



Robust-LSTM: a novel approach to short-traffic flow prediction based on signal decomposition

Erdem Doğan¹

Accepted: 17 March 2022 / Published online: 5 April 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Intelligent transport systems need accurate short-term traffic flow forecasts. However, developing a robust short-term traffic flow forecasting approach is a challenging task due to the stochastic character of traffic flow. This study proposes a novel approach for short-term traffic flow prediction task, namely Robust Long Short Term Memory (R-LSTM) based on Robust Empirical Mode Decomposing (REDM) algorithm and Long Short Term Memory (LSTM). Short-term traffic flow data provided from the Caltrans Performance Measurement System (PeMS) database were used in the training and testing of the model. The dataset was composed of traffic data collected by 25 traffic detectors on different freeways' main lanes. The time resolution of the dataset was set to 15 min, and the Hampel preprocessing algorithm was applied for outlier elimination. The R-LSTM predictions were compared with the state-of-the-art models, utilizing RMSE, MSE, and MAPE as performance criteria. Performance analyses for various periods show that R-LSTM is remarkably successful in all time periods. Moreover, developed model performance is significantly higher, especially during midday periods when traffic flow fluctuations are high. These results show that R-LSTM is a strong candidate for short-term traffic flow prediction, and can easily adapt to fluctuations in traffic flow. In addition, robust models for short-term predictions can be developed by applying the signal separation method to traffic flow data.

Keywords Short-term prediction · Traffic flow · LSTM · Signal decomposition

1 Introduction

ITS aims to increase the efficiency of transportation infrastructures by using technology and intelligent systems. Achieving this goal depends primarily on the success of gathering traffic information from the field on time. However, the information at the present time is not sufficient for most control decisions. Because ITS also requires information on future traffic for driver warning systems and various control decisions. Therefore, further studies are essential to develop advanced models that can predict traffic parameters accurately.

Traffic flow is a crucial parameter for road network control decisions. However, traffic flow is a random

phenomenon, and developing a reliable short-term prediction model is a challenging task due to the randomness of the traffic flow. Over the past years, different approaches have been attempted to develop a robust short-term traffic model using various approaches, such as time series methods (Han et al. 2004; Zeng et al. 2008; Wang et al. 2017; Doğan 2018, 2020a), Support Vector Machine (SVM) (Zhao-sheng et al. 2006; Yang and Lu 2010; Zhang et al. 2011; Feng et al. 2018), genetic algorithm (Abdulhai et al. 2002; Vlahogianni et al. 2005; Xu et al. 2016). However, there is still a lack of a robust approach. Artificial Neural Network (ANN) is a promising tool for developing such an approach and has been used in many studies (Vlahogianni et al. 2005; Hu et al. 2008; Sun and Liu 2008; Doğan 2020b; Yao et al. 2020). However, the exploding and vanishing gradient problem caused by the structure of ANNs blocked the development of more effective approaches. Hochreiter and Uergen Schmidhuber (1997) have overcome this difficulty with the Long Short-Term Memory (LSTM).

✉ Erdem Doğan
edogan@kku.edu.tr

¹ Department of Civil Engineering, Engineering and Architecture Faculty, Kirikkale University, Yahsihan, 71451 Kirikkale, Turkey

Although LSTM is an effective approach for revealing patterns in a time series, the traffic flow signal on a section of road is an overlap of upstream signals, i.e., sub-signals, from different directions. Using this superposition signal directly has the potential to prevent the development of more robust prediction models. However, most of the approaches developed in recent years have proposed more complex and hybrid structures to detect patterns in the data set and have ignored this phenomenon (e.g., Xiao and Yin 2019; Yang et al. 2019; Lu et al. 2020; Zhaowei et al. 2020). These models provide successful results, but their complex nature makes them difficult to understand and apply to new problems. To overcome this problem, firstly, traffic flow data should be decomposed into sub-signals, and separate models should be developed for each signal. Thus, it will be easier to detect patterns in the flow, and the model complexity will be reduced.

This study aims to develop an approach that can adapt to different flow conditions by decomposing short-term traffic flow into sub-signals and presents the Robust-LSTM (R-LSTM) approach that combines the REMD method (enhanced version of EMD) and the LSTM deep network. In the literature, a few studies that use the decomposition method and LSTM together stand out. For example, searching the Northern Atlantic Emission Index (Yuan et al. 2019), web service proposal (Singh et al. 2019), and short-term rail passenger forecast (Chen and Li 2019). However, this approach has not been explored sufficiently for the traffic flow prediction problem. Yu et al. (2019) used the time series decomposition method in the deep learning network that includes an LSTM network. However, the model presented in the study has many methods and their relationships. Therefore, it is not clear that the reason for the high performance stated in the results is due to the decomposition process. In another study, the researchers decomposed traffic flow as periodic and volatility and also used LSTM and hybrid models to predict traffic flow (Chen et al. 2020). Analysis showed that decomposing approach has more accurate results. However, in the study, traffic flow is only separated by time series methods according to periodicity and volatility. However, in the study, traffic flow was only separated by time series methods according to periodicity and volatility. In this study, REMD, which is a signal processing method, is used instead of the time series method, and the signal is divided into sub-signals. Thus, a more accurate short-term traffic flow forecasting model is developed.

The main achievements, including contributions, may be summarized as follows:

- A low-complexity methodology short-term traffic flow method, R-LTSM, was developed that combines

REMD's accurate signal decompose and LSTMs accurate signal patterns capture ability.

- The effectiveness of RLSTM has been proven by performing detailed analyses with a substantial amount of datasets and comparing R-LSTM performances with the latest models.
- Analyses that contributed to the determination of appropriate IMF values for traffic flow signals were performed, and the results were discussed.

The remainder of the article is organized as follows: Information about the structure of REMD and R-LSTM is given in the methodology section. Descriptive statistics of the datasets and the model training parameters are discussed in the experimental setup section. Information about determining suitable models and comparisons with state-of-the-art models are in the experimental result section. The last section includes a discussion of the results and suggestions.

2 Methodology

2.1 Traffic data pre-processes

Traffic data are collected by several types of detectors placed on roads. As a result of the malfunction of these devices, errors, deficiencies, or inconsistencies occur in the data. For this reason, the data should be pre-processed before proceeding with the model development.

Hampel filter has been effectively applied to studies in various areas (Allen 2009; Allen et al. 2010; Pearson et al. 2015; Ghaleb et al. 2018). Therefore, the Hampel filter was applied to the time-dependent traffic flow datasets before the decomposition process. The Hampel filter identifies an outlier in time series and replaces it with an appropriate value (Hampel 1974). Thus, the generalization capability and performance of the model developed by using these series increase.

Hampel filter processing steps include calculating local median Eq. (2) and standard deviation Eq. (3) values and determining outliers by using a threshold value (θ).

Traffic data are collected by several types of detectors placed on roads. As a result of the malfunction of these devices, errors, deficiencies, or inconsistencies occur in the data. For this reason, the data should be pre-processed before proceeding with model development.

Let $X(t) = \{x_t \in \mathbb{R}^+ | t \leq N \wedge N \in \mathbb{N}^+\}$ be the main dataset and γ be the sliding window length. Then, the local dataset,

$$X_j = \{x_{j-\gamma} + x_{j-\gamma+1}, x_{j-\gamma+2}, \dots, x_j, \dots, x_{j+\gamma-2}, x_{j+\gamma-1}, x_{j+\gamma}\} \quad (1)$$

$$m_j = \text{median}\{X_j\} \quad (2)$$

$$\sigma_j = 1.4826 \cdot \text{median}\{|x_{j-\gamma} - m_i|, \dots, |x_{j+\gamma} + m_j|\} \quad (3)$$

Finally, x_j is tested using Eq. (4). If inequality is valid, then the Hampel filter declares x_j as an outlier and replaces it with m_j

$$|x_{j-\gamma} - m_i| > \theta \cdot \sigma_i \quad (4)$$

After the outliers were extracted, the datasets were standardized by using Eq. (5) for LSTM model training. Let $\hat{X}$ be mean and σ be the standard deviation of the X , thus the standardized value of x_j ,

$$z_j = \frac{x_j - \hat{X}}{\sigma} \quad (5)$$

2.2 Robust empirical mode decomposition

Empirical mode decomposition (EMD) was developed by Huang et al. (1998) for the decomposing of mixed signals and has been applied successfully to many theoretical and practical subjects. EMD performance is significantly dependent on the adjustment of EMD parameters. Sifting stop criterion (SSC) is one of these parameters and performs an important role in improving EMD performance. Therefore, efforts were made to develop an effective SSC method (Huang et al. 1998; Rilling et al. 2003; Qiwei et al. 2007). However, these methods require a predefined threshold to stop sifting, and expertise is required to fix this threshold properly. Such approaches were referred to as hard sifting stopping criteria (H-SSC). To overcome this drawback, an intuitive soft sifting stopping criterion (S-SSC) was proposed to determine the number of shifting (Liu et al. 2017b). Unlike H-SSC, S-SSC does not require any predefined threshold value and is more accurate than H-SSC. Robust empirical mode decomposition (REMD) is a signal mode decomposition method using S-SSC. Thus, REMD can perform signal decomposition more effectively than the traditional EMD method.

Algorithm 1 (Alg1) shows the decomposition steps of the signal X into IMFs plus residual. The REMD begins by assuming X as a residual signal (Alg1 Line 2). Next, this residual signal is checked for monotonicity. If X is not monotonic, then this signal is decomposed. In the decomposition process, first, the local average is determined by the Local Mean Function (LMF) whose steps are given in Alg 2, and the new IMF is extracted arithmetically from the previous IMF. This process continues until the S-SSC decides to terminate the process. In the end, the new IMF is

obtained from the previous IMF. These steps are repeated until the signal monotonicity is achieved.

Algorithm 1: REMD(X)

```

1:  $i \leftarrow 0$ 
2:  $residual \leftarrow X$ 
3: while  $No == \text{monotonic\_func}(residual)$  do
4:    $IMF \leftarrow residual$ 
5:   while  $No == S\text{-}SSC(IMF)$  do
6:      $m \leftarrow LMF(IMF)$ 
7:      $IMF \leftarrow IMF - m$ 
8:   end while
9:    $residual \leftarrow residual - IMF$ 
10:   $y(i) \leftarrow IMF$ 
11:   $i \leftarrow i + 1$ 
12: end while

```

The sifting stopping criterion dictates the number of shifting iterations in the process and EMD decomposition is easily affected by the parameters of the shifting stop criterion. Therefore, EMD performance benefits from the S-SSC that determines the iteration number with an adaptive approach.

Algorithm 2: LMF(IMF)

```

1:  $mins \leftarrow \text{findMinima}(IMF)$ 
2:  $maxs \leftarrow \text{findMaxima}(IMF)$ 
3:  $e^L \leftarrow \text{cubicCurve}(mins)$ 
4:  $e^U \leftarrow \text{cubicCurve}(maxs)$ 
5:  $m \leftarrow (e^L + e^U)/2$ 

```

The sifting stopping criterion dictates the number of shifting iterations in a sifting process, and EMD decomposition is easily affected by the parameters of the scan stop criterion. Therefore, EMD performance benefits from a soft shifting stop criterion that determines the iteration number with an adaptive approach.

The S-SSC steps used by this study were explained in detail in the previous two studies (Liu et al. 2017a; Peng et al. 2019). Therefore, the steps of the S-SSC are briefly described in this section. Let $h_{ik}(n)$ be the signal of i th IMF after k th shifting and $m_{ik}(n)$ is the mean of $h_{ik}(n)$ where, $n = 1, 2, \dots, N_s$ and N_s are the total number of points in h_{ik} . Therefore, S-SSC steps are described briefly, as follows.

1. Calculate the $h_{ik}(n)$ and $m_{ik}(n)$
2. Determine the Root Mean Square (RMS) and Excess Kurtosis (EK) by using Eqs. (7) and (8), respectively.
3. Calculate three consecutive objective function value ($f_{ik}, f_{ik} + 1, f_{ik} + 2$) using Eq. (6).

4. If $f_{ik} + 1 > f_{ik}$ and $f_{ik} + 2 > f_{ik} + 1$, stop the sifting process and use $(k - 1)$ th results. if not, the sifting process continues until the k reaches a predefined maximum iteration value.

$$f_{ik} = [(\text{RMS}(m_{ik}(n)) + |\text{EK}(m_{ik}(n))|)], m_{ik} \in \mathbb{N} \quad (6)$$

$$\text{RMS} = \sqrt{\frac{1}{N_s} \sum_{n=1}^{N_s} (m_{ik}(n))^2} \quad (7)$$

$$\text{EK} = \frac{\frac{1}{N_s} \sum_{n=1}^{N_s} (m_{ik}(n) - \bar{m})^4}{\left[\frac{1}{N_s} \sum_{n=1}^{N_s} (m_{ik}(n) - \bar{m})^2 \right]^2} - 3 \quad (8)$$

where $\bar{m}$ is the arithmetic mean of $m(n)$.

2.3 Naïve-LSTM and SVR (baseline models)

An LSTM (Naïve-LSTM) network is a type of deep neural network consisting of input, output, and hidden layers as seen in Fig. 1. Hidden layers are formed by LSTM units and their connections. The LSTM unit has customized vectors and gates to capture correlations within time series. LSTM overcomes the problem of vanishing/exploding gradients. To overcome this problem, LSTM uses input, output, and forget gates (Han et al. 2004). Thus, the weight of relevant information from previous iterations increases, others are forgotten. These advanced features make the LSTM method a strong candidate for developing a short-term traffic prediction model.

An LSTM network consists of sequentially connected LSTM units, as shown in Fig. 1. The current step (t) LSTM unit accepts the previous step (t) unit information vector (c_{t-1}) and the prediction value as input variables. It also processes the data point value in the current step (x_t). The input vectors pass through the sigmoid (σ) and tangent hyperbolic ($\tanh$) functions. These vectors create the vectors h_t (LSTM prediction for time t) and c_t (cell state for time t) for the next step as a result of the pairwise multiplication and addition operations. This process is applied to every point in the data set. As a result, after each step, the LSTM network becomes a better-trained prediction model.

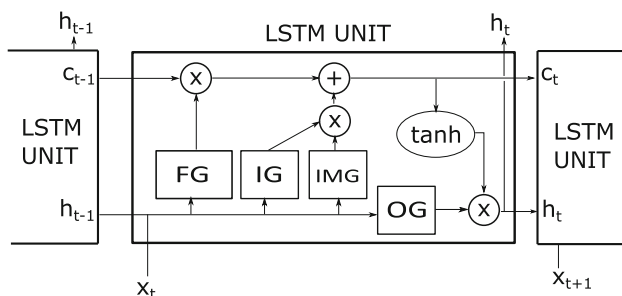


Fig. 1 LSTM unit and connections

The LSTM network is briefly introduced in this manuscript, but the LSTM principles can be explored in detail from the original study (Hochreiter and Jürgen Schmidhuber 1997).

Support vector machines (SVM) is a supervised learning algorithm, and it is developed for classification operations, initially (Vapnik 2013). In the following years, Smola and Schölkopf (2004) demonstrated that this method can also be utilized to develop regression models. SVM regression (SVR) is similar to linear regression. However, linear regression uses all the data in the dataset to develop the model. On the other hand, the SVR model decides whether to use an observation point in the model by considering a certain ε distance (Fig. 2).

SVR aims to find a linear function, $f(x)$, given in Fig. 2. For this, the algorithm aims to minimize the coefficient vector (w) of the $f(x)$. This optimization objective function and constraints are given in Eq. (9),

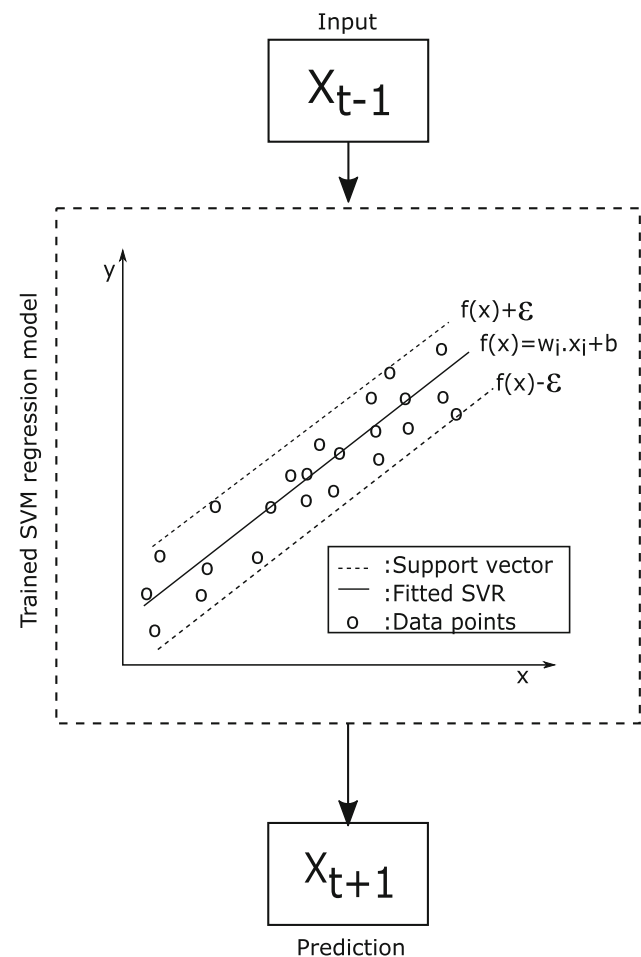


Fig. 2 Linear support vector regression model

$$\begin{aligned} \min \quad & \frac{1}{2}w^2 \\ \text{subject to} \quad & \begin{cases} y_i - w_i \cdot x_i - b \leq \varepsilon \\ w_i \cdot x_i + b - y_i \leq \varepsilon \end{cases} \end{aligned} \quad (9)$$

where x and y are training data pairs, b is the linear function constant and ε is the soft margin value. SVR is introduced briefly, but detailed information can be obtained from many studies on SVR in the literature (Drucker et al. 1997; Smola and Schölkopf 2004; Awad and Khanna 2015).

2.4 R-LSTM

R-LSTM is an approach that combines the REMD process and the LSTM network. The approach consists of four main steps, as illustrated in Fig. 3. (1) Data preprocessing, (2) decomposition, (3) LSTM model training, (4) traffic flow predicting.

Data preprocessing details are as discussed in Sect. 2.1. In this first step, the Hampel filter is applied to the traffic flow dataset and the filtered set is divided into training and test sets as seen in Fig. 3. Thanks to data preprocessing, the dataset is cleared of outliers and missing data points are completed by the filter. This process avoids the reduction in model performance by the impact of inappropriate data samples. Then, the dataset is divided into two sets: the training set and the test set.

The REMD process detailed in Sect. 2.2 decomposes the dataset and generates N new data signals, namely the IMF, and the residual signal. This process is applied to the training and test sets separately. The IMF and residual signals generated from the training set are used to train individual LSTM models. As a result, $N + 1$ trained LSTM (see Fig. 3, LSTM-1*, LSTM-2*, etc.) models are obtained for traffic flow prediction.

In the final step, the trained models are utilized to predict traffic flow. Let the $x(t)$ be the element of the IMF sub-signal set for period t . Therefore, $x_1(t) \in \text{IMF}_1, x_2(t) \in \text{IMF}_2, \dots, x_N(t) \in \text{IMF}_N$. These sub-signal samples are given as input to the trained LSTM models, and $x_1(t+1), x_2(t+1), \dots, x_N(t+1)$ values are obtained. These values are signal predictions, and for the prediction of the traffic flow for period t , as seen in Fig. 3, arithmetic addition is performed and the process is terminated.

2.5 Prediction errors criteria and evaluation periods

Various error criteria were used to measure the errors of the models developed in this study. Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE) are frequently used in similar studies. The equations of these criteria are given in Eqs. (10)–(12).

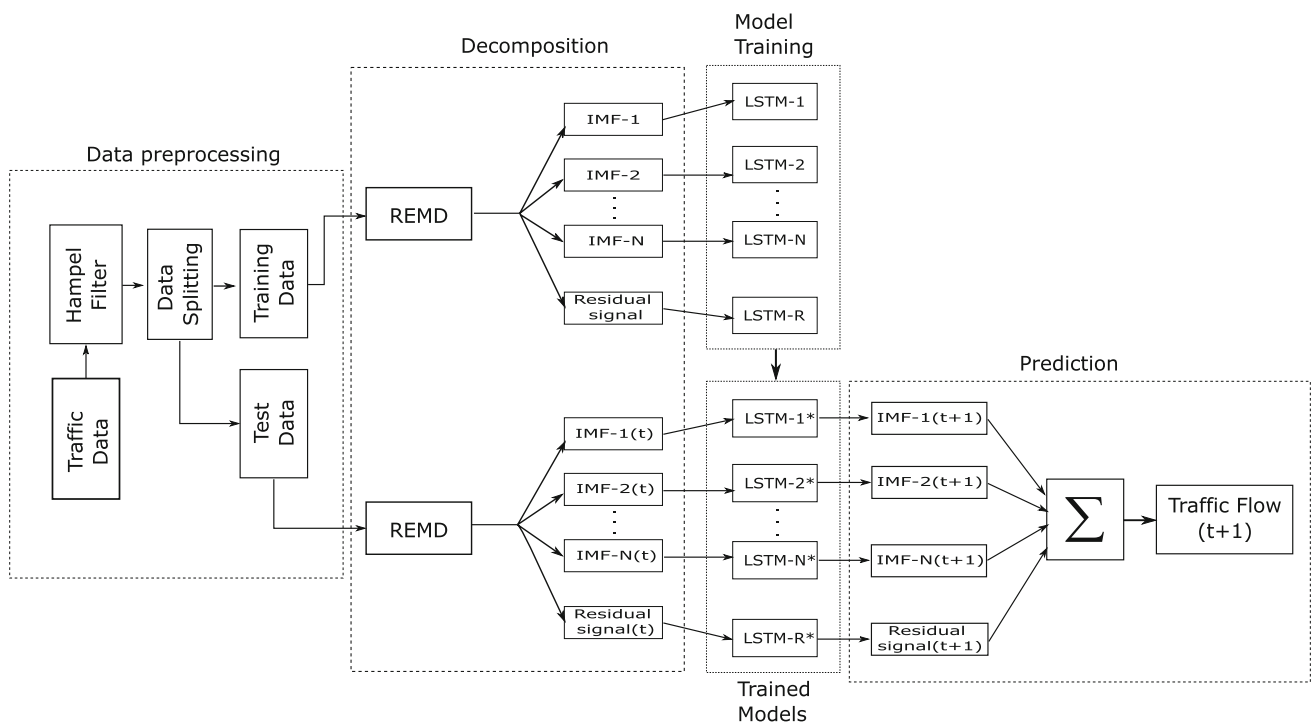


Fig. 3 R-LSTM working mechanism

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \hat{x}_i}{x_i} \right)^2 \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_i - \hat{x}_i)^2} \quad (11)$$

$$\text{MSE} = \frac{1}{n} \sum_{j=1}^n (x_i - \hat{x}_i)^2 \quad (12)$$

where n is the total number of the test sample, x_i is the actual traffic flow, and $\hat{x}_i$ is the predicted flow.

Model performances diverge under different flow conditions. To interpret this situation, the test set was divided into three sub-periods, and each sub-period was examined separately (Fig. 4). In addition, the predictive capabilities of the models were calculated and compared for all samples in the test data. These time sub-periods are explained as follows:

- **Fluctuation period:** The traffic flow and fluctuation are highest in this period. The period includes the peak hour volumes occurring in the morning and evening hours. (Fig. 4, red box).
- **Linear period:** In this period, the amount of traffic flow tends to decrease or increase linearly. (Fig. 4, green box).
- **Weekend period:** The traffic flow pattern on the weekend differs from week to week. For this reason, examining it separately allows examining the model performance in detail. The boundaries of this period are depicted by the blue box in Fig. 4.

3 Experimental setup

3.1 The traffic flow data

The traffic flow data sets used in this study were obtained from Caltrans Performance Measurement System (PeMS) database. PeMS provides an easy-to-access traffic data resource. The PeMS system processes raw data, and the user has the option of using processed data or raw data. PeMS collects data with the help of numerous detectors on the highway and its connections. Therefore, the desired amount of instant data can be obtained from the field.

In this study, traffic station data in District 3 were selected for developing prediction models, and the IDs of these traffic stations are presented in Table 1. The data were collected from the main lanes of highways with continuous flow conditions. These data are presented in PEMS as a series with a time interval of 5 min. The time interval for these series was converted to 15 min for use in this study. The highways where traffic counting stations are located have different lanes. To avoid the complexity of this difference in comparisons, the total traffic flow was divided by the number of lanes. Traffic flow data sets obtained through 25 selected traffic detector stations from different regions were used for the training and testing of R-LSTM and other approaches. Datasets with different statistical properties were preferred to better interpret the performance of the models. Descriptive statistics for the datasets are in Table 1.

The data sets contain traffic flow data covering 7 weeks (22 Apr 2019—09 Jun 2019), and each dataset has 4704 samples (1 sample for each 15 min). Traffic flow oscillation is similar at consecutive weeks. Therefore, the model should be tested for one or more weeks. For training, 4032

Fig. 4 Sample traffic flow (1-week) and comparison periods

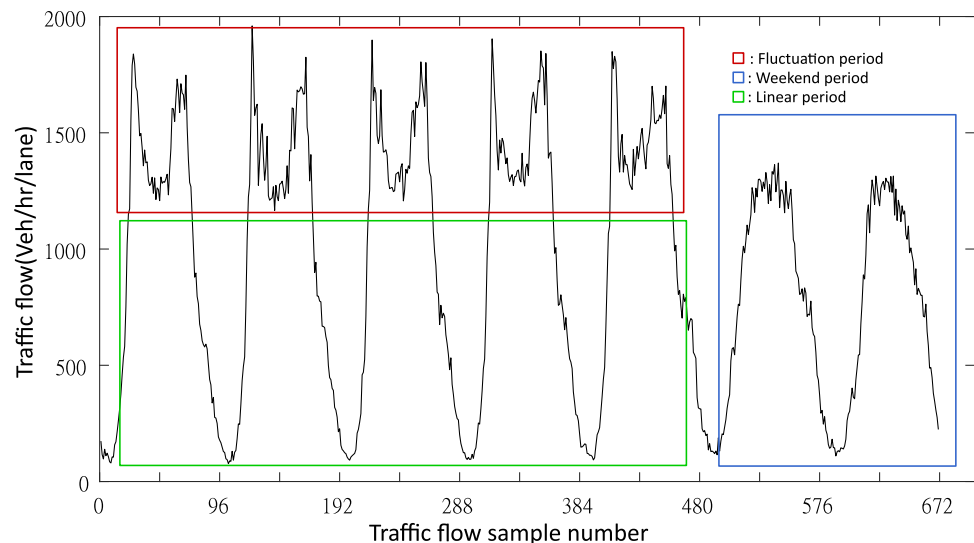


Table 1 Descriptive statistics of datasets based on traffic stations

Station ID	Mean	SD	Max	Min
311903	280	123	546	102
312010	609	331	1369	43
312103	1235	631	2202	115
312132	1106	626	2193	76
312139	324	180	735	33
312205	1055	431	1810	290
312233	969	532	2212	79
312264	534	278	1139	0
312326	911	548	2071	29
312346	717	391	1615	47
312517	652	358	1475	47
312520	760	341	1311	111
312523	783	385	1491	82
312562	702	305	1337	156
312564	955	414	1800	124
312694	1126	540	1965	130
312745	761	407	1456	61
312757	877	438	1541	90
312771	949	435	1690	107
312807	771	399	1493	66
312865	771	398	1590	78
312895	979	536	1778	79
312896	820	444	1671	62
312900	634	338	1342	72
311903	697	383	1538	25

samples, and the remaining 672 traffic flow samples were used for testing, i.e., the test data covers a one-week period.

3.2 Model training parameters

Determination of the strength and effectiveness of the R-LSTM model is important. Therefore, the Naive-LSTM and SVR models were used as baseline models for comparison. Naive-LSTM is a model that does not use REMD and processes the raw dataset. SVR is an approach stated in studies that produced successful results and has been applied to the traffic flow problem.

The proposed R-LSTM and baseline models were developed using Matlab Software. The number of suitable N-LSTM and R-LSTM hidden layer neurons and the maximum number of steps were determined by the trial method as 200 and 250, respectively. Adam optimization was used for N-LSTM and R-LSTM training, and Sequential minimal optimization (SMO) was used for SVR. The training parameters used for these processes are presented in Table 2.

Table 2 LSTM and SVR training parameters

Model	Parameter	Value
LSTM (Adam optimization)	Gradient decay factor	0.9
	Epsilon	$e-8$
	Initial learning rate	0.01
SVR (SMO optimization)	Gap tolerance	e^{-3}
	Kernel offset	0
	Alpha	1

4 Experimental result

In this section, firstly, the test results of the R-LSTM models developed using different IMF amounts were examined, and the results were discussed. Then, the appropriate R-LSTM model and baseline models were compared considering the evaluation periods.

4.1 Determining the appropriate IMF number

Assigning an appropriate number of IMF to the process improves the R-LSTM performance significantly. Therefore, the effect of the IMF number on R-LSTM model prediction errors was analyzed, and appropriate IMF numbers were determined in this section. Figure 5 illustrates the effect of the number of IMF on R-LSTM prediction performances. The red horizontal line within the boxes in the figure shows the error mode. These modes are the lowest at the IMF = 1 for all criteria. In addition, the increment in the number of IMFs increases the average error. The lowest and highest error values can be seen from the lower and whiskers. For MAPE, the lowest error occurred for a dataset when the IMF = 2 and the error is around 2%. However, this value is slightly less than the IMF = 1 MAPE, and it is not a significant difference. The upper whiskers indicate the highest prediction error. MAPE upper whiskers are following an exponential trend with the IMF increase. In other words, the prediction accuracy of the model was significantly reduced for some data sets. There is also an upward trend in RMSE and MSE, but the trend is linear. These different trends are explained by the fact that the MAPE metric is sensitive to errors in predictions for small traffic flows.

MSE is a sensitive measure to large errors. In Fig. 5, MSE upper mustaches have a low error at IMF = 1 than for other IMFs. In addition, the blue box representing 50% of the data sets shows small errors at IMF = 1 than the others. RMSE is the root of MSE and therefore less susceptible to large errors, and gives good insight into average error. The range of the upper and lower whiskers is remarkable for

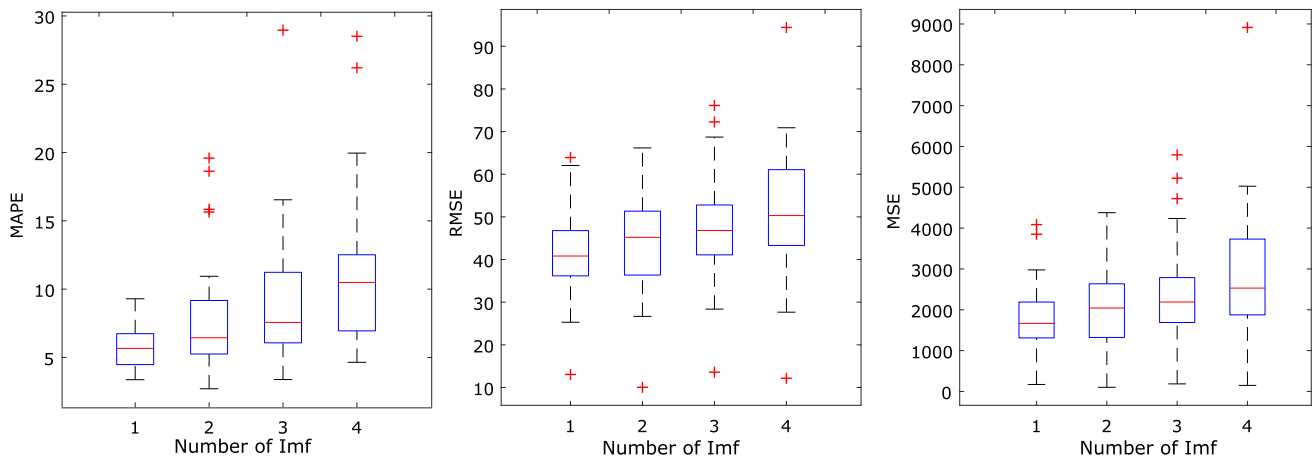


Fig. 5 R-LSTM model error values using different IMF numbers

RMSE. The range is the difference between upper and lower mustaches limit, and IMF = 1 range is close to other IMF ranges. This observation shows that the error distributions of IMFs are similar for RMSE. These analyses demonstrate that using one IMF is appropriate for short-traffic flow datasets used in this study. In this case, two LSTM models should be developed, and the prediction function is as follows (Also check Eq. (9)).

Increasing the number of IMFs in the REMD process decomposes the traffic flow data into more sub-signals. The IMF and residual signals produced by this REMD procedure are shown in Fig. 6. Increasing the number of IMFs smooths the residual signals. However, an excessive increase in the number of IMFs over-smooths the residual signal and distorts the signal pattern. This fact can be observed in Fig. 6 for IMF = 3 and 4. As a result, choosing a high number for the IMF number can affect model performance negatively.

4.2 Comparison with baseline models

The R-LSTM is compared with selected baseline models in this section. Baseline models were developed using two state-of-the-art approaches, N-LSTM and SVR. First, all data points in the test set were considered for error calculations, i.e., whole test period is considered. Subsequently, the data set was divided into sub-periods described in Sect. 2.3, and the predictions of the models were analyzed.

Table 3 shows the mean error criteria of the models examined. According to the RMSE criteria, R-LSTM estimates are approximately two times more accurate than SVR, and approximately three times according to MSE criteria. According to the MAPE criteria, R-LSTM is about 1.5% more accurate. Moreover, the MSE criterion shows that R-LSTM has about two times low error than N-LSTM.

Consequently, R-LSTM is superior to SVR, and utilizing REMD enhances the prediction performance of N-LSTM, remarkably.

Figure 7 illustrates the errors of the three models for the datasets. The R-LSTM is more accurate than other models, except for two datasets according to the MAPE, and one according to the other criteria. The error trends for the datasets of the three models are similar, but the MSE proves that the R-LSTM made fewer major errors than other models. For example, the MSE errors of SVR and N-LSTM in Fig. 7 are significantly large for some data sets (e.g., dataset no: 4,7,9).

R-LSTM shows remarkable performance for the whole test period. However, a reliable model should also have satisfactory performance over different time periods. Thus, error analysis of models was also performed for sub-periods mentioned in Sect. 2.3.

The fluctuation period is approximately in the middle of the day and includes traffic rush hours. Therefore, it is crucial to enhance the model performance for this period. In Fig. 8, all criteria show that the R-LSTM flow predictions are closer to actual flow than other models. The smallest error of R-LSTM in MAPE is about 2%, for N-LSTM and SVR at about 3.5 and 5, respectively. These results indicate that the R-LSTM's adaptation to traffic flow fluctuation is better than other models. RSME and MSE also point out that this conclusion is acceptable. Another remarkable point is the error range of R-LSTM for MSE. While SVR upper whisker limit approaches 15000, for R-LSTM it is around 4000. This proves that SVR and N-LSTM make relatively high errors, especially in large fluctuations.

The R-LSTM performance is also satisfactory for the linear period but two cases are remarkable. In Fig. 9, MAPE upper whisker of the N-LSTM is about 13% while it

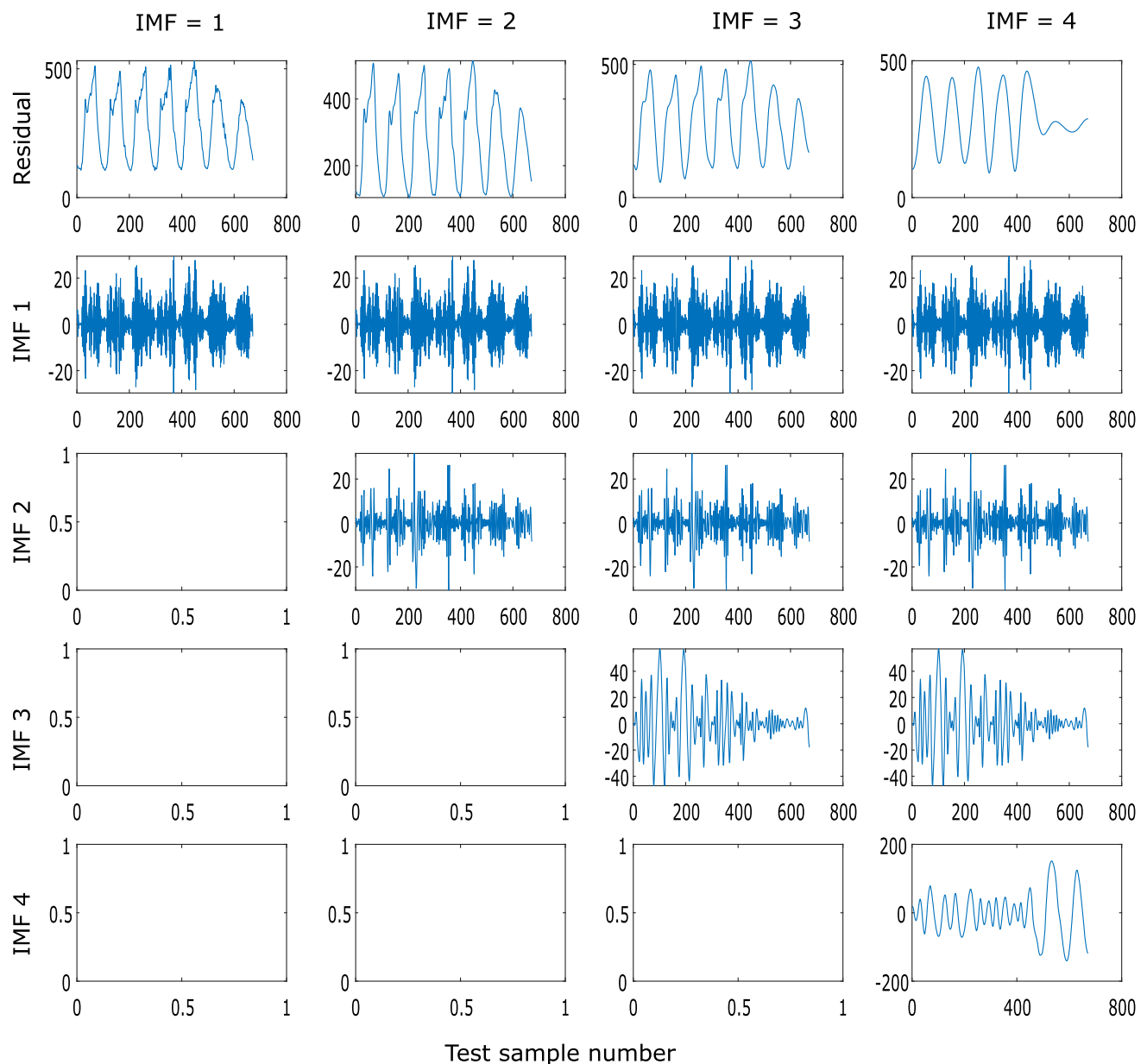


Fig. 6 Residual and IMF signals of test data after decomposition (Station ID: 311903)

Table 3 Average errors of models

Model	MAPE	RMSE	MSE
SVR	10.32	60.80	4176
N-LSTM	7.81	49.12	2702
R-LSTM	6.58	37.41	1563

is 16% for R-LSTM. In other words, N-LSTM is better at small flow predictions for some datasets than R-LSTM. The ranges of these two models are also close at MSE, but the R-LSTM mode is about 500 lower than the N-LSTM

mode. Based on the RMSE criteria, the R-LSTM is significantly better than the others. The SVR approach produces large errors for all criteria. The NLSTM performance is satisfactory for large and small flow values. However, it appears that a slightly better result was achieved with the R-LSTM. This result indicates that N-LSTM can also be effective in this period.

The traffic flow pattern on weekends is different from weekdays. Also, traffic flow may be higher or lower than weekdays depending on the classification of the road. Figure 10 illustrates the model prediction errors for the weekend. It is also clear that the proposed model produced small errors for this period. The MAPE range of R-LSTM

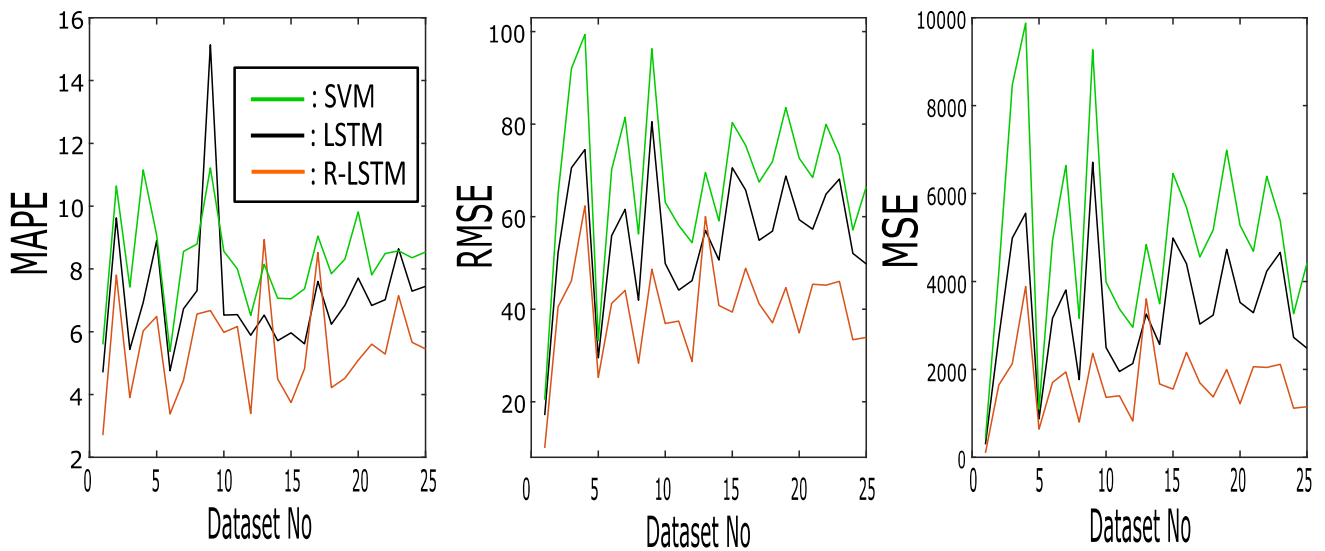


Fig. 7 Performance comparison of models for datasets

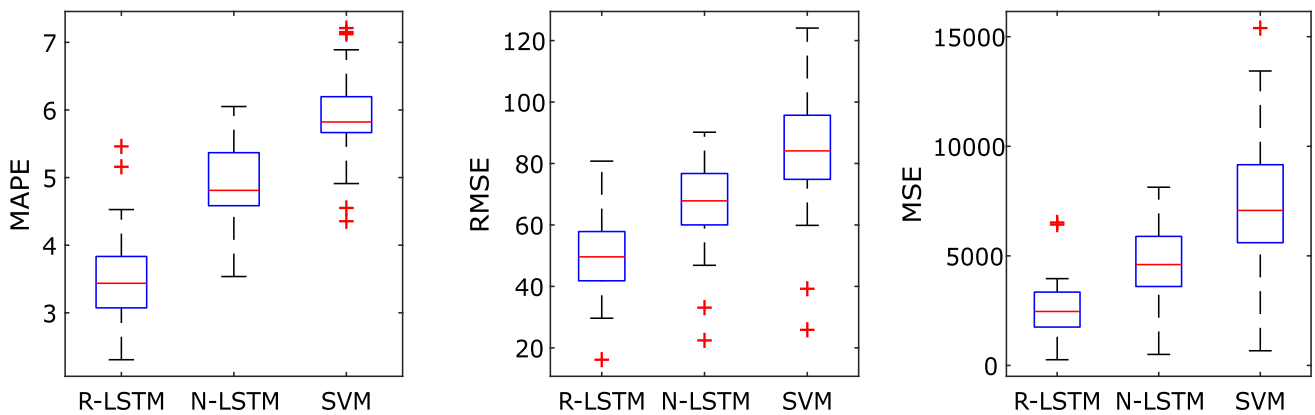


Fig. 8 Models performance criteria for fluctuation period

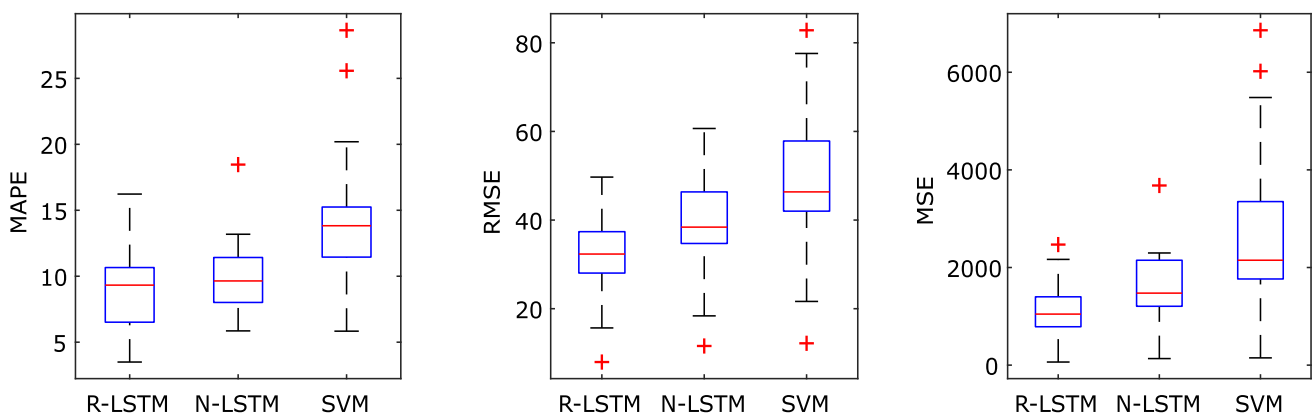


Fig. 9 Performance comparison of models for datasets linear period

for the weekend is about 5%, and this is lower than the MAPE of the linear period. The difference between R-LSTM and N-LSTM modes is large in this period,

similar to the fluctuation period. SVR predictions are also high for this period. These results prove that the R-LSTM adapts better than others to fluctuations in traffic flow.

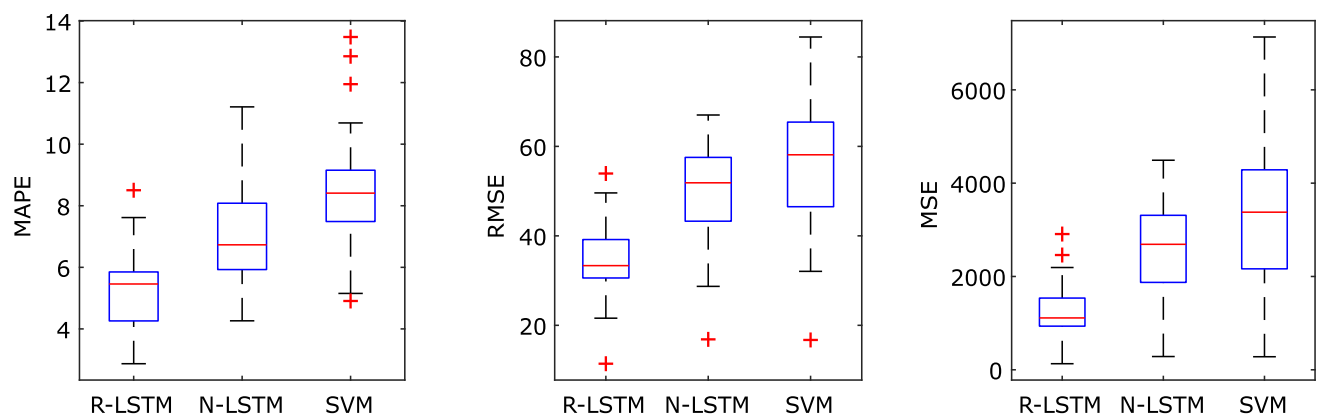


Fig. 10 Performance comparison of models for datasets weekend period

5 Conclusion

A robust short-term traffic flow forecasting model can make a significant contribution to traffic management systems. However, a widely accepted approach has not yet been developed, as a significant number of factors influence the short-term value of traffic flow. Separating these factors from traffic flow data will contribute to the increase in model performance.

In this study, a novel approach called R-LSTM is proposed, which decomposes traffic flow using REMD into sub-signals (IMFs) and predicts short-term traffic flow utilizing IMFs and LSTM models. To our knowledge, this is the first study to combine a decomposition algorithm and LSTM for short-term traffic flow prediction. In previous studies, traffic data are usually preprocessed, but these operations are applied only to remove outliers, complete missing data, or smooth the data. In this study, the pattern capture capability of the LSTM is enhanced by the utilization of REMD in addition to the traditional data pre-processes.

R-LSTM performance evaluation was performed with data from 25 traffic counting stations and MSE, RMSE and MAPE criteria. In addition, the R-LSTM performance was compared with the N-LSTM and SVR prediction performances. Moreover, model performances for traffic flow with different characteristics during the week, i.e. sub-periods performances, were also examined. The results show that using traffic flow sub-signals as model inputs significantly in the training process improves LSTM model performance, and R-LSTM is a promising alternative to short-term traffic flow prediction tasks. Valuable information was obtained by examining the sub-periods. In the fluctuation period, traffic flow oscillates remarkably, and the models have larger errors than the other periods. However, R-LSTM is superior in MSE criteria, especially in the fluctuation period. The adaptation of SVR to traffic oscillation is lower than other models. The models' predictions

in linear, and weekend periods are more accurate than the fluctuation period. Consequently, the fluctuation period is the most challenging part for models, and it gives an overall picture of a model's performance.

Despite the success of the proposed model, there are some limitations to the study. Data sets with different statistical properties were selected to train the models. However, more robust analysis results can be obtained by increasing the number of data sets. Another shortcoming is related to hyperparameters. During model training, hyperparameters are fixed to certain values. Therefore, the sensitivity of the models to hyperparameters should be discussed in further study. There remain some issues that should be explored in future research. First, the robustness of the model will be revealed more clearly by completing the mentioned deficiencies. In addition, performing analyses for datasets with missing data, discussing the effectiveness of different decomposition methods together with LSTM, are further studies that will contribute to the advancement of the model. Consequently, the traffic flow signal decomposition enhances the quality of the LSTM model and provides more accurate short-term traffic flow predictions.

Funding No funding was received for conducting this study.

Data availability The datasets generated during and/or analyzed during the current study are available in the Caltrans Performance Measurement System (PeMS) repository, https://pems.dot.ca.gov/?dnode=Clearinghouse&type=station_5min&district_id=3&submit=Submit

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Abdulhai B, Porwal H, Recker W (2002) Short-term traffic flow prediction using neuro-genetic algorithms. *ITS J-Intell Transp Syst J* 7:3–41
- Allen DP (2009) A frequency domain Hampel filter for blind rejection of sinusoidal interference from electromyograms. *J Neurosci Methods* 177:303–310
- Allen DP, Stegemöller EL, Zadikoff C et al (2010) Suppression of deep brain stimulation artifacts from the electroencephalogram by frequency-domain Hampel filtering. *Clin Neurophysiol* 121:1227–1232
- Awad M, Khanna R (2015) Support vector regression. In: *Efficient learning machines*. Springer, pp 67–80
- Chan KY, Dillon TS, Singh J, Chang E (2011) Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg–Marquardt algorithm. *IEEE Trans Intell Transp Syst* 13:644–654
- Chen C, Li J (2019) Short-term urban rail transit passenger flow forecasting based on empirical mode decomposition and LSTM. In: *2019 international conference on modeling, analysis, simulation technologies and applications (MASTA 2019)*. Atlantis Press
- Chen L, Zheng L, Yang J et al (2020) Short-term traffic flow prediction: From the perspective of traffic flow decomposition. *Neurocomputing* 413:444–456. <https://doi.org/10.1016/j.neucom.2020.07.009>
- Doğan E (2018) D-200 Karayolu için Dönemsel ARIMA Yöntemi ile Kısa Dönemli Trafik Akım Tahmin Modellerinin Geliştirilmesi. *Sakarya Üniv Fen Bilimleri Enstitüsü Dergisi* 22:390–400
- Doğan E (2020a) Analysis of the relationship between LSTM network traffic flow prediction performance and statistical characteristics of standard and nonstandard data. *J Forecast*. <https://doi.org/10.1002/for.2683>
- Doğan E (2020b) Short-term traffic flow prediction using artificial intelligence with periodic clustering and elected set. *Promet-Traffic Transp* 32:65–78
- Drucker H, Burges CJC, Kaufman L et al (1997) Support vector regression machines. *Adv Neural Inf Process Syst* 9:155–161
- Feng X, Ling X, Zheng H et al (2018) Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction. *IEEE Trans Intell Transp Syst* 20:2001–2013
- Ghaleb FA, Kamat MB, Salleh M et al (2018) Two-stage motion artefact reduction algorithm for electrocardiogram using weighted adaptive noise cancelling and recursive Hampel filter. *PLoS ONE* 13:e0207176
- Hampel FR (1974) The influence curve and its role in robust estimation. *J Am Stat Assoc* 69:383–393
- Han C, Song S, Wang C (2004) A real-time short-term traffic flow adaptive forecasting method based on ARIMA model. *Acta Simul System Sin* 7:43
- Hochreiter S, Unger Schmidhuber JJ (1997) Long short-term memory. *Neural Comput* 9:1735–1780
- Hosseini SH, Moshiri B, Rahimi-Kian A, Araabi BN (2012) Short-term traffic flow forecasting by mutual information and artificial neural networks. In: *2012 IEEE international conference on industrial technology*. IEEE, pp 1136–1141
- Hu C, Xie K, Song G, Wu T (2008) Hybrid process neural network based on spatio-temporal similarities for short-term traffic flow prediction. In: *2008 11th International IEEE conference on intelligent transportation systems*. IEEE, pp 253–258
- Huang NE, Shen Z, Long SR et al (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond Ser A Math Phys Eng Sci* 454:903–995
- Liu Z, Jin Y, Zuo MJ, Feng Z (2017a) Time-frequency representation based on robust local mean decomposition for multicomponent AM-FM signal analysis. *Mech Syst Signal Process* 95:468–487
- Liu Z, Zuo MJ, Jin Y et al (2017b) Improved local mean decomposition for modulation information mining and its application to machinery fault diagnosis. *J Sound Vib* 397:266–281. <https://doi.org/10.1016/j.jsv.2017.02.055>
- Lu S, Zhang Q, Chen G, Seng D (2020) A combined method for short-term traffic flow prediction based on recurrent neural network. *Alexandria Eng J*. <https://doi.org/10.1016/j.aej.2020.06.008>
- Mingheng Z, Yaobao Z, Ganglong H, Gang C (2013) Accurate multisteps traffic flow prediction based on SVM. *Math Probl Eng* 2013:1–8
- Pearson RK, Neuvo Y, Astola J, Gabbouj M (2015) The class of generalized hampel filters. In: *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, pp 2501–2505
- Peng D, Liu Z, Jin Y, Qin Y (2019) Improved EMD with a soft sifting stopping criterion and its application to fault diagnosis of rotating machinery. *J Mech Eng* 55:122–132. <https://doi.org/10.3901/JME.2019.10.122>
- Qiwei X, Bo X, Jianping L et al (2007) EMD algorithm based on bandwidth and the application one economic data analysis. In: *2007 15th European Signal Processing Conference*. IEEE, pp 2419–2423
- Rilling G, Flandrin P, Gonçalves P (2003) On empirical mode decomposition and its algorithms. In: *IEEE-EURASIP workshop on nonlinear signal and image processing*. NSIP-03, Grado (I), pp 8–1
- Singh VP, Pandey MK, Singh PS, Subbiah K (2019) An empirical mode decomposition (EMD) enabled long sort term memory (LSTM) based time series forecasting framework for web services recommendation. In: *FSDM*. pp 715–723
- Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14:199–222
- Sun X, Liu T (2008) Short-term Traffic Flow Forecasting Based on a Hybrid Neural Network Model and SARIMA Model. *J Transport Syst Eng Inf Technol* 2008:5
- Vapnik V (2013) *The nature of statistical learning theory*. Springer Science & Business Media, Berlin
- Vlahogianni EI, Karlaftis MG, Golias JC (2005) Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. *Transport Res Part c: Emerg Technol* 13:211–234
- Wang Y, Li L, Xu X (2017) A piecewise hybrid of ARIMA and SVMs for short-term traffic flow prediction. In: *International conference on neural information processing*. Springer, pp 493–502
- Xiao Y, Yin Y (2019) Hybrid LSTM neural network for short-term traffic flow prediction. *Information* 10:105
- Xu C, Li Z, Wang W (2016) Short-term traffic flow prediction using a methodology based on autoregressive integrated moving average and genetic programming. *Transport* 31:343–358
- Yang B, Sun S, Li J et al (2019) Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing* 332:320–327. <https://doi.org/10.1016/j.neucom.2018.12.016>
- Yang Y, Lu H (2010) Short-term traffic flow combined forecasting model based on SVM. In: *2010 International Conference on Computational and Information Sciences*. IEEE, pp 262–265
- Yao R, Zhang W, Zhang L (2020) Hybrid methods for short-term traffic flow prediction based on ARIMA-GARCH model and wavelet neural network. *J Transport Eng Part A Syst* 146:4020086
- Yu Y, Zhang W, Qi Y, Guo H (2019) Short-Term Traffic Flow Prediction for Hybrid Time Series Decomposition Analysis and

- LSTM Neural Network. In: CICTP 2019. American Society of Civil Engineers, pp 2350–2362
- Yuan S, Luo X, Mu B et al (2019) Prediction of North Atlantic Oscillation index with convolutional LSTM based on ensemble empirical mode decomposition. *Atmosphere* 10:252
- Zeng D, Xu J, Gu J et al (2008) Short-term traffic flow prediction using hybrid ARIMA and ANN models. In: 2008 workshop on power electronics and intelligent transportation system, IEEE, pp 621–625
- Zhang T, Zhang J, Yang X (2011) Short-term highway traffic flow prediction based on mixed AGO-SVM. *J Transp Syst Eng Inf Technol* 11:157–162
- Zhao-sheng Y, Yuan W, Qing G (2006) Short-term traffic flow prediction method based on SVM. *J Jilin Univ (eng Technol Ed)* 6:9
- Zhaowei Q, Haitao L, Zhihui L, Tao Z (2020) Short-term traffic flow forecasting method with MB-LSTM hybrid network. *IEEE Trans Intell Transport Syst* 2020:5

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.