EMD-SVR: A Hybrid Machine Learning Method to Improve the Forecasting Accuracy of Highway Tollgates Traveling Time to Improve the Road Safety

Atilla Altıntaş(✉) and Lars Davidson

Division of Fluid Dynamics, Department of Mechanics and Maritime Sciences, Chalmers University of Technology, 412 96 Gothenburg, Sweden
altintas@chalmers.se

Abstract. Tollgates are known as the bottleneck of the highways, which cause long waiting queues in rush-hour times of the day. This brings many undesirable consequences such as higher carbon emission and road safety issues. To avoid this scenario, traffic control authorities need accurate travel time forecasts at tollgates to take effective action to monitor potential traffic load and improve traffic safety. Accurate forecasting of the traffic travel time will help traffic regulators to prevent arising problems by taking action. The main objective of this study is to improve the short-term forecasting (minutes) of the traffic flow on highway tollgates by improving a novel hybrid forecasting method that combines Empirical Mode Decomposition with Support Vector Regression (EMD-SVR). Results claim that compared with SVR, the new proposed hybrid prediction model, EMD-SVR, can effectively improve prediction accuracy. Better forecasting of the traffic load will provide safer roads but will also lower the carbon emissions caused by longer traveling times.

Keywords: Empirical Mode Decomposition · SVR · Machine learning · Forecasting

1 Introduction

A number of methodologies have recently been developed for forecasting purposes which can be divided into traditional mathematical statistics and machine learning methods. Regression analysis [2] and time series analysis [8] are some of the examples of traditional mathematical statistics methods. References [10,14] are examples of machine learning algorithm applications to predict traffic load. A review and comparison of the methods are given in Ref. [20].

Recently, Empirical Mode Decomposition (EMD) [12] has become a useful tool to improve forecasting methodologies in many areas from solar and

This work is supported by the Chalmers University, Transport Area of Advance.

A. L. Martins et al. (Eds.): INTSYS 2020, LNICST 364, pp. 241–251, 2021.
https://doi.org/10.1007/978-3-030-71454-3_15

wind energy to financial time series ([3–5,15–17,19]). EMD is a method which decomposes a complex time-series data into its frequency components i.e. so-called intrinsic mode functions (IMFs) ([3,5,12]). EMD divides data into its IMFs, which are a number of high to low-frequency components, where the high-frequency component corresponds to the short term changes, and the low-frequency component corresponds to the long term changes. By using different combination of the frequency components of the data enable us to predict short or long term predictions much more accurately compared to using original (i.e. raw) data. EMD separates data into its components (IMFs) and in this way reduces the complexity of the data and separate trends into different scales which results in a higher accuracy forecasting.

The general approach to EMD-based hybrid prediction methods is to individually predict each IMF and then sum these predicted values to obtain a final prediction. However, the time series of separate IMFs can be categorized in different characteristics and thereby achieve advancement in the time series analysis techniques.

While hybrid methods with EMD come forward as a more effective approach with its higher prediction accuracy, yet there is not a consensus in the literature on which IMFs should be included in the forecasting process.

It is reported that for varying combinations of IMFs a varying prediction accuracy is obtained [6,7,9,11,13,21,22,24]. It is suggested that IMFs that has lower frequency carries the characteristics of the original data, regarded as representing the mean tendency trend [7,9,11,22,24]. The authors associated the higher frequency IMFs with a large amount of noise, which results in a lack of accuracy on the prediction of the wind data. References [9,11,21,24] claimed that the elimination of the IMFs which have high frequency resulted in improved predictive accuracy. References [9,11] carried on an analysis by eliminating the first IMF. Quite the opposite, Ref. [13] excluded the residue from the prediction and reported that omitting the residue is not showing a significant effect on prediction results. As a different approach, instead of removing the high-frequency IMFs from the calculation Refs. [7,22,23] decomposed them separately and reconstructed them. A detailed review could be found in Ref. [6].

Lin et al. [14] have studied travel time and volume predictions with SVR by including scaling methods and they achieved accurate forecasting for the rush hours [4]. In a previous study by Altıntaş et al. [3], it has been shown that the EMD method is superior to conventional filter-based mode decomposition methods. In this study we improve the tollgate traffic travel time predictions by using the EMD-SVR method. We have obtained higher prediction by using selected IMFs as input for the SVR regression model.

The paper is organized as follows. First, the theory and method is given followed by the application of the method to traffic travel time data. The results are summarized and addressed in the following section, and some concluding remarks are given in the final section.

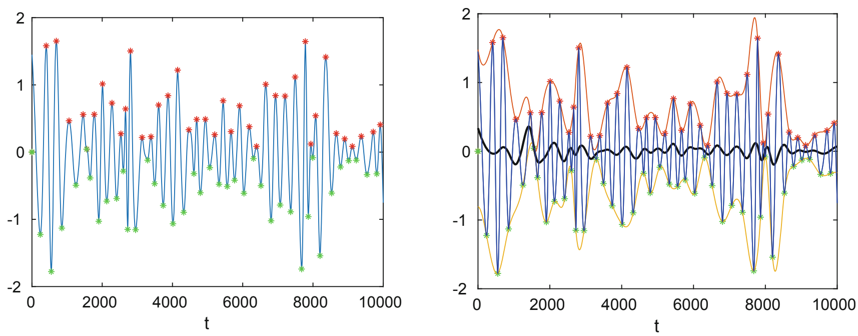
2 Theory and Method

2.1 Scale Decomposition by Empirical Mode Decomposition

EMD is constructed on the premise that any data signal consists of various simple intrinsic modes of oscillations, the original signal being a superposition of these oscillations. Each mode is referred to as an IMF [12] that satisfies the subsequent two conditions: (i) the local extrema and zero-crossing numbers must be equal or differ by one at the most; (ii) the mean of the curve that is constructed by connecting the maxima and minima should be zero.

EMD Algorithm. For a continuous times series $X(t)$, an algorithm could be written as follows to apply the EMD. Fluctuations will be obtained by subtracting the data from its time averaging (therefore the time history data will oscillate around zero).

i. All the maxima and minima will be obtained, see Fig. 1(a).



(a) All local maxima (red points), and local minima (green points).

(b) Construction of the mean curve by applying a cubic spline.

Fig. 1. Finding maxima, minima, and constructing a curve. (Color figure online)

ii. An envelope will be constructed for both maxima and minima and a mean curve of these two envelope curves, i.e. $m_{11}(t)$, see Fig. 1(b).

iii. First IMF will be constructed from the original data, i.e. $h_{10} = X(t)$. The first index in h_{ij} represents the number of the IMF in construction, the second represents the number of the iteration. As an example, the first iteration to find the IMF 1 represented as, $h_{11}(t) = h_{10}(t) - m_{11}(t)$.

iv. The steps (i), (ii), (iii) will be done recursively, $h_{1k}(t) = h_{1(k-1)}(t) - m_{1k}(t)$. The stopping criteria is: for $0 \leq t \leq T$

$$sd_n = \sum_{t=0}^{T} \left(\frac{\left| h_{n(k-1)}(t) - h_{nk}(t) \right|^2}{h^2_{n(k-1)}(t)} \right)$$

Empirically a number $sd_n < \epsilon$ is defined as a stopping criterion where ϵ is a number between 0.1 and 0.3.

v. When the first IMF, i.e. $h_{1k}(t)$ is found, it is subtracted from $h_{10}(t)$ to obtain $h_{20}(t)$. The process then restarts from (i) to find the second IMF.

vi. Set $c_i(t) = h_{ik}(t)$, where $c_i(t)$ is the ith. IMF. All the IMFs has been obtained when subtraction at step (v) gives a monotonic or constant data (residue).

As a result, a set of IMFs are obtained. As an example, for the signal $X(t)$ given as in (Fig. 2):

$$X(t) = 4cos(10t) + 2cos(t) + 3.$$

The resulting IMFs of the EMD process will be the frequency components of the raw signal $X(t)$. The highest frequency component, $4cos(t)$, is the first IMF, the second IMF is, $2cos(t)$, finally the residual is 3 (Fig. 3).

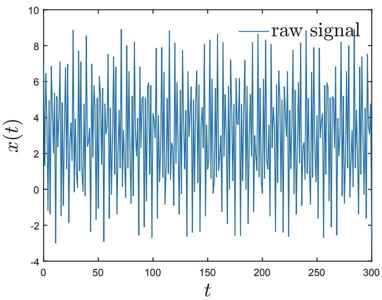


Fig. 2. $x(t) = 4cos(10t) + 2cos(t) + 3$.

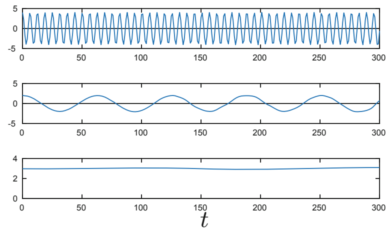


Fig. 3. IMF1 $= 4cos(10t)$, IMF2 $= 2cos(t)$, residual $= 3$.

2.2 SVR Method

The Support Vector Regression (SVR) is an algorithm for machine learning, which is a variant of Support Vector Machine (SVM). ([18]). SVR has widely been applied to forecasting problems. For a time-series data,

$$D = (X_i, y_i), 1 \leq i \leq N,$$

where X_i represents the ith element and y_i corresponds the target output data. The SVR function, f, is a linear function which is issued to formulate the nonlinear relation between input and output data as: $f(X_i) = \omega^T \phi(X_i) + b$, where ω, b and $\phi(X_i)$ are the weight vector, bias and function that maps the input vector X into a higher dimensional feature space, respectively. ω and b are obtained by solving the optimization problem:

$$min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) \tag{1}$$

subject to:

$$\begin{aligned} y_i - \omega^T(\psi(x)) - b &\leq \epsilon + \xi_i \\ \omega^T(\psi(x)) + b - y_i &\leq \epsilon + \xi_i \\ \xi_i, \xi_i^* &\geq 0. \end{aligned} \tag{2}$$

The first term of Eq. 1 measures the flatness of the function. The parameter C balances the trade-off between the complexity of the model and its generalization ability. The cost of error is measured by the variables, ξ_i and ξ_i^*.

The final SVR function is obtained as:

$$y_i = f(X_i) = \sum_{i=1}^{N}((\alpha_i - \alpha_i^*)K(X_i, X_j)) + b \tag{3}$$

where $K(X_i, X_j)$ is the Kernel function [18] and α_i and α_i^* are the Lagrange multipliers.

2.3 Application to Traffic Travel Time Data

The data are provided by the Knowledge Discovery and Data Mining (KDD) 2017 web site [1]. The data consist of the travel time of vehicles for the period of 19th June to 24th October 2016 for six routes. These are from intersection A to tollgates 1 and 2, and from intersection B and C to tollgates 1 and 3 (Fig. 4).

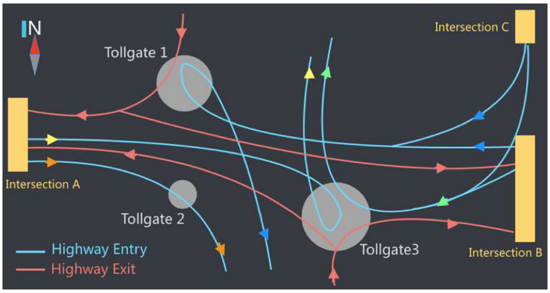


Fig. 4. Road map. [1]

The data consist of a list of records of actual vehicles including intersection ID, tollgate ID, vehicle ID, the time point when the vehicle enters the road, trajectory and travel time which is the total time taken from intersection to the tollgate (Table 1). The travel time data of the vehicles are averaged over 20 min of time-windows (Table 2). There are missing records which means that the time-window has no vehicle recorded, taken as zero average.

The prediction is performed for route B-3 (see Fig. 4) for the morning rush hours (08:00–10:00). The prediction is for every 20 min time-window between 08:00–10:00 and therefore there are six time-windows to forecast. The previous

20 min time-window has been used as the feature to forecast the next 20 min time-window, i.e., time-window 07:40–08:00 enters the process to forecast the next time-window, 08:00–08:20. The data that have been used in the predictions are given in Fig. 5. The data set is split into two data sets as training for the period of 19-07-2016 to 18-10-2016 (first 13 weeks) and test for the period of 18-10-2016 to 24-10-2016 (last 1 week), respectively.

Table 1. Original data from 19th June to 24th October 2016 for six routes.

Field	Type	Description
intersection_id	string	intersection ID
tollgate_id	string	tollgate ID
vehicle_id	string	vehicle ID
starting_time	datetime	minute the vehicle enters the route
travel_seq	string	trajectory of the link traces
travel_time	float	total travel time (in seconds).

Table 2. The data used in this study is given in the table below. The travel time data is averaged over 20 minutes of time-windows.

Field	Type	Description
intersection_id	string	intersection ID
tollgate_id	string	tollgate ID
time_window	string	20-minute time-window, e.g., [2016-19-07 07:40:00, 2016-19-07 08:00:00]
average_travel_time	float	the average travel time of vehicles in the time-window

In this study, we use SVR method as the baseline method and we compare it with the EMD-SVR hybrid method. The data are scaled by Min-Max scaling method to an interval of [-1, 1] before the SVR process. The radial basis function (RBF) is chosen as the kernel function, then Kernel function written as:

$$K(X_i, X_j) = exp(-\gamma \|X_i - X_j\|^2), \tag{4}$$

where the parameter γ, intuitively defines the degree to which the effect of a single example of training reaches. In this study parameters are set to $\gamma = 0.96$, $C = 1.0$, $\epsilon = 0.1$ and are used for all the predictions.

In this text, original data represents the data used in SVR method (first 13 weeks of the data). Original data prediction is the prediction made with SVR. Real data is the test data that never entered any prediction method, traffic travel time data of the last 7 days (last week).

In the EMD-SVR hybrid method, EMD used as a preprocessor to SVR. EMD splits data into IMFs and each IMF is a feature (input) for SVR. IMFs are frequency modes that are obtained by applying EMD to the original data. The sum of all IMFs is equal to the original data. In Fig. 6, the original data of average travel time 07:40–08:00 and its IMFs' obtained by EMD are given. The data set has been split into its IMFs. The number of IMFs has been limited to four, the fourth IMF is including the residual. Each four IMF has been an input for SVR. The combinations of the outputs are the predictions. That process is repeated for all six time-window predictions in the EMD-SVR hybrid method. A process of EMD combined with SVR is given in Fig. 7.

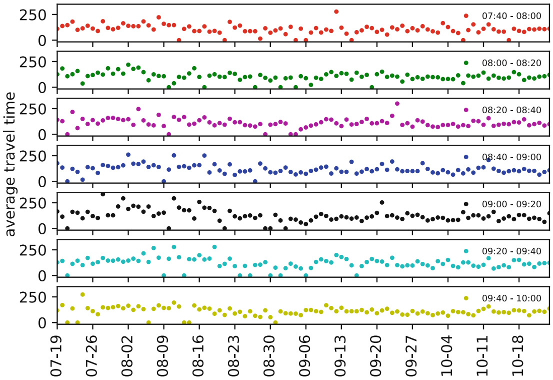


Fig. 5. The rush-hour average travel time data for the route B-3 for the dates 19-07-2016 to 24-10-2016. The data for the time-window 07:40–8:00 are also added since they were used to predict the first rush-hour window, 08:00–8:20.

3 Results and Discussion

The predictions for the test data which is the last week of 14 weeks data is obtained for both the SVR and the EMD-SVR method. We would like to clarify that all the parameters in both SVR and EMD are kept the same for all predictions. SVR is performed by using the original data as the feature. In EMD-SVR, EMD used as a preprocessor to SVR that splits original data into its IMFs. In EMD-SVR, each IMF is a feature for SVR instead the original data.

A total of six 20-min time-window predictions for the rush hours, 08:00–0:00, for the dates between 18-10-2016 and 24-10-2016 are given in Figs. 8(a) to 8(f). Mean square (MSE) and root mean square error (RMSE) are given in Table 3, where the best approximation is given in a separate column and also highlighted in red.

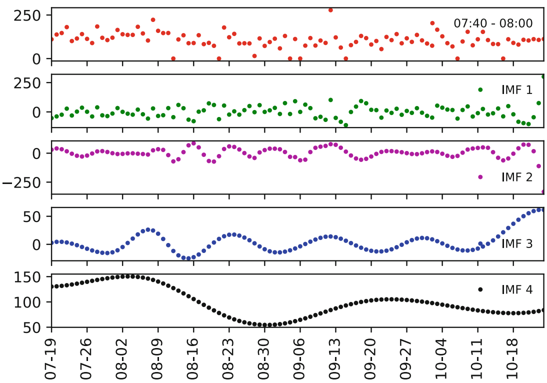


Fig. 6. The original data for the average travel time for the window 07:40–8:00 and the intrinsic mode functions (IMFs) by applying empirical mode decomposition (EMD). The upper signal is the original data, and the subsequent four signals are the IMFs obtained by applying EMD to the original data.

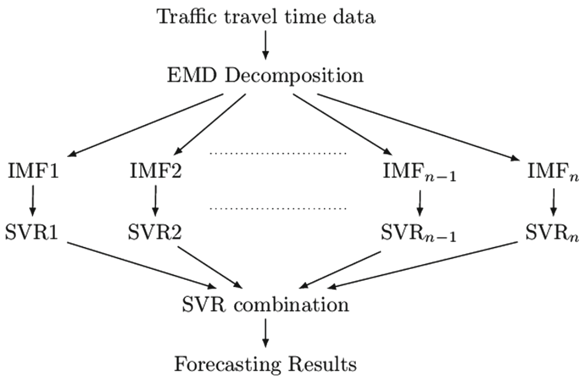
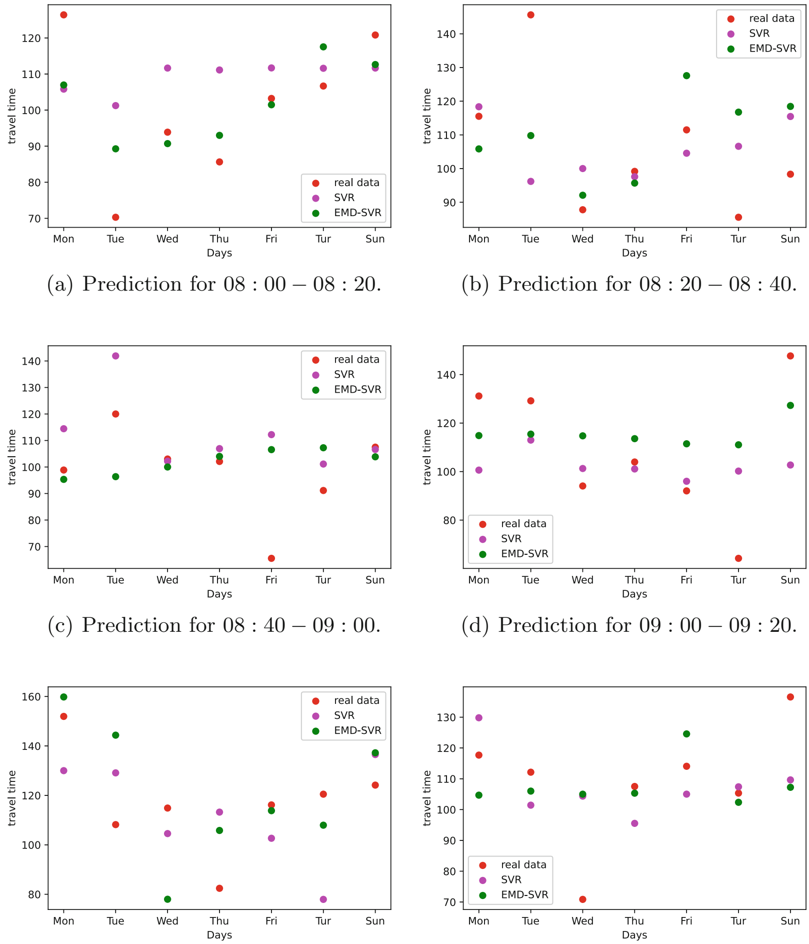


Fig. 7. Process schema of EMD-SVR method.

For the travel time windows 08:00–8:20, 08:20–8:40 and 09:20–9:40, IMF 1 gives a better approximation compared to the original data and all the other combinations of IMFs (see Figs. 8(a), 8(b), 8(e), respectively and Table 3). For the time windows 08:40–9:00 and 09:00–9:20, a combination of IMF 3 + IMF 4 and IMF2 + IMF 3, respectively, agree better with real data than all the other IMF combinations and original data (see Fig. 8(c), Fig. 8(d), respectively and Table 3). For the time-window 09:40–0:00, the original data approximate the real data better than all the other IMFs and their combinations. However, IMF 1 approximates the real data with less than an MSE error of 1% difference compared to the original data prediction.



(a) Prediction for 08 : 00 – 08 : 20.

(b) Prediction for 08 : 20 – 08 : 40.

(c) Prediction for 08 : 40 – 09 : 00.

(d) Prediction for 09 : 00 – 09 : 20.

(e) Prediction for 09 : 20 – 09 : 40.

(f) Prediction for 09 : 40 – 10 : 00.

Fig. 8. Traffic travel time predictions for the rush hours, 8:00–0:00, for the 20 minutes time interval. The data is the date between 18-10-2016 to 24-10-2016 (Monday to Sunday).

Table 3. The travel time predictions errors for the rush hours, 08:00–0:00, for 20 min time-window (see Figs. 8(a) - 8(f)). MSE = mean square error, RMSE = root mean square error.

	Original data		IMF 1		IMF 2		IMF 3		IMF 4		IMF 3 + IMF 4		IMF 2 + IMF 3		Best Approximation
	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE	
08.00-08:20	0.2000	0.4472	0.0783	0.2796	0.2749	0.5243	0.9813	0.9906	0.1853	0.4304					IMF 1
08.20-08:40	0.2346	0.4844	0.2108	0.4592	0.2190	0.4679	0.7614	0.8726	0.2655	0.5153					IMF 1
08.40-09:00	0.2014	0.4488	0.2419	0.4919	0.2591	0.5091	0.2093	0.4575	0.2022	0.4497	0.1690	0.4112			IMF 3 + IMF 4
09.00-09:20	0.1787	0.4228	0.3306	0.5749	0.2405	0.4904	0.5427	0.7367	0.9815	0.9907			0.1542	0.3927	IMF 2 + IMF 3
09.20-09:40	0.1912	0.4373	0.1676	0.4094	0.8584	0.9265	0.8328	0.9126	0.3163	0.5624					IMF 1
09.40-10:00	0.1836	0.4285	0.1850	0.4301	0.3050	0.5523	0.2112	0.4595	0.3367	0.5803					Original data

4 Conclusion

In this study, an EMD-based decoupling procedure is applied as a preprocessor to SVR to improve the travel time forecasting. First, IMFs are obtained by applying EMD to the original data, each IMF is used as a feature for SVR instead of the original data. The prediction results are compared for the combination of the IMFs and the original data. The KDD Cup 2017 data have been used for the rush hours 08:00–0:00. The data set has been split into 20 min time windows and a previous 20 min time-window has been used as a feature in the forecasting. All the parameters are kept for all predictions.

As a result, for five out of six 20 min time-windows, IMF or IMF combinations approximate the real data better than using original data in the prediction process. The time-window that gives better results for original data, namely 09:40–0:00, only gives 1% better agreement compared to IMF 1. Therefore we claim that the EMD based signal decomposition could be beneficial in forecasting studies to obtain better approximations.

References

1. Kdd2017. https://tianchi.aliyun.com/competition/information.htm?spm=5176.100067.5678.2.ru0ea4&raceId=231597. Accessed 15 Mar 2017
2. Alam, I., Farid, D., Rossetti, R.J.F.: The prediction of traffic flow with regression analysis. In: Abraham, A., Dutta, P., Mandal, J., Bhattacharya, A., Dutta, S. (eds.) Emerging Technologies in Data Mining and Information Security. Advances in Intelligent Systems and Computing, vol. 813, pp. 661–671. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-1498-8_58
3. Altıntaş, A., Davidson, L., Peng, S.: A new approximation to modulation-effect analysis based on empirical mode decomposition. Phys. Fluids **31**(2), 025117 (2019)
4. Andersson, L.E., Aftab, M.F., Scibilia, F., Imsland, L.: Forecasting using multivariate empirical mode decomposition-applied to iceberg drift forecast. In: 2017 IEEE Conference on Control Technology and Applications (CCTA), pp. 1097–1103. IEEE (2017)
5. Barnhart, B., Eichinger, W.: Empirical mode decomposition applied to solar irradiance, global temperature, sunspot number, and CO2 concentration data. J. Atmos. Solar Terr. Phys. **73**(13), 1771–1779 (2011)
6. Bokde, N., Feijóo, A., Villanueva, D., Kulat, K.: A review on hybrid empirical mode decomposition models for wind speed and wind power prediction. Energies **12**(2), 254 (2019)
7. Dejun, L., Hui, L., Zhonghua, M.: One hour ahead prediction of wind speed based on data mining. In: 2010 2nd International Conference on Advanced Computer Control, vol. 5, pp. 199–203. IEEE (2010)
8. Ghosh, B., Basu, B., O'Mahony, M.: Multivariate short-term traffic flow forecasting using time-series analysis. IEEE Trans. Intell. Transp. Syst. **10**(2), 246–254 (2009)
9. Guo, Z., Zhao, W., Lu, H., Wang, J.: Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renew. Energy **37**(1), 241–249 (2012)
10. Hong, W.C., Dong, Y., Zheng, F., Lai, C.Y.: Forecasting urban traffic flow by SVR with continuous ACO. Appl. Math. Model. **35**(3), 1282–1291 (2011)

11. Hu, J., Wang, J., Zeng, G.: A hybrid forecasting approach applied to wind speed time series. Renew. Energy **60**, 185–194 (2013)
12. Huang, N.E., et al.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 454, pp. 903–995. The Royal Society (1998)
13. Liang, Z., Liang, J., Zhang, L., Wang, C., Yun, Z., Zhang, X.: Analysis of multi-scale chaotic characteristics of wind power based on Hilbert-Huang transform and Hurst analysis. Appl. Energy **159**, 51–61 (2015)
14. Lin, A.Y., Zhang, M., Selpi: Using scaling methods to improve support vector regression's performance for travel time and traffic volume predictions. In: Rojas, I., Pomares, H., Valenzuela, O. (eds.) ITISE 2017. Contributions to Statistics, pp. 115–127. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-96944-2_8
15. Monjoly, S., André, M., Calif, R., Soubdhan, T.: Hourly forecasting of global solar radiation based on multiscale decomposition methods: a hybrid approach. Energy **119**, 288–298 (2017)
16. Nava, N., Di Matteo, T., Aste, T.: Financial time series forecasting using empirical mode decomposition and support vector regression. Risks **6**(1), 7 (2018)
17. Premanode, B., Vongprasert, J., Toumazou, C.: Noise reduction for nonlinear non-stationary time series data using averaging intrinsic mode function. Algorithms **6**(3), 407–429 (2013)
18. Qiu, X., Suganthan, P.N., Amaratunga, G.A.: Short-term electricity price forecasting with empirical mode decomposition based ensemble kernel machines. Procedia Comput. Sci. **108**, 1308–1317 (2017)
19. Ren, Y., Suganthan, P., Srikanth, N.: Ensemble methods for wind and solar power forecasting-a state-of-the-art review. Renew. Sustain. Energy Rev. **50**, 82–91 (2015)
20. Salotti, J., Fenet, S., Billot, R., El Faouzi, N.E., Solnon, C.: Comparison of traffic forecasting methods in urban and suburban context. In: 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 846–853. IEEE (2018)
21. Sivanagaraja, T., Veluvolu, K.: A hybrid approach for short-term forecasting of wind speed. Sci. World J. **2013**, 548370 (2013). https://doi.org/10.1155/2013/548370
22. Sun, C., Yuan, Y., Li, Q.: A new method for wind speed forecasting based on empirical mode decomposition and improved persistence approach. In: 2012 10th International Power & Energy Conference (IPEC), pp. 659–664. IEEE (2012)
23. Yu, C., Li, Y., Zhang, M.: Comparative study on three new hybrid models using Elman Neural Network and Empirical Mode Decomposition based technologies improved by Singular Spectrum Analysis for hour-ahead wind speed forecasting. Energy Convers. Manag. **147**, 75–85 (2017)
24. Zhang, K., Qu, Z., Wang, J., Zhang, W., Yang, F.: A novel hybrid approach based on cuckoo search optimization algorithm for short-term wind speed forecasting. Environ. Progress Sustain. Energy **36**(3), 943–952 (2017)