Wavelet Convolutional Neural Network for Forecasting Malaysian PM_{10} Time Series Data

Mohd Aftar Abu Bakar[1](✉), Noratiqah Mohd Ariff[1], and Mohd Shahrul Mohd Nadzir[2]

[1] Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, UKM, 43600 Bangi, Selangor, Malaysia
aftar@ukm.edu.my

[2] Department of Earth Sciences and Environment, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, UKM, 43600 Bangi, Selangor, Malaysia

Abstract. Hourly particulate matter time series data from eight air quality monitoring stations in Peninsular Malaysia were forecast by using the Convolutional Neural Network (CNN) algorithm. Instead of using the original time series, which are time-domain sequence data, this study used the time-frequency domain sequence data retrieved by wavelet transformation. Air pollutants' concentration considered for this study is the particulate matter with a diameter of 10 microns or less, PM_{10}. The transformation used in this study is the Morlet wavelet transform, which is continuous wavelet transformation (CWT). Different time steps for the time series dependencies were considered to assess the PM_{10} dependencies on its past values. The results were compared with the results from the CNN algorithm using the original time series. It is shown that the Wavelet Convolutional Neural Network algorithm improves the forecast accuracy of the PM_{10} time series.

Keywords: Convolution Neural Network · Wavelet Transform · Time Series Forecasting · Air Quality

1 Introduction

Air quality is determined by measuring air pollutants' concentration such as suspended particles with a diameter of fewer than 10 μm (PM10), suspended particles with a diameter of fewer than 2.5 μm (PM2.5), sulphur dioxide (SO2), nitrogen (NO2), ozone (O3) and others. High concentrations of toxic elements and gases released into the atmosphere will cause air pollution. Urban air pollution has been a major concern for the last fifty years, affecting human health, well-being, and the environment. Studies show that the rapid development of industry and population growth at the end of the 20th century poses a severe threat to air quality [1].

Many air quality monitoring stations have been set up to monitor the air quality in Malaysia. From the data collected by the air quality monitoring stations, an air quality

N. Wahi et al. (Eds.): ICMSS 2022, ACSR 98, pp. 205–213, 2023.
https://doi.org/10.2991/978-94-6463-014-5_20

model can be built, which is helpful for air quality management. Many air quality models have been built using outdated time series models, where many assumptions were required and with lower prediction accuracy. Nowadays, many deep learning techniques have been proposed which is more flexible and with better accuracy [2, 3]. According to Azid et al. [4], artificial neural networks (ANN) can improve the decision-making process and propose a problem-solving tool for improving Malaysian atmospheric management. Thus, in this study, the convolutional neural network (CNN) model was proposed for modelling air quality in Malaysia.

CNN can be defined as involves pooling many small filters from the input data and a feed-forward artificial neural network algorithm. Its ability to discover the internal structure and extract deep features make it useful for pattern identification, which is useful for analyzing time series data.

In this study, the continuous wavelet transform (CWT) of the air quality time series data will be used as the input for the CNN algorithm to forecast the hourly PM10 time series for several areas in Malaysia. Wavelet analysis is a technique that can be used to analyze time-frequency characteristics of time domain signals. It has been used extensively in many areas, such as in engineering [5, 6], finance [7] and signal processing [8]. Several studies have shown that combining the time series model with wavelet analysis could improve the time series model accuracy [9, 10].

2 Convolutional Neural Network (CNN)

CNN technique extracts feature from datasets through convolution kernels and pooling operation. The convolutional layer combines several local filters with a sequential input to identify the sequence. Each feature map corresponding to the local filter can be generated by applying a filter to all sequential input. Then the pooling layer is used to extract the most significant features with fixed lengths from each feature map. The convolutional and pooling layers can be combined and arranged to perform analysis.

The convolutional layer is a linear process that aims to extract local patterns in the time dimension and find local dependencies in the raw sequence. The raw sequential input, S and the sequential filter, FS are defined as follows:

$$S = [s_1, s_2, s_3, \ldots, s_L] \tag{1}$$

$$FS = [w_1, w_2, w_3, \ldots, w_K] \tag{2}$$

where $s_i \in R$ is a single sequential data point arranged by time and $w_j \in R^{m\times 1}$ is one of the filter vectors. L is the length for the raw sequential input S and K is the number of total filters in the convolutional layer. A convolutional operation is defined as a multiplication operation between filter vector w_j and network vector symbol $s_{i:i+m-1}$.

$$s_{i:i+m-1} = s_i \oplus s_{i+1} \oplus s_{i+2} \oplus \ldots \oplus s_{i+m-1} \tag{3}$$

where $\oplus$ is networking operation and $s_{i:i+m-1}$ represents a continuous period starting from i. Bias $b \in R$ required to be considered in convolutional operations. Therefore, the final computational equation is given as follow:

$$c_i = f\left({w_j}^T \cdot s_{i:i+m-1} + b\right) \tag{4}$$

where w_j^T represents a change in the order of the filter matrix and f is the nonlinear activation function. The i index represents the timestep's index and j is for the filter's index.

The activation functions are used to improve the model's ability to learn more complex functions that can further improve forecasting performance. Applying an appropriate activation function can speed up the rate of mapping and improve the model's illustration ability. Rectified Linear Units (ReLU) has been selected for this study since it has been used in many applications, including for air quality predictive models, given its advantages over other activation functions in improving the model's accuracy [11].

In general, several filters are arranged in a convolutional layer to extract some input data's key features efficiently. Based on the example, there are as many K filters with m-sized window measurements in the convolutional layer. Based on the equations above, each vector w_j represents a filter and the value c_i represents a window activation.

A convolutional operation applied to the entire sequential input is executed by applying a filtering window based on a fixed timestep. Thus, the feature map corresponding to the filter can be defined in vector form as follows:

$$F_j = [c_1, c_2, c_3, \ldots, c_{L-m+1}] \tag{5}$$

where j is the filter's index and elements in F_j corresponds to $\{s_{1:m}, s_{2:m}, \ldots, s_{l-m+1:L}\}$.

The pooling function is similar to sub-sampling because the pooling function samples the convolutional layer's output based on a specific pooling measure, p. This means that the pooling layer can effectively compress the feature map's length to further reduce the number of model parameters. The feature compression vector $F_{j\text{-compress}}$ can be obtained based on the max-pooling applied in the air quality predictive model.

$$F_{j-\text{compress}} = [h_1, h_2, h_3, \ldots, h_{\frac{L-m}{p}+1}] \tag{6}$$

where $h_j = \max(c_{(j-1)p}, c_{(j-1)p+1}, \ldots, c_{jp-1})$ [12].

Figure 1 illustrates the One-Dimensional CNN architecture for time series data used in this study. The CNN model considered here is a univariate time series model which only considered the time series lag values (previous observation) as the input variable without including other variables. The PM_{10} concentrations time series will be modelled separately for each station without considering the correlation between other air quality variables or between those air quality monitoring stations. This study considers different timesteps for building the CNN model to evaluate the model's accuracy in different timesteps. There are seven timesteps considered, of which the shortest is 24 h (a day) and the longest is 168 h (seven days). The CNN model applied in this study was restricted to three convolutional layers and one pooling layer only. The CNN models were trained up to 30 cycles (epoch) for optimizing the CNN hyperparameters.

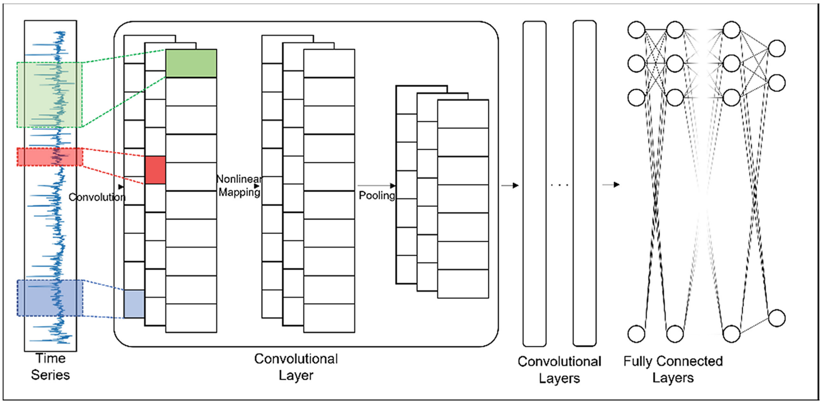

Fig. 1. One-Dimensional CNN architecture for time series data.

3 Continuous Wavelet Transform

Time series data can be decomposed into several resolutions using wavelet transform, where each resolution represents a contribution of oscillations from different frequencies [13]. One of the common applications of wavelet in deep learning forecasting is for transforming the time series into several resolutions, where one-step prediction will be performed at each resolution independently and will be summed to obtain the final prediction [14]. Air quality time series have various frequency levels similar to other time series data types influenced by human activities or weather conditions [15]. Therefore, the prediction may be improved by accounting for those important properties at certain frequencies extracted from the time series [16].

There are two types of wavelet transformations, which are discrete wavelet transformation (DWT) and continuous wavelet transformation (CWT). In this study, the decomposition of the time series was conducted before constructing the CNN predictive model using the CWT. CWT is chosen here since it is more suited for time series. This technique can reveal the time series characteristics under multi-temporal scales with higher resolutions [17], while DWT is more suited for data compression [18].

The CWT of a signal $x(t)$ at time a and scale b can be defined as

$$W_x(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-b}{a}\right)dt, \tag{7}$$

where $\psi(t)$ is the wavelet function. Here, $W_x(a, b)$ are known as the wavelet coefficients, which provide information about the signal, $x(t)$, at scale a and around time b.

In this study, the Morlet wavelet has been used as the wavelet function, defined as

$$\psi(t) = \exp\left(-\frac{t^2}{2\sigma^2}\right)\exp(i\omega t), \tag{8}$$

where σ and ω are parameters that control the size of the wavelet envelope and oscillations, respectively. The Morlet wavelet is a modulated Gaussian function, and its integral is approximately zero for $\sigma\omega_0 > 5$. The Fourier transform of the Morlet wavelet is

$$\psi(\omega) = \exp\left(-\frac{(\omega - 2\pi)^2}{2}\right), \tag{9}$$

which provides good localization in the frequency domain (Carmona et al. 1997).

4 Analysis and Results

In this study, hourly PM_{10} time series from eight air quality monitoring stations in Peninsular Malaysia, shown in Fig. 2, have been fitted to the Wavelet CNN (WCNN) and CNN algorithms. The hourly PM_{10} time series from 12 am of 5th July 2017 to 11 pm of 31st March 2019 were used to train the model, while the hourly PM_{10} time series from 12 am of 11th April 2019 to 11 pm of 30th June 2019 were used for model accuracy assessment. Seven different timesteps were also considered for this study to assess the PM_{10} time series dependencies on its past values.

Figure 3 shows the hourly PM_{10} time series and WCNN forecast plots for all stations considered in this study. In this figure, the forecast values based on the WCNN algorithms

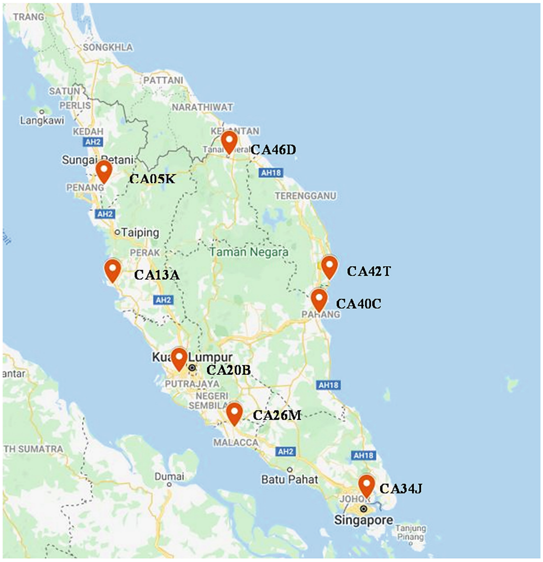

Fig. 2. Air quality monitoring stations considered in this study.

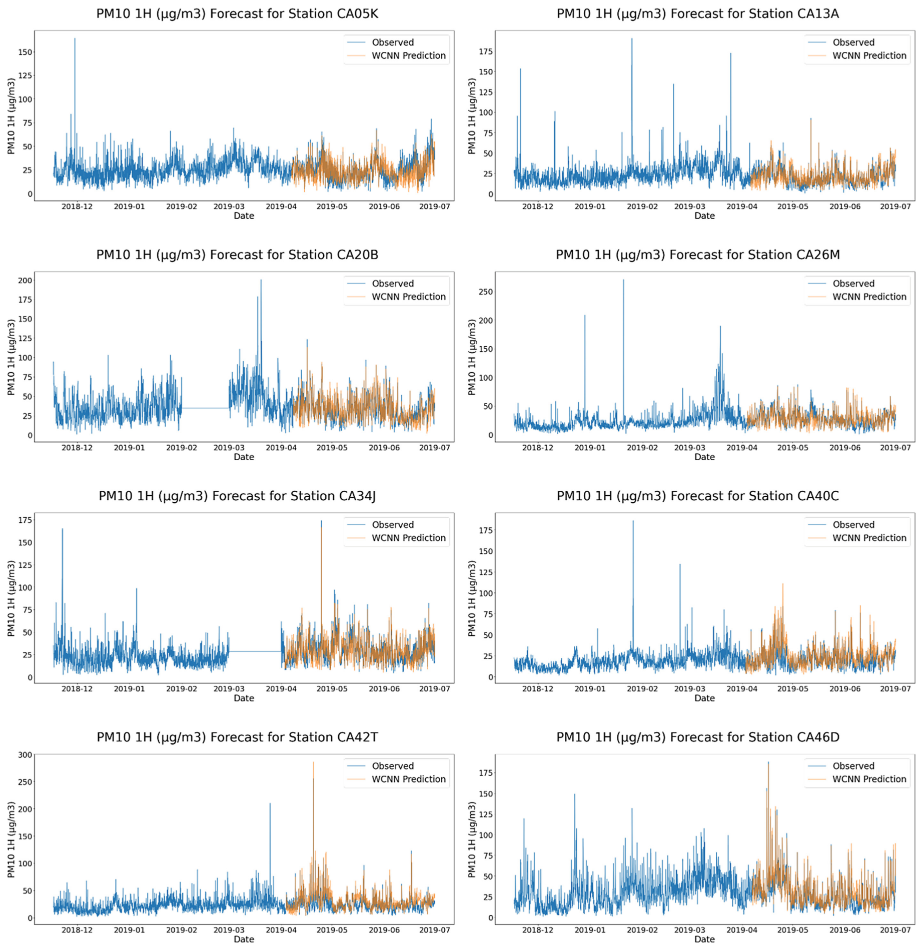

Fig. 3. Hourly PM_{10} time series and forecast based on WCNN technique.

were only plotted on the testing set period, from 12 am on 11th April 2019 to 11 pm on 30th June 2019. Since seven different timesteps were considered in this study, each station's forecast values were plotted using the best timesteps, as provided in Table 1. Based on those plots, it can be deduced that the WCNN did quite well in forecasting the hourly PM_{10} values for all the stations. The figure also shows that the WCNN technique can forecast those extreme events, which happen on those actual values.

Table 1 lists the RMSE values for the testing sets based on WCNN and CNN techniques for seven different time steps. The RMSE values for the WCNN technique are lower compared to the RMSE from CNN techniques for all stations and all timesteps. This shows that the WCNN has better accuracy in predicting the future values of PM_{10} for all stations than the CNN technique. This is because the time series' wavelet decomposition provides extra information related to the cyclical patterns. This information was fed into the CNN algorithms, resulting in a better prediction of the time series.

Table 1. RMSE values of hourly PM_{10} ($\mu g/m^3$) testing sets based on WCNN and CNN techniques

Station	Timestep													
	24		48		72		96		120		144		168	
	WCNN	CNN	WCNN	CNN	WCNN	CNN	WCNN	CNN	WCNN	CNN	WCNN	CNN	WCNN	CNN
CA05K	6.9111	***7.0128***	6.6335	7.0451	6.3038	7.4909	6.9718	7.1786	6.5117	7.3366	**6.2703**	7.2895	6.9982	7.2096
CA13A	6.3137	***6.0560***	5.9307	6.1069	5.5151	6.4379	4.9819	6.2994	**4.7949**	7.0416	4.8623	6.6614	4.9165	6.5854
CA20B	6.5947	8.0278	6.6320	8.0001	6.4552	***7.9584***	6.0048	8.0025	6.8831	8.5906	7.6015	8.1334	**6.0012**	8.0611
CA26M	6.9883	7.7908	6.9950	***7.5001***	**6.7373**	7.5735	7.4476	7.9342	7.1211	8.2608	7.8158	7.8967	8.0426	8.0412
CA34J	6.3454	***8.7361***	**5.9437**	8.8241	6.0324	8.8624	6.2159	8.9797	6.1779	8.9620	6.8351	9.1149	7.2512	9.1308
CA40C	6.8452	7.3199	**6.8103**	***7.1500***	7.3687	7.1510	8.8551	7.2397	8.0910	7.3762	8.0125	7.3776	8.6405	7.3147
CA42T	8.8112	***10.3166***	9.5278	10.3512	**7.4529**	10.5409	8.8520	10.4821	7.5482	10.7649	8.7759	10.9788	9.4896	12.7121
CA46D	7.4417	9.5730	6.5687	***9.4341***	6.8356	9.7124	5.5584	9.7747	6.1520	9.8111	**5.0086**	9.9508	5.4881	10.4437

The table also shows that the best timesteps length of previous values to be used by each station varies. The stations with a better prediction for shorter previous time steps have a shorter dependency on their past values than those with longer dependencies on their previous values. This means that those stations with longer dependencies on their previous values will be expecting much more predictable PM_{10} values. In comparison, those stations with shorter dependencies on their previous values will have more volatile PM_{10} values. These different dependencies of hourly PM_{10} time series on its previous values for each station could be caused by several factors such as wind speed, humidity, temperature, land terrain and activities in that area [19–21].

5 Conclusion

This study shows that the WCNN improved the forecast accuracy of the hourly PM_{10} time series compared to the CNN technique. This also shows that the wavelet transformation of the original time series gives extra information for modelling the time series, especially on the cyclical characteristics of the time series. Furthermore, this study also shows that selecting suitable timesteps of past data plays an important role in forecasting air quality data. It was found that each station considered in this study require different lengths of timesteps used as the input data for producing better predictive models.

References

1. Ramanathan, V., Feng, Y.: Air pollution, greenhouse gases and climate change: Global and regional perspectives. Atmos. Environ. **43**(1), 37–50 (2009)
2. Lai, G., Chang, W., Yang, Y., Liu, H.: Modeling long- and short-term temporal patterns with deep neural networks. In: SIGIR 2018. The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, Michigan, pp. 95–104 (2018)
3. Zou, Y., Yu, L., Tso, G.K.F., He, K.: Risk forecasting in the crude oil market: a multiscale convolutional Neural Network approach. Physica A **541**, 123360 (2020)
4. Azid, A., Juahir, H., Latif, M., Zain, S., Osman, M.: Feed-forward artificial neural network model for air pollutant index prediction in the southern region of Peninsular Malaysia. J. Environ. Prot. **4**(12A), 1–10 (2013)

5. Moradi, L., Mohammadi, F., Baleanu, D.: A direct numerical solution of time-delay fractional optimal control problems by using Chelyshkov wavelets. J. Vibr. Control **25**(2), 310–324 (2019)
6. Bakar, M.A.A., Green, D.A., Metcalfe, A.V.: Comparison of spectral and wavelet estimators of transfer function for linear systems. East Asian J. Appl. Math. **2**(3), 214–237 (2012)
7. Mohammed, S.A., Bakar, M.A.A., Ariff, N.M.: Volatility forecasting of financial time series using wavelet based exponential generalized autoregressive conditional heteroscedasticity model. Commun. Stat. Theory Methods **49**(1), 178–188 (2020)
8. Mallat, S.: A Wavelet Tour of Signal Processing. Elsevier, Amsterdam (1999)
9. Yong, N.K., Awang, N.: Wavelet-based time series model to improve the forecast accuracy of PM10 concentrations in Peninsular Malaysia. Environ. Monit. Assess. **191**(2), 64 (2019)
10. Mohammed, S.A., Bakar, M.A.A., Ariff, N.M.: Analysis of relationships between Malaysia's Islamic and conventional stock markets using wavelet techniques. In: Proceeding of The 2018 UKM FST Postgraduate Colloquium, AIP Conference Proceedings, vol. 2111, no. 1, p. 020018. AIP Publishing, Melville (2019)
11. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning, Omnipress, Madison, WI (2010)
12. Wang, F., Yu, Y., Zhang, Z., Li, J., Zhen, Z., Li, K.: Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting. Appl. Sci. **8**(8), 1286 (2018)
13. Gottman, J.M.: Time-Series Analysis. Cambridge University Press, Cambridge (1981)
14. Aussem, A., Murtagh, F.: Combining neural network forecasts on wavelet-transformed time series. Connect. Sci. **9**, 113–122 (1997)
15. Davis, W.T., Godish, T., Fu, J.S.: Air Quality. CRC Press, Boca Raton (2015)
16. Yoo, Y., Baek, J.G.: A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network. Appl. Sci. **8**(7), 1102 (2018)
17. Sang, Y.F.: A review on the applications of wavelet transform in hydrology time series analysis. Atmos. Res. **122**, 8–15 (2013)
18. Nason, G.P., Silverman, B.W.: The discrete wavelet transform in S. J. Comput. Graph. Stat. **3**(2), 163–191 (1994)
19. Harrison, R.M., Deacon, A.R., Jones, M.R., Appleby, R.S.: Sources and processes affecting concentrations of PM10 and PM2.5 particulate matter in Birmingham (U.K.). Atmos. Environ. **31**(24), 4103–4117 (1997)
20. Elminir, H.K.: Dependence of urban air pollutants on meteorology. Sci. Total Environ. **350**(13), 225–237 (2005)
21. Mohd Nadzir, M.S., et al.: Risk assessment and air quality study during different phases of COVID-19 lockdown in an urban area of klang valley, Malaysia. Sustainability **13**(21), 12217 (2021)