CHEMICAL ENGINEERING TRANSACTIONS

VOL. 83, 2021

A publication of AIDIC

The Italian Association of Chemical Engineering
Online at www.cetjournal.it

Guest Editors: Jeng Shiun Lim, Nor Alafiza Yunus, Jiří Jaromír Klemeš
Copyright © 2021, AIDIC Servizi S.r.l.
ISBN 978-88-95608-81-5; **ISSN** 2283-9216

DOI: 10.3303/CET2183060

Convolution Recurrent Neural Network for Daily Forecast of PM_{10} Concentrations in Brunei Darussalam

Effa Nabilla Aziz[a,*], Asem Kasem[a], Wida Susanty Haji Suhaili[a], Peijiang Zhao[b]

[a]School of Computing and Informatics, Universiti Teknologi Brunei, BE1410, Brunei Darussalam
[b]Big Data Analytics Laboratory, National Institute of Information & Communications Technology, 184-8795, Tokyo, Japan
p20190005@student.utb.edu.bn

PM_{10} is a particulate matter with an aerodynamic diameter less than or equal to 10 µm. It is one of the primary pollutants contributing to the ambient air quality level. Air quality monitoring in Brunei Darussalam is using only the PM_{10} concentrations to measure the nation's daily Pollutant Standard Index (PSI). This study sheds light on a data-centric landscape of air pollution prediction in Brunei Darussalam, highlights potential uses of forecasting daily PM_{10} concentrations, and presents comparisons of prediction models built using several methods, namely: moving average, linear regression, recurrent neural network (RNN), long short-term memory (LSTM), LSTM with 1-D convolutions, and convolutional recurrent neural network (CRNN). This study is using daily PM_{10} concentrations obtained from the air quality monitoring stations located at every district in Brunei Darussalam for a period of 15 y (2005–2019). The results of the analysis of the daily prediction performance on shows that the CRNN approach provides the most accurate prediction among compared methods. The mean value of RMSE, MAE and SMAPE for the CRNN model are 3.414, 2.293 and 0.125. The results from the CRNN model can be used as part of early-warning application with the ability to provide health advisory such as wearing face masks or limit outdoor activities, or to show safer routes to school or work from heavy polluted areas, to mitigate the negative impacts of haze pollution on the citizens. Future work would include multiple days predictions, the inclusion of air quality data from neighbouring countries in Southeast Asia to account for transboundary air pollution, and the inclusion of other pollutants concentrations: carbon monoxide (CO), fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), sulphur dioxide (SO_2) and ozone (O_3).

1. Introduction

Haze pollution has been a recurrent issue in Southeast Asia region, and the sources may come from localized or transboundary pollution. Transboundary haze episodes are mostly caused by long-range transport of biomass fires from slash-and-burn activities in Indonesia (Dotse et al., 2017) during dry seasons, with the prevailing southern monsoon wind going upwards, affecting several countries in the region, including Brunei Darussalam, Malaysia, Singapore, Indonesia, and Southern Thailand. In Brunei Darussalam, most of the localized haze episodes were resulted from peat fires where the peatlands area is being drained or deforested. This have released carbon emission that accumulated from dead plant matter, and lead to peatland fires and resulted in smoke haze and greenhouse gases into the atmosphere. Particulate matter (PM_{10}) is the only pollutant used in determining the severity of haze and air quality index in Brunei Darussalam.

Several studies have shown significant relationship between cardiovascular and respiratory morbidity rate with the level of PM_{10} concentrations. Newell et al. (2017) analysed an increase rate of cardiovascular (0.27 %) and respiratory (0.56 %) morbidity in China correlated with the increase of PM_{10} concentrations in China. During the catastrophic 1997–1998 haze episodes in Brunei Darussalam, the PM_{10} concentrations were linked to the increase cases of respiratory morbidity, such as are: asthma, influenzas, acute upper respiratory infections, pneumonia, bronchitis, emphysema, and conjunctivitis (Anaman and Ibrahim, 2003), and the drop of visibility (Yadav et al., 2003). Brunei have also lost 3.75 % number of tourists, with an estimation of economic loss of BND 1 million (Anaman and Looi, 2000) and Brunei Airport was forced to close due to the poor visibility and most flights were delayed or cancelled (Limin et al., 2006). Anaman (2001) surveyed several

Paper Received: 15/07/2020; Revised: 30/08/2020; Accepted: 01/09/2020
Please cite this article as: Aziz E.N., Kasem A., Haji Suhaili W.S., Zhao P., 2021, Convolution Recurrent Neural Network for Daily Forecast of PM_{10} Concentrations in Brunei Darussalam, Chemical Engineering Transactions, 83, 355-360 DOI:10.3303/CET2183060

households during the 1998 haze episodes and discovered that the average household spent about BND 15 on face masks to reduce the negative effects of the haze, in which for 65, 000 households in Brunei Darussalam would have estimation cost of BND 1 million for the daily usage of the face masks alone.

Predictions on PM_{10} concentrations have been widely studied due to its potential negative impacts on human well-being, economic, and biodiversity, and its important role in climate change (IPCC, 2013). This led to many governments and stakeholders in Brunei Darussalam to monitor PM_{10} concentration closely and focus research on PM_{10} predictions. Many studies conducted on particulate matter have used mathematical statistical models (Donkelaar et al., 2010) to make long-term prediction by using climate model or a satellite remote sensing. Several machine learning methods (Saeed et al., 2017) were used to improve accuracy in short-term prediction with the inclusion of meteorological data, which can influence the predicted values. Another potential way to predict particulate matter concentrations is using deep learning techniques, for example hybrid of convolutional neural network and long short-term memory methods (Yang et al., 2020) were used to predict hourly particulate matter concentrations in Seoul, South Korea; and to date, prediction method study on daily PM_{10} concentration in Brunei is using hybrid framework of genetic algorithm, random forests and back propagation neural networks (Dotse et al., 2017) with using 5 y of air quality data (2009-2013).

The literature review above identified some research gaps. Firstly, prediction study on PM_{10} concentration using hybrid method of convolution neural network and recurrent neural network, and evaluation on performance models using several neural network-based methods have not been considered in the literature on data-centric landscape of air pollution prediction in Brunei Darussalam. Prediction methods using recent data on PM_{10} concentration in Brunei have also not been used, and the use of recent data is essential especially in forecasting study as it may improve the accuracy and applicability of the prediction models. Lastly, there are few studies on negative impacts of haze episodes in Brunei using recent haze episodes, as most of the studies are based on the catastrophic 1997-1998 haze episodes (Anaman and Ibrahim, 2003) in Brunei. In summary, the main gap in the literature study is a thorough analysis on PM_{10} concentration prediction using recent air quality data in Brunei, and comparing the model performance while considering the spatial and temporal distribution of the data that add significant contribution to the framework.

This study aims to utilize convolution recurrent neural network as proposed in a $PM_{2.5}$ prediction study in Japan (Zhao and Zettsu, 2018), using longer PM_{10} dataset of 15 y (2005-2019) provided by the Department of Environment, Park and Recreation (JASTRE), gathered from four stations in every district, to predict the daily PM_{10} concentrations in Brunei Darussalam. This study also conducted other prediction methods such as linear regression, recurrent neural network, and long short-term memory, and evaluate all methods using performance metrics. Results from the prediction model can aid the governments and organizations in monitoring the air quality, acts as early warning advisory to the public. A summarized data in Table 1 highlights the total number of days with PM_{10} exceeding 50 μgm^{-3} (above the level for good air quality guidelines issued by the Ministry of Health in 2013 health advisory). Table 1 shows that 2013 and 2015 had the highest number of exceedance days (with a total of 36 d in 2013 and 74 da in 2015).

Table 1: Number of days PM_{10} concentrations exceeded 50 μgm^{-3} (above MOH guideline for good air quality)

District Year	Anggerek Brunei-Muara	Bukit Bendera Tutong	Mumong Belait	Taman Batang Duri Temburong	Sum of Number of Days Exceeding 50 μgm^{-3}
2005	7	1	6	1	15
2006	8	-	20	20	48
2007	-	-	3	-	3
2008	2	1	2	-	5
2009	8	13	20	12	53
2010	-	-	2	-	2
2011	4	5	8	3	20
2012	-	7	5	-	12
2013	3	7	21	5	36
2014	-	-	6	-	6
2015	13	26	32	3	74
2016	-	3	6	-	9
2017	-	-	-	-	-
2018	-	1	1	-	2
2019	13	14	25	3	55
Total	58	78	157	47	

The contribution and novelty of this study to the research literature are:

1. Utilize the new proposed prediction method that combines both convolution and recurrent neural network (Zhao and Zettsu, 2018) with the inclusion of the spatial and temporal distribution of the dataset, to predict daily PM_{10} concentrations.
2. All prediction methods used in this study can be used as real-time daily forecasting, notably convolution recurrent neural network model, reducing the expense of using tapered element oscillating microbalance instrument to monitor the pollutant concentration.
3. This prediction method is using larger dataset of air quality concentrations from Brunei (2005-2019), being the first prediction study to develop prediction methods using 15 y of recent air quality dataset.

2. Data and methods

2.1 Study area and data

Brunei Darussalam (latitude 4.5353° N, longitude 114.7277° E) and its four administrative districts are shown in Figure 1, indicating the locations of Anggerek air quality monitoring station in Brunei Muara district (latitude 4.9329° N, longitude 114.9415 E), Bukit Bendera air quality monitoring station in Tutong district (latitude 4.8102° N, longitude 114.6601° E), Mumong air quality monitoring station in Belait district (latitude 4.5751° N, longitude 114.2330° E) and Taman Batang Duri air quality monitoring station in Temburong district (latitude 4.5786° N, longitude 115.1215° E). The PM_{10} data used in this study is gathered from four air quality monitoring stations for every district and provided by JASTRE.

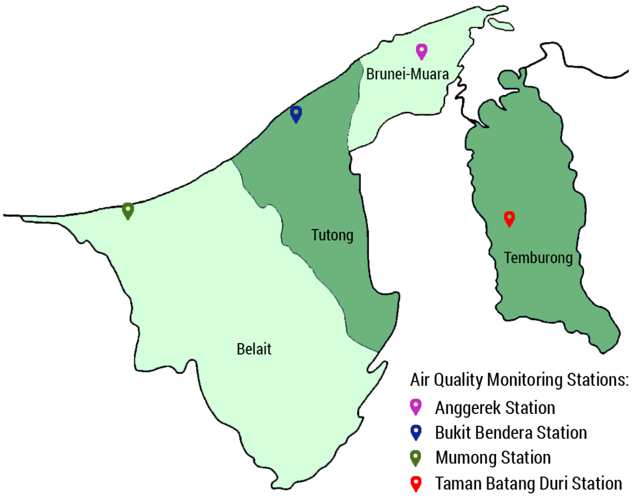

Figure 1: Map showing the locations of the air quality monitoring stations in each district in Brunei Darussalam.

The mean daily PM_{10} concentrations and meteorological datasets were pre-processed and restructured to the following format: year, month, day, latitude, and longitude of the air quality monitoring station, and the mean daily PM_{10} concentration of the station. The available data was divided into two subsets of training and test sets. 13 y data (2005–2017) were used for training the models, and 2 y data (2018–2019) were used for testing the trained models.

2.2 Methods

This study experimented with several methods to forecast PM_{10}, including common mathematical methods, and Neural Networks–based methods, and also leveraged an approach of utilizing Convolutional Recurrent Neural Network (CRNN) as proposed by Zhao and Zettsu (2018). A naïve forecast is used as baseline model, in which the previous timestep (i.e. day) of PM_{10} values are used as the next-day forecasts, without any attempt to model the data. The predicted results by naïve forecast are used to compare with the forecast results by other models. Moving average forecast is formed by taking the average value of PM_{10} over a number of previous timesteps, and the averaging window is then shifted forward throughout the PM_{10} data. Linear regression is also used by estimating a linear function of the parameters in a previous time window of PM_{10} data (e.g. window size is 7 d). Recurrent neural network (RNN) models were also used, including simple RNN, bidirectional long short-term memory (LSTM), and LSTM with 1-D convolutional layer. The methods were applied on each station's data individually as multiple time-series prediction tasks. The approach proposed by Zhao and Zettsu (2018) allows to combine spatio-temporal data from multiple locations and model it by using CRNN model. Figure 2 shows an example of how convolutional networks process the spatial features of the data.

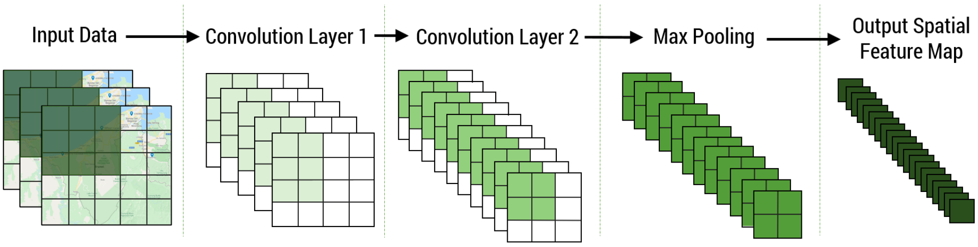

Figure 2: The processes of feeding spatial environment data based on coordinates to the convolutional neural network to generate a spatial feature map.

Missing data within the grid, i.e. where there are no stations, are approximated using inverse distance weighting (IDW) for its simplicity. Other methods that take terrain into account (Wang et al., 2019) may be also used.

2.3 Model performance evaluation

Performance evaluation is used to calculate the error between the predicted and the real observed data on the test set, which can be measured using different types of performance metrics. In this study, the following metrics were used: mean absolute error (MAE), root mean squared error (RMSE), and symmetric mean absolute percentage error (SMAPE).

MAE measures the average of absolute differences between predicted data and the real observed data. MAE value close to zero suggests both predicted and observed data has good agreement, as shown in Eq(1):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_p - y_o| \qquad (1)$$

RMSE calculates the standard deviation of predicted data from the real observed data, as shown in Eq(2):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_p - y_o)^2}{n}} \qquad (2)$$

SMAPE calculates the absolute differences between predicted data and the real observed data and square the result, as shown in Eq(3):

$$SMAPE = \frac{100\,\%}{n}\sum_{i=1}^{n}\frac{|y_p - y_o|}{|y_o| + |y_p|} \qquad (3)$$

where, in Eq(1) to Eq(3), y_p is the predicted data, y_o is the real observed data, and n is the total number of samples.

3. Results and Discussions

The computational experiments for this study were conducted using Python and TensorFlow. The parameters for the predictive models were learnt by training using the 15 y dataset (2005–2017) of mean daily PM_{10} concentrations (80 % of the data), and the remaining 20 % of the data (2018–2019) were used for evaluation to avoid biased optimistic evaluation due to over-fitting.

For Moving Average method, it was found that a two-days averaging window gave the best performance on the test set among the few experimented sizes (2-10). For the other methods, model fitting was run individually on each station’s training dataset, using a variety of parameters. All experiments, except for CRNN, used 500 epochs (very similar results were obtained with 80 and 300 epochs as well). For the CRNN method, the CNN part ran for 800 epochs and the RNN part ran for 700 epochs.

The simple RNN network used two layers of 40 neurons each; the LSTM network used two bidirectional layers of 32 LSTM units each, and the LSTM with 1D-Conv used two unidirectional layers of 32 LSTM units, preceded with a Conv1D layer of 32 units (kernel size is 5).

Table 2 presents the performance metrics of each method, averaged for all stations, with performance results of the CRNN method showing the least error in prediction, across all metrics.

Table 2: Mean performance metrics of the models at the four stations

Metric	RMSE			MAE			SMAPE		
Epoch	80	300	500	80	300	500	80	300	500
Naïve Forecast	5.294			3.241			0.156		
Moving Average	5.867			3.662			0.172		
Linear Regression	5.217	5.180	5.169	3.245	3.226	3.216	0.167	0.155	0.154
Simple RNN	5.419	5.168	5.195	3.471	3.228	3.228	0.164	0.155	0.155
LSTM	5.393	5.353	5.270	3.293	3.349	3.330	0.155	0.160	0.159
LSTM 1D-Conv	5.575	5.784	5.917	3.386	3.494	3.548	0.160	0.165	0.165
CRNN	3.414			2.293			0.125		

Figure 2 shows an example of how one model has fit the training data and is also able to generalize and predict the data in the test set for the station in Belait district.

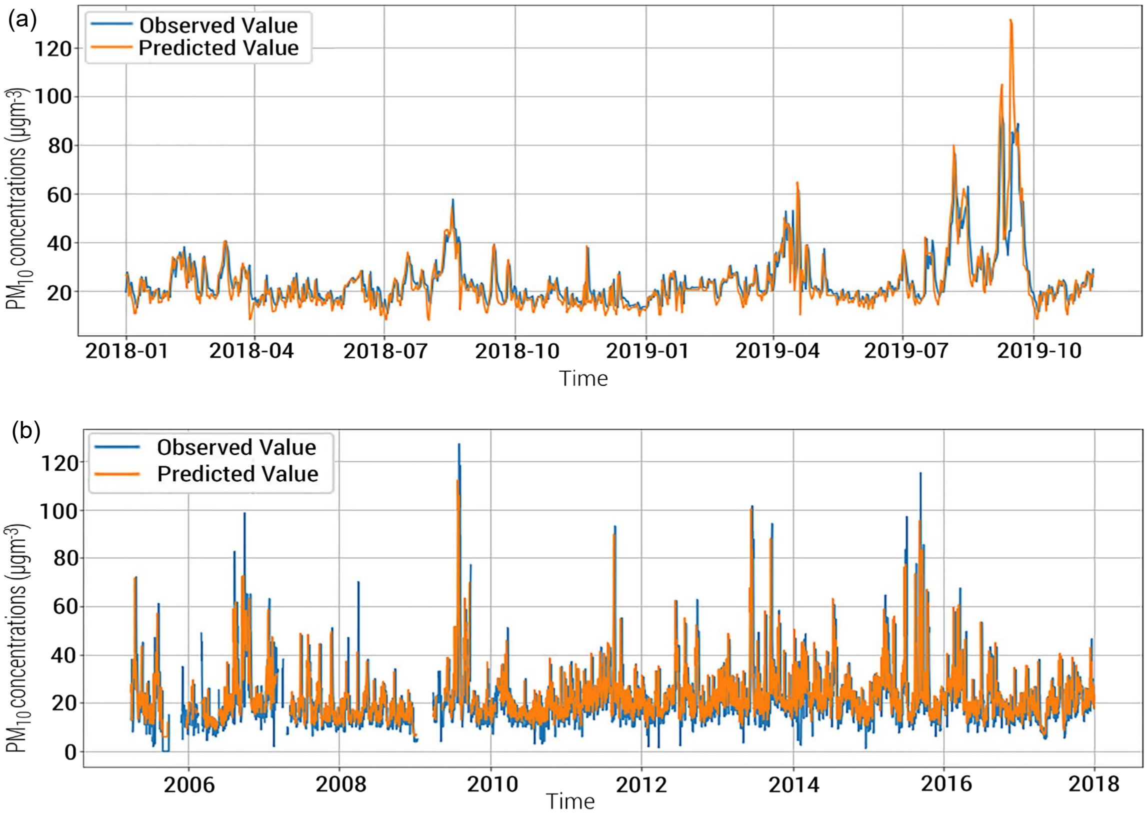

Figure 3: Example of time-series graph of the observed and predicted values of PM$_{10}$ concentrations at Mumong station, Belait district, for (a) testing sets (2018-2019), and (b) training sets from 2005 to 2017.

Overall, based on the obtained metrics from all methods, it seems that all models have a relatively small error for one-day prediction, this finding suggests that probably any of them can be used in real forecast. It is also interesting to note that most of the methods, except for CRNN, achieved a slightly worse error than the naïve forecasting. For practical applications, it is believed that multi-days forecast is necessary, and the accuracy of prediction will degrade much more compared to only single future timestep case, and it is expected that the CRNN method will prevail as the suitable method to be used.

4. Conclusions

This study conducted several prediction methods to forecast daily PM$_{10}$ concentrations in four districts of Brunei Darussalam and evaluated the performance of these methods using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Symmetric Mean Absolute Percentage Error (SMAPE) metrics. Convolution Recurrent Neural Network (CRNN) model has shown the most satisfactory forecasting results across all metrics, with MAE, RMSE, and SMAPE values of 3.414, 2.293, and 0.125. Since air pollution is a recurrent issue in Brunei Darussalam especially during the dry seasons, the results from the forecasting model can be used as part of an early-warning application with provides health advisory for citizens, such as wearing facemasks or limiting outdoor activities, or to show safer routes to school or work from heavily polluted areas, and to mitigate the negative impacts of haze pollution on the residents. This application aids the government, especially the Ministry of Health, as PM_{10} concentrations have been shown to have negative impacts on human well-being and many studies have shown significant relationship between PM_{10} concentrations with cardiovascular and respiratory diseases. Future work should include running the CRNN model for multi-timesteps predictions, and the use of air quality data from neighbouring countries in Southeast Asia region to account for transboundary air pollution prediction. One possible approach similar to what has been conducted by Zhao and Zettsu (2019) is "Convolution Recurrent Neural Networks Based Dynamic Transboundary Air Pollution Prediction". Besides, the inclusion of other meteorological parameters such as rainfall, humidity, and temperature, and other pollutants concentrations, such as carbon monoxide (CO), fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), sulphur dioxide (SO_2) and ozone (O_3) is expected to improve the accuracy of predictions, and the applicability of its results.

Acknowledgments

The authors acknowledge the support from JASTIP-Net 2019 Grant, and Universiti Teknologi Brunei's Internal Grant "UTB/GSR/1/2019(10)" and thank them for supporting the research activities required to complete this study. A great appreciation is extended to the Department of Environment, Park and Recreation (JASTRE) for providing the daily PM_{10} concentrations data.

References

Anaman K.A., 2001, Urban householders' assessment of the causes, responses, and economic impact of the 1998 haze-related air pollution episode in Brunei Darussalam, ASEAN Economic Bulletin, 18(2), 193–205.

Anaman K.A., Ibrahim N., 2003, Statistical estimation of dose-response functions of respiratory diseases and societal costs of haze-related air pollution in Brunei Darussalam, Pure and Applied Geophysics, 160(1), 279–293.

Anaman K.A., Looi C.N., 2000, Economic impact of haze related air pollution of the tourism industry in Brunei Darussalam, Economic Analysis and Policy, 30(2), 133–143.

Donkelaar A.V., Martin R., Brauer M., Kahn R., Levy R.C., Verduzco C., Villeneuve P.J., 2010, Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application, Environment Health Perspectives, 118(6), 847-855.

Dotse S.Q., Petra M.I., Dagar L., Silva L.C.D., 2017, Application of computational intelligence techniques to forecast daily PM10 exceedances in Brunei Darussalam, Atmospheric Pollution Research, 9(2), 358–368.

IPCC, 2013, Fifth assessment report of the intergovernmental panel on climate change, Cambridge University Press, Cambridge, United Kingdom.

Limin S.H., Rieley J.O., Jaya S., Gumiri S., 2006, The impact of forest fires and resultant haze on terrestrial ecosystems and human health in central Kalimantan, Indonesia, Tropics, 15(3), 321–326.

MOH, 2013, Health advisory during haze episodes, Ministry of Health Brunei Darussalam <www.moh.gov.bn/SiteCollectionDocuments/Haze/health-advisory-2013.pdf> accessed 19.05.2020.

Newell K., Kartsonaki C., Lam K.B.H., Kurmi O.P., 2017, Cardiorespiratory health effects of particulate ambient air pollution exposure in low-income and middle-income countries: a systematic review and meta-analysis, Lancet Planetary Health, 1(9), 368–380.

Saeed S., Hussain L., Awan I.A., Idris A., 2017, Comparative analysis of different statistical methods for prediction of PM2.5 and PM10 concentrations in advance for several hours, International Journal of Computer Science and Network Security, 17(11), 45–52.

Wang X., Klemeš J.J., Fan W., Dong X., 2019, An overview of air-pollution terrain nexus, Chemical Engineering Transactions, 72, 31–36.

Yadav A.K., Kumar K., Kasim A.M., Singh M.P., 2003, Visibility and incidence of respiratory diseases during the 1998 haze episode in Brunei Darussalam, Pure and Applied Geophysics, 160(1), 265–277.

Yang G., Lee H., Lee G., 2020, A hybrid deep learning model to forecast particulate matter concentration levels in Seoul, South Korea, Atmosphere 2020, 11(4), 348–367.

Zhao P., Zettsu K., 2018, Convolution recurrent neural networks for short-term prediction of atmospheric sensing data, 2018 IEEE International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 30th July–3rd August 2018, Halifax, NS, Canada.

Zhao P., Zettsu K., 2019, Convolution recurrent neural networks based dynamic transboundary air pollution prediction, 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA), 15th March–18th March 2019, Suzhou, China.