Air Quality Forecasting with Hybrid LSTM and Extended Stationary Wavelet Transform

Yongkang Zeng
China Jiliang University
Xiang Ma
China Jiliang University
Ning Jin
China Jiliang University
Xiaokang Zhou
Shigakukan University: Shigakukan Daigaku
Ke Yan (keddiyan@gmail.com)
National University of Singapore - Kent Ridge Campus: National University of Singapore

Research Article

Keywords: Air Quality Forecasting, Nested Long Short Term Memory, Wavelet Transform

Posted Date: May 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-357905/v1

Version of Record: A version of this preprint was published at Building and Environment on January 1st, 2022. See the published version at https://doi.org/10.1016/j.buildenv.2022.108822.

Air Quality Forecasting with Hybrid LSTM and Extended Stationary Wavelet Transform

Yongkang Zeng[1], Xiang Ma[1], Ning Jin[1, *], Xiaokang Zhou[2, 3] and Ke Yan[1, 4, *]

[1] Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province, College of Information Engineering, China Jiliang University, Hangzhou, China 310018.

[2] Faculty of Data Science, Shiga University, Hikone, Japan 5228522.

[3] RIKEN Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan 1030027.

[4] National University of Singapore, Singapore 117566.

* Corresponding Author: jinning1117@cjlu.edu.cn (Ning Jin) and keddiyan@gmail.com (K. Yan).

ABSTRACT Artificial intelligence (AI) technology-enhanced air quality forecasting is one of the most promising directions in the field of smart environment development. Despite recent advances in this area, two difficulties remain unsolved. First, multiple factors influence forecasting results, such as weather conditions, fuel usage and traffic conditions. These factors are usually unavailable in air quality sensor data. Second, traditional predicting models typically use the most recent training data, which neglects the historical data. In this study, we propose a hybrid deep learning model that embraces the merits of the stationary wavelet transform (SWT) and the nested long short term memory networks (NLSTM) to improve the prediction quality in the problem of hour-ahead air quality forecasting. The proposed method decomposes the original $PM_{2.5}$ data into several more stationary sub-signals with different resolutions using an extended SWT algorithm. A framework that leverages several NLSTM recurrent neural networks is constructed to output forecasting results for different sub-signals, respectively. The final forecasting result is obtained by combining all sub-signal forecasting results using the inverse wavelet transform. Experiments on real-world data show that, accuracy-wise, our proposed method outperforms most of the existing prediction models in the literature. And the resulting forecasting curves of the proposed method are much closer to the real values without any lags, comparing with existing prediction models.

Key Words: Air Quality Forecasting; Nested Long Short Term Memory; Wavelet Transform.

I. INTRODUCTION

Air quality measurements and forecasting remain as popular research topics in sustainable smart environmental design, urban area development and pollution control, especially for those fast-developing countries, such as China and India (Zhang et al. 2012). Recently, the problem of industrial pollution gas emissions becomes worse in those developing countries, threatening the health of a tremendously large amount of people (Bellinger et al. 2017). The accurate prediction of air quality indices (AQIs) in the short-term future helps decision makers take necessary actions to alleviate the air pollution situations. For example, the hour-ahead prediction of $PM_{2.5}$ can be useful for the city pollution central control system to send pre-caution messages and issue further preventive actions if necessary.

Traditional air quality indicators usually include $PM_{2.5}$ and PM_{10}, which are defined according to the pollution particle sizes. The indicator values can be influenced by various potential factors, such as the location, weather, transportation and industry. Those factors are usually not available in the sensor dataset (data collected by remote sensors), making the short-term forecasting difficult for traditional model-based methods (Gu et al. 2018; Ando et al. 2000; Yuan et al. 2015). More recent studies show that artificial intelligence (AI) enhanced models have the advantages for univariate time series data forecasting.

Existing AI enhanced forecasting methods include machine learning models, such as the support vector regression (SVR) (Liu et al. 2017), and deep learning models, such as the recurrent neural network (RNN) and long short term memory (LSTM) neural network. RNN is probably the first formal attempt of neural networks that specifically deals with time series data, where the hidden layers are used to store historical information to perform the forecasting tasks (Shi et al. 2017). The main shortcoming of the RNN is that it becomes problematic for RNN to memorize a long history of data. LSTM is a special form of RNN that has better performance in memorizing information and forecasting (Kong et al. 2017). The original neurons in RNN are replaced by memory cells in LSTM. Each memory cell consists of three gates, namely, forget gate, memory gate and output gate to dispose, memorize and transfer information, respectively.

Although existing AI enhanced forecasting models have demonstrated their capability of forecasting time series data, particularly, for the air quality forecasting problem, two obstacles exist in the current literature (Kolehmainen et al. 2001). First, there are many internal factors that affect the air quality data, which are usually unavailable in the training process. The air quality data thus becomes extremely non-stationary, making it almost not possible for a single ML model to capture the moving pattern. Second, compared with other forecasting problems, such as stock index forecasting (Zhuo 2018), energy consumption forecasting (Yan et al. 2019), solar irradiance forecasting (Yan et al. 2020), a more extended history of the air-quality data is required to be retained in

the neural network for better forecasting accuracy. The AI-enhanced forecasting model is demanded to have more complex internal structure memorizing longer time historical states.

In this study, we propose a novel forecasting model that integrates an extended SWT method (ESWT) and the nested long short-term memory (NLSTM) neural network for $PM_{2.5}$ air hour-ahead quality forecasting. The NLSTM model is a recently developed deep learning technique, dealing with time-series data forecasting. The NLSTM model nests an additional LSTM cell in the original memory cell of a LSTM neural network to enhance the memory competence of the original LSTM structure (Moniz and Krueger 2018). We further extend the original NLSTM neural network by combining the deep learning process with an extended SWT method (ESWT-NLSTM). Ordinary SWT proposes to decompose the low frequency sub-signal that neglects the high frequency sub-signal, which is harder to predict. The proposed ESWT algorithm extends the original decomposition process into two steps: The original data is first decomposed into high frequency component and low frequency component. Secondly, The SWT is applied to each component separately. The original air quality data is decomposed into sub-signal wavelets that have more stationary wavelet forms for forecasting. Each sub-signal wavelet is then attached with an NLSTM neural network to produce forecasting result. The final forecasting output is produced by combining all individual forecasting results through inverse wavelet transform (IWT). A real-world $PM_{2.5}$ air quality dataset collected in Beijing, China is explored in the experimental section, evaluating the performance of the proposed method. The results show that the proposed method outperforms state-of-art forecasting methods in terms of four metrics, including MAE, RMSE, MAPE and R^2.

In summary, the main contributions of this study to the literature in the fields of smart environment design and data-driven forecasting include the following points.

1) NLSTM in smart environment design. As a recently invented deep learning technology, NLSTM is proposed by Moniz and Krueger in 2018 (Moniz and Krueger 2018). To the best of our knowledge, this is the first study considering NLSTM in air quality prediction in smart environment design.

2) Hybrid structure of deep learning techniques. The original NLSTM is further extended to adapt the volatile air quality time series signal by decomposing the original data into more stationary sub-signals using an extended SWT method (ESWT). The hybrid forecasting model combining ESWT and NLSTM is found more suitable for air quality forecasting, while the original data is volatile and univariate.

3) Accuracy improvement. In the experimental phase, we compared the proposed ESWT-NLSTM method with state-of-art forecasting methods, including decision tree, random forest, support vector regression (SVR), multi-layer perception (MLP), conventional LSTM extensions and empirical mode decomposition (EMD) embedded LSTM extensions. The ESWT-NLSTM outperforms all compared method using four different error metrics.

4) Forecasting lag reduction. According to the experimental results obtained, our method has significantly reduced the forecasting lag due to the decomposition operation using wavelet transform. The lag reduction is extremely useful in real-time air quality monitoring and forecasting.

II. Related works

Time series data forecasting is one of the hot topics in the fields of machine learning and AI. Zhou et al. (Zhou et al. 2019) introduced an attention-LSTM model to forecast solar energy output considering seasonal changes of irradiance. However, the solar irradiance dataset the study employed is a combined sequence of daily trend caused by solar movement and local fluctuations caused by other factors. The method proposed in (Zhou et al. 2019) did not fully consider the characteristics of the original data. Wang et al. (Wang et al. 2019) proposed to use the gated recurrent unit (GRU) neural network to perform air quality forecasting in internet of things (IoT). Gated recurrent unit (GRU) extends the original LSTM, which solves the problem of slow convergence and maintains the prediction accuracy of LSTM. The main shortcoming of GRU is that it does not fully extract the effective features indicating the potential relationship in temporal space for time series data. Hu et al. (Hu et al. 2019; Yan et al. 2020) compared different machine learning models for tunnel surface settlement forecasting. The tunnel settlement data in the study is relatively small and the proposed machine learning method is more suitable than the comparative models. Yan et al. (Yan et al. 2019; Yan et al. 2018) combined convolutional neural network (CNN) with LSTM to predict energy consumption for individual households. Putu Sugiartawan et al. (Sugiartawan et al. 2017) proposed a deep learning model combining wavelet transform and LSTM to predict the growth in the number of visitors and tourism investments. Similarly, Zhang et al. (Zhang et al. 2020) proposed an EMD-LSTM hybrid model combining empirical mode decomposition (EMD) and LSTM for air quality forecasting. In (Zhang et al. 2020), the purpose of the EMD function is used similar to that of the wavelet transform. It decomposes unstable time series data into multiple data components of different frequencies to achieve more accurate prediction. In Section IV, we compared the performance of our method with that of EMD-LSTM with the real-world air quality data. Huang and Wu (Huang and Wu 2008) proposed a GA-based support vector machine (SVM) model for financial time series data forecasting. Emamgholizadeh et al. (Emamgholizadeh et al. 2014) proposed an artificial neural network (ANN) model with adaptive neuro-fuzzy inference system for ground water level prediction. The forecasting accuracy was high. Wang et al. (Wang et al. 2016) employed Elman Recurrent Random Neural Networks (ERNN) model to forecast stock market indices. The proposed ST-ERNN has greatly improved the forecasting accuracy compared to the existing methods.

AI methodology-based predictions of air quality and air pollution receive increasing attention due to climate change and urbanization. Various AI models have been proposed and implemented for air quality prediction. Alimissis et al. (Alimissis et al. 2018) explored two interpolation methods for urban air pollution modeling. Among the two compared methods, the artificial neural network model has the superior performances on all five tested pollutants. Elangasinghe et al. (Elangasinghe et al. 2014) extracted key information from daily available meteorological parameters and built a physical-based artificial neural network air pollutant prediction model that can fully capture the time variation of air pollutant concentrations under specific circumstances. The main

shortcoming of Elangasinghe et al.'s work is that only part of the historical data is used for training. Pardo and Malpica (Pardo and Malpica 2017) used a double-layered LSTM neural network to predict the air quality model of Madrid. Deeper LSTM networks may improve the forecasting performance but also result in higher computational complexity. Song et al. (Song et al. 2019) proposed a combined model of LSTM and Kalman filtering to predict concentration of several components that affect air quality. In (Bai et al. 2019), a stacked auto-encoder model emphasizing on seasonality is proposed to forecast $PM_{2.5}$ values in an hourly manner, throwing alarm messages whenever the forecasted value is above a threshold. The main shortcoming for the work in (Bai et al. 2019) is that the threshold is dynamic, instead of fixed, in most of the situations, which is hard to be modeled by a singular mathematical function. In (Zhu et al. 2017), Zhu et al. proposed two hybrid time series forecasting models, EMD-SVR and EMD-ARIMA, to forecast the air quality indicators. The experimental results showed that the two EMD integrated models outperform the conventional methods, such as ARIMA, support vector regression (SVR), general regression neural network (GRNN), EMD-GRNN, Wavelet-GRNN and Wavelet-SVR. A hybrid model that integrates graph convolutional networks and LSTM (GC-LSTM) is proposed in (Qi et al. 2019) to combine historical observation data from different sites as a series of time-space maps for $PM_{2.5}$ air quality forecasting. The time-space mapping includes the information of historical meteorological factors, spatial items, and other time series attributes. Wen et al. (Wen et al. 2019) proposed a spatiotemporal convolution-long short-term memory neural network extension (C-LSTME) model to forecast the concentration of air quality indicators. The model incorporates the historical air quality data of the current site and K nearest neighbor (KNN) sites into the model to cover the spatial and temporal nature of the data. The combination of CNN and LSTM-NN is employed to extract advanced spatio-temporal features. In addition, the meteorological data and aerosol data are integrated to improve the forecasting results. However, the constructions of both spatial and temporal model require a large amount of data records, which significantly increases the time complexity of the proposed method.

III. Methodology

In this section, first, we introduce the dataset and the necessary pre-processing step for the hybrid deep learning framework for air quality forecasting, including zero-mean normalization and wavelet transform. By wavelet transform, the data is decomposed into more stationary sub-signals. Second, each sub-signal is attached with a nested LSTM (NLSTM) neural network for forecasting. Last, the overall forecasting framework is described combining wavelet transform, NLSTM neural networks and inverse wavelet transform (IWT).

A. Overview of the proposed model

The flowchart of the proposed method is shown in Figure 1. The entire air-quality forecasting process is composed of two phases, the data pre-processing phase and the prediction phase.

In the data pre-processing phase, the original univariate $PM_{2.5}$ dataset is first normalized using zero-mean normalization method. Then, ESWT is applied to decompose the dataset into multiple sub-signals by two steps. On the basis of the decomposed sub-signals, the time series for each data sample is created and the whole dataset is finally divided into training set, validation set and test set. After this phase, the original univariate $PM_{2.5}$ data is transformed into multiple well-processed and properly-divided sub-signals.

In the prediction phase, each sub-signal is considered an independent dataset and assigned to one NLSTM to preform hour-ahead prediction. Afterwards, with inverse wavelet transform (IWT), all of the sub-signal prediction results are combined to reconstruct the complete result. At last, the final prediction result is produced after conducting the inverse process of zero-mean normalization. The forecasting performance is evaluated by calculating the prediction error.

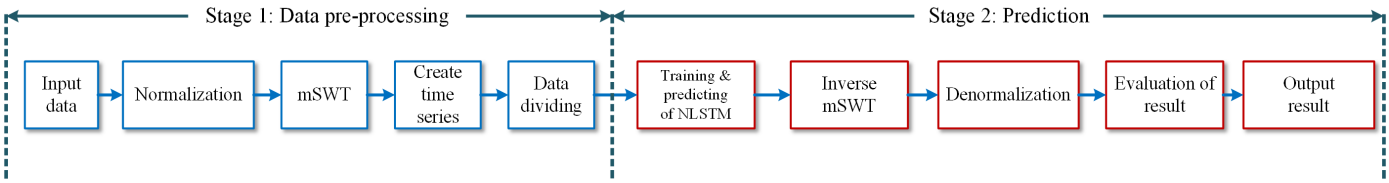


Figure 1: Flowchart of the proposed model, which is divided into two parts: data pre-processing and prediction

B. Data pre-processing phase

1) Data normalization

The model proposed in this paper uses the zero-mean normalization (z-score normalization) method to normalize the recorded $PM_{2.5}$ concentration data. Zero-mean normalization is also called standard deviation standardization. The processed data has a mean of 0 and a standard deviation of 1. And the normalized data is calculated using the normalization formula:

$$x^* = \frac{x - \bar{x}}{\sigma} \tag{1}$$

where $\bar{x}$ is the mean of the original data, and σ is the standard deviation of the original data. The normalization helps the machine learning algorithms better measuring the distance between the standard deviation and mean of the processed data samples.

Similarly, the inverse normalization formula can be derived as:

$$x = \sigma x^* + \bar{x} \tag{2}$$

2) The extended stationary wavelet transform (ESWT)

The original $PM_{2.5}$ concentration time series data is very volatile and fast-changing, making its characteristics very difficult to

capture by the deep learning models. The explanation of such feature is caused by its mixture of factors of different temporal resolution. The raw data includes components of both low and high frequency, which can also be explained as seasonal longer-term trend and shorter-term fluctuations. Therefore, ESWT is introduced to decompose the original data to separate the temporal features for better forecasting, and the ESWT is based on the regular SWT.

Unlike the Fourier transform, which uses infinitely long sine functions with different frequencies as the basis function, the wavelet transform uses finite-length, attenuating wavelets as the basis function. Through the window adjustment method (Huang et al. 2002), the input signal is decomposed into low-frequency signals that reflect the overall trend of data changes and high-frequency signals that fluctuate sharply.

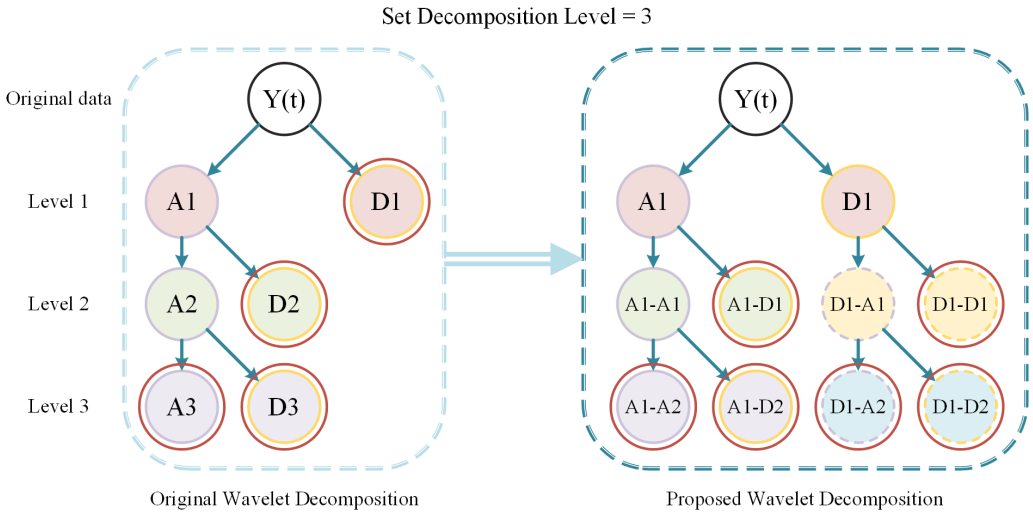
The original $PM_{2.5}$ data is decomposed into 'sub-signals' of different dimensions by discrete stationary wavelet transform (SWT) decomposition, using Daubechies Wavelet as basis function. Compared with the original data, these sequence groups have the same sizes by up-sampling after filtering, resulting in more stationary signals and fewer singular value points.

By wavelet transform decomposition of level m, a data sequence is decomposed into:

$$y(t) = A_{mt} + \sum_{i=1}^{m} D_{it} \tag{3}$$

In equation (3), A_{mt} is an approximate information set, indicating the overall trend characteristics of the original data and D_{it} is a high-frequency information set, and represents a small high-frequency fluctuation, that is, a noise portion of the original data.

The regular SWT only composes further decomposition upon the lower frequency component A_{mt}, because the method is mostly used for filtering the noises. Therefore, the major focus of the SWT is to make the low frequency component, which is the major signal, more stable and recognizable. However, in the experiments in deep learning driven time series forecasting, the main obstacle is that the high frequency component is too volatile and is difficult to be learned by forecasting models.



(a) (b)

Figure 2: The decomposition of SWT and ESWT. (a) Original SWT process. (b) Proposed modified SWT.

The proposed ESWT method decomposes the high frequency components together with the low frequency components. Therefore, the time series information within the high frequency signal can be more effectively and accurately expressed with the decomposition. The decomposing process is demonstrated in Figure 2.

In the proposed model, we used ESWT with decompose level of 3, decomposing the original raw data y(t) into six sub-signals in two steps. First, the original air quality data is decomposed into A1 and D1. Then, both A1 and D1 is further decomposed into A1-A2, A1-D2, A1-D1 and D1-A2, D1-D2, D1-D1, respectively. The original SWT method decomposing the original data into A3, D3, D2, D1 is employed as comparative models in Section IV to demonstrate the superiority of the ESWT method.

3) Dividing of the dataset

With wavelet transform, the original data is decomposed into six sub-signal datasets. In time series forecasting, the data in the history is employed to predict the data in the future.

Each of the datasets is divided into data series X and Y. Y is the sequence of data samples to be predicted. X is the sequence of history time windows corresponding to Y, and each time window consists of history data samples to predict the corresponding data sample in Y. The length of the history time window in X is called time step L, which means how many history data samples are utilized to predict the next data sample. For example, the ith element in X and Y should be:

$$\begin{cases} X_{(i)} = \{x_i, x_{i+1}, \dots, x_{i+L}\} \\ Y_{(i)} = x_{i+L+1} \end{cases} \tag{4}$$

Then, the datasets are divided into training set, validation set and test set. The experiment utilized the first 22,800 (950 days) data samples as the training set, 1200 data samples (50 days) as validation set, 1200 data samples (50 days) of the remaining data as the testing dataset. The proportion between training, validation and testing dataset is approximately 19:1:1.

After dividing and fitting the decomposed datasets into a proper shape, they are ready to be learned and predicted by the deep

learning models in the prediction phase.

C. Prediction phase

1) NLSTM

The recently proposed nested LSTM model nests one more memory cell of LSTM to the original LSTM cell to improve the prediction performance by memorize additional information in the historical data (Ma et al. 2018). The external storage cell is free to selectively read and write the relevant long-term information of the internal cell, which in overall improves the robustness of the original LSTM neural network structure.

In LSTM, the output gate follows a principle that information that is not relevant to the current time step is still worth remembering.

As shown in Figure 3, the structure of a common LSTM memory cell is as follows:

$$i_t = \sigma_i(x_t W_{xi} + h_{t-1} W_{hi} + b_i) \quad (5)$$

$$f_t = \sigma_f(x_t W_{xf} + h_{t-1} W_{hf} + b_f) \quad (6)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma_c(x_t W_{xc} + h_{t-1} W_{hc} + b_c) \quad (7)$$

$$o_t = \sigma_o(x_t W_{xo} + h_{t-1} W_{ho} + b_o) \quad (8)$$

$$h_t = o_t \odot \sigma_h(c_t) \quad (9)$$

Consider equation (9), the update of the memory cell state c_t is made by adding two parts, that is

$$\tilde{h}_{t-1} = f_t \odot c_{t-1} \quad (10)$$

$$\tilde{x}_t = i_t \odot \sigma_c(x_t W_{xc} + h_{t-1} W_{hc} + b_c) \quad (11)$$

Using another LSTM cell to replace the c_t calculation equation in the ordinary LSTM model constitutes a nested LSTM cell, that is, an NLSTM cell. The outer LSTM is called the outer memory, and the inner LSTM is called the inner memory. In an ordinary LSTM cell, the memory cell status c_t is updated as follows:

$$c_t = \tilde{h}_{t-1} + \tilde{x}_t \quad (12)$$

In the NLSTM cell, this process is replaced by an inner LSTM cell, where $\tilde{x}_t$ and $\tilde{h}_{t-1}$ are used as short-term and long-term memory inputs, respectively. The structure of the inner LSTM cell is as follows:

$$\tilde{i}_t = \tilde{\sigma}_i(\tilde{x}_t \tilde{W}_{xi} + \tilde{h}_{t-1} \tilde{W}_{hi} + \tilde{b}_i) \quad (13)$$

$$\tilde{f}_t = \tilde{\sigma}_f(\tilde{x}_t \tilde{W}_{xf} + \tilde{h}_{t-1} \tilde{W}_{hf} + \tilde{b}_f) \quad (14)$$

$$\tilde{c}_t = \tilde{f}_t \odot \tilde{c}_{t-1} + \tilde{i}_t \odot \tilde{\sigma}_c(\tilde{x}_t \tilde{W}_{xc} + \tilde{h}_{t-1} \tilde{W}_{hc} + \tilde{b}_c) \quad (15)$$

$$\tilde{o}_t = \tilde{\sigma}_o(\tilde{x}_t \tilde{W}_{xo} + \tilde{h}_{t-1} \tilde{W}_{ho} + \tilde{b}_o) \quad (16)$$

$$\tilde{h}_t = \tilde{o}_t \odot \tilde{\sigma}_h(\tilde{c}_t) \quad (17)$$

In this way, the update mode of the outer memory cell status c_t becomes:

$$c_t = \tilde{h}_t \quad (18)$$

With the structure of outer and inner memory, NLSTM networks form a hierarchy of memory, remembering longer term information compared with traditional LSTM neural networks. Therefore, in this study, the NLSTM neural network is implemented for better prediction performance on air quality forecasting.

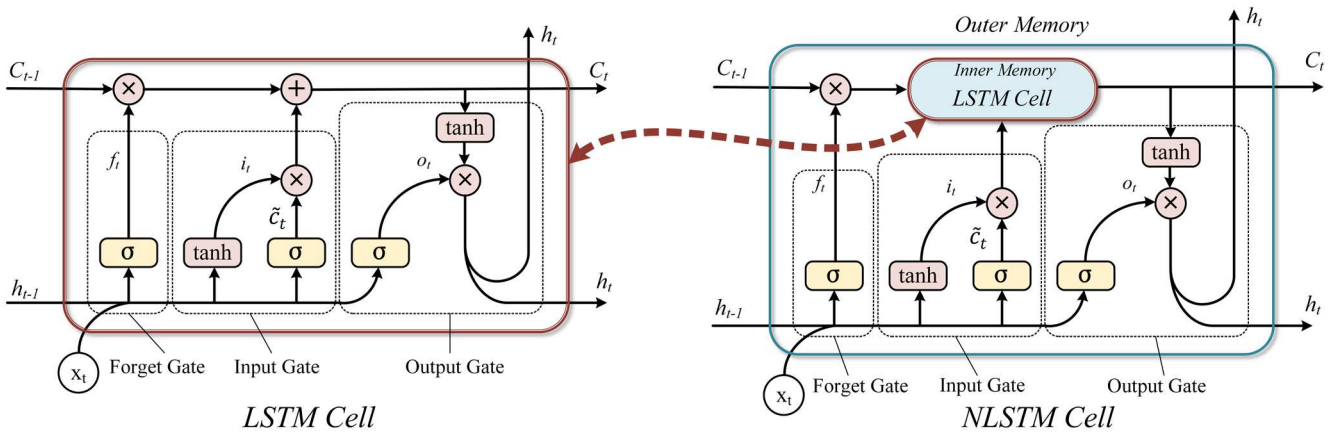


Figure 3: Internal structure of the NLSTM memory cell.

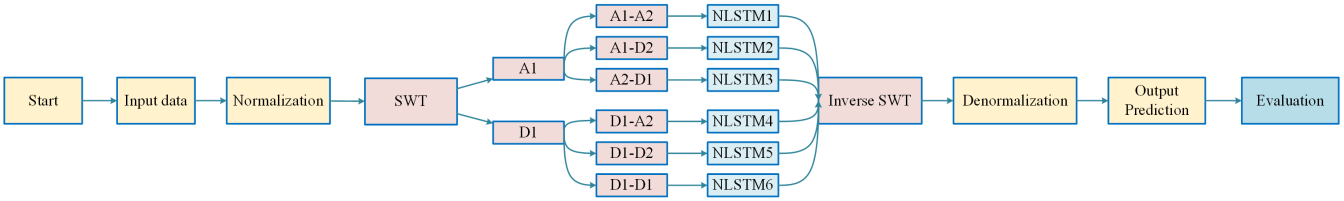


Figure 4: The flowchart of the proposed ESWT-NLSTM framework

2) Hybrid ESWT-NLSTM forecasting model

The overall flowchart of the proposed Hybrid ESWT-NLSTM forecasting model is depicted in Figure 4.

In the proposed framework, ESWT is applied to decompose the univariate $PM_{2.5}$ data into six sub-signals, which are named A1-A2, A1-D2, A1-D1, D1-A2, D1-D2 and D1-D1. The six sub-signals are considered independent datasets and they are learned and predicted by six NLSTM models separately. The six NLSTM neural networks produce the prediction of A1-A2, A1-D2, A1-D1, D1-A2, D1-D2 and D1-D1, respectively, and the final prediction of the original $PM_{2.5}$ data is generated by combining the prediction results of the six sub-signals.

IV. Experimental process and results

In the experiment, the original univariate $PM_{2.5}$ dataset is first to be decomposed into six sub-signals shown in Figure 5. The six sub-signals are utilized to train six individual NLSTM neural networks. At last, the prediction results for each sub-signals are produced and combined into the final result.

A. Data description

The air quality data is collected from 12 observing stations around Beijing from year 2013 to 2017, downloaded from UCI Machine Learning Repository (Zhang et al. 2017). The data consists of 12 variables, including concentration of $PM_{2.5}$, PM_{10} and so on. In this study, only the univariate $PM_{2.5}$ data series is utilized.

The original sensor data is collected in a time step of every hour, across 1,461 days. Thus, the total number of data samples for each observing station is 35,064.

B. Evaluation metrics

To better quantify and evaluate the accuracy of the experimental prediction results, four evaluation metrics, namely, average absolute error (MAE), root mean square error (RMSE), average absolute percentage error (MAPE) and r-square (R^2), are used to evaluate the forecasting performance of the proposed model by calculating the gap between the actual value and the predicted value. The specific formulas of the above mentioned four metrics are listed below:

$$\mathrm{MAE}(f,\hat{f}) = \frac{1}{n}\sum_{i=1}^{n}\left|f_i - \hat{f}_i\right| \tag{18}$$

$$\mathrm{RMSE}(f,\hat{f}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\left|f_i - \hat{f}_i\right|\right)^2} \tag{19}$$

$$\mathrm{MAPE}(f,\hat{f}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|f_i - \hat{f}_i\right|}{f_i} \tag{20}$$

$$R^2(f,\hat{f}) = 1 - \frac{\sum_{i=1}^{n}(f_i - \hat{f}_i)^2}{\sum_{i=1}^{n}(f_i - \bar{f})^2} \tag{21}$$

where f refers to the actual data value; and $\hat{f}$ is the predicted value.

Evaluation metrics including MAE, RMSE and MAPE evaluate the forecasting performance by calculating the level of error. Amongst the metrics, MAE is the average of absolute errors between the real value and the predicted value, RMSE measures the deviation between the actual value and the predicted value but it is more sensitive by outliers. MAPE measures the relative level of absolute error in a proportional approach. R^2 is the coefficient of determination that evaluates the fitting effectiveness.

Therefore, lower the values of MAE, RMSE and MAPE reflect lower error level, meaning the prediction is more accurate. The closer the value of R^2 is to 1, the better the fitting effect.

C. Experimental setup

The experiments were conducted in a server equipped with Intel i7-8700K CPU and NVIDIA GeForce GTX 1080 GPU. The software environment is based on the Windows 10 operating system with Python 3.6.6 installed. Multiple Python expansion packages are installed, including Numpy, Sklearn, Tensorflow and Keras. The proposed deep learning model is constructed using Keras with Tensorflow backend. With data pre-processing, the original $PM_{2.5}$ univariate data series is decomposed into six sub-signals. The six sub-signals are learned and predicted by six NLSTM models, respectively. The six NLSTM models share the completely same inner structure, which consists of one NLSTM layer with 64 cells and a fully connected layer.

The RMSprop optimizer is adopted to optimize the deep learning model during the training process. The default learning rate is used. The loss function is set to be mean square error (MSE). The loss values of the six sub-signals in the training process is shown in Figure 6.

To further justify the capability and efficiency of the proposed model, multiple state-of-art time series forecasting models are compared, including decision tree, random forest, support vector regression (SVR), multi-layer perception (MLP), conventional LSTM extensions and empirical mode decomposition (EMD) embedded LSTM extensions. Amongst the compared models, the machine learning models are implemented with sklearn package. The deep learning models are constructed using Keras and the inner structures are similar to the proposed model, which consist of one LSTM or extended LSTM layer with 64 cells and a fully connected layer. By building the same inner structure for each deep learning model, the difference between the LSTM extensions and the superiority of the proposed ESWT-NLSTM is better reviewed. The forecasting performance of different models is evaluated with the aforementioned evaluation metrics and then compared with the proposed method.

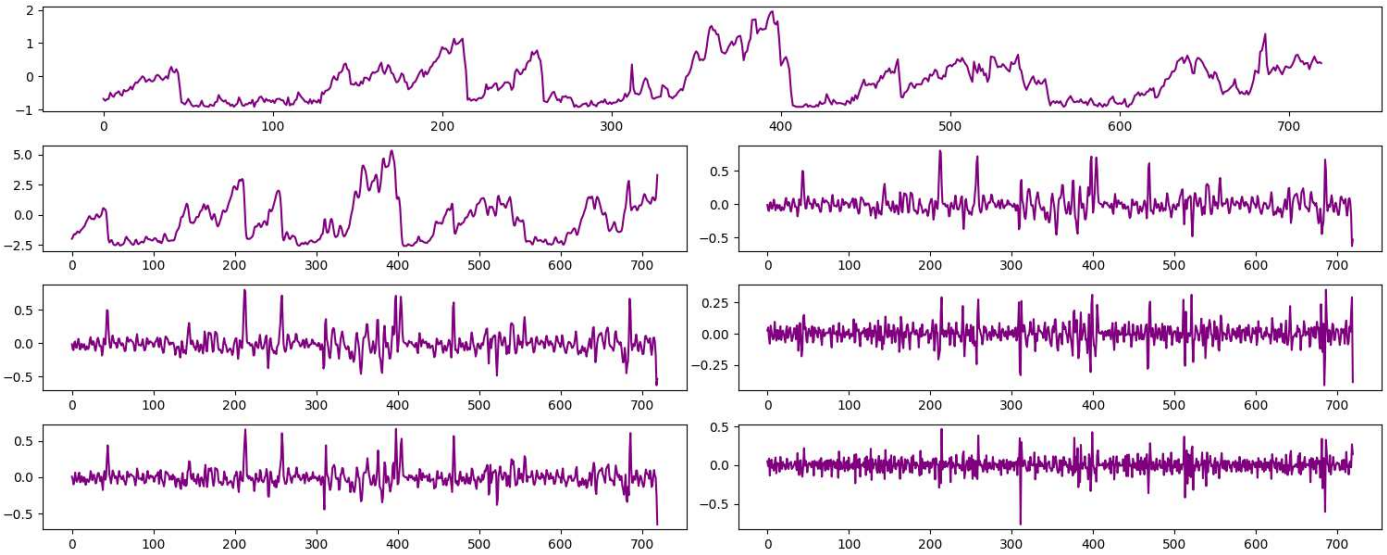


Figure 5: Decomposed wavelets (720 hours) of the original $PM_{2.5}$ data points. The figure on the top is the original data. The figures in the left column from top to bottom are A1-A2, A1-D2, A1-D1, respectively. The figures in the right column from top to bottom are D1-A2, D1-D2, D1-D1, respectively.

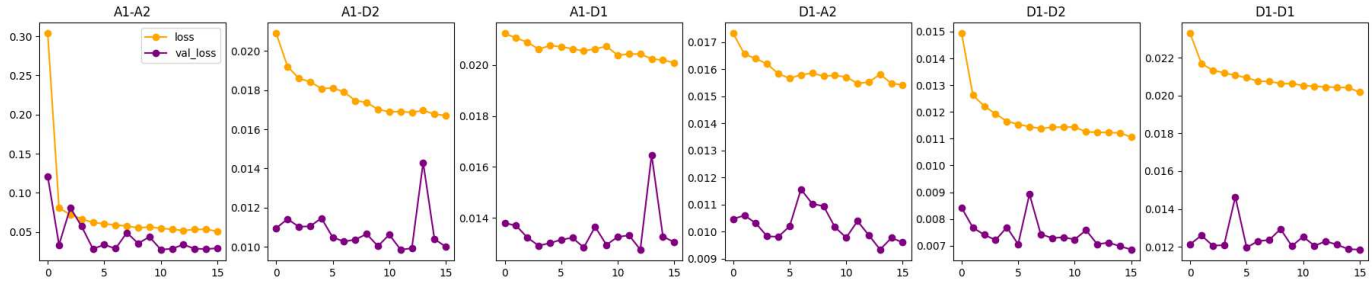


Figure 6: Training loss changes of A1-A2, A1-D2, A1-D1, D1-A2, D1-D2 and D1-D1. The 'loss' label refers to the loss of the training procedure and 'val_loss' label refers to the loss of the validation set.

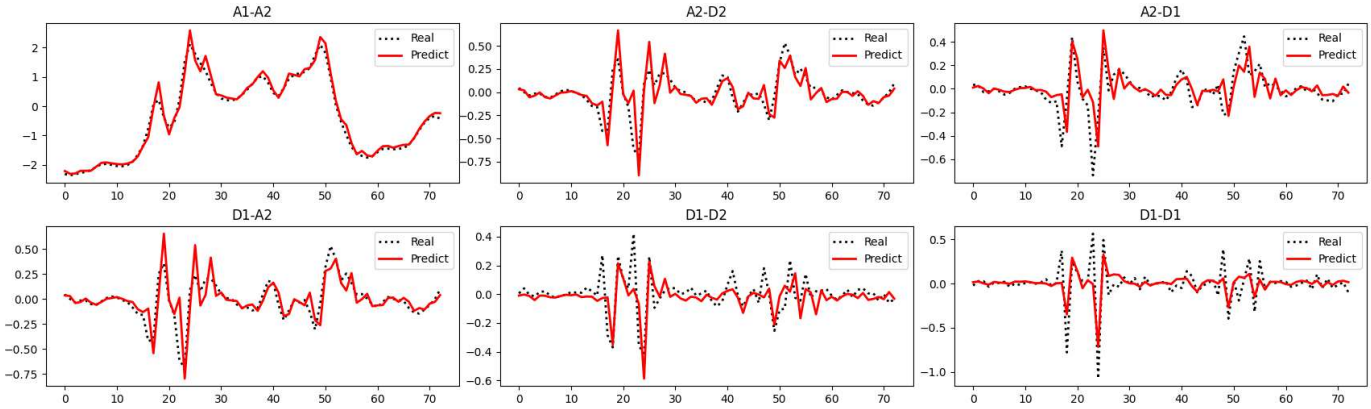


Figure 7: The prediction result of each component compared with the real value. The red line is the predicted result and the black dotted line is the actual data.

D. The result

A comparative study between the prediction results of each component and the actual value is shown in Figure 7. The combination of all sub-signal predictions shown in Figure 7 is depicted in Figure 8.

In Figure 8, a prediction performance comparison between the proposed methods and existing machine learning and deep learning methods, including decision tree, random forest, support vector regression (SVR), multi-layer perception (MLP), conventional LSTM extensions and empirical mode decomposition (EMD) embedded LSTM extensions is shown.

The absolute error of the proposed method and the comparative models are compared in Figure 9.

Based on the aforementioned evaluation metrics, the evaluation result on forecasting performance of the proposed forecasting models and comparative models is made and listed in Table 1.

According to Figure 8 and table 1, the forecasting performance of proposed ESWT-NLSTM neural networks is remarkably improved compared to the comparative models. The absolute error indices are cut down sharply and R^2 index shows that the fitting

effect is greatly improved. The lines in Figure 8 show that the proposed model can successfully and correctly predict most of the peaks and troughs in air quality data, while the comparative models show lagging effect of different degree. The boxes in Figure 9 show that the absolute error of the proposed method is lower in average and is distributed in a smaller range, which means the predicted values can reflect the actual value more accurately but also more steadily.

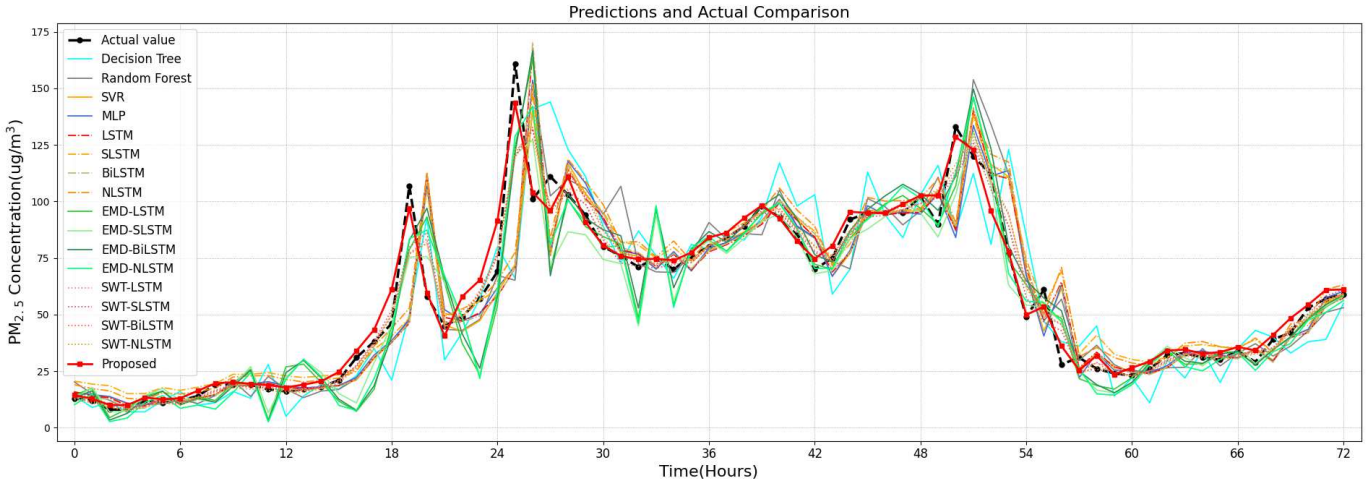


Figure 8: Comparing the final prediction result with the actual value. For better visualization effect, only 72 data samples are depicted.

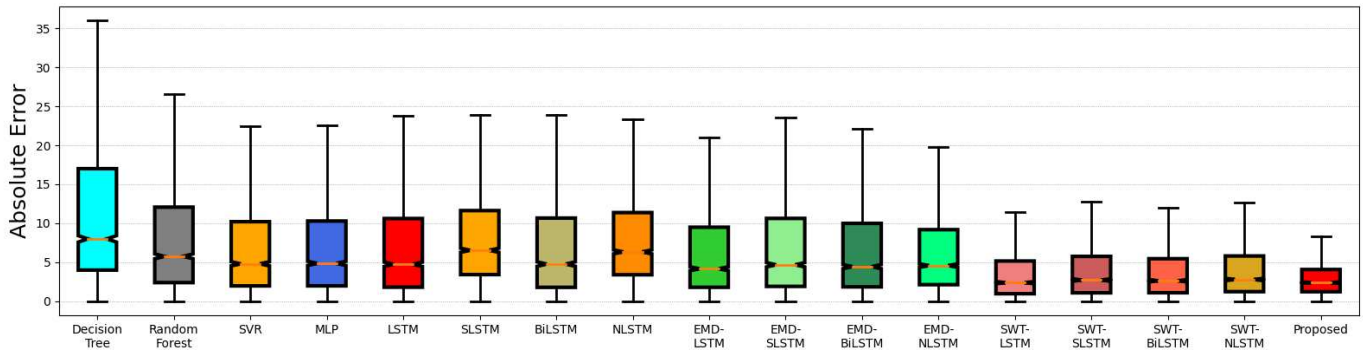


Figure 9: Absolute error of the proposed method and the compared models.

TABLE 1: Performance comparison between proposed model and multiple comparative methods.

Algorithms	MAE	RMSE	MAPE(%)	R^2
Decision Tree	14.495	24.970	46.94	0.798
Random Forest	10.388	18.646	37.62	0.887
SVR	9.138	17.466	33.03	0.901
MLP	9.146	17.395	33.40	0.902
LSTM (Greff et al. 2016)	9.232	17.395	31.66	0.902
SLSTM (Ballesteros et al. 2016)	10.437	17.761	44.32	0.898
NLSTM (Moniz and Krueger 2018)	9.290	17.968	31.60	0.895
BiLSTM (Graves et al. 2013)	10.291	18.048	40.51	0.894
EMD-LSTM (Rilling et al. 2003)	7.413	11.994	24.82	0.953
EMD-SLSTM	8.092	12.909	23.25	0.946
EMD-NLSTM	7.248	11.292	22.75	0.959
EMD-BiLSTM	7.962	13.233	25.96	0.943
SWT-LSTM (Yan et al. 2019)	4.531	8.900	17.42	0.974
SWT-SLSTM	5.101	9.579	18.17	0.970
SWT-NLSTM	4.782	9.049	17.16	0.973
SWT-BiLSTM	5.008	9.168	16.76	0.973

PROPOSED	**3.456**	**5.579**	**11.61**	**0.990**

Compared with the results of air quality forecasting with the existing machine learning methods, including decision tree, random forest, support vector regression (SVR), multi-layer perception (MLP). Our method shows lower prediction errors among all compared methods.

Compared with the results of air quality forecasting with the conventional LSTM extensions, including the original LSTM model (Greff et al. 2016), Stacked LSTM (SLSTM) model (Ballesteros et al. 2016), bidirectional LSTM (BiLSTM) model (Graves et al. 2013), and nested LSTM (NLSTM) model (Moniz and Krueger 2018), the performance of the proposed ESWT-NLSTM model shows superiority upon the compared methods without the ESWT on the purpose of detrending and denoising, in terms of all four evaluation error metrics. In Table 1, it is also noted that the MAPE value of the proposed WNLSTM method is only around half of those for the traditional methods, including (Greff et al. 2016), (Ballesteros et al. 2016), (Graves et al. 2013) and (Moniz and Krueger 2018).

In Table 1, the results of empirical mode decomposition (EMD) (Rilling et al. 2003) embedded LSTM models, including EMD-LSTM model (Bedi and Toshniwal 2018), EMD-SLSTM model, EMD-BiLSTM model, and EMD-NLSTM model are listed. As one of the most popular signal decomposition method, EMD is used in many situations to decompose the original data, similar to the wavelet transform (Rilling et al. 2003; Bedi and Toshniwal 2018). Referring Table 1, the prediction performance of the proposed WNLSTM model outperforms all hybrid models using EMD for data pre-processing for various LSTM extensions. In particular, the performance comparison between WNLSTM and EMD-NLSTM shows that wavelet transform has the obvious advantage over EMD and is more suitable for time series prediction of datasets like $PM_{2.5}$ concentration.

Compared with the LSTM extensions with regular SWT method shown in Figure 2(a), the proposed method also shows significant superiority by additionally decomposing the high frequency component. The decomposition and separately conducted learning of the higher frequency sub-signals improves the forecasting performance in predicting the more volatile part in the original data and thus improves the overall prediction accuracy. Therefore, according to Figures 8-9 and Table 1, the level of error is reduced and the time lag between the actual data and predicted data is also shortened.

Based on the comprehensive comparative results list in Table 1, the combination of ESWT and NLSTM proves to be the best fit for the target air quality forecasting problem. The modified SWT outperforms the state-of-art EMD and SWT method by taking the high frequency into consideration. And because of the nested memory cell structure, the NLSTM-based hybrid deep learning framework outperforms other extensions of LSTM neural networks, such as the stacked LSTM and bidirectional LSTM networks. The superior performance shows better memorizing and handling longer-term history information for NLSTM networks compared with traditional LSTM extensions.

Figures 10-13 illustrates the forecasting results of the methods used in Table 1 respectively. For better visualization purposes, only 72 data samples are shown. In particular, the performance comparison between the proposed method and machine learning methods listed in Table 1 is showed in Figure 10. The performance comparison between the proposed method and LSTM extensions is showed in Figure 11. Figure 12 shows the comparison between the proposed method and LSTM extensions combining with EMD. Figure 13 shows the comparison between the proposed method and LSTM extensions combining with SWT.

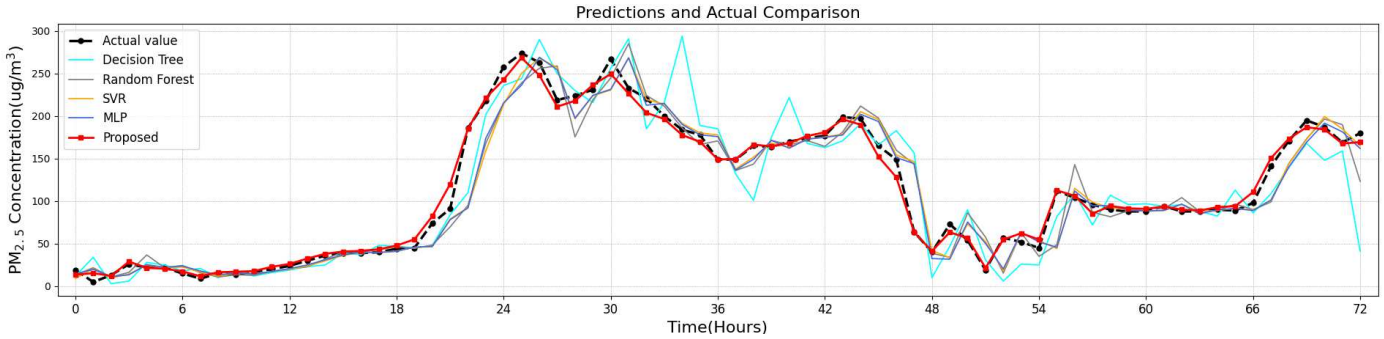


Figure 10: Comparing the prediction result of proposed method and conventional methods (72 hours).

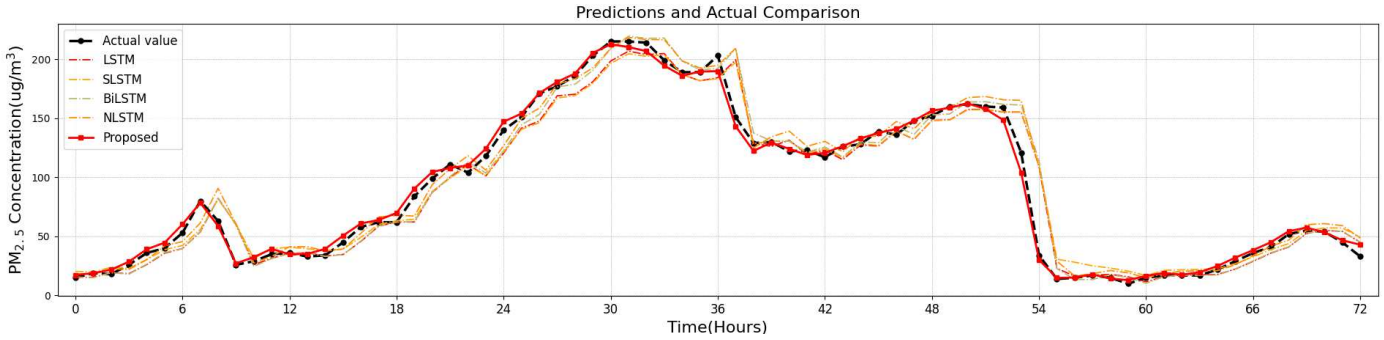


Figure 11: Comparing the prediction result of proposed method and LSTM and extensions (72 hours).

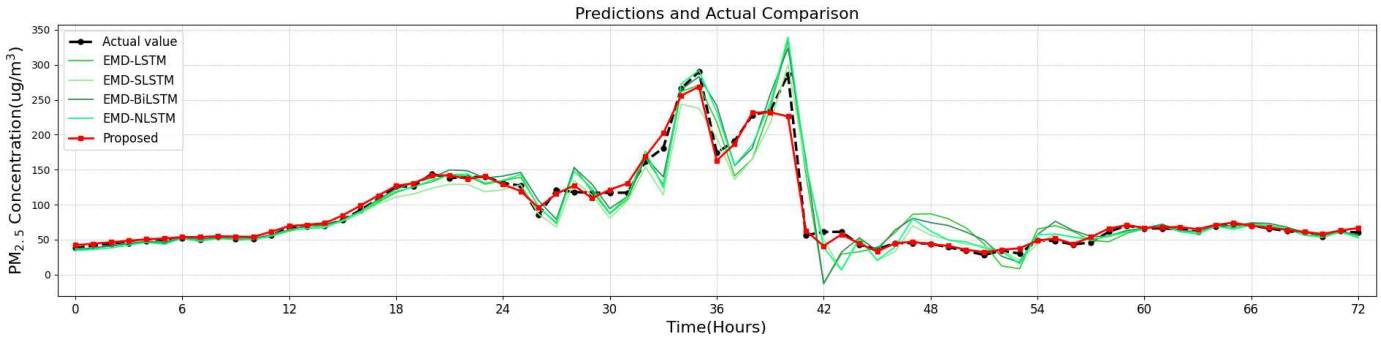


Figure 12: Comparing the prediction result of proposed method and LSTM networks combining EMD (72 hours).

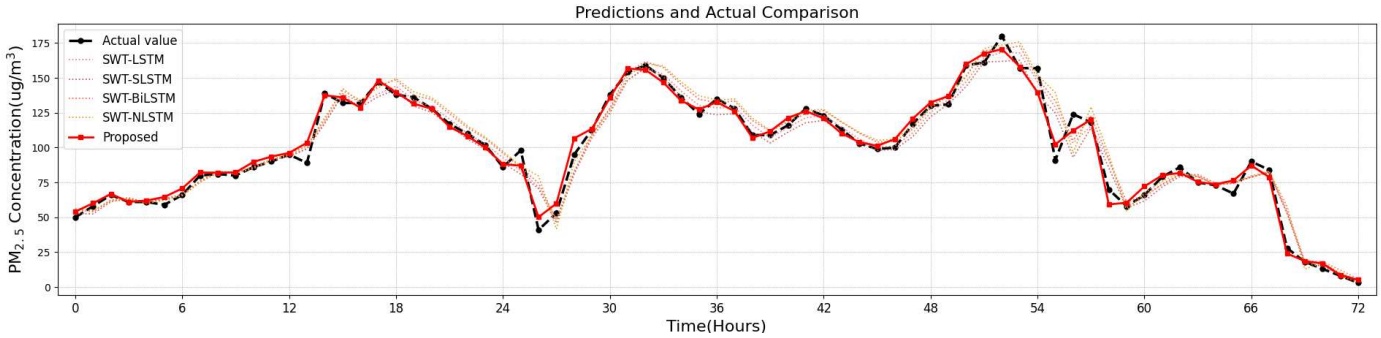


Figure 13: Comparing the prediction result of proposed method and LSTM networks combining SWT (72 hours).

From Figures 10-13, the existing methods not only show higher prediction errors, but also demonstrate obvious lagging effects compared with the proposed ESWT-NLSTM method. For peak values in Figures 10-13, the prediction peaks of other methods are obviously lagging, compared with the proposed method. EMD, as one decomposition method, is originally proposed to deal with the lagging problem. However, the EMD cannot completely separate signals of high and low frequencies for extremely sensitive and drastic changes, which can be notably found in Figure 12. According to Figures 10-13, the lagging effect in air quality prediction using the proposed mothed is seriously reduced, resulting in less forecasting error. Learning more stationary sub-signals separately reduced the sensitivity of transient fluctuations for more stable and robust performance results.

Table 2 lists the prediction performance of the sub-signals decomposed by ESWT and SWT using NLSTM. The performance in predicting the high frequency component D1 is compared. As described, the difference between the proposed modified SWT and the regular SWT is that the ESWT also decompose the high frequency D1 into D1-A2, D1-D2 and D1-D1 and the decomposed sub-signals are learned by NLSTM to improve the forecasting performance on D1. According the results listed in Table 2, the NLSTM combining with ESWT can predict D1 much more accurately, while NLSTM cannot predict D1 well without decomposition.

TABLE 2: Performance comparison between NLSTM combining proposed ESWT and regular SWT. The value is normalized thus it does not reflect the actual magnitude of error.

Component	MAE	RMSE	R^2
ESWT-A1-A2	0.114	0.185	0.993
ESWT-A1-D2	0.069	0.118	0.747
ESWT-A1-D1	0.076	0.137	0.471
ESWT-D1-A2	0.071	0.118	0.745
ESWT-D1-D2	0.057	0.100	0.382
ESWT-D1-D1	0.073	0.134	0.361
SWT-A3	0.114	0.185	0.993
SWT-D3	0.080	0.124	0.721
SWT-D2	0.075	0.134	0.492
SWT-D1	0.094	0.179	-0.008
SWT-D1	0.094	0.179	-0.008
ESWT-D1	0.050	0.091	0.739

Therefore, according to Figure 8 and Figure 13, the SWT embedded models are insensitive to peaks and troughs. The incapability of predicting peaks and troughs is a major shortcoming, because air quality monitoring occasionally involves sending warnings and alarms according to peaks and troughs. Compared with SWT-NLSTM, the proposed method is more accurate in predicting high

frequency component D1. Therefore, the proposed method is more sensitive to minor changes, and is able to predict more detailed subtle changes and peaks.

According to Figures 8-13, the drawbacks of the existing methods include:

1) Insufficient learning capability. The forecasting issue of the unstable $PM_{2.5}$ data requires very high-level learning capability to capture the temporal dependency. The instability is hinted from both temporal domain and frequency domain. Traditional methods, including SVR and deep learning models, such as the original LSTM neural network, failed to learn enough useful features in the temporal domain. Therefore, there are obvious lagging effects and prediction errors in the results. Compared to the conventional methods, the proposed model employs ESWT to stabilize and decompose the original data and thus reduces the predict complexity. Combining the modified SWT method and the upgraded version of LSTM neural network, which is the NLSTM neural network, the propose model is able to better capture the useful features in both temporal domain and frequency domain, consequently producing better quality forecasting results.

2) Unsuitable data decomposing strategy. Compared to SWT, EMD employs global waves to decompose the original data sequence, instead of local wavelets. The property of EMD makes it more suitable for decomposing periodic and more stable data, such as the combinations of multiple sine waves or cosine waves. When decomposing unstable and irregularly curves, such as the $PM_{2.5}$ data, as shown in Figure 12, EMD cannot effectively segregate the seasonal trend and local fluctuations, making the prediction very unstable for harmonic waves. Therefore, EMD is less competitive against SWT in this case, making the models employing EMD to decompose the air quality data less effective compared to the proposed model.

3) Insensitivity to peaks. The regular SWT decomposing technique lacks further analysis towards the higher frequency sub-signals. Therefore, models combining with regular SWT are insensitive to high frequency fluctuations due to the lack of ability to accurately predicting them. According to Figure 8 and Figure 13, compared with the proposed ESWT, the results of models combining regular SWT are too smooth and less accurate when predicting minor fluctuations. The SWT embedded models also tend to fail to predict the peaks and troughs in the $PM_{2.5}$ data, making the forecasting less timely and less meaningful.

From the experimental results shown in Tables 1 and 2, it is evident that the proposed ESWT-NLSTM model outperforms all compared machine learning and deep learning methods in the literature. Compared with other models, not only the prediction error is lower, but also the lagging effect is seriously reduced. Compared with LSTM models combining regular SWT, the proposed model is more sensitive to peaks and fluctuations. Therefore, we believe that the proposed model is suitable for real-world air quality forecasting applications, such as the pollution alarm system that helps people's health.

V. CONCLUSION

This study proposed a hybrid deep learning framework for hour-ahead $PM_{2.5}$ prediction. The proposed method employs the cutting-edge deep learning technology, namely, nested LSTM neural network to combine with the modified SWT based on the state-of-art method SWT to enhance the prediction performance. The original $PM_{2.5}$ dataset has been decomposed into sub-signals, which are more stationary wavelet forms. Individual NLSTM neural networks have been developed and applied to the corresponding sub-signals. The inverse wavelet transform is adopted to produce the final forecasting result.

In the experimental result phase, a real-world dataset collected by weather stations located around Beijing, China is utilized. A comprehensive comparative study with various existing methods in the literature has been conducted. By outperforming most of the existing technique in the related field, we demonstrate that the proposed ESWT-NLSTM framework is effective and suitable for real-world applications for both $PM_{2.5}$ value forecasting and $PM_{2.5}$ varying trend forecasting.

The main limitation of this work is that we only perform the prediction for $PM_{2.5}$ values in the given dataset. Although the forecasting process and results for other air-quality indices are similar to the $PM_{2.5}$, the deep learning technique actually can perform transit learning from one index to another. In this study, we did not take that advantage in the $PM_{2.5}$ prediction, which is one of the future work directions of this study. Additional future work directions of this study include extending the proposed framework to other areas of time-series data analysis, such as wind speed forecasting (Liu et al. 2018), irradiance forecasting (Yan et al. 2020) and energy consumption forecasting (Yan et al. 2019).

Declarations

Ethics approval and consent to participate. Ethics Committee approval was received from the research ethics committees in the College of Information Engineering of China Jiliang University and National University of Singapore.

Consent for publication. All authors declare that they have the consent regarding the publication of this manuscript

Availability of data and materials. The data used in this paper is publicly available at UCI machine learning knowledge base. URL: http://archive.ics.uci.edu/ml/

Competing Interest. No competing interest.

Funding. This work was supported by the Ministry of Education (MOE) Singapore, Tier 1 Grant for National University of Singapore (NUS) under grant number R296000208133.

Authors' contributions. Conceptualization, K.Y.; methodology, N. J.; software, X. M..; validation, Y.Z.; formal analysis, X.M.; investigation, K.Y.; resources, K.Y.; original draft preparation, K.Y.; writing, review and editing, K.Y.; visualization, X.Z.; supervision, K.Y., N. J. and X.Z.; project administration, K.Y.; funding acquisition, K.Y.

Acknowledgements. Authors from the original publication of the dataset (Zhang et al. 2017) are appreciated.

REFERENCES

Alimissis, A., Philippopoulos, K., Tzanis, C. G., & Deligiorgi, D. (2018). Spatial estimation of urban air pollution with the use of artificial neural network models. Atmospheric Environment, 191, 205-213.

Ando, B., Baglio, S., Graziani, S., & Pitrone, N. (2000). Models for air quality management and assessment. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 30(3), 358-363.

Bai, Y., Li, Y., Zeng, B., Li, C., & Zhang, J. (2019). Hourly PM2. 5 concentration forecast using stacked autoencoder model with emphasis on seasonality. Journal of cleaner production, 224, 739-750.

Ballesteros, M., Goldberg, Y., Dyer, C., & Smith, N. A. (2016, November). Training with Exploration Improves a Greedy Stack LSTM Parser. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (pp. 2005-2010).

Bedi, J., & Toshniwal, D. (2018). Empirical mode decomposition based deep learning for electricity demand forecasting. IEEE Access, 6, 49144-49156.

Bellinger, C., Jabbar, M. S. M., Zaïane, O., & Osornio-Vargas, A. (2017). A systematic review of data mining and machine learning for air pollution epidemiology. BMC public health, 17(1), 907.

Elangasinghe, M. A., Singhal, N., Dirks, K. N., & Salmond, J. A. (2014). Development of an ANN–based air pollution forecasting system with explicit knowledge through sensitivity analysis. Atmospheric pollution research, 5(4), 696-708.

Emamgholizadeh, S., Moslemi, K., & Karami, G. (2014). Prediction the groundwater level of bastam plain (Iran) by artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS). Water resources management, 28(15), 5433-5446.

Graves, A., Jaitly, N., & Mohamed, A. R. (2013, December). Hybrid speech recognition with deep bidirectional LSTM. In 2013 IEEE workshop on automatic speech recognition and understanding (pp. 273-278). IEEE.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems, 28(10), 2222-2232.

Gu, K., Qiao, J., & Lin, W. (2018). Recurrent air quality predictor based on meteorology-and pollution-related factors. IEEE Transactions on Industrial Informatics, 14(9), 3946-3955.

Hu, M., Li, W., Yan, K., Ji, Z., & Hu, H. (2019). Modern machine learning techniques for univariate tunnel settlement forecasting: A comparative study. Mathematical Problems in Engineering, 2019.

Huang, S. C., & Wu, T. K. (2008). Integrating GA-based time-scale feature extractions with SVMs for stock index forecasting. Expert Systems with Applications, 35(4), 2080-2088.

Huang, S. J., Yang, T. M., & Huang, J. T. (2002). FPGA realization of wavelet transform for detection of electric power system disturbances. IEEE Transactions on Power Delivery, 17(2), 388-394.

Kolehmainen, M., Martikainen, H., & Ruuskanen, J. (2001). Neural networks and periodic components used in air quality forecasting. Atmospheric Environment, 35(5), 815-825.

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Transactions on Smart Grid, 10(1), 841-851.

Liu, B. C., Binaykia, A., Chang, P. C., Tiwari, M. K., & Tsao, C. C. (2017). Urban air quality forecasting based on multi-dimensional collaborative support vector regression (svr): A case study of beijing-tianjin-shijiazhuang. PloS one, 12(7).

Liu, H., Mi, X., & Li, Y. (2018). Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. Energy Conversion and Management, 159, 54-64.

Ma, X., Li, Y., Cui, Z., & Wang, Y. (2018). Forecasting transportation network speed using deep capsule networks with nested lstm models. arXiv preprint arXiv:1811.04745.

Moniz, J. R. A., & Krueger, D. (2018). Nested lstms. arXiv preprint arXiv:1801.10308.

Pardo, E., & Malpica, N. (2017, June). Air quality forecasting in Madrid using long short-term memory networks. In International

Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 232-239). Springer, Cham.

Qi, Y., Li, Q., Karimian, H., & Liu, D. (2019). A hybrid model for spatiotemporal forecasting of PM2. 5 based on graph convolutional neural network and long short-term memory. Science of the Total Environment, 664, 1-10.

Rilling, G., Flandrin, P., & Goncalves, P. (2003, June). On empirical mode decomposition and its algorithms. In IEEE-EURASIP workshop on nonlinear signal and image processing (Vol. 3, No. 3, pp. 8-11). NSIP-03, Grado (I).

Shi, H., Xu, M., & Li, R. (2017). Deep learning for household load forecasting—A novel pooling deep RNN. IEEE Transactions on Smart Grid, 9(5), 5271-5280.

Song, X., Huang, J., & Song, D. (2019, May). Air Quality Prediction based on LSTM-Kalman Model. In 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC) (pp. 695-699). IEEE.

Sugiartawan, P., Pulungan, R., & Sari, A. K. (2017). Prediction by a hybrid of wavelet transform and long-short-term-memory neural network. International Journal of Advanced Computer Science and Applications, 8(2), 326-332.

Wang, B., Kong, W., Guan, H., & Xiong, N. N. (2019). Air quality forecasting based on gated recurrent long short term memory model in Internet of Things. IEEE Access, 7, 69524-69534.

Wang, J., Wang, J., Fang, W., & Niu, H. (2016). Financial time series prediction using elman recurrent random neural networks. Computational intelligence and neuroscience, 2016.

Wen, C. , Liu, S. , Yao, X. , Peng, L. , Li, X. , & Hu, Y. , et al. (2019). A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. Science of the Total Environment, 654(MAR.1), 1091-1099.

Yan, K., Dai, Y., Xu, M., & Mo, Y. (2020). Tunnel Surface Settlement Forecasting with Ensemble Learning. Sustainability, 12(1), 232.

Yan, K., Li, W., Ji, Z., Qi, M., & Du, Y. (2019). A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households. IEEE Access, 7, 157633-157642.

Yan, K., Shen, H., Wang, L., Zhou, H., Xu, M., & Mo, Y. (2020). Short-Term Solar Irradiance Forecasting Based on a Hybrid Deep Learning Methodology. Information, 11(1), 32.

Yan, K., Wang, X., Du, Y., Jin, N., Huang, H., & Zhou, H. (2018). Multi-step short-term power consumption forecasting with a hybrid deep learning strategy. Energies, 11(11), 3089.

Yuan, X., Wood, E. F., & Ma, Z. (2015). A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development. Wiley Interdisciplinary Reviews: Water, 2(5), 523-536.

Zhou, H., Zhang, Y., Yang, L., Liu, Q., Yan, K., & Du, Y. (2019). Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism. IEEE Access, 7, 78063-78074.

Zhang, L., Liu, P., Zhao, L., Wang, G., Zhang, W., & Liu, J. (2020). Air quality predictions with a semi-supervised bidirectional LSTM neural network. Atmospheric Pollution Research.

Zhang, S., Guo, B., Dong, A., He, J., Xu, Z., & Chen, S. X. (2017). Cautionary tales on air-quality improvement in Beijing. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 473(2205), 20170457.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., & Baklanov, A. (2012). Real-time air quality forecasting, part I: History, techniques, and current status. Atmospheric Environment, 60, 632-655.

Zhu, S., Lian, X., Liu, H., Hu, J., Wang, Y., & Che, J. (2017). Daily air quality index forecasting with hybrid models: A case in China. Environmental pollution, 231, 1232-1244.

Zhuo, Y. (2018, June). Research on Stock Index Forecasting Based on Machine Learning. In 2018 6th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2018). Atlantis Press.