IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 6, 2018, accepted May 29, 2018, date of publication June 22, 2018, date of current version July 30, 2018.

*Digital Object Identifier 10.1109/ACCESS.2018.2849820*

# Adaptive Deep Learning-Based Air Quality Prediction Model Using the Most Relevant Spatial-Temporal Relations

**PING-WEI SOH[1], JIA-WEI CHANG[2], AND JEN-WEI HUANG[2]**
[1]Institute of Computer and Communication Engineering, National Cheng Kung University, Tainan 701, Taiwan
[2]Department of Electrical Engineering, National Cheng Kung University, Tainan 701, Taiwan

Corresponding author: Jen-Wei Huang (jwhuang@mail.ncku.edu.tw)

**ABSTRACT** Air pollution has become an extremely serious problem, with particulate matter having a significantly greater impact on human health than other contaminants. The small diameter of fine particulate matter (PM2.5) allows it to penetrate deep into the alveoli as far as the bronchioles, interfering with a gas exchange within the lungs. Long-term exposure to particulate matter has been shown to cause the cardiovascular disease, respiratory disease, and increase the risk of lung cancers. Therefore, forecasting air quality has also become important to help guide individual actions. This paper aims to forecast air quality for up to 48 h using a combination of multiple neural networks, including an artificial neural network, a convolutional neural network, and a long-short-term memory to extract spatial-temporal relations. The proposed predictive model considers various meteorology data from the previous few hours as well as information related to the elevation space to extract terrain impact on air quality. The model includes trends from multiple locations, extracted from correlations between adjacent locations, and among similar locations in the temporal domain. Experiments employing Taiwan and Beijing data sets show that the proposed model achieves excellent performance and outperforms current state-of-the-art methods.

**INDEX TERMS** Dynamic time warping(DTW), convolutional neural network(CNN), long-short-term memory(LSTM), spatio-temporal analysis, big data, air quality forecast.

## I. INTRODUCTION

Increasing attention has been given to air quality degeneration, with particulate matter (PM) having a significant egregious impact on human health. The small diameter of fine particulate matter (PM2.5) allows it to penetrate deep into the alveoli as far as the bronchioles, interfering with gas exchange within the lungs. Xing showed that long term exposure to particulate matter increased the risk of the cardiovascular disease, respiratory disease, and lung cancer [1]. With increasing public health consciousness, many cities have established air quality monitoring locations. However, most services only show the current air quality and do not forecast air quality. Air quality prediction is essential to help guiding individual actions limiting PM2.5 exposure, e.g., choosing outdoor or indoor activities.

However, accurate air quality forecasting is hindered by a complex array of factors [2]–[4], including emissions, traffic patterns, and meteorological conditions. Meteorologists are still substantially limited to provide reliable wind pattern predictions, which can vary considerably in direction and strength every hour [5], and there are insufficient sensors deployed to provide emission data from factories or vehicles. Recent studies have shown it is critical that time and space be explicitly considered to analyze air quality [6]–[8]. Particulate matter has high cyclicality and is easily affected by space, stagnating or diffusing to pollute surrounding environments. If PM is only analyzed in the time domain, this may neglect impacts and relationships between other regions; whereas considering only spatial relationships may omit PM diffusion from over time. Therefore, time and spatial relations must be simultaneously considered to accurately model PM diffusion.

Data mining provides new methods to analyze air quality in the absence of physical models [9]–[11], and may identify hidden information in the collected data. Furthermore, prediction speed is far quicker once a model is trained than for traditional physical models. Therefore, we propose a model to provide 48 hours air quality index (AQI) predictions every hour at every monitoring location. As shown in Figure 1, we forecast 48-hours predictions from the current time, $t_c$,

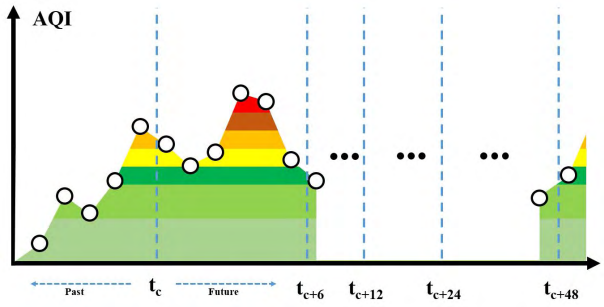



**FIGURE 1.** Air quality index (AQI) forecast format.

using historical data. In particular, the peaks and valleys are the most important segments for the future predictions.

This study proposes a general predictive model for air quality forecasts called spatial-temporal deep neural network (ST-DNN) that incorporates various information from monitoring locations, including PM2.5, PM10, temperature, wind speed, wind direction, average wind speed, average wind direction, relative humidity, and data related to the elevation space. The model was trained using current and previous few hours air pollutant and meteorological condition data. The proposed scheme does not consider forecast data from external sources. We developed a method to integrate the relevant data based on geographical and temporal correlations among monitoring locations. We first found the most relevant spatial-temporal relations among locations, then combined multiple neural network architectures using a convolutional neural network (CNN) [12] and long short term memory (LSTM) [13]. Target and similar location spatial-temporal features were used to increase the predictive model sensitivity and explicitly consider terrain impacts for pollutant propagation. Thus, the proposed model uses (i) temporal information based on target location historical data, (ii) spatial relationships based on related locations' data, i.e., locations with high spatial or temporal similarity, and (iii) terrain information for the area around the locations.

To validate the proposed model, we performed experiments using two real-world datasets: 76 locations in 23 cities in Taiwan [14] and locations from Beijing dataset [15]. The experimental results confirm that the proposed methods achieve excellent performance, superior to current baselines and several state-of-the-art methods. The main contributions of this study are as follows.

- We propose a framework to mine spatial-temporal data for a given location to provide a predictive model.
- We develop a deep learning model combining multiple neural networks to incorporate air quality correlations among similar locations and temporal dependency at a given location. Spatial and temporal predictions are combined dynamically based on the trained neural network.
- The proposed system has been deployed throughout Taiwan, providing access to fine grain information regarding air quality via a public website [16] and Facebook chatbot [17].

The remainder of this paper is organized as follows. Section II describes related works, and Section III defines the problem to be addressed. Section IV presents an overview of the proposed system, describes the proposed method to mine spatial-temporal relationships, and provides the proposed spatial-temporal prediction model framework and details. Section V compares the proposed model performance with previous state-of-the-art algorithms based on real-world datasets, and Section VI introduces real applications developed to provide convenient and publicly accessible PM2.5 forecasts. Finally, Section VII summarizes and concludes the paper.

## II. RELATED WORKS

Qin *et al.* [18] proposed mining environmental spatial-temporal relationships using an a-priori pattern mining algorithm. They shifted one of the time sequences to create specific gaps in each time sequence to generate high frequency candidates for rule generation. The resulting rules reveal the appearance of pollutants with delays in different locations. However, his method requires repeatedly running the rule generation process through numerous combinations, which is time consuming.

Zheng *et al.* [5] proposed a framework that considered temporal as well as spatial relationships. The framework was divided into four components: temporal predictor, spatial predictor, prediction aggregator, and inflection predictor. The temporal predictor considered only historical data of the target location using linear regression. The spatial predictor considered global data using the mean and median in the region around a location, using an artificial neural network (ANN). The ANN excluded information from other locations, hence the results were insensitive to surroundings conditions and global trends. The predictor aggregator used a regression tree with three inputs to combine temporal and spatial predictors with local meteorological data. The inflection predictor identified sudden drops in target feature value, i.e., PM2.5, by finding situations where specific feature thresholds were surpassed. The proposed framework pre-trained the temporal predictor and spatial predictor, and then trained the predictor aggregator to combine the results. However, this can over-fit the data, since the same features are adopted as delimiters in the predictor aggregator; and the spatial predictor considers the mean and median values in a large region, which can reduce sensitivity. Zheng also separated the wind directions into 8 classes and did not consider terrain information, although PM2.5 diffusion depends strongly on wind, which is greatly influenced by terrain. Therefore, Zheng's model may not be suitable for undulating terrain.

Soh *et al.* [19] previously proposed a $k$-nearest neighbor by DTW distance (kNN-DTWD) method to consider time sequence similarities for different locations and then compared surrounding locations using the $k$-nearest neighbor by Euclidean distance (kNN-ED). DTW [20], [21] is a well-known method to calculate similarity between two time

series, hence we applied DTW to calculate location temporal distances and used k-nearest neighbor (kNN) to identify locations with the most similar temporal behavior [19]. This differs from the approach of Zheng *et al.* [5] in that we were seeking to enhance prediction sensitivity over short durations rather than longer periods. Experimentally, kNN-DTWD outperformed kNN-ED on average. However, for some special cases, e.g., flat areas with numerous locations, the kNN-ED method proved advantageous. The kNN-DTWD method was well suited selection locations with similar behavior, and the kNN-ED method was best avoided where there was a mountain between the locations.

This work combines kNN-ED benefits for featureless landscapes with kNN-DTWD benefits for complex landscapes, allowing the model to derive optimal combinations of based on the training data, as detailed in the following sections.

## III. PROBLEM DEFINITION

We first identify the locations with the most influential spatial-temporal relationships to the target location and then predict sequences for the target location based on time sequence features that include spatial information. Hence, the locations are fixed, and the time sequences vary according to their positions. Spatial features could also impact the sequences, e.g., a mountain between two locations. Therefore, we use both temporal and spatial relationships in the predictive model. To extract the features for prediction, we first define related spatial and temporal relationship parameters. Suppose we have a set of locations: $\mathrm{L} = \{l_1, l_2, \ldots, l_n\}$ and a set of features: $\mathrm{F} = \{f_1, f_2, \ldots f_m\}$. Each location has geographical information, such as latitude and longitude, hence we define the location coordinate (LC) as

*Definition 1:* Location Coordinate.

$$Lc_i = (l_i, x_i, y_i), \quad l_i \in L \tag{1}$$

where $x_i$ and $y_i$ are the latitude and longitude at location $l_i$, respectively. Since related location features could improve prediction, we define the distance between two location coordinates as

$$\begin{aligned} Ds_{q,c} &= dist_{location}(Lc_q, Lc_c) \\ &= dist_{location}((l_q, x_q, y_q), (l_c, x_c, y_c)), \\ &\qquad l_q, l_c \in L, \quad q \neq c. \end{aligned} \tag{2}$$

to find the most closely related locations in the spatial domain; and the spatial relationships sequence (SRS) set as

*Definition 2:* Spatial Relations Sequence Set.

$$SRS = \{Ds_{1,2}, Ds_{1,3}, \ldots, Ds_{n-1,n}\}, \quad Ds_{i,i} = 0, \quad 0 < i < n+1 \tag{3}$$

where the matrix elements are calculated from (2); $n$ is the number of locations; and the diagonal elements, $Ds_{i,i}$, are all zero. The most relevant locations are $SRS\_cand(l_i,k)$, the set of $k$ locations with the smallest spatial distance to $l_i$.

We consider the features of these relevant locations in the spatial domain, since locations with similar feature sequences may help to improve prediction. We define the feature sequence interval (FSI) for a specific location as

*Definition 3:* Feature Sequence Interval with Location.

$$\begin{aligned} S(l_i, f_j, t_{st,ft}) = \{&e(l_i, f_j, t_{st}), e(l_i, f_j, t_{st+1}), \ldots, \\ &e(l_i, f_j, t_{ft})\}, \quad l_i \in L,\ f_j \in F,\ st < ft. \end{aligned} \tag{4}$$

where $l_i$ has feature $f_j$ that varies from start, $st$, to finish, $ft$, time ($st < ft$); and $e(l_i,f_j,t_k)$ represents the measured value of $f_j$ at $t_k$.

We can express the distance between feature sequences for any two locations as

$$\begin{aligned} &Dt_{q,c,t_{st,ft}} \\ &\quad = dist_{sequence}(S(l_q, f_{target}, t_{st,ft}), S(l_c, f_{target}, t_{st,ft})), \\ &\qquad l_q, l_c \in L, \quad q \neq c. \end{aligned} \tag{5}$$

to obtain the most related locations in the temporal domain. Before using the measure, we must select a feature $f_{target}$, as the target prediction sequence: this study chose PM2.5, but other targets could be employed as required. Then we can use (5) to calculate the temporal relations sequence (TRS) set,

*Definition 4:* Temporal Relations Sequence Set.

$$TRS_{t_{st,ft}} = \{Dt_{1,2,t_{st,ft}}, Dt_{1,3,t_{st,ft}}, \ldots, Dt_{n-1,n,t_{st,ft}}\} \tag{6}$$

We then select candidates $TRS\_cand(l_i,k)$, the set of $k$ locations with the least difference from location $l_i$. To consider both relationships simultaneously, we define the spatial-temporal relations (STR) set, i.e., the set of locations most strongly related to $l_i$ as

*Definition 5:* Spatial-Temporal Relations Set.

$$\begin{aligned} &STRS\_cand(l_i, k) \\ &\quad = SRS\_cand(l_i, k) \cup TRS\_cand(l_i, k), \quad l_i \in L. \end{aligned} \tag{7}$$

where we use the union $SRS\_cand(l_i,k)$ and $TRS\_cand(l_i,k)$ rather than the intersection, to provide a larger number of relationships for the model to learn; and since some location behaviors may differ from adjacent locations, the intersection would have fewer candidates (or none), resulting in the loss of useful target features. We define the spatial-temporal predictor (STP) using $STRS\_cand(l_i,k)$ as

*Definition 6:* Spatial-Temporal Sequence Prediction

$$\begin{aligned} &M(STRS\_cand(l_i, k))[t_{t_{lb},t_c}] \\ &\quad = S(l_i, f_{target}, t_{st',ft'}), \quad t_{lb} < t_c < st' \leq ft'. \end{aligned} \tag{8}$$

to build the model $M$ to predict a target feature sequence, where $M$ returns a sequence set, $S$, of the target features for the period $t_{st'}$ to $t_{ft'}$. $S$ is generated from the most similar time series compared with $t_{lb}$ to $t_c$, where $t_{lb}$ is the lookback time. Section IV proposes the solution method for this prediction problem.

## IV. PREDICTION MODEL FRANEWORK

To find the most relevant relationships between locations we apply spatial-temporal analysis to explore sequence delays and interactions between locations using historical

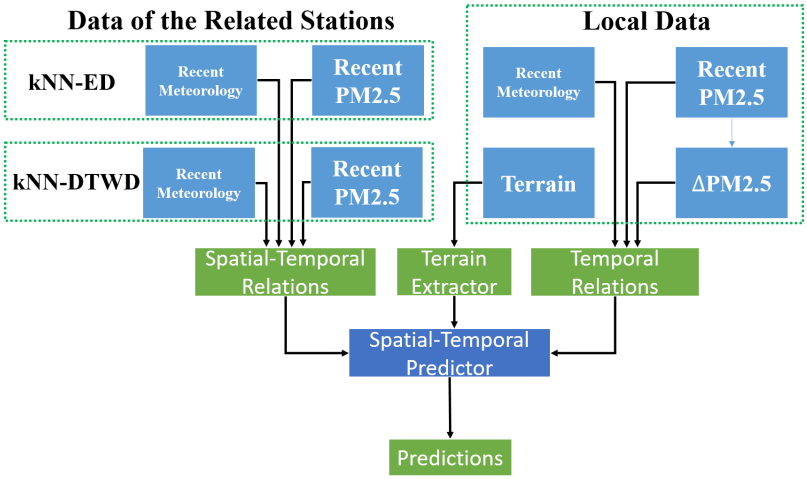



**FIGURE 2.** Predictive model (ST-DNN) framework.

temporal patterns, and consider location feature trends for potential factors. We consider adjacent locations or locations with similar temporal patterns because they have high correlations with the target location. After importing processed data into the system, we determine the most related locations to the target location using the proposed kNN-ED and kNN-DTWD relationship extractors, and then generate training datasets from the top $k$ related locations. Finally, we train the deep learning based model and compare prediction performances. Figure 2 shows the proposed predictive model framework comprises four main components as follows.

- The temporal relationships extractor (TRE) obtains air quality features from the target location meteorological data over the previous few hours using LSTM model.
- The spatial-temporal relationships extractor (SRE) uses ANN model to obtain air quality feature data from related locations selected by kNN-ED or kNN-DTWD.
- The terrain extractor (TE) obtains terrain information in the vicinity of the target location and uses a CNN to extract interactions between terrain and air quality features.
- The merge layer provides the STP to combine the discrete component outcomes. In some cases, predictions based on historical target location data are more relevant, whereas in other cases, such as windy days, spatial data should be given a higher weight. The full connected ANN layer can learn the weights from the training data.

The following sections introduce the methods to mine spatial-temporal relationships and the ST-DNN predictive model.

### *A. MINING SPATIAL-TEMPORAL RELATIONSHIPS FROM RELATED LOCATIONS*

#### 1) k-NEAREST NEIGHBOR BY EUCLIDEAN DISTANCE (kNN-ED)

We calculated Euclidean distance between the locations using their geographical coordinates, as shown in Figure 3.

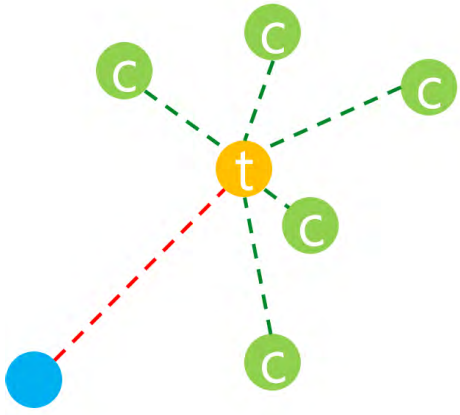



**FIGURE 3.** Example derivation (kNN-ED, $k = 5$).

The 20 selected candidates, shown in green in Figure 3, are denoted as $SRS\_cand(l_i, k)$ and used to train the predictive model. The procedure is shown in Algorithm 1.

#### 2) k-NEAREST NEIGHBOR BY DTW DISTANCE (kNN-DTWD)

We used the DTW algorithm to calculate the distance between two sequences by minimizing the errors in shifting and scaling between the sequences., as shown in Figure 4. Although the sequences are dissimilar using Euclidean distance, DTW can restore sequence distortions by mapping the data points to corresponding intervals. Thus, DTW identified the most strongly related temporal relationships to the target location and calculated the time series feature distances between locations. The distances were sorted and the top $k$ most similar locations chosen as candidates to predict the target location sequence. We refer to this method as kNN-DTWD. Figure 5 shows that the PM2.5 time series exhibit various shift and scale differences.S7 (purple) has delays compared with S1 (red), and S2 (orange), as highlighted within the red circle.

We calculated the degree of similarity between $TRS\_cand(l_i, k)$, i.e., the candidate set, members using conventional

**Algorithm 1** Geographical Relationship Set Generator (kNN-ED)

**Input**: Target station $l_i$; Set of Locations' coordinate $Lc$, where $l_i \notin Lc$;
Number of candidates $k$;
**Output**: Set of Locations by $SRS\_cand(l_i, k)$;
Let $SRS\_cand \leftarrow \emptyset$; **for** *each* $lc \in Lc$ **do**
Calculate distances between $l_i$ and $lc$: $ED(l_i, lc)$;
$SRS\_cand \cup \{lc, ED(l_i, lc)\}$;
**end**
Sort $SRS\_cand$ by $ED(l_i, lc)$;
**if** $k \leq Size\ of\ SRS\_cand$ **then**
$SRS\_cand(l_i, k) \leftarrow$ first $k^{th}$ of $SRS\_cand$
**end**
**else**
$SRS\_cand(l_i, k) \leftarrow SRS\_cand$
**end**

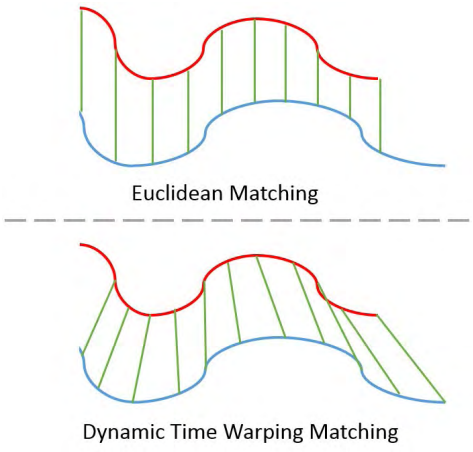



**FIGURE 4.** Euclidean and DTW matching.

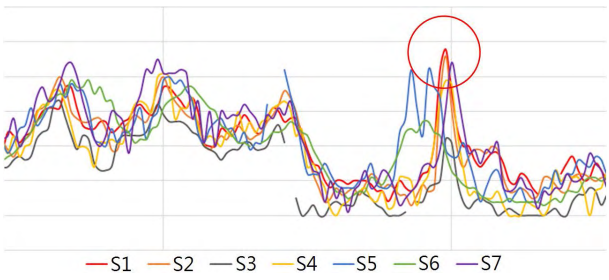



**FIGURE 5.** Example PM2.5 time series data (seven locations).

DTW, hence eliminating time shift and scaling effects, and identified $TRS\_cand(l_i, k)$ candidates, i.e., those with the strongest temporal relationship to the target location.

Unfortunately, this method cannot calculate the degree of similarity between time series with missing data. We tackle this problem as follows. (i) Two selected sequences were divided into a plurality of common interval sequences, choosing the shortest interval threshold, $l_{min}$, to filter short, i.e., not meaningful, sequences. (ii) We then apply DTW to the filtered sequences and convert the calculated values to unit similarity as shown in Figure 6. (iii) Finally, $TRS\_cand(l_i, k)$ for the predictive model is generated using Algorithm 2.

**Algorithm 2** Temporal Similarity Set Generator (kNN-DTWD)

**Input**: Target station $l_i$; Set of Locations' $L$, where $l_i \notin L$;
Number of candidates $k$; Target feature $f_{target}$;
Time Interval $t_{st,ft}$;
**Output**: Set of Locations by $TRS\_cand(l_i, k)$;
Let $TRS\_cand \leftarrow \emptyset$;
**for** each $l \in L$ **do**
Calculate similarity between $l_i$ and $l$;
$dist_{dtw} \leftarrow$
$DTWDsim(S(l_i, f_{target}, t_{st,ft}), S(l_j, f_{target}, t_{st,ft}), l_{min})$;
$TRS\_cand \cup \{l, dist_{dtw}\}$; **end** Sort $TRS\_cand$ by $dist_{dtw}$;
**if** $k \leq Size\ of\ TRS\_cand$ **then**
$TRS\_cand(l_i, k) \leftarrow$ first $k^{th}$ of $TRS\_cand$
**end**
**else**
$TRS\_cand(l_i, k) \leftarrow TRS\_cand$
**end**

Figure 6 is explained in details as follow. When two time series intervals have non-missing values simultaneously and the common interval length exceeds $l_{min}$ (i.e., $l_a > l_{min}$, Figure 7) then the common interval the interval is included for DTW similarity. In contrast, $l_b < l_{min}$ (Figure 7), and hence is ignored in the DTW calculation. Although the ignored cause some loss of information, but this method effectively removes most noise related errors.

The average unit distance is defined as

$$d_\Delta = \frac{\sum_1^n d_i}{\sum_1^n l_i} \tag{9}$$

where $d_i$ is the distance in a common interval, and $l_i$ is the length of that interval; which combines multiple fragment sequences to facilitate overall similarity identification.

Figure 8 shows two interval distances, $d_1$ and $d_2$, calculated by DTW; and their common interval lengths, $l_1$ and $l_2$, respectively, recorded for calculating the average unit distance.

### B. PREDICTION MODEL DESIGN

The proposed ST-DNN model combines target location temporal information, and related location spatial-temporal and terrain information (see Figure 2). The data flow includes target and related location historical data, i.e., pollutants, meteorological conditions, and target features and their trends over the previous few hours. These data were input to the LSTM, adaptive temporal extractor (ASE), and ANN. We used a matrix of 121 square sections for terrain data, i.e., 11×11 coordinate lines at 500 m intervals, where the central square in the grid represents the local location.


**FIGURE 6.** Similarity measure procedure.


**FIGURE 7.** Shortest interval threshold.

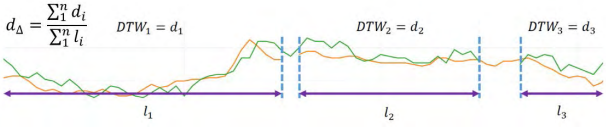



**FIGURE 8.** Average unit distance calculation example.

Thus, 120 unobserved points, with AQI calculated by inverse distance weighting (IDW) [22]. We convolve the relative elevation with the unobserved point AQIs to reduce AQI impact at higher elevation and provide this matrix as input to the CNN. CNN inputs can be adjusted later to increase the resolution, e.g., LASS open source data.

We set $l_{min} = 6$ hour in kNN-DTWD [23], [24], i.e., the minimum time interval for meteorological forecasts. Pollutants, meteorological conditions, and target feature(s) of locations with high similarity (determined using kNN-ED and kNN-DTWD) were input to the LSTM and ASE without pretraining. A two-layer ANN was employed to combine TRE, SRE, and TE. This final prediction was the deviation between the target feature value at $t_c$ and some future time $t_{c+h}$, where $t_{c+h} < ft$.

Figure 9 shows the model structure. Air quality and meteorological condition data sources are input to LSTM and ASE, and terrain related data are input to CNN. The models are merged via side by side concatenation, and the variables are passed to the following layer. The model is trained hourly over the subsequent 48 hours, since the current status varies with respect to its effect on future time intervals. Thus, we pair the inputs with target feature deviations in the various time intervals to train multiple models with the same structure corresponding to the different time intervals. The advantage of this structure is that the input sizes are constant, regardless of the location and time interval.

### 1) TEMPORAL RELATIONS

The TRE uses historical target location features as inputs to predict the future time series. The input time series for PM2.5 and other concentrations are continuous and coherent, and can be divided into low frequency (trends) and high frequency (rapid changes) information.

Since LSTM models historical time series behavior, we consider the TRE LSTM to obtain target location time series trends; whereas the ANN uses current data only, and hence is sensitive to rapid changes. Thus, the LSTM and ANN provide low and high frequency information, respectively, from the sequences.

The LSTM TRE models trends for PM2.5 and PM10 concentrations as well as local meteorological data (wind speed, wind direction, humidity, and temperature) over the previous six hours; and the ASE TRE increases the model sensitivity using the same features as LSTM TRE. Previous studies have verified the relevance of these features with regard to air quality [11].

### 2) SPATIAL-TEMPORAL RELATIONS

Pollutant dispersal means that air quality at one location can be spatially correlated with that at other sites. SRE uses historical spatial-temporal neighborhood location features inputs since air quality at a given location is affected by local emissions as well as emissions in surrounding areas. Therefore, we devised a SRE to predict target location air quality based on AQIs and meteorological data from other locations. Partitioning spaces into regions using circles of various diameter overlooks terrain impacts, e.g., a mountain between locations.

Partitioned region data are often mean or mode values, which can be highly inaccurate, particularly in areas with few locations. Thus, the SRE for a location requires data mining from locations in the spatial-temporal neighborhood using kNN-ED and kNN-DTWD, including AQIs and meteorological conditions (wind speed and direction) for the previous 6 hours. Similar to the target location, spatial-temporal neighborhood time series features are also continuous and coherent. Thus, we included the ANN SRE to increase model sensitivity by considering spatial-temporal neighborhood impacts.

### 3) TERRAIN EXTRACTOR

The relationships between locations vary due to various barriers and altitude differences. Therefore, terrain related data were included to enhance location correlations. Terrain data in the vicinity of the locations were captured using a matrix of 121 square sections; i.e., $11 \times 11$ coordinate lines at 500 m intervals. We adopted the approach of Ferrero *et al.* [25] to define the relationships between terrain and PM2.5. The elevation of each point, *elev*, was normalized as

$$H_s = \frac{elev - elev_{st}}{elev_{st}} \tag{10}$$

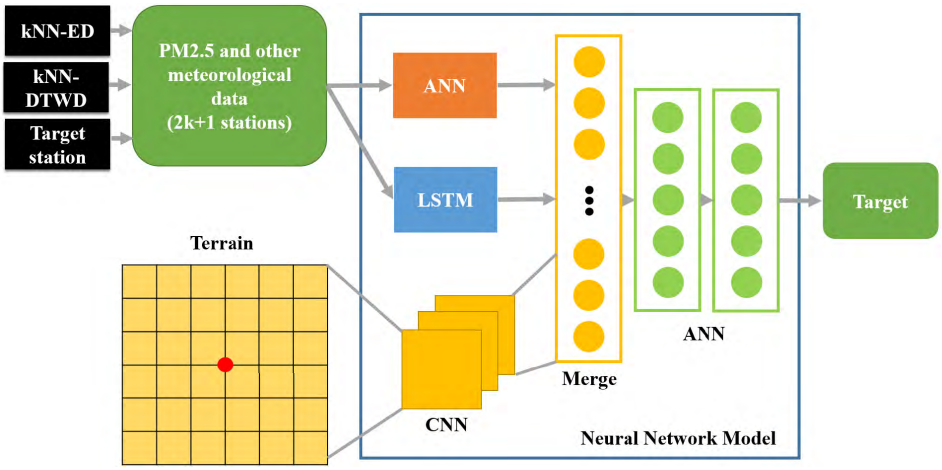



**FIGURE 9.** Proposed prediction model (ST-DNN) structure.

and transformed to the relative elevation,

$$elev_{rel} = \frac{1}{e^{H_s}} \tag{11}$$

where $H_s$ is the standardize elevation, to decrease the impact of higher altitudes, as shown in Figure 10. Figure 11 shows that PM2.5 distribution is strongly related to elevation.

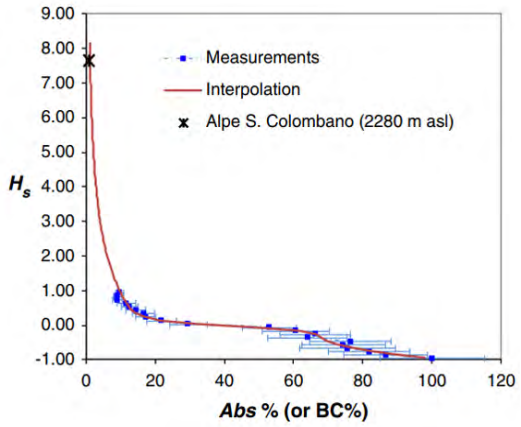



**FIGURE 10.** Relative elevation function design [25].

We then calculated AQIs for each location using IDW and multiplied this by $elev_{rel}$ to reduce location impact at higher elevations. Thus, we could extract relationships between locations that would otherwise have been obscured, such as the impact of wind direction and wind speed for locations adjacent to mountains. The CNN was designed to extract useful information particularly in the spatial domain. We used the CNN to include spatial correlations between locations and extract the obscured terrain relationships.

#### 4) MERGE LAYER

We concatenated the TRE, SRE, and TE outcomes, and passed these to the ANN. Thus, the model applied local and

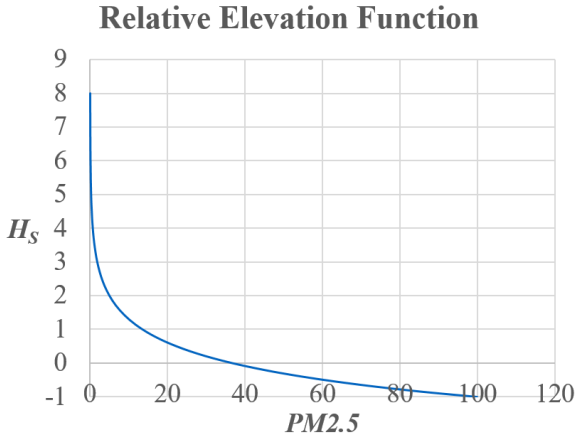



**FIGURE 11.** Relative elevation function of PM2.5 distribution.

global inputs for prediction. In some cases, local information is more relevant than global, e.g., when air circulation between locations is weak. On the other hand, global dispersion may be a major factor determining air quality when wind speed is high. Thus, we looked for meteorological condition trends at a given location, such as wind speed, wind direction, humidity, temperature, etc., to weight prediction calculations provided by the three components.

## V. EXPERIMENTS

### A. DATASETS

We chose PM2.5 as the predictive feature because it is the most widely reported metric and also the most difficult air pollutant to predict. Similar architectures can be applied to predict other pollutants.

#### 1) TAIWAN DATASET

We collected air quality and meteorological data every hour from 76 locations in 23 Taiwanese cities. Each location recorded PM2.5 and PM10, but not all locations

| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Air Pollution Banding | Low | Low | Low | Moderate | Moderate | Moderate | High | High | High | Very High |
| $PM_{2.5}$ concentration ($\mu g/m^3$) | 0-11 | 12-23 | 24-35 | 36-41 | 42-47 | 48-53 | 54-58 | 59-64 | 65-70 | >=71 |
| Accompanying health messages for the general population | Enjoy your usual outdoor activities | | | Enjoy your usual outdoor activities | | | Anyone experiencing discomfort such as sore eyes, cough or sore throat should consider reducing activity, particularly outdoors. | | | Reduce physical exertion, particularly outdoors, especially if you experience symptoms such as cough or sore throat. |

**FIGURE 12.** TWEPA PM2.5 index for the Taiwan dataset [26].

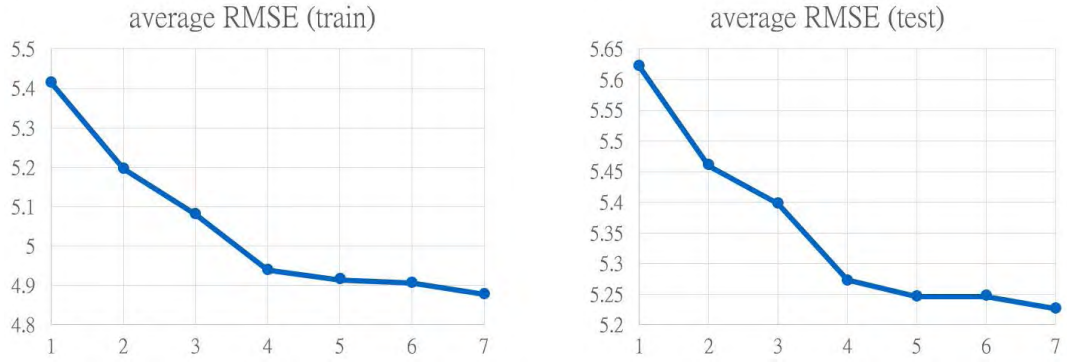



**FIGURE 13.** Influence of past $h$ hours (x-axis: $hour(s)$, y-axis: $\mu g/m^3$).

recorded levels of other pollutants or meteorological data. More than 28 million instances were collected from January 2014 to September 2017 by Taiwan Environmental Protection Administration (TWEPA). PM2.5 was measured using Met-One BAM-1020 and other measurement instruments are shown in [27]. Data was partitioned into a training data (Jan. 2014 to Sep. 2016) and testing set (Oct. 2016 to Sep. 2017) at 2:1 ratio, based on seasonal cycles in Taiwan, where the testing set covers all four seasons.

Figure 12 shows the TWEPA PM2.5 index over the study period, and TWEPA forecast data was not included. The proposed PM2.5 prediction used PM2.5, PM10, wind speed, wind direction, temperature, and humidity features.

#### 2) BEIJING DATASET

The Beijing dataset focused on Beijing city [15], excluding weather forecasts because they were based on physical models. We modified the proposed model to predict min-max outputs to ensure a fair comparison. The dataset contained records from May 2014 to Apr. 2015. Months 5, 6, 8, 9, 11, 12 of 2014 and 2, 3 of 2015 were adopted as training data and the balance as testing data, covering the last month of each season, i.e., 7, 10 of 2014 and 1, 4 of 2015.

### B. METRICS AND GROUND TRUTH

Air quality prediction were compared with ground truth results obtained at each location, and the mean absolute error (MAE) [28]–[31] was adopted to evaluate prediction performance,

$$e = \frac{\sum_1^n |\hat{y}_i - y_i|}{n} \tag{12}$$

where $\hat{y}_i$ and $y_i$ are the prediction and ground truth for the $i_{th}$ hour, respectively; and $n$ is the number measurements within a time interval.

We calculated MAE for 1-6, 7-12, 13-24, and 25-48 hours, which are common time frames in conventional weather forecasting. Lower absolute error indicates higher prediction accuracy.

### C. COMPARATIVE MODELS AND PARAMETER SETTINGS

We compared the proposed ST-DNN method predictions with a number of baselines.

1) Baselines that feed all features into a single model, e.g., linear regression (LR_ALL) and neural network (ANN_ALL), without treating various features differently.
2) Baselines using kNN-ED, kNN-DTWD, and both methods (referred to as adaptive baselines), to identify the top $k$ related locations.
3) Zheng *et al.* [5] proposed a predictive model that considered local spatial data similar to the approach adopted in the current study but used average neighbor values over a specified region.

To determine suitable lookback time, $h$, for the models, we evaluated different $h$ for predicting the next hour, $t + 1$, as shown in Figure 13. Hence, we chose $h = 6$ hours

as the minimum time interval for meteorological forecasting [23], [24]. We also used an independent LSTM for each feature with $h = 6$ hours, and chose $k = 3$ for kNN-ED and kNN-DTWD in the proposed ST-DNN model. We used a 5Ã-5 filter for the CNN with one convolutional layer. The CNN did not include a max pooling layer since we did not extract the highest concentrations but the most strongly related concentrations. We chose a linear activation function to consider negative correlations. We used a total of 4276 parameters for the ST-DNN models for the Taiwan dataset.

### D. PERFORMANCE OF PREDICTIONS

#### 1) TAIWAN DATASET

We first checked if the input components to the ST-DNN model were significant, examining all Adaptive ANN (A), LSTM (L), and CNN (C) combinations to identify the best models, as shown in Figure 14. For the first hour prediction, we found that ST-DNN models with all components (A+L+C) outperformed all other models, and the CNN only model outperformed all other combinations for 2-6 hour predictions. Therefore, we include only A+L+C and C models in further discussions.

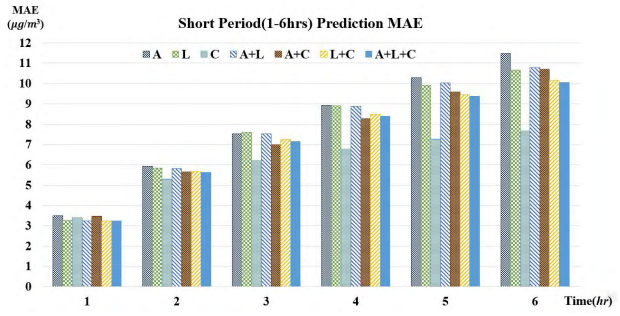



**FIGURE 14.** Proposed ST-DNN prediction model with different component combinations for the Taiwan Datase.

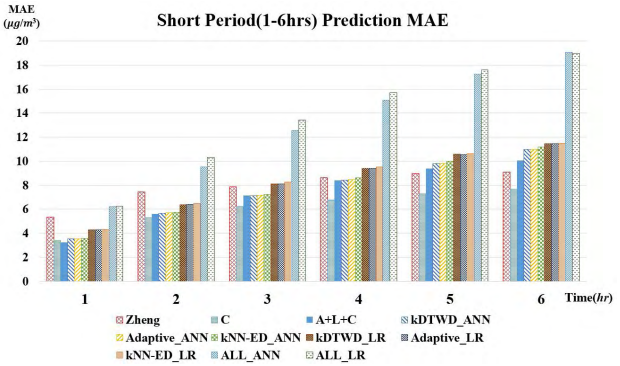



**FIGURE 15.** Short period (1-6 hour) prediction comparisons for the Taiwan dataset.

Figure 15 compares overall model performance at all locations. The proposed ST-DNN(A+L+C) model exhibits superior performance to all models to predict the next hour, whereas the ST-DNN(C) has superior prediction performance for further hours. The Zheng model [5] is superior to the other considered models except ST-DNN(C) for 4-6 hour predictions, since that model partitions regions at 30, 100, and 300 km, which helps detect diffusion from distant locations. This phenomenon happens as Zheng partitions regions with distances of 30km, 100km and 300km, which help to detect diffusions from other long distanced places. The proposed ST-DNN methods choose nearby or recently similar locations that have insufficient diffusion data. Although ST-DNN(C) has less information, it surpasses all other model performances (see Figure 15). In contrast, feeding all the data into ANN and Linear Regression models provided poorer prediction performance due to interference between locations. Generally, kNN-DTWD based models show superior performance to those based on kNN-ED, and Adaptive methods perform similarly to kNN-DTWD based models.

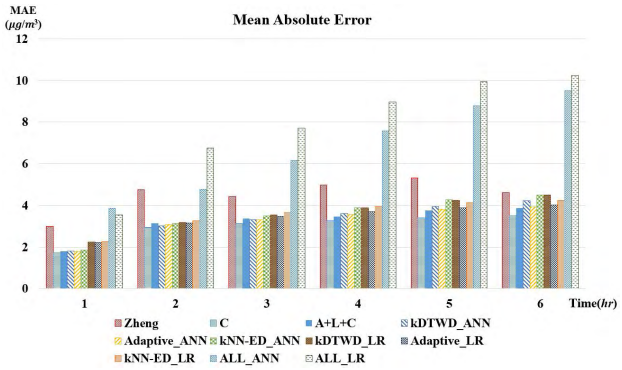



**FIGURE 16.** Eastern area short period (1-6 hour) prediction performance.

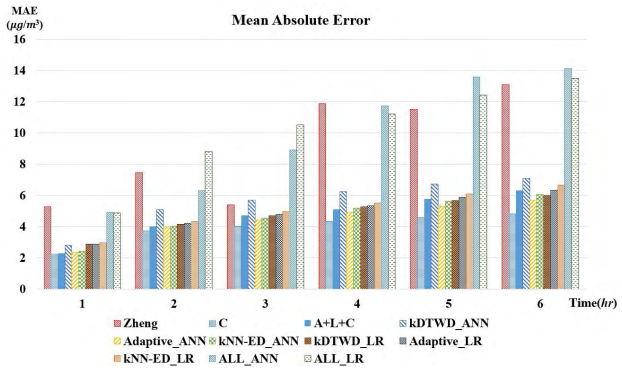



**FIGURE 17.** Northeastern area short period (1-6 hour) prediction performance.

Figure 16 and Figure 17 compare overall model performance in different cities. Most model trends are similar, with ST-DNN(C) outperforming all other models in some cities, including Hua-Tung (Figure 16) and ILan (Figure 17), which have sparse locations and complicated terrain. CNN extracted neighborhood elevation and determined diffusion delays (direction and time) for target features. For example, when predicting $t+3$, higher weight filters were farther from the target location. Thus, CNN may be a good approach for further study.

However, CNN also has limitations where reliable data for each coordinate are unavailable. This study used IDW to interpolate location data for each coordinate. The Zheng model also showed poor performance in these areas due to the complex terrain, with mean and mode in partitions leading to inaccurate estimation. However, the Zheng model exhibits superior performance in the southern area, particularly for 3-6 hour. The southern area is somewhat flatter, as shown in Figure 18, and hence a given location may be significantly affected by dispersion from other cities.

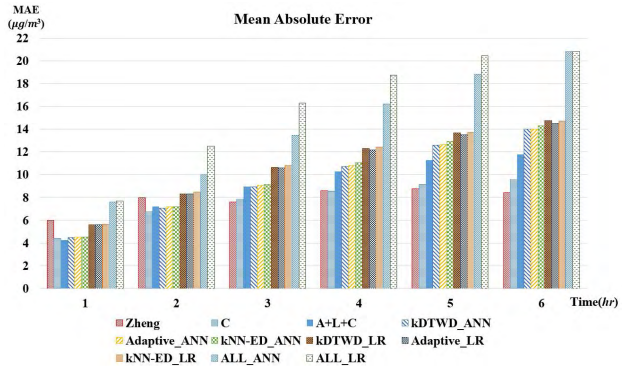



**FIGURE 18.** Southern area short period (1-6 hour) prediction performance.

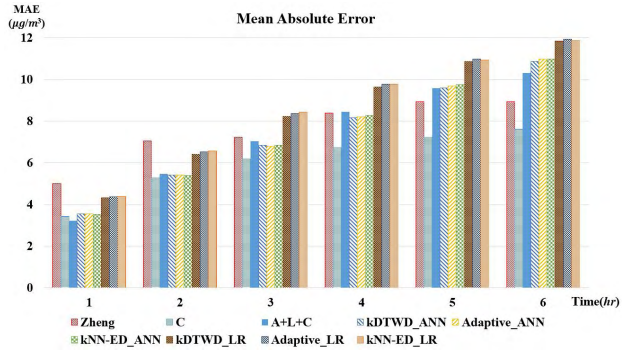



**FIGURE 19.** Flatland city short period (1-6 hour) prediction performance.

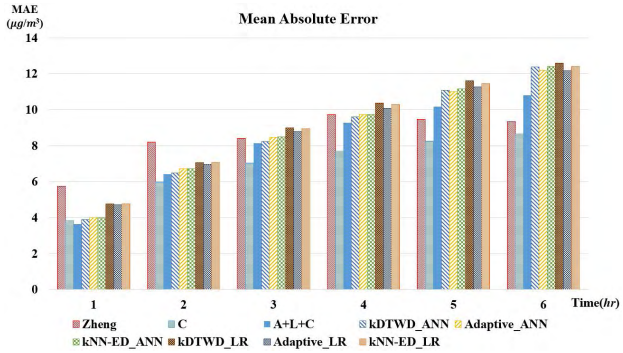



**FIGURE 20.** Flatland suburbs short period (1-6 hour) prediction performance.

Figures 19–23 group the Taiwan dataset prediction results by different terrain: flatland city, flatland suburb, basin city,

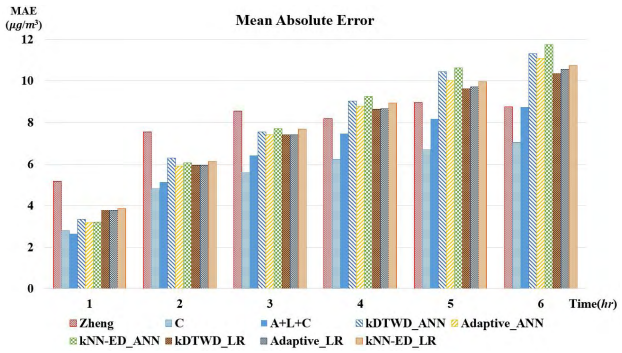



**FIGURE 21.** Basin city short period (1-6 hour) prediction performance.

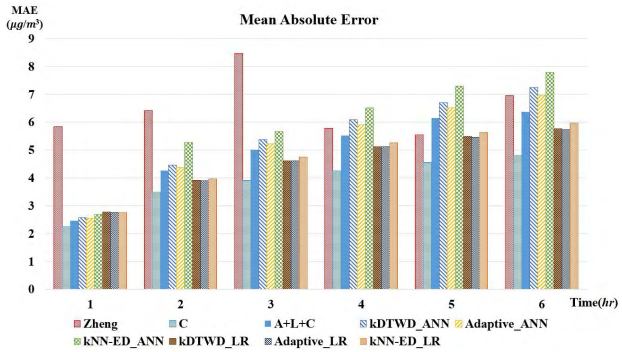



**FIGURE 22.** Mountain city short period (1-6 hour) prediction performance.

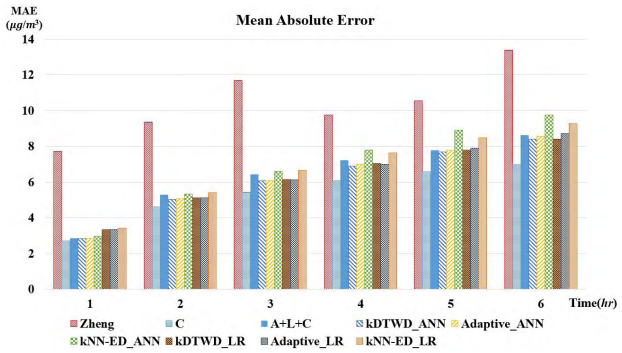



**FIGURE 23.** Island short period (1-6 hour) prediction performance.

mountain city, and island, respectively. ST-DNN(C) exhibits the best flatland performance (Figures 19 and 20) for all locations, with the Zheng model superior to other models for 5-6 hour, except ST-DNN(C), due to the lack of distant location data for the other models. However, the Zheng model performs poorly for complex terrain (Figures 21–23).

Overall, ST-DNN(A+L+C) provides superior prediction for the immediate next hour, whereas ST-DNN(C) provides superior prediction for 2-6 hour. This dataset shows that different places should use different models considering the impact of distances, delays, and terrains among locations and surroundings. The proposed approach always provides superior performance to adaptive baseline models verifying that considering individual feature trends is more appropriate

than combining inputs. The proposed models also outperform the Zheng, confirming the importance of observations at individual locations, rather than using neighborhood averages. Thus, the proposed ST-DNN(C) model is a suitable approach for further study.

#### 2) BEIJING DATASET

We modified the models somewhat for the Beijing dataset to ensure robust prediction. We neglect forecast data, since these were based on physical models, and omitted the CNN component from the proposed ST-DNN model since elevations were not available in Beijing.

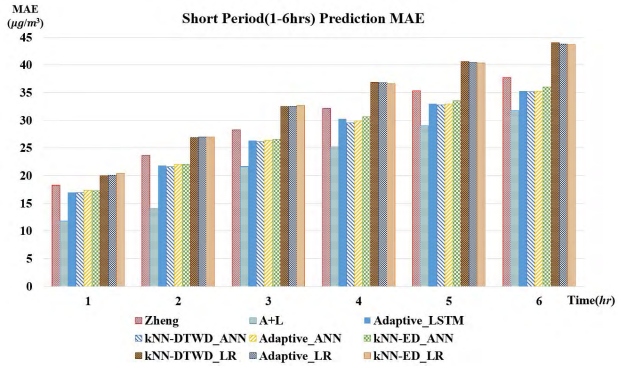



**FIGURE 24.** Short period (1-6 hour) prediction performance for the Beijing dataset.

Figure 24 shows that the proposed Adaptive_LSTM, kNN-DTWD_ANN, and Adaptive_ANN models outperform all other models. Linear regression models exhibit the poorest prediction among the considered models. The Zheng model exhibits intermediate performance between ANN and linear regression, because using TRE with region mean or mode results in loss of sensitivity to other locations. There are also mountainous regions in the north, northwest, and west of Beijing City, and partitioning across mountain regions also reduces precision.

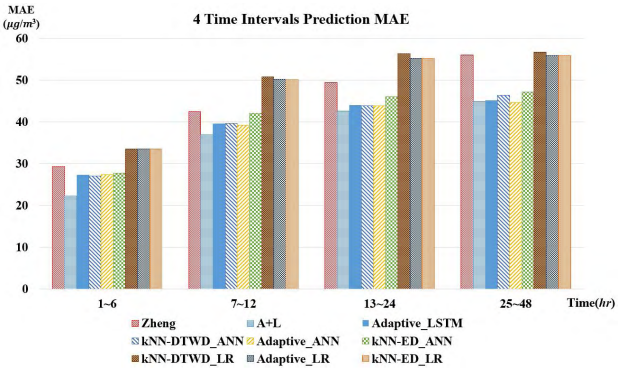



**FIGURE 25.** Extended time interval prediction performance for the Beijing dataset.

Figure 25 shows that overall model performances are similar to short time performances (Figure 24).

### E. BEHAVIOUR OF PROPOSED MODEL

#### 1) ANALYSIS OF LSTM AND CNN

We used the Taiwan dataset Tainan training data to investigate patterns between different delays and identify LTSM prediction improvements, comparing PM2.5 variations from $t_{c+0}$ versus $t_{c+1}$, $t_{c+0}$ versus $t_{c+2}$ and $t_{c+0}$ versus $t_{c+3}$ as shown in Figures 26–28, respectively. Very short prediction (Figure 26) exhibits almost linear then a sharp rise and subsequent sharp drop; whereas this linear relationship disperses for longer prediction times (Figures 27 and 28). Thus, LSTM only helps improve first hour prediction.

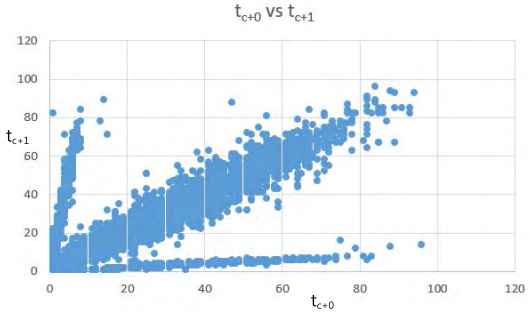



**FIGURE 26.** Example relationships between $t_{c+0}$ and $t_{c+1}$.

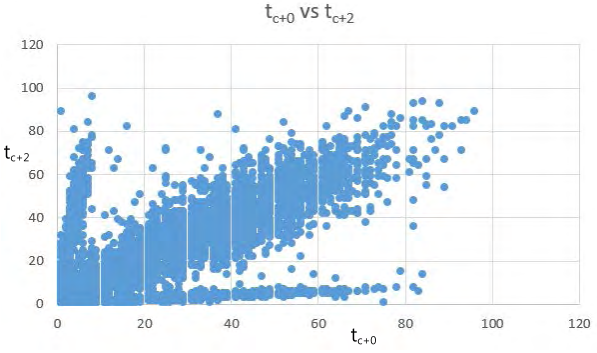



**FIGURE 27.** Example relationships between $t_{c+0}$ and $t_{c+2}$.

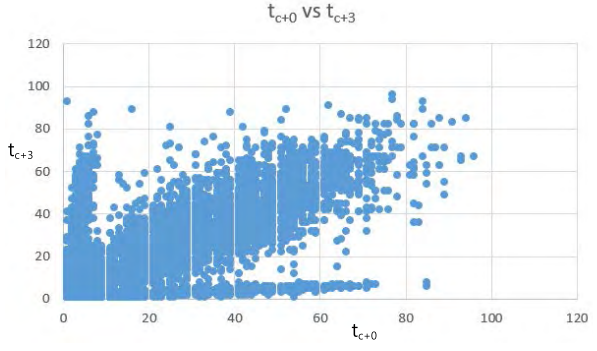



**FIGURE 28.** Example relationships between $t_{c+0}$ and $t_{c+3}$.

Limitations of DTW mean that kNN-DTWD chooses the most similar candidate locations but neglects long time delay. However, the delay interval is important for long term predictions. The proposed method could be improved by shifting

delayed sequences for DTW identified candidates and choosing the most relevant locations. Comparing different components combinations in ST-DNN, CNN always improves model performance.

**TABLE 1. Proposed model with CNN (ST-DNN(C)) with and without elevation.**

| ST-DNN(C) | Hours | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| + elevation | **3.402** | **5.308** | **6.213** | 6.782 | **7.269** | **7.668** |
| - elevation | 3.410 | 5.310 | 6.215 | **6.774** | 7.270 | 7.669 |

Including relative elevation helps reduce location interference and provides excellent prediction performance for 1-6 hour with ST-DNN(C). Table 1 compares performances with and without relative elevations, confirming the observed improvement. Lu *et al.* [32] showed that fine particulate matter concentrations decrease at higher altitude, hence relative elevation is important and should be considered in further studies. CNN can also extract the delay factor from surrounding target features, improving prediction performance.

In particular, Taiwan includes significant many mountainous and hilly terrain regions. Since CNN with terrain factors can learn the propagation patterns from PM2.5 variations of the specific location and its neighbor locations, the proposed ST-DNN(C) model was superior to LSTM, which only considered time series variation at the given location. Hence, the experimental results confirm that CNN provides superior prediction performance in complex terrain.

Thus, the proposed ST-DNN(A+L+C) model provided superior performance for first hour predictions, but the proposed ST-DNN(C) model provides superior longer time frame predictions through 2-6 hour. Overall, ST-DNN has two main characteristics: selecting spatial-temporal candidates using kNN-ED and kNN-DTWD increased model sensitivity; and including terrain information using CNN incorporates elevation impacts.

### 2) INFLUENCE OF *k*

Figure 29 compares kNN-ED and kNN-DTWD with ANN to investigate the performance effects of $k$. kNN-ED MAE increases as $k$ increases, whereas kNN-DTWD MAE decreases. Thus, kNN-DTWD outperforms kNN-ED, which only considers spatial adjacent locations, omitting terrain, whereas kNN-DTWD chooses temporal similar locations, with similar responses, relaxing geographical impact restrictions.

## VI. REAL APPLICATIONS FOR THE PROPOSED MODEL

Figure 30 shows the web user interface of the proposed system [16], where icons on the map represent monitoring stations and the number associated with an icon denotes its PM2.5 concentration. The color of the icons indicates the air quality at that location based on TWEPA PM2.5

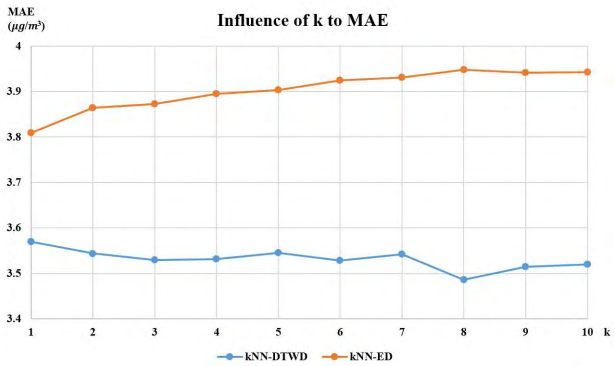



**FIGURE 29. Nearest neighbor ($k$) effect on prediction performance.**

standards [26]. In addition, users can see the predictions of PM2.5 for the next 1-6 hour. More details can be shown by clicking on a specific station, which opens a pop-up chart showing recent trends in the air quality and meteorological conditions. Clicking on the analysis of historical data allows users to select results for past 1, 3, or 7 days. Clicking on the replay icon opens a timeline, which allows users to check the air quality at any time in a particular location to observe the PM2.5 diffusion.

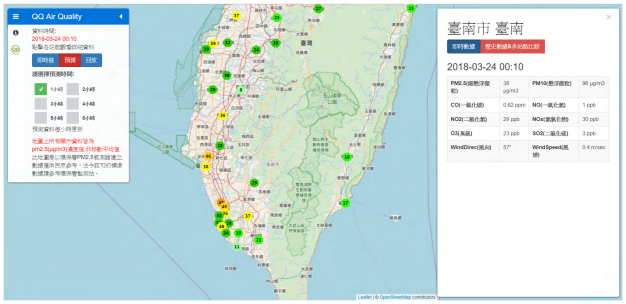



**FIGURE 30. Website of QQ air quality.**

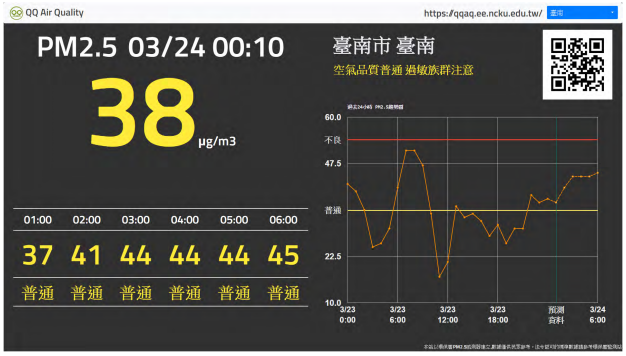



**FIGURE 31. Dashboard of QQ air quality.**

Figure 31 shows the dashboard interface of the real-time PM2.5 and the PM2.5 predictions for the next 1-6 hour. The dashboard will automatically update information every hour. In addition, users can choose any specific station on

their demands. This service aims to help people prevent the exposure of unhealthy air.

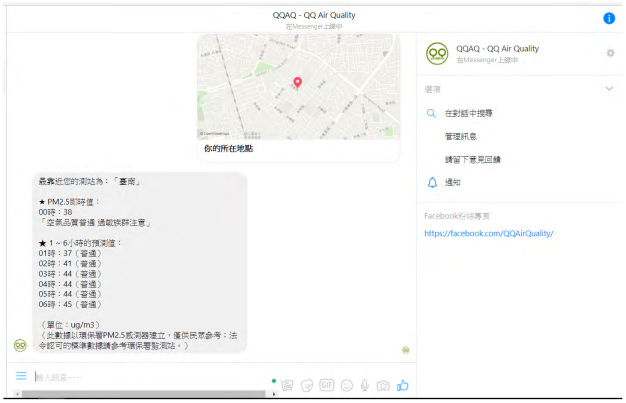



**FIGURE 32.** Facebook chatbot of QQ air quality.

In addition, we develop the Facebook chatbot to send the PM2.5 forecasts to users who subscribed the daily report of specific stations, as shown in Figure 32. Users can query the current PM2.5 by typing the station name or sending current GPS location. This service benefits people to easily use it on mobile phone or PC, which reminds people the air quality before going outside.

## VII. CONCLUSION

This study proposed an air quality forecasting system using data driven models, ST-DNN, to predict PM2.5 over 48 hours. The proposed method is also generally applicable to other pollutants, etc.

The proposed ST-DNN shows that including an LSTM module enhanced first hour predictions, with CNN module inclusion being more useful for longer time frame predictions, since CNN can extract the temporal delay factor from surrounding target features by learning spatial information.

We evaluated the proposed models using real-world Taiwan and Beijing datasets. Relevant location selection was verified to be important, with inclusion of all locations causing increased model noise and hence poorer prediction performance. the proposed methods outperformed all baselines and comparative models considered.

Future research will improve the proposed model performances, and consider specific Airbox sensor source models as features to tune and mitigate noise due to machine differences [33], [34]. We will also consider more chemical features that affect PM2.5 components [35]. For long period predictions, we will consider concentric circles for different distance partitions or clusters to emphasize air pollution propagation delay effects.

Ultimately, we intend to detect air pollution sources, including domestic and transboundary pollution. To control the air pollution, we must first understand how it is generated and propagated. Only then can we devise effective solutions to reduce pollution sources. Therefore, future pollution studies will be greatly dependent on the proposed model.

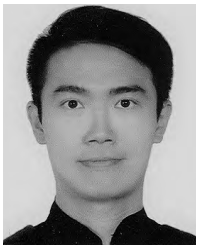
## ACKNOWLEDGMENT

The authors are grateful to the Taiwan Environmental Protection Administration for providing the monitoring data used in this study.

## REFERENCES

[1] Y.-F. Xing, Y.-H. Xu, M.-H. Shi, and Y.-X. Lian, "The impact of $PM_{2.5}$ on the human respiratory system," *J. Thoracic Disease*, vol. 8, no. 1, pp. E69–E74, 2016.
[2] H.-J. Chu, C.-Y. Lin, C.-J. Liau, and Y.-M. Kuo, "Identifying controlling factors of ground-level ozone levels over southwestern Taiwan using a decision tree," *Atmos. Environ.*, vol. 60, pp. 142–152, Dec. 2012.
[3] A. P. K. Tai, L. J. Mickley, and D. J. Jacob, "Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: Implications for the sensitivity of $PM_{2.5}$ to climate change," *Atmos. Environ.*, vol. 44, no. 32, pp. 3976–3984, 2010.
[4] C.-M. Liu, C.-Y. Young, and Y.-C. Lee, "Influence of Asian dust storms on air quality in Taiwan," *Sci. Total Environ.*, vol. 368, nos. 2–3, pp. 884–897, 2006.
[5] Y. Zheng *et al.*, "Forecasting fine-grained air quality based on big data," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, 2015, pp. 2267–2276.
[6] Y. Hwa-Lung and W. Chih-Hsin, "Retrospective prediction of intraurban spatiotemporal distribution of $PM_{2.5}$ in Taipei," *Atmos. Environ.*, vol. 44, no. 25, pp. 3053–3065, 2010.
[7] B. S. Beckerman *et al.*, "A hybrid approach to estimating national scale spatiotemporal variability of $PM_{2.5}$ in the contiguous united states," *Environ. Sci. Technol.*, vol. 47, no. 13, pp. 7233–7241, 2013.
[8] H.-J. Chu, H.-L. Yu, and Y.-M. Kuo, "Identifying spatial mixture distributions of $PM_{2.5}$ and $PM_{10}$ in Taiwan during and after a dust storm," *Atmos. Environ.*, vol. 54, pp. 728–737, Jul. 2012.
[9] H.-W. Chen, C.-T. Tsai, C.-W. She, Y.-C. Lin, and C.-F. Chiang, "Exploring the background features of acidic and basic air pollutants around an industrial complex using data mining approach," *Chemosphere*, vol. 81, no. 10, pp. 1358–1367, 2010.
[10] A. Kurt and A. B. Oktay, "Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7986–7992, 2010.
[11] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 1436–1444.
[12] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[14] *Air Quality Index Historical Data*. Accessed: Aug. 22, 2016. [Online]. Available: http://taqm.epa.gov.tw/taqm/tw/YearlyDataDownload.aspx
[15] Y. Zheng *et al.* (2015). *Forecasting Fine-Grained Air Quality Based on Big Data*. [Online]. Available: http://research.microsoft.com/apps/pubs/?id=246398
[16] *QQ Air Quality(QQAQ)*. Accessed: Jun. 6, 2018. [Online]. Available: https://qqaq.ee.ncku.edu.tw
[17] *QQAQ Facebook Fan Page*. Accessed: Jun. 6, 2018. [Online]. Available: https://www.facebook.com/QQAirQuality/
[18] S. Qin, F. Liu, C. Wang, Y. Song, and J. Qu, "Spatial-temporal analysis and projection of extreme particulate matter ($PM_{10}$ and $PM_{2.5}$) levels using association rules: A case study of the Jing-Jin-Ji region, China," *Atmos. Environ.*, vol. 120, pp. 339–350, Nov. 2015.
[19] P.-W. Soh, K.-H. Chen, J.-W. Huang, and H.-J. Chu, "Spatial-temporal pattern analysis and prediction of air quality in taiwan," in *Proc. Int. Conf. Ubi-Media Comput. (UMedia)*, Aug. 2017, pp. 1–6.
[20] T. Rakthanmanon *et al.*, "Searching and mining trillions of time series subsequences under dynamic time warping," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 262–270.
[21] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowl. Inf. Syst.*, vol. 7, no. 3, pp. 358–386, 2005.
[22] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proc. 23rd ACM Nat. Conf.*, 1968, pp. 517–524.
[23] *METAPP*. Accessed: Jun. 6, 2018. [Online]. Available: http://www.metapp.org.tw/index.php/weatherknowledge/37-typhoon/83-2009-01-22-08-04-48

[24] H. L. Chang, "Evaluation and application of the short-range (0-6hr) pqpfs from an ensemble prediction system based on laps," Ph.D. dissertation, Graduate Inst. Atmos. Phys., Nat. Central Univ., Taoyuan, Taiwan, 2014.
[25] L. Ferrero, G. Mocnik, B. S. Ferrini, M. G. Perrone, G. Sangiorgi, and E. Bolzacchini, "Vertical profiles of aerosol absorption coefficient from micro-Aethalometer data and Mie calculation over Milan," *Sci. Total Environ.*, vol. 409, no. 14, pp. 2824–2837, 2011.
[26] *PM2.5 Index*. Accessed: Aug. 22, 2016. [Online]. Available: http://taqm.epa.gov.tw/taqm/en/fpmi.htm
[27] *TWEPA Instruments*. Accessed: Jun. 6, 2018. [Online]. Available: https://taqm.epa.gov.tw/taqm/tw/b0102-3.aspx
[28] Y. Pan, Y. Liu, B. Xu, and H. Yu, "Hybrid feedback feedforward: An efficient design of adaptive neural network control," *Neural Netw.*, vol. 76, pp. 122–134, Apr. 2016.
[29] J. de Jesús Rubio, "Stable Kalman filter and neural network for the chaotic systems identification," *J. Franklin Inst.*, vol. 354, no. 16, pp. 7444–7462, 2017.
[30] Y. Pan and H. Yu, "Biomimetic hybrid feedback feedforward neural-network learning control," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 6, pp. 1481–1487, Jun. 2017.
[31] J. de Jesús Rubio, "Error convergence analysis of the SUFIN and CSUFIN," *Appl. Soft Comput.*, to be published. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494618301881
[32] S.-J. Lu, D. Wang, X.-B. Li, Z. Wang, Y. Gao, and Z.-R. Peng, "Three-dimensional distribution of fine particulate matter concentrations and synchronous meteorological data measured by an unmanned aerial vehicle (UAV) in yangtze river delta, China," *Atmos. Meas. Techn. Discuss.*, pp. 1–19, Mar. 2016. [Online]. Available: https://www.atmos-meas-tech-discuss.net/amt-2016-57/
[33] L.-J. Chen, Y.-H. Ho, H.-H. Hsieh, S.-T. Huang, H.-C. Lee, and S. Mahajan, "ADF: An anomaly detection framework for large-scale $PM_{2.5}$ sensing systems," *IEEE Internet Things J.*, vol. 52, no. 2, pp. 559–570, Aug. 2018.
[34] N. Moustafa, G. Creech, E. Sitnikova, and M. Keshk. (2017). "Collaborative anomaly detection framework for handling big data of cloud computing." [Online]. Available: https://arxiv.org/abs/1711.02829
[35] W. M. Hodan and W. R. Barnard, "Evaluating the contribution of $PM_{2.5}$ precursor gases and re-entrained road emissions to mobile source $PM_{2.5}$ particulate matter emissions," MACTEC Federal Programs, Research Triangle Park, NC, USA, 2004.



**PING-WEI SOH** received the B.S degree in electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 2015, and the M.S degree from the Institute of Computer and Communication Engineering, National Cheng Kung University, in 2018. His research interests include spatial-temporal and air pollution issues.

**JIA-WEI CHANG** received the Ph.D. degree from the Department of Engineering Science, National Cheng Kung University in 2017. He is currently an Adjunct Assistant Professor with the Department of Engineering Science and a Post-Doctoral Fellow Researcher with the Department of Electrical Engineering, National Cheng Kung University. His research interests include natural language processing, artificial intelligence, data mining, and e-learning technologies.

**JEN-WEI HUANG** received the B.S. and Ph.D. degrees in electrical engineering from National Taiwan University, Taiwan, in 2002 and 2009, respectively. He was a Visiting Scholar with the IBM Almaden Research Center from 2008 to 2009, an Assistant Professor with Yuan Ze University from 2009 to 2012, and a Visiting Scholar with the University of Chicago in 2016. He is currently an Assistant Professor with the Department of Electrical Engineering, National Cheng Kung University, Taiwan. He majors in computer science and is familiar with data mining. His research interests include data mining, mobile computing, and bioinformatics. Among these, social network analysis, spatial-temporal data mining, and multimedia information retrieval are his special interests. In addition, some of his research is on data broadcasting, privacy preserving data mining, e-learning, and Fin-tech.