

PREDICTING DAILY PM_{2.5} USING WEIGHTED LONG SHORT-TERM MEMORY NEURAL NETWORK MODEL

Hong Fan^{1*}, Mei Yang¹, Fei Xiao², Kang Zhao¹

¹ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; yangmei2012@whu.edu.cn (M. Y.); hfan3@whu.edu.cn

² School of Public Health, Southern Medical University, Guangzhou, 510515 Guangdong, China; geoxiaofei@163.com

* Correspondence: hfan3@whu.edu.cn; Tel.: +86-18627716767

Commission III, WG III-IVc

KEY WORDS: PM_{2.5} prediction; long short-term memory neural network; multi-layer perception; spatiotemporal dependency

ABSTRACT:

Over the past few decades, air pollution has caused serious damage on public health, thus making accurate predictions of PM_{2.5} crucial. Due to the transportation of air pollutants among areas, the PM_{2.5} concentration is strongly spatiotemporal correlated. However, the distribution of air pollution monitoring sites is not even, making the spatiotemporal correlation between the central site and surrounding sites varies with different density of sites, and this was neglected by most existing methods. To tackle this problem, this study proposed a weighted long short-term memory neural network extended model (WLSTME), which addressed the issue that how to consider the effect of the density of sites and wind condition on the spatiotemporal correlation of air pollution concentration. First, several the nearest surrounding sites were chosen as the neighbour sites to the central station, and their distance as well as their air pollution concentration and wind condition were input to multi-layer perception (MLP) to generate weighted historical PM_{2.5} time series data. Second, historical PM_{2.5} concentration of the central site and weighted PM_{2.5} series data of neighbour sites were input into LSTM to address spatiotemporal dependency simultaneously and extract spatiotemporal features. Finally, another MLP was utilized to integrate spatiotemporal features extracted above with the meteorological data of central site to generate the forecasts future PM_{2.5} concentration of the central site. Daily PM_{2.5} concentration and meteorological data on Beijing–Tianjin–Hebei from 2015 to 2017 were collected to train models and evaluate the performance. Experimental results with 3 other methods showed that the proposed WLSTME model has the lowest RMSE (40.67) and MAE (26.10) and the highest p (0.59). This finding confirms that WLSTME can significantly improve the PM_{2.5} prediction accuracy.

1. INTRODUCTION

Over the past few decades, rapid economic growth worldwide has caused serious air pollution, which has elicited extensive attention worldwide. PM_{2.5} (particulate matter with a diameter less than 2.5 μm), as an important component of air pollutant, has been confirmed to be related with cardiopulmonary and other systemic diseases by penetrating the respiratory system (Manuel, 2013, Polezer, 2018). According to a recent WHO study (WHO, 2016), in 2012, about 90% people breathe the air that does not comply with the WHO Air Quality Guidelines, and about 3 million deaths worldwide are attributable to outdoor air pollution.

Considering the proven negative effect of air pollution, the public should be provided with accurate forecasts of daily PM_{2.5} concentration to help control air pollution and combat health problems. Many machine learning models have been applied widely for their great ability of handling nonlinear relationships and they generally provide satisfactory performance. Representative models include artificial neural network (Xiao, 2015, Ordieres, 2005), recursive neural network (Biancofiore, 2017), support vector regression (Hou, 2014), and hybrid model (Zhou, 2014, Qin, 2014). Recently, long short-term memory neural network (LSTM) (Hochreiter, 1997), has been used extensively for processing time series data due to its capability of simulating long and short-term tendency simultaneously. Li (2017) constructed a method of using the historical concentrations of all stations as the inputs of LSTM layer and integrating auxiliary variables by a fully-connected layer. Qin

(2019) proposed a combined model for forecasting a city's PM_{2.5} concentration. The model used convolutional neural network to automatically extract features of input data of all stations, and LSTM to consider the time dependency. Apart from those studies considering all stations, many studies addressed the spatial correlation by considering the most related sites. K Nearest neighbours method was extensively employed to determine the K nearest sites to the target site (Soh 2018, Zhao 2019, Wen 2019), and their meteorological conditions and air quality data were input to neural networks to capture spatial correlation. However, the distribution of monitoring stations is extremely uneven, causing the density of sites is quite different for different areas. The lower the station density, the farther the geographical distance of the selected K nearest sites to the target site, therefore the lower their affect degree. However, most of the existing studies neglected the importance of the station density. Considering that air pollutants are transported based on wind, some studies further combined the geographical distance with wind condition, and generated weights of each surrounding site to represent their affect degree to the target site. For example, a space-time support vector regression model (STSVR) was proposed in Yang (2018), which constructed a Gauss vector weight to combine the distance of surrounding sites and wind direction effects. Through integrating the weights and PM_{2.5} concentration of the surrounding sites, the spatial dependence was introduced into SVR model. Similarly, through a linear combination between the wind direction and geographical distance, Li and Gong (2014) proposed a wind field distance definition to represent the spatial correlation degree between two sites. Further, Li and Fan (2017) introduced the wind speed into the definition of wind field distance. Both of these studies

enhanced the interpolation accuracy, and indicated that the combination between wind and geographical distance have great potential for determining the spatial dependence.

To further deeply simulate the wind and geographical distance impacts on spatial correlation, the current study proposed a weighted LSTM neural network extended model (WLSTME) to forecast daily PM_{2.5} concentration. First, the K nearest stations were chosen as K neighbours of the target station, and local MLP was utilized to calculate the weighted PM_{2.5} concentration time series based on the historical PM_{2.5} concentration observation, geographical distance and historical wind condition of the K neighbours. Second, the weighted PM_{2.5} concentration series data of the selected neighbour stations and the historical PM_{2.5} observation of the target station were fed into LSTM layer, which simultaneously simulated spatial and temporal dependencies and extract spatiotemporal features. Finally, another MLP network was used to integrate the obtained spatiotemporal features with auxiliary variables, including meteorological variables and time stamp data, and generate the prediction of the next day's PM_{2.5} concentration of the target station. Through the nonlinear combination by MLP, the proposed model can more effectively simulate the impact of station density and wind on spatial dependency, and avoid the remote sites' interference to the prediction accuracy. Daily average PM_{2.5} concentration and meteorological data of Beijing, Tianjin, and Hebei in China collected from 2015 to 2017 were employed as experimental data here. We conducted comparison experiments with other three methods, and the results demonstrated the effectiveness of our model in predicting daily PM_{2.5} concentration.

1. MATERIALS AND METHODS

1.1 Study Area and Data

2.1.1 Study Area: Beijing-Tianjin-Hebei (BTH) region of China contains Beijing, Tianjin, and 11 cities of Hebei Province, and it has become one of the most economical active area in China at present. According to (CSY 2018), in 2017, the region GDP of BTH contributed 9.77% of the total of China, with the region population occupied 8.09%. What followed is serious air pollution and its damage to public health which cannot be ignored. In 2012, to bring the new air quality standards into force, the General Office of the Ministry of environmental protection of China listed 74 cities as typical cases to develop measurements of new indicators such as PM_{2.5}, CO, O₃ and so on. (MEP 2018) reported the comparison of the annual average air quality composite index of these 74 cities in 2017. Among the top twenty most polluted cities, 9 cities belong to Hebei Province, and Tianjin is fifteenth, while Beijing is nineteenth. In BTH region, averagely 44% days in 2017 didn't meet the standards, and additionally, there are over half over-polluted days when PM_{2.5} is the primary pollutant. Thus, this paper took BTH region as study area, and collected their data from Jan/01/2015 to Dec/31/2017.

2.1.2 Data collection and pre-process: Daily average PM_{2.5} concentration data of 110 monitoring stations in BTH area were collected from <http://beijingair.sinaapp.com/> and the locations are shown in 6. 1. Due to critical failure or temporary power cutoff, missing values for a long or short periods happened in air pollution monitoring stations (junninen, 2004). We dropped out data of 20 stations where daily average PM_{2.5} concentration had over 10% missing days. The rest missing values were interpolated day by day by inverse distance interpolation method. Finally, we replaced the values which lied outside three-sigma interval with the mean of totally 15 days before and after this day.

There are 98640 records of 90 stations remained. We clustered these stations into 12 groups according to their geographical locations and sorted them by latitude. The number of stations and average daily PM_{2.5} concentration of each group are listed in supplementary materials. Results show that the lower the latitude, the higher average PM_{2.5} is, which complies with related results of spatial distribution research (Li, 2016).

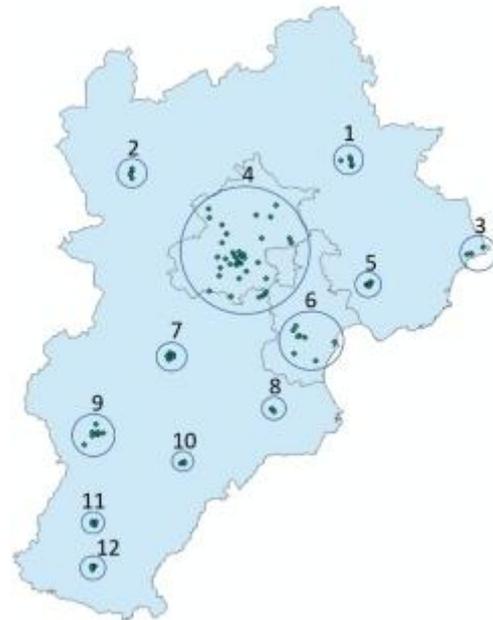


Figure 1 Location of monitoring stations.

Auxiliary variables mainly include two parts: meteorological factors and time stamp data. Meteorological factors we chose are temperature, wind speed, wind direction, dew point temperature, mean sea level pressure and total column water vapor, and they were downloaded from ECMWF (<http://apps.ecmwf.int/datasets/data/cams-realtime/levtype=sfc/>). Additionally, we downloaded MOD11A1 data from NASA (<https://modis.gsfc.nasa.gov/>) as basic temperature data, with negative values seemed as invalid. The temperature data from the two sources were both standardized and zero-centered respectively, and we conducted a linear regression: standardized MODIS data = $-0.14 * \text{standardized ECMWF data} + 0.97$, with $R=0.899$, which shows good feasibility of replacing MODIS data with ECMWF data. Thus, the invalid data of MODIS were substituted with the predicted value according to ECMWF data. Time stamp data include what month and what day is the day. Moreover, considering the relationships between latitude with PM_{2.5} concentration, the latitude is contained in this study, too.

All variables excluding time stamp data were zero-centered and standardized, and time stamp data were one-hot encoded. For example, month was encoded as a 11-dimensions vectors, with the i th element equalled 1 or 0 to denote whether this day is the i th month or not, and specifically, December is encoded as (0,0,0,0,0,0,0,0,0,0,0).

2.1.3 Dataset partition: Considering that our data is special in that it has strong time sequence, which means that we cannot use data at time $t+1$ to predict value at time t , so ordinary random division is not suit for this case. Instead, we divided our data into three parts according to sequence of time: data of 2015 and 2016 are training set and validation set respectively, which are used for constructing the best architecture of models, and data of 2017 is test set, which is used to evaluate the performance of various models. The reason why we divided data by year rather than by

month is that many studies (Yan et al. (2018); Ye et al. (2018)) have shown that PM2.5 concentration has a strong seasonal difference, thus each dataset must include at least one year to cover seasonal influence.

1.2 Methods

Spatial correlation is negatively related with distance between sites. The nearer to the central site, the higher the affect degree of surrounding sites. The density of stations affects the distance between surrounding sites and central site, and further affects the spatial correlation. We proposed a WLSTME model to take the density of stations into consideration by a nonlinear combination between distance, wind and PM2.5 concentration.

1.2.1 WLSTME neural network: Figure 2 show the overall framework of WLSTME model. The **input** originated from two parts (see green frames in Figure 2). The historical PM2.5 concentration and wind of the central and surrounding sites at past r days; and auxiliary variables of central site. The output was the prediction of the next day's PM2.5 concentration of the central site. The architecture of WLSTME consists of three parts: using MLP to combine PM2.5 with distance and wind to generate weighted PM2.5 for each neighbour site; using stateful LSTM to extract spatiotemporal features; and using MLP to integrate auxiliary variables. The processes of each part are introduced as follows.

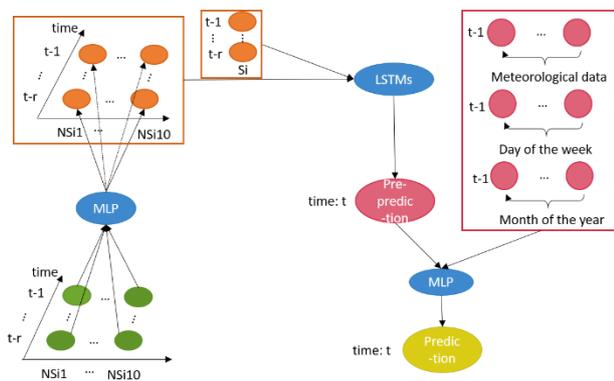


Figure 2 Framework of WWLSTME model.

First, we generated weighted PM2.5 series data for each neighbour site. neighbour sites were defined as the K nearest surrounding sites to the central site. Since pollutants are transported among areas based on wind, air pollution of central site is spatially correlated with that of neighbour sites. However, the distribution of monitoring stations is not even. Consequently, the distance between neighbour sites and central site is different for different central site. For example, the density of stations in south area is much sparser than Beijing as Figure 1 shows. Thus, for central sites in south area, the selected neighbour sites were more distant, and the spatial correlation was lower than that for sites in Beijing. Based on the above consideration, the geographical distance of the selected neighbour sites should be considered into model.

MLP can theoretically approximate any Borel measurable function with arbitrary precision (Hornik, 1989). This study proposed an MLP layer to integrate the distance and wind of neighbour sites with its PM2.5 to generate weighted PM2.5 series data for each neighbour site. The structure of MLP is as Figure 2 shows. PM2.5 and V_{jt} represent the PM2.5 concentration and wind speed of the j th neighbour site at time t , respectively; d_{ij}

represents the distance between the central site i and its j th neighbour site; θ_{ijt} represents the angle between the wind direction of j th neighbour site at time t and the edge between i and j . H_1, \dots, H_n are neurons of the hidden layer, and $WPM_{2.5}$ is the weighted PM2.5 concentration which is calculated by the following equations.

$$H_s = g(\omega_{1s}^1 PM_{2.5_{jt}} + \omega_{2s}^1 d_{ij} + \omega_{3s}^1 v_{jt} + \omega_{4s}^1 \theta_{ijt}) \quad (1)$$

$$WPM_{2.5_{jt}} = \sum_{s=1}^n \omega_s^2 H_s \quad (2)$$

Where g is the activation function used for the nonlinear transformation of inputs. ω is the weight between the neuron of previous layer and the next layer.

Next, we extracted spatiotemporal features from the pollution data of central site and neighbour sites. The weighted PM2.5 series data of neighbour sites and PM2.5 concentration observation data of the central site were merged as a 2D matrix, with each column represented the historical PM2.5 concentration of the central site or weighted PM2.5 concentration of a neighbour site. The size was $r \times (k+1)$, where r was time lag, and K was the number of selected neighbour sites. LSTM is a special recurrent neural network, with its recurrent neuron simultaneously captures long and short dependencies in time series data. LSTM has been used in many fields, such as financial market predictions (Fischer, 2017), epileptic seizures (Tsiouris, 2018), and reservoir operation (Zhang, 2018). All of the LSTM models used in these fields exhibited better performance than many other machine learning methods. The LSTM model used in our model was a two-layer stateful LSTM, which used the state of the current batch of LSTM samples as the initial state of the next batch of samples. It is more suitable for processing long-term time series data than the other models. The structure of the recurrent memory cell of LSTM model is shown in Figure 3.

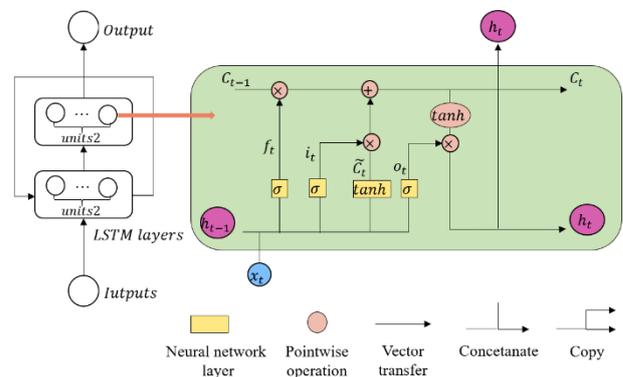


Figure 3. The illustration of two-layers LSTM model. x_t and h_t are the inputs and outputs at time t respectively.

Three key gates, namely, forget gate (f_t), input gate (i_t), and output gate (o_t), of LSTM are designed to control the memory of new information and to forget old information. The values of the three gates are updated with time by the following equations.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (5)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \quad (6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (7)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (8)$$

Where x_t and h_t are the inputs and outputs at time t, respectively; σ and \tanh are widely used activation functions; and W and b are the weight matrix and bias vector, respectively.

Finally, auxiliary variables were introduced to the WLSTME model to promote prediction accuracy. The auxiliary variables considered in this study included meteorological data (temperature, wind speed, dew point temperature, mean sea level pressure and total column water vapor) and time stamp data (day of week and month of the year) and latitude of the central site at time T. We integrated the auxiliary variables with the spatiotemporal features extracted by LSTM and input them into MLP to output the prediction of the next day's PM2.5 concentration of the central site. The structure of MLP was the same as Figure 2 (b), however, the input and output were substituted by the spatiotemporal features and PM2.5 concentration prediction, respectively.

2.2.3. Evaluation methods : The experimental data employed in this study was the air pollution and meteorological data of 110 stations in BTH region collected from 01 Jan., 2015 to 31 Dec., 2017. The data of 2015, 2016 and 2017 were set as train, valid and test set, respectively. Three criteria, namely, mean absolute error (MAE), root mean square error (RMSE), and total accuracy index (p), were used in the experiments to evaluate the effectiveness of our model. Smaller RMSE and MAE result in lower prediction error of the model. Higher p value result in better prediction accuracy of the model. The formulas of these criteria are defined as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^*|, \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2}, \quad (10)$$

$$p = 1 - \frac{\sum_{i=1}^n |y_i - y_i^*|}{\sum_{i=1}^n y_i} \quad (11)$$

where n is sample size, y_i is the observation of the ith sample, and y_i^* is the corresponding prediction.

3. Results and Discussions

Several parameters, including time lags (r), number of considered neighbour stations (K), number of neurons in the hidden layer in the first MLP, LSTM, and the final MLP (denoted by units 1, 2, and 3 respectively), and batch size of training need to be determined before building the WLSTME model. The best parameters were determined through a trial-and-error method with a low RMSE in the training and validation sets, where r=10, s=10, unit1=10, unit2=10, unit3=20, and batch size=7*90.

We conducted comparative experiments between the proposed WLTMSE model and other three models that also consider spatiotemporal dependency, including geographically weighted regression (GWR) model, LSTME model and STSVR model (Yang,2018). GWR is able to consider the spatial heterogeneity and spatial correlation among different sites. LSTME refers to a neural network which has the same architecture as our model, except for the first MLP network, which means that the historical PM2.5 data of S neighbour stations at past T days are directly

merged with the data of the central station and formed the inputs of the LSTM layers. STSVR is an improved spatiotemporal form of SVR by integrating the wind and PM2.5 concentration of spatially correlated neighbour sites into SVR model.

Tab. 2 shows the performance of GWR, LSTME, STSVR, and WLSTME models on the test set. It can be found that WLSTME and STSVR exhibited better forecasting performance than GWR and LSTME, with lower RMSE, lower MAE, and higher PP. These results indicate that

Model	RMSE	MAE	p
GWR	49.34	35.84	0.43
LSTME	51.81	36.58	0.42
STSVR	42.60	27.80	0.56
WLSTME	40.67	26.10	0.59

TABLE I. THE PREDICTION PERFORMANCE OF GWR, LSTME, STSVR AND WLSTME MODELS ON THE TEST SET.

Table 2 shows the performance of GWR, LSTME, STSVR, and WLSTME models on the test set. It can be found that WLSTME and STSVR exhibited better forecasting performance than GWR and LSTME, with lower RMSE, lower MAE, and higher p. These results indicate that the geographic distance and wind condition of neighbour stations have great significance on the spatiotemporal correlation, and integrating the distance with wind condition to evaluate the spatiotemporal dependency can considerably promote the prediction accuracy. Furthermore, in comparison with STSVR which integrated wind direction and distance in a Gauss vector weight definition, the proposed WLSTME model showed lower RMSE, lower MAE, and higher p, indicating that the MLP network can better simulate the sophisticated influence of geographic distance and wind on the spatiotemporal correlation than the Gaussian form.

4. Conclusions

In this study, we developed a WLSTME model to predict the daily average PM2.5 concentration of a specific station with the uneven distribution of monitoring sites were considered. First, MLP was used to combine historical wind speed and wind direction with corresponding days' PM2.5 data of neighbour stations and generate wind-weighted PM2.5. Second, the wind-weighted PM2.5 of the neighbour stations in the past 10 days were merged with the historical PM2.5 data of the central station, and input into LSTM layer to address the spatial and temporal dependence simultaneously and extract spatiotemporal features. Finally, another MLP was used to conduct bias adjustment by integrating the spatiotemporal features with the central site's meteorological data and time stamp data.

Comparative experiments were conducted on the data of 2017 in the BTH area. RMSE, MAE, and p were used to quantify the forecasting performance and calculated in all subsets. All results showed in BTH region, WLSTME performed higher predictive accuracy and reliability than STSVR, LSTME, and GWR. The main reason for this superiority is that WLSTME not only addresses spatial and temporal dependencies simultaneously but also exhibited an exceptional capability to simulate the impact of wind and geographical distance on such dependencies.

In the future, the focus should be put on the prediction of the sudden increase in PM2.5, especially in winter when all models performed poorly. The inherent laws of winter could be different from those of other seasons, and a specific model can enhance

accuracy. Additionally, other humanistic and economic factors, such as the influence of government policies and the number of factories in the area should be introduced. Finally, more sophisticated methods for taking the density of sites into consideration can be investigated.

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (Grant No. 41471323).

REFERENCES

- Biancofiore, F.; Busilacchio, M.; Verdecchia, M.; Tomassetti, B.; Aruffo, E.; Bianco, S.; Tommaso, S. D.; Colangeli, C.; Rosatelli, G.; Carlo, P. D., Recursive neural network model for analysis and forecast of PM10 and PM2.5. *Atmospheric Pollution Research* 2017, 8 (4), 652-659.
- Fischer, T.; Krauss, C., Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research* 2017.
- Hochreiter, S.; Schmidhuber, J., Long Short-Term Memory. *Neural Computation* 1997, 9 (8), 1735-1780.
- Hou, W.; Li, Z.; Zhang, Y.; Hua, X.; Ying, Z.; Li, K.; Li, D.; Peng, W.; Yan, M., Using support vector regression to predict PM10 and PM2.5. 2014, 17 (1), 012268.
- Hornik, K., Multilayer feedforward networks are universal approximators. 1989, 2 (5), 359-366.
- Junninen, H.; Niska, H.; Tuppurainen, K.; Ruuskanen, J.; Kolehmainen, M., Methods for imputation of missing values in air quality data sets. *Atmospheric Environment* 2004, 38 (18), 2895-2907.
- Li, J., Fan, Z., Deng, M. (2017). A method of spatial interpolation of air pollution concentration considering wind direction and speed. *J. Geo-inf. Sci. (in Chinese)*, 19(3), 382-389.
- Li, L.; Gong, J.; Zhou, J., Spatial interpolation of fine particulate matter concentrations using the shortest wind-field path distance. *Plos One* 2014, 9 (5), e96111.
- Li, X.; Peng, L.; Yao, X.; Cui, S.; Hu, Y.; You, C.; Chi, T., Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution* 2017, 231 (Pt 1), 997-1004.
- Manuel, A. L. G.; Santibañez, D. A.; Sergio, I. E.; Patricia, M. C.; Seguel, R., A five-year study of particulate matter (PM2.5) and cerebrovascular diseases. *Environmental Pollution* 2013, 181 (1), 1-6.
- Ordieres, J. B.; Vergara, E. P.; Capuz, R. S.; Salazar, R. E., Neural network prediction model for fine particulate matter (PM2.5) on the US-Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environmental Modelling & Software* 2005, 20 (5), 547-559.
- Polezer, G.; Tadano, Y. S.; Siqueira, H. V.; Godoi, A. F. L.; Yamamoto, C. I.; André P. A. D.; Pauliquevis, T.; Andrade, M. D. F.; Oliveira, A.; Saldiva, P. H. N., Assessing the impact of PM 2.5 on respiratory disease using artificial neural networks. *Environmental Pollution* 2018, 235, 394-403.
- Qin, D.; Yu, J.; Zou, G.; Yong, R.; Zhao, Q.; Zhang, B., A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM2.5 Concentration. *IEEE Access* 2019, 7, 20050-20059.
- Qin, S.; Liu, F.; Wang, J.; Sun, B., Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models. *Atmospheric Environment* 2014, 98, 665-675.
- Soh, P.; Chang, J.; Huang, J., Adaptive Deep Learning-Based Air Quality Prediction Model Using the Most Relevant Spatial-Temporal Relations. *IEEE Access* 2018, 6, 38186-38199.
- Tsiouris, K. M., Pezoulas, V. C., Zervakis, M., Konitsiotis, S., Koutsouris, D. D., & Fotiadis, D. I. (2018). A long short-term memory deep learning network for the prediction of epileptic seizures using eeg signals. *Comput. Biol. Med.*, 99, 24-37. doi: <https://doi.org/10.1016/j.combiomed.2018.05.019>.
- Wen, C.; Liu, S.; Yao, X.; Peng, L.; Li, X.; Hu, Y.; Chi, T., A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Science of The Total Environment* 2019, 654, 1091-1099.
- World Health Organisation, Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease. Working Papers 2016.
- Xiao, F.; Qi, L.; Zhu, Y.; Hou, J.; Jin, L.; Wang, J., Artificial neural networks forecasting of PM 2.5 pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment* 2015, 107, 118-128.
- Yang, W.; Min, D.; Feng, X.; Hang, W., Prediction of hourly PM 2.5 using a space-time support vector regression model. *Atmospheric Environment* 2018, 181, 12-19.
- Zhou, Q.; Jiang, H.; Wang, J.; Zhou, J., A hybrid model for PM2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network. *Science of the Total Environment* 2014, 496 (2), 264-274.
- Zhao, J.; Deng, F.; Cai, Y.; Chen, J., Long short-term memory - Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. *Chemosphere* 2019, 220, 486-492.
- Zhang, D.; Lin, J.; Peng, Q.; Wang, D.; Yang, T.; Sorooshian, S.; Liu, X.; Zhuang, J., Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology* 2018, 565, 720-736.