be well trained (very small RMSE values) and could make reliable and accurate multi-step-ahead (t+1 up to t+10) $PM_{2.5}$ forecasts in the testing stages for all the 73 stations. That means multi-site (regional) and multi-horizon forecasting can be achieved by exactly one model (i.e. the proposed CNN-BP model), hitting a new milestone.

From the perspective of pollution sources, thermal power plants would be the main source of $PM_{2.5}$ emission in the two most polluted regions, i.e. R2 and R3. Both regions suffer larger RMSE values because changes in $PM_{2.5}$ concentration are more dramatic and irregular here (Figure 5, Table 1). R1 and R5 have moderate $PM_{2.5}$ concentrations and RMSE values. The primary source of $PM_{2.5}$ emission in R1 (the economic center of Taiwan) would be a great number of moving vehicles, resulting in high traffic loads. It is noticed that R5 (the surrounding islands) does not produce the smallest RMSE values among the five regions (Figure 5) even though this region has neither large industrial facilities nor many vehicles. In consideration of the geographical locations of these surrounding islands, we speculate that the primary source of $PM_{2.5}$ emissions in R5 would be transboundary transmissions, especially the monsoon blows from China. As for R4, $PM_{2.5}$ concentration is relatively low here and its RMSE values are also the smallest in both stages (Table 1, Figure 5). We also notice that there is no major industry in R5 and $PM_{2.5}$ concentration does not change much throughout the investigative period. Here are two interesting findings. First, the sources of $PM_{2.5}$ emission in R1−R4 are associated primarily with local emissions from industrial and/or human activities, which reveals $PM_{2.5}$ concentrations of these four regions are closely related to the six air quality factors ($PM_{2.5}$, PM_{10}, SO_2, CO, NO_2 and O_3). Second, transboundary transmission would be the source of $PM_{2.5}$ emission in R5 (surrounding islands), which implies $PM_{2.5}$ concentrations here are closely related to meteorological factors (ambient temperature and relative humidity).

Furthermore, the CNN-BP model is compared with the LSTM model, a dynamic model that preserves the previous state in forecasting. The parameter setting of the LSTM model is shown in Table 2. Figure 6 shows the forecast performance of CNN-BP and LSTM models at horizons t+6 and t+10 for the whole of Taiwan. The forecast results indicate that the CNN-BP model is significantly superior (higher R^2 and lower RMSE values) to the LSTM model. The main reason could be that the

LSTM model capable of preserving the previous state of a time series (single station) encountered the over-fitting problem, while the samples of the large region with multiple stations (73 stations) we investigated were time-discontinuous among various stations, which led to poor forecast accuracy.

In brief, the results demonstrate that the CNN-BP model not only performs better than the BPNN, RF and LSTM models for multi-step-ahead $PM_{2.5}$ forecasts but is also able to model different $PM_{2.5}$ mechanisms (local emission and transboundary transmission) for the five regions (R1−R5) and the whole Taiwan. In other words, we extract data features from multiple stations to make multi-site multi-horizon forecasts using only a single CNN-BP model. Therefore, the model's applicability is largely increased. Moreover, forecast accuracy is significantly improved by learning more similar data features from samples of other stations, rather than just of a single station.

4.3.2. $PM_{2.5}$ forecasts at a station with high $PM_{2.5}$ concentrations

We further investigate the three ANN models for $PM_{2.5}$ forecasting at the Nantzu air quality monitoring station (see Nantzu Station in R3 of Figure 2) that suffers high $PM_{2.5}$ concentrations (maximum=94 μg/m^3, mean=46.81 μg/m^3, standard deviation=17.45μg/m^3). Figure 7 displays the comparative results of the three ANN models at horizon t+10 for this station regarding the errors between the observed and forecasted $PM_{2.5}$ concentrations in the testing stages spanning between 3/1/2017 and

18/1/2017 (24 hours x 16 days = 384 hours). The results show that the absolute errors of peaks exceed 50 μg/m^3 for the RF model and the BPNN model but is less than 30 μg/m^3 for the CNN-BP model (Figure 7). Moreover, it is easy to tell that the patterns (384 time series) of forecast errors created by all three models are similar and the absolute errors of the CNN-BP model are significantly smaller than those of RF and BPNN models. This supports that the CNN-BP model not only can efficiently handle heterogeneous data with large time-lags but also can effectively characterize the $PM_{2.5}$ trend and features of each sample using the filter in the CNN. This also explains why the CNN-BP model can catch the variation in $PM_{2.5}$ concentration more precisely than RF and BPNN models.

Furthermore, it is noticed from Table 4 that the CNN-BP model has the lowest MAE values in both training (9.46 μg/m^3) and testing (9.18 μg/m^3) stages, followed by the RF model (10.35 μg/m^3 in training, and 10.40 μg/m^3 in testing), then by the BPNN model (12.98 μg/m^3 in training, and 12.78 μg/m^3 in testing) at the Nantzu Station. The results clearly demonstrate that the CNN-BP model serves as a better predictor than the RF and the BPNN models for long-term (e.g. 10 hours) $PM_{2.5}$ forecasting.

4.3.3. $PM_{2.5}$ forecasts for the whole of Taiwan

We also investigate the reliability and accuracy of the constructed CNN-BP model with a recent snapshot of $PM_{2.5}$ concentration. Figure 8 presents the observations and the forecasts obtained from the CNN-BP model at horizons t+6 and t+10 for the whole Taiwan upon a snapshot at 2 am on 21th January in 2018, where the Kriging method is implemented to make a two-dimensional visualization of the observations and the forecasts through spatial interpolation. The color scale of Figure 8 refers to the Indicator Table announced by the TW EPA (https://www.hpa.gov.tw/Pages/ashx/File.ashx?FilePath=~/File/Attach/3007/File_3697.pdf). $PM_{2.5}$ concentration higher than 54.5 μg/m^3 is considered harmful to the human body (EPA, 2019). The results of Figure 8 show that the CNN-BP model, in general, can well forecast $PM_{2.5}$ concentrations at both t+6 and t+10. It appears that $PM_{2.5}$ concentrations are much higher in central (R2) and southern (R3) regions. According to Figures 8(b)-8(e), the model does suitably catch the variations of $PM_{2.5}$ concentrations at both t+6 and t+10 under the conditions of good and moderate $PM_{2.5}$ concentrations while slightly underestimating in certain areas of southern region (R3) under the condition of unhealthy $PM_{2.5}$ concentrations.

Figure 9 gives the results of the RMSE values between the observed and forecasted $PM_{2.5}$ concentrations associated with Figure 8. The southern region (R3) has the largest forecast errors (15.45 μg/m^3 at t+6, and 18.02 μg/m^3 at t+10), followed by the central region (R2, 9.97 μg/m^3 at t+6, and 7.70 μg/m^3 at t+10). Besides, the

RMSE values of the eastern region (R4) are 2.53 μg/m^3 and 2.3 μg/m^3 at t+6 and t+10, respectively, while the RMSE values of the surrounding islands (R5) are 4.70 μg/m^3 and 4.26 μg/m^3 at t+6 and t+10, respectively. The relatively low forecast errors in R4 and R5 would be a consequence that the CNN-BP model can easily catch the trends of $PM_{2.5}$ concentrations under conditions of low concentrations (Table 1). As for the whole Taiwan, the RMSE values are 9.36 μg/m^3 and 10.68 μg/m^3 at horizons t+6 and t+10, respectively. In sum, the results of the recent case (2 am on 21th January, 2018) support the generalizability and reliability of our proposed CNN-BP model.

5. Conclusion

Fine particulate matter (e.g. $PM_{2.5}$) is a complicated air pollutant because it involves a great variety of pollution sources. To model the nonlinear and dynamic multivariate time series of $PM_{2.5}$ concentrations, we propose a deep learning framework hybriding CNN and BPNN for sharing the features extracted from air quality- and meteorological-related time series data to make multi-site (73 stations) multi-horizon (one to ten hours) $PM_{2.5}$ forecasts concurrently. The main contributions of the proposed approach (CNN-BP) are three-fold. Firstly, the CNN-BP model can adequately characterize the $PM_{2.5}$ concentrations into a function of air quality and meteorological variables based on a large number of high-dimensional hourly observed datasets at various stations. Secondly, the CNN-BP model can combine the essential features of CNN and BPNN to significantly improve the forecast accuracy of $PM_{2.5}$ concentrations. Thirdly, the CNN-BP model can effectually produce $PM_{2.5}$ forecasts for multiple stations at multiple horizons simultaneously.

This study evaluated the proposed CNN-BP models with three types of machine learning models (static BPNN and RF, and dynamic LSTM). The results demonstrated that the CNN-BP model performed the best, in terms of the smallest RMSE and the highest R^2 values for the whole of Taiwan and the five regions (R1−R5). The accuracy and reliability of $PM_{2.5}$ forecasts increased significantly for the CNN-BP model. We also demonstrated that the CNN-BP model could more adequately handle heterogeneous inputs with relatively large time-lags to tackle the curse of dimensionality and could more effectively and deeply learn and extract useful information (knowledge) from high-dimensional datasets (input-output patterns), as compared with BPNN, RF and LSTM models. From the standpoint of a monitoring station (Nantzu Station) representative of high $PM_{2.5}$ concentrations, the CNN-BP model could create more precise and stable multi-step-ahead $PM_{2.5}$ forecasts. Therefore, the proposed CNN-BP model can significantly contribute to improving the reliability and accuracy of long-term $PM_{2.5}$ forecasting. In light of methodological transferability, future research can extend the CNN-BP methodology from one single pollutant ($PM_{2.5}$ in this study case) to multi-pollutant (e.g. $PM_{2.5}$, PM_{10}, O_3, etc.) forecasting as well as from deterministic forecasting to probabilistic forecasting by means of post-processing techniques, for instance, Kalman filtering, Generalized Likelihood Uncertainty Estimation (GLUE), and Bayesian methods (Djalalova et al., 2015; Kamińska, 2018; Pucer et al., 2018). Besides, for a longer lead time (e.g. daily forecast), it is very difficult to make accurate forecasts based solely on hourly datasets. Therefore, future work can be extended to daily forecasting in consideration of a collaboration with physical based models.

Acknowledgment

This study is supported by the Ministry of Science and Technology, Taiwan (MOST: 108-2119-M-002-017-A). The datasets provided by the Environmental Protection Administration of 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.

Reference:

Ausati, S., & Amanollahi, J. (2016). Assessing the accuracy of ANFIS, EEMD-GRNN, PCR, and MLR models in predicting PM2. 5. *Atmospheric environment*, *142, 465-474.* https://doi.org/10.1016/j.atmosenv.2016.08.007

Bai, Y., Zeng, B., Li, C., & Zhang, J. (2019). An ensemble long short-term memory neural network for hourly PM2. 5 concentration forecasting. *Chemosphere*, *222,* *286-294.* https://doi.org/10.1016/j.chemosphere.2019.01.121

Borovykh, A., Bohte, S., & Oosterlee, C. W. (2017). Conditional time series forecasting with convolutional neural networks. arXiv preprint arXiv:1703.04691.

Chan, C. C., Chuang, K. J., Chien, L. C., Chen, W. J., & Chang, W. T. (2006). Urban air pollution and emergency admissions for cerebrovascular diseases in Taipei, Taiwan. *European heart journal, 27(10), 1238-1244.* https://doi.org/10.1093/eurheartj/ehi835

Chang, F.J., Kao, L.S., Kuo, Y.M., Liu, C.W., 2010, “Artificial Neural Networks for Estimating Regional Arsenic Concentrations in a Blackfoot Disease Area in Taiwan”, *Journal of Hydrology, 388: 65-76.* https://doi.org/10.1016/j.jhydrol.2010.04.029

Chen, X., Kopsaftopoulos, F., Wu, Q., Ren, H., & Chang, F. K. (2019). A Self-Adaptive 1D Convolutional Neural Network for Flight-State Identification. *Sensors, 19(2), 275.* https://doi.org/10.3390/s19020275

Cheng, Y., Zhang, H., Liu, Z., Chen, L., & Wang, P. (2019). Hybrid algorithm for short-term forecasting of PM2. 5 in China. *Atmospheric environment*, *200*, 264-279. https://doi.org/10.1016/j.atmosenv.2018.12.025

Djalalova, I., Delle Monache, L., & Wilczak, J. (2015). PM2. 5 analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. . *Atmospheric Environment, 108, 76-87.* https://doi.org/10.1016/j.atmosenv.2015.02.021

Du, S., Li, T., Yang, Y., & Horng, S. J. (2018). Deep Air Quality Forecasting Using Hybrid Deep Learning Framework. arXiv preprint arXiv:1812.04783.

Du, X., Kong, Q., Ge, W., Zhang, S., & Fu, L. (2010). Characterization of personal exposure concentration of fine particles for adults and children exposed to high ambient concentrations in Beijing, China. *Journal of Environmental Sciences, 22(11), 1757-1764.* https://doi.org/10.1016/S1001-0742(09)60316-8

Elbayoumi, M., Ramli, N. A., & Yusof, N. F. (2015). Development and comparison of regression models and feedforward backpropagation neural network models to predict seasonal indoor PM2. 5-10 and PM2. 5 concentrations in naturally ventilated schools. *Atmospheric Pollution Research, 6(6), 1013-1023.* https://doi.org/10.1016/j.apr.2015.09.001

EPA. (2019) Open dataset platform. Available at: https://taqm.epa.gov.tw/taqm/en/b0101.aspx

EPA. (2019) $PM_{2.5}$ indicator and activity recommendation table. Available at: https://www.hpa.gov.tw/Pages/ashx/File.ashx?FilePath=~/File/Attach/3007/File_3697.pdf)

EPM. (2015) EPA-Environmental Policy Monthly. Available at:

https://www.epa.gov.tw/DisplayFile.aspx?FileID=9BBCE15769A6EF24&P=8905d3c6-79af-49fd-b608-6ec92bcdf740

Feng, R., Zheng, H. J., Gao, H., Zhang, A. R., Huang, C., Zhang, J. X., & Fan, J. R. (2019). Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China. *Journal of Cleaner Production.* https://doi.org/10.1016/j.jclepro.2019.05.319

Feng, R. Zheng, H., Zhang, A., Huang, C., Gao, H., & Ma, Y. (2019). Unveiling tropospheric ozone by the traditional atmospheric model and machine learning, and their comparison: A case study in hangzhou, China. *Environmental Pollution*, 252, Part A, 366-378.

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. *Atmospheric Environment*, *107, 118-128.* https://doi.org/10.1016/j.atmosenv.2015.02.030

Fernando, H. J., Mammarella, M. C., Grandoni, G., Fedele, P., Di Marco, R., Dimitrova, R., & Hyde, P. (2012). Forecasting PM10 in metropolitan areas: Efficacy of neural networks. *Environmental pollution, 163, 62-67.* https://doi.org/10.1016/j.envpol.2011.12.018

Gao, M., Yin, L., & Ning, J. (2018). Artificial neural network model for ozone concentration estimation and Monte Carlo analysis. *Atmospheric Environment, 184, 129-139.* https://doi.org/10.1016/j.atmosenv.2018.03.027

Glorot, X., Bordes, A., & Bengio, Y. (2011, June). Deep sparse rectifier neural networks. *In Proceedings of the fourteenth international conference on artificial intelligence and statistics(pp. 315-323).*

Hao, Y., & Liu, Y. M. (2016). The influential factors of urban PM2. 5 concentrations in China: a spatial econometric analysis. *Journal of cleaner production, 112, 1443-1453*. https://doi.org/10.1016/j.jclepro.2015.05.005

Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. *In Neural networks for perception (pp. 65-93).* https://doi.org/10.1109/IJCNN.1989.118638

Hernandez, G., Berry, T. A., Wallis, S., & Poyner, D. (2017). Temperature and humidity effects on particulate matter concentrations in a sub-tropical climate during winter. *International Proceedings of Chemical, Biological and Environmental Engineering*, V01, 102.

Hien, P. D., Bac, V. T., Tham, H. C., Nhan, D. D., & Vinh, L. D. (2002). Influence of meteorological conditions on PM2. 5 and PM2. 5− 10 concentrations during the monsoon season in Hanoi, Vietnam. *Atmospheric Environment, 36(21), 3473-3484.* https://doi.org/10.1016/S1352-2310(02)00295-9

Ho, T. K. (1995). Random Decision Forests (PDF). *Proceedings of the 3rd*

International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282. Archived from the original (PDF) on 17 April 2016. Retrieved 5 June 2016. https://doi.org/10.1109/ICDAR.1995.598994

Hsiao, T. C., Chen, W. N., Ye, W. C., Lin, N. H., Tsay, S. C., Lin, T. H., & Wang, S. H. (2017). Aerosol optical properties at the Lulin Atmospheric Background Station in Taiwan and the influences of long-range transport of air pollutants. *Atmospheric Environment*, *150, 366-378*. https://doi.org/10.1016/j.atmosenv.2016.11.031

Hsu, C. Y., Chiang, H. C., Lin, S. L., Chen, M. J., Lin, T. Y., & Chen, Y. C. (2016). Elemental characterization and source apportionment of PM10 and PM2. 5 in the western coastal area of central Taiwan. *Science of the Total Environment, 541, 1139-1150.* https://doi.org/10.1016/j.scitotenv.2015.09.122

Hsu, S. C., Liu, S. C., Jeng, W. L., Chou, C. C., Hsu, R. T., Huang, Y. T., & Chen, Y. W. (2006). Lead isotope ratios in ambient aerosols from Taipei, Taiwan: Identifying long-range transport of airborne Pb from the Yangtze Delta. *Atmospheric Environment*, *40(28), 5393-5404*. https://doi.org/10.1016/j.atmosenv.2006.05.003

Huang, C. J., & Kuo, P. H. (2018). A deep cnn-lstm model for particulate matter (Pm2. 5) forecasting in smart cities. *Sensors, 18(7), 2220.* https://doi.org/10.3390/s18072220

Jiang, P., Dong, Q., & Li, P. (2017). A novel hybrid strategy for PM2. 5 concentration analysis and prediction. *Journal of environmental management*, 196, 443-457. https://doi.org/ 10.1016/j.jenvman.2017.03.046

Jin, J., Dundar, A., & Culurciello, E. (2014). Flattened convolutional neural networks for feedforward acceleration. *arXiv preprint arXiv:1412.5474.*

Kamińska, J. (2018). Probabilistic Forecasting of Nitrogen Dioxide Concentrations at an Urban Road Intersection. *Sustainability, 10(11), 4213.* https://doi.org/10.3390/su10114213

Kamińska, J. A. (2018). The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: a case study in Wrocław. *Journal of Environmental Management, 217, 164-174.* https://doi.org/10.1016/j.jenvman.2018.03.094

Karambelas, A., Holloway, T., Kiesewetter, G., & Heyes, C. (2018). Constraining the uncertainty in emissions over India with a regional air quality model evaluation. *Atmospheric Environment, 174, 194-203.* https://doi.org/10.1016/j.atmosenv.2017.11.052

Karginova, N., Byttner, S., & Svensson, M. (2012). Data-driven methods for classification of driving styles in buses. *SAE Technical Paper,* No. 2012-01-0744.

Kumar, D. (2018). Evolving Differential evolution method with random forest for prediction of Air Pollution. *Procedia Computer Science, 132, 824-833.*

https://doi.org/10.1016/j.procs.2018.05.094
Kong, L., Xin, J., Liu, Z., Zhang, K., Tang, G., Zhang, W., & Wang, Y. (2017). The PM2. 5 threshold for aerosol extinction in the Beijing megacity. *Atmospheric environment, 167, 458-465*. https://doi.org/10.1016/j.atmosenv.2017.08.047
Lai, H. C., Ma, H. W., Chen, C. R., Hsiao, M. C., & Pan, B. H. (2019). Design and application of a hybrid assessment of air quality models for the source apportionment of PM2. 5. *Atmospheric Environment*, *212*, 116-127. https://doi.org/10.1016/j.atmosenv.2019.05.038
Li, L., Lei, Y., Wu, S., Chen, J., & Yan, D. (2017). The health economic loss of fine particulate matter (PM2. 5) in Beijing. Journal of cleaner production, 161, 1153-1161. https://doi.org/10.1016/j.jclepro.2017.05.029
Li, R., Mei, X., Wei, L., Han, X., Zhang, M., & Jing, Y. (2019). Study on the contribution of transport to PM2. 5 in typical regions of China using the regional air quality model RAMS-CMAQ. *Atmospheric Environment, 214, 116856.* https://doi.org/10.1016/j.atmosenv.2019.116856
Li, X., Feng, Y. J., & Liang, H. Y. (2017, July). The impact of meteorological factors on PM2. 5 variations in Hong Kong. In IOP Conference Series: *Earth and Environmental Science (Vol. 78, No. 1, p. 012003). IOP Publishing.* https://doi.org/10.1088/1755-1315/78/1/012003
Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2017). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. arXiv preprint arXiv:1707.01926.
Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R news*, 2(3), 18-22.
Liu, Y., Cao, G., Zhao, N., Mulligan, K., & Ye, X. (2018). Improve ground-level PM2. 5 concentration mapping using a random forests-based geostatistical approach. *Environmental pollution, 235, 272-282.* https://doi.org/10.1016/j.envpol.2017.12.070
Lou, C., Liu, H., Li, Y., Peng, Y., Wang, J., & Dai, L. (2017). Relationships of relative humidity with PM 2.5 and PM 10 in the Yangtze River Delta, China. *Environmental Mmonitoring and Assessment*, 189(11), 582. https://doi.org/10.1007/s10661-017-6281-z
Loy-Benitez, J., Vilela, P., Li, Q., & Yoo, C. (2019). Sequential prediction of quantitative health risk assessment for the fine particulate matter in an underground facility using deep recurrent neural networks. *Ecotoxicology and Environmental Safety, 169, 316-324.* https://doi.org/10.1016/j.ecoenv.2018.11.024
Ma, J., Ding, Y., Cheng, J. C., Jiang, F., & Wan, Z. (2019). A temporal-spatial interpolation and extrapolation method based on geographic Long Short-Term Memory neural network for PM2. 5. *Journal of Cleaner Production, 237, 117729*. https://doi.org/10.1016/j.jclepro.2019.117729

Ma, J., Ding, Y., Cheng, J. C., Jiang, F., Tan, Y., Gan, V. J., & Wan, Z. (2020). Identification of high impact factors of air quality on a national scale using big data and machine learning techniques. *Journal of Cleaner Production, 244, 118955.* https://doi.org/10.1016/j.jclepro.2019.118955

Mahajan, S., Liu, H. M., Tsai, T. C., & Chen, L. J. (2018). Improving the accuracy and efficiency of PM2. 5 forecast service using cluster-based hybrid neural network model. *IEEE Access, 6, 19193-19204.* https://doi.org/ 10.1109/ACCESS.2018.2820164

Mihăiţă, A. S., Dupont, L., Chery, O., Camargo, M., & Cai, C. (2019). Evaluating air quality by combining stationary, smart mobile pollution monitoring and data-driven modelling. *Journal of Cleaner Production, 221, 398-418.* https://doi.org/10.1016/j.jclepro.2019.02.179

Mishra, D., Goyal, P., & Upadhyay, A. (2015). Artificial intelligence based approach to forecast PM2. 5 during haze episodes: A case study of Delhi, India. *Atmospheric Environment, 102, 239-248*. https://doi.org/10.1016/j.atmosenv.2014.11.050

Moisan, S., Herrera, R., & Clements, A. (2018). A dynamic multiple equation approach for forecasting PM2. 5 pollution in Santiago, Chile. *International Journal of Forecasting, 34(4), 566-581.* https://doi.org/10.1016/j.ijforecast.2018.03.007

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. *Atmospheric environment, 134, 168-180.* https://doi.org/10.1016/j.atmosenv.2016.03.056

Nurkiewicz, T. R., Porter, D. W., Hubbs, A. F., Stone, S., Moseley, A. M., Cumpston, J. L., . . . Frisbee, J. C. (2011). Pulmonary particulate matter and systemic microvascular dysfunction. *Research report (Health Effects Institute), (164), 3-48*.

Prechelt, L. (1998). Early Stopping-but when? *In Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 1998; pp. 55–69, ISBN 978-3-642-35288-1, 978-3-642-35289-8.* https://doi.org/10.1007/978-3-642-35289-8_5

Pucer, J. F., Pirš, G., & Štrumbelj, E. (2018). A Bayesian approach to forecasting daily air-pollutant levels. . *Knowledge and Information Systems, 57(3), 635-654.* https://doi.org/10.1007/s10115-018-1177-y

Qiu, H., Yu, I. T., Wang, X., Tian, L., Tse, L. A., & Wong, T. W. (2013). Differential effects of fine and coarse particles on daily emergency cardiovascular hospitalizations in Hong Kong. *Atmospheric environment, 64, 296-302.* https://doi.org/10.1016/j.atmosenv.2012.09.060

Romero, A., Ballas, N., Kahou, S. E., Chassang, A., Gatta, C., & Bengio, Y. (2015). Imagenet classification with deep convolutional neural networks. *In International Conference on Learning Representations.* https://doi.org/10.1145/3065386

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., De'Donato, F., . . . Scortichini, M. (2019). Estimation of daily PM10 and PM2. 5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environment international, 124, 170-179.* https://doi.org/10.1016/j.envint.2019.01.016

Tai, A. P., Mickley, L. J., & Jacob, D. J. (2010). Correlations between fine particulate matter (PM2. 5) and meteorological variables in the United States: Implications for the sensitivity of PM2. 5 to climate change. *Atmospheric Environment*, 44(32), 3976-3984. https://doi.org/10.1016/j.atmosenv.2010.06.060

Tang, G., Zhao, P., Wang, Y., Gao, W., Cheng, M., Xin, J., ... & Wang, Y. (2017). Mortality and air pollution in Beijing: The long-term elationship. *Atmospheric environment*, *150*, 238-243. https://doi.org/10.1016/j.atmosenv.2016.11.045

Tsai, Y. I., & Kuo, S. C. (2005). PM2. 5 aerosol water content and chemical composition in a metropolitan and a coastal area in southern Taiwan. *Atmospheric Environment*, *39*(27), 4827-4839. https://doi.org/10.1016/j.atmosenv.2005.04.024

Wang, H. W., Li, X. B., Wang, D., Zhao, J., & Peng, Z. R. (2020). Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach. *Journal of Cleaner Production, 253, 119841.* https://doi.org/10.1016/j.jclepro.2019.119841

Wallace, J., & Kanaroglou, P. (2009). The effect of temperature inversions on ground-level nitrogen dioxide (NO2) and fine particulate matter (PM2. 5) using temperature profiles from the Atmospheric Infrared Sounder (AIRS). *Science of the Total Environment, 407(18), 5085-5095.* https://doi.org/10.1016/j.scitotenv.2009.05.050

Widiana, D. R., You, S. J., Yang, H. H., Wang, L. C., Tsai, J. H., & Chen, H. M. (2019). Air Pollution Profiles and Health Risk Assessment of Ambient Volatile Organic Compounds above a Municipal Wastewater Treatment Plant, Taiwan. *Aerosol and Air Quality Research, 19(2), 375-382.* https://doi.org/10.4209/aaqr.2018.11.0408

Yang, G., Huang, J., & Li, X. (2018). Mining sequential patterns of PM2. 5 pollution in three zones in China. *Journal of cleaner production, 170, 388-398.* https://doi.org/10.1016/j.jclepro.2017.09.162

Yuan, C. S., Sau, C. C., & Chen, M. C. (2004). Influence of Asian dusts on the physicochemical properties of atmospheric aerosols in Taiwan district—Using the Penghu Islands as an example. *China Particuology, 2(4), 144-152.* https://doi.org/10.1016/S1672-2515(07)60045-1

Zhang, L., Lin, J., Qiu, R., Hu, X., Zhang, H., Chen, Q., . . . Wang, J. (2018). Trend analysis and forecast of PM2. 5 in Fuzhou, China using the ARIMA model.

Ecological indicators, *95, 702-710.* https://doi.org/10.1016/j.ecolind.2018.08.032

Zhao, J., Deng, F., Cai, Y., & Chen, J. (2019). Long short-term memory-Fully connected (LSTM-FC) neural network for PM2. 5 concentration prediction. *Chemosphere, 220, 486-492.* https://doi.org/10.1016/j.chemosphere.2018.12.128

Zhou, L., Chen, X., & Tian, X. (2018). The impact of fine particulate matter (PM2. 5) on China's agricultural production from 2001 to 2010. *Journal of Cleaner Production*, *178*, 133-141. https://doi.org/10.1016/j.jclepro.2017.12.204

Zhou, Y., Chang, F. J., Chang, L. C., Kao, I. F., Wang, Y. S., & Kang, C. C. (2019). Multi-output support vector machine for regional multi-step-ahead PM2. 5 forecasting. *Science of the Total Environment, 651, 230-240*. https://doi.org/10.1016/j.scitotenv.2018.09.111

Zhou, Y., Chang, F.J., Chang, L. C., Kao, I. F., & Wang, Y. S. (2019). Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *Journal of Cleaner Production, 209*, 134-145. https://doi.org/10.1016/j.jclepro.2018.10.243

Credit Author Statement

Pu-Yun Kow: Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Roles/Writing - original draft. **Yi-Shin Wang:** Data curation, Investigation, Project administration. **Yanlai Zhou:** Methodology. **I-Feng Kao:** Investigation, Methodology. **Maikel Issermann:** Formal analysis. **Li-Chiu Chang:** Methodology, Project administration, Resources, Supervision. **Fi-John Chang:** Funding acquisition, Methodology, Project administration, Supervision, Writing - review & editing.

*Declaration of Interest Statement

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: