

## Travel Time Prediction Utilizing Hybrid Deep Learning Models

Dhivya Bharathi<sup>1</sup> , Juan Manuel González Sopeña<sup>1</sup> ,  
Siobhan Clarke<sup>2</sup>, and Bidisha Ghosh<sup>1</sup> 

Transportation Research Record  
1–10

© National Academy of Sciences:  
Transportation Research Board 2023



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/03611981231182964

[journals.sagepub.com/home/trr](https://journals.sagepub.com/home/trr)



### Abstract

Travel time prediction is vital to the development and maintenance of advanced intelligent transportation system technologies. The travel time on a road segment is dependent on various factors like dynamic traffic demands, incidents, weather conditions, and geometric factors. However, uncertainties associated with prediction performance consistency may reduce the effectiveness of such systems. To tackle these challenges, this paper proposes a hybrid deep learning algorithm-based methodology by integrating variational mode decomposition, multivariate long short-term memory, and quantile regression to predict estimates of travel time ranges instead of single-point predictions. Travel time data collected from loop detectors on motorways near the city of Dublin, Republic of Ireland were modeled. The proposed method was evaluated using various design scenarios and was found to perform efficiently in comparison with conventional deep learning algorithms.

### Keywords

data and data science, advanced traffic management systems, intelligent transportation systems, Traffic Prediction, traveler information systems

Travel time information in real time is the most sought-after data among travelers as it is very useful for making trip-related decisions such as route choice and departure time. It is also useful for practitioners wanting to interpret the efficiency of road segments and in managing traffic using intelligent transportation system (ITS) applications. However, travel time may vary significantly over space and time as a consequence of variations in traffic demand, capacity, incidents, roadwork, adverse weather, driving behavior, and congestion. As a result, being able to depend on extensive traffic data and recent technologies to precisely predict travel times is essential. There is plenty of literature in the domain of travel time prediction, which can be broadly classified as inductive approaches (i.e., data-driven methods) and deductive approaches (i.e., traffic flow theory-based methods). As the present study has proposed data-based modeling and prediction, the following paragraphs brief the reported studies based on inductive approaches.

Numerous studies have reported predicting travel times based on naïve methods, statistical methods, and artificial intelligence (AI)-based methods. Naïve methods (1) predict travel time by averaging over time and space

selectively. Statistical approaches like time series (2, 3) and regression methods (4, 5) predict travel time based on correspondences among the identified limited independent variables. However, these methods are largely dependent on the correspondence between a limited amount of training and testing data. AI-based techniques such as artificial neural networks (6, 7), support vector machines (SVMs) (8, 9), recurrent neural networks (RNNs) (10, 11), and convolutional neural networks (CNNs) (12, 13) are widely used prediction techniques when there is a large amount of data available for various applications in traffic. In light of this, RNNs and CNNs have gained greater research attention in recent times, owing to their ability to model complex temporal dependencies in data. Therefore, we adopted a multivariate long short-term memory (LSTM) neural network to

<sup>1</sup>Department of Civil, Structural and Environmental Engineering, Trinity College Dublin, Republic of Ireland

<sup>2</sup>School of Computer Science and Statistics, Trinity College Dublin, Republic of Ireland

### Corresponding Author:

Dhivya Bharathi, [bhaskard@tcd.ie](mailto:bhaskard@tcd.ie)

develop a travel time prediction method in this study. Hybrid methodologies generally delve into mode decomposition algorithms to disintegrate the original traffic data sequence into multiple subsignals. Further, hybrid models integrate one or more AI-based methods in prediction methodology to tackle the nonlinearity and non-stationarity of the traffic system. Popular mode decomposition algorithms include empirical mode decomposition (EMD) (14), empirical ensemble mode decomposition (EEMD) (15), and wavelet transform.

A few recent studies have explored hybrid models for prediction in the domain of traffic engineering. Zheng et al. proposed an EMD-based hybrid modeling framework by integrating SVM and LSTM for traffic flow prediction (16). Tian explored hybrid models by integrating EEMD with SARIMA models to perform traffic flow prediction (17). Xiu et al. developed a hybrid methodology by combining EEMD bidirectional gated recursive units (GRUs) to predict the passenger flow in the metro system, and reported a superior performance when compared with a single GRU model (18). Although most of the research on hybrid models has been based on EMD and EEMD, Sopena et al. developed a hybrid modeling framework using variational mode decomposition (VMD) with a feed-forward neural network (FFNN) for the purpose of traffic flow prediction (19). This study reported the superior performance of VMD when compared with other mode decomposition techniques using FFNN. However, the use of hybrid models like VMD has not been investigated in relation to travel time prediction. As travel time can render higher variations owing to its dynamic behavior and can be affected by various factors, adopting a hybrid model like VMD might be expected to better capture variations at different scales. Thus, the present study proposed a VMD-based hybrid modeling methodology for travel time prediction. In this study, we integrated the multivariate LSTM technique, a special type of RNN with a VMD algorithm, because LSTM has proven to be an excellent tool in time series prediction. Furthermore, to date, the combination of VMD and LSTM has not been explored. Therefore, the current methodology comprising a VMD integrated multivariate LSTM technique, was expected to improve the accuracy of the forecast while reducing the computational complexity of the prediction algorithm.

Point forecasts (i.e., a singular number that represents an estimate of an unknown variable value at a future date) cannot provide any information with respect to the uncertainty associated with the forecasts themselves, thus affecting the reliability of the prediction system. To overcome this issue, the present study utilized quantile regression (QR), a nonparametric method to identify the probabilistic estimates of prediction, known as prediction intervals (PIs). Overall, the study contributions include

1. Adopting a novel methodology consisting of a mode decomposition algorithm to decompose the time series data to capture the speed dynamics at different frequencies;
2. Formulating a hybrid prediction methodology with multiple deep learning models to predict the decomposed speed time series data to improve prediction accuracy when compared with traditional deep learning models; and
3. Providing an interval estimate unlike traditional models that fuses a QR-based loss function with an LSTM technique, which essentially equips the methodology by yielding reliability bounds.

To summarize, the present study proposed a travel time prediction methodology based on LSTM, a special type of RNN integrated with a mode decomposition algorithm, VMD and QR, a nonparametric approach to estimate PIs.

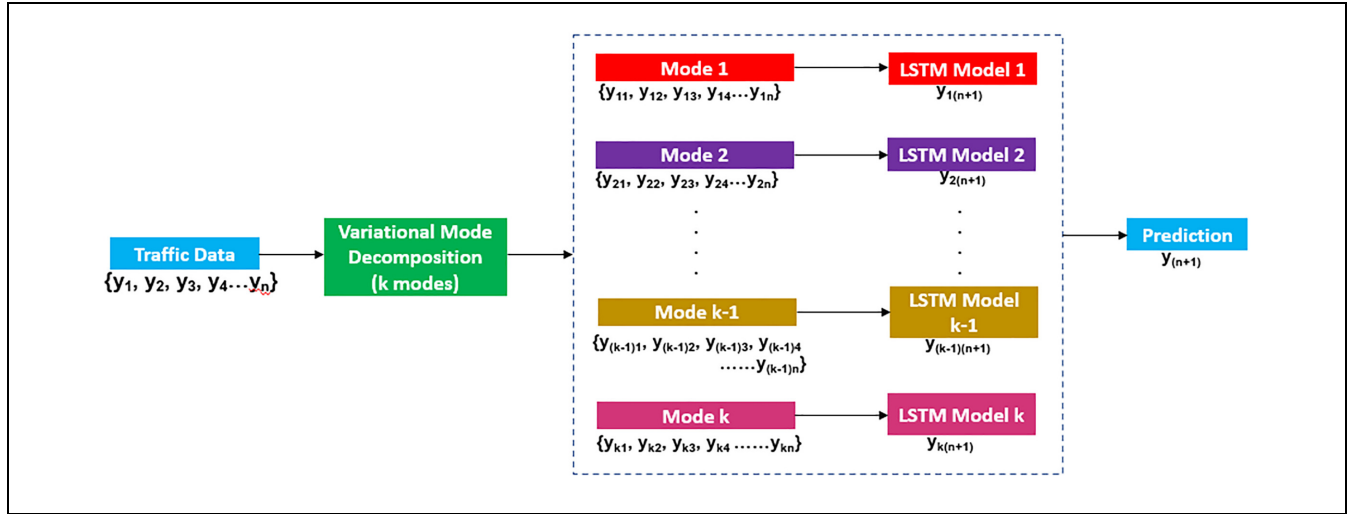
The remainder of this paper is organized as follows: the following section details the methodology; data collection and processing are then described, followed by presentation of the results. The final section presents our conclusions from this work.

## Methodology

The present study focused on developing a hybrid deep learning model-based prediction framework to forecast the probability estimates of predicted travel time. This methodology integrated three different techniques—VMD, LSTM, and QR—to build a multi-input, single-output model while considering traffic flow and speed as inputs to predict travel time. Let  $f(t) = y_1, y_2, y_3 \dots y_n$  be the observations of the speed time series, and  $g(t) = x_1, x_2, x_3 \dots x_n$  the observations of the traffic flow time series. The speed time series was decomposed into multiple band-limited intrinsic mode functions (IMFs), as shown in Figure 1, using VMD, and dedicated LSTM models integrated with QR loss function were built for each of these modes to predict the upper and lower bounds of the predicted travel time. In addition, the decomposed signals were reconstructed to provide predicted travel time outputs. The following section briefs the background details of VMD, LSTM, and QR.

### Variational Mode Decomposition

VMD is a nonrecursive signal processing method designed for decomposing complex nonstationary signals (20). The decomposition process is performed by a constrained variational problem to determine the bandwidth of each mode. This process involves three steps: 1) the Hilbert transform is used to obtain the unilateral frequency spectrum for each mode, 2) an exponential tuned



**Figure 1.** Research schema.

to the estimated center frequencies is used to shift every mode's frequency spectrum to baseband, and 3) the bandwidth of each mode is identified using the  $H^1$  Gaussian smoothness of the demodulated signal. Thus, the constrained variational problem is defined as

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (1)$$

$$s.t \sum_{k=1}^K u_k = f(t), \quad (2)$$

where

$\{u_k\}$  is set of all modes,

$\{\omega_k\}$  is set of respective center frequencies,

$k$  is number of predefined modes,

$\delta(t)$  is Dirac function,

$j$  is an imaginary number. This is a complex valued analytic signal,

$*$  denotes a convolution, and

$\|\cdot\|_2^2$  denotes a squared  $L^2$ -norm.

The present study adopted the number of predefined modes ( $k$ ) as 3, based on mode decomposition analysis. As suggested by Dragomiretskiy and Zosso, this constrained variational problem can be transformed into an unconstrained problem introducing a quadratic penalty term and Lagrangian multipliers,  $\lambda$ , as follows (21):

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha' \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| y(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), y(t) - \sum_{k=1}^K u_k(t) \right\rangle \quad (3)$$

This equation can be solved using a sequence of iterative suboptimizations known as the alternate direction method of multipliers (22, 23). By doing so, the modes,  $u_k$ , and their respective center frequencies,  $\omega_k$ , are then updated simultaneously using the following expressions:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha'(\omega - \omega_k)^2} \quad (4)$$

$$w_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (5)$$

The modes are solved in the spectral domain and can be transformed back into the time domain by taking the real part of the inverse Fourier transform of the signal. In Equation 4, value  $\alpha'$  represents a penalty term, defined by the user, which will define the shape of the modes.

### Long Short-Term Memory

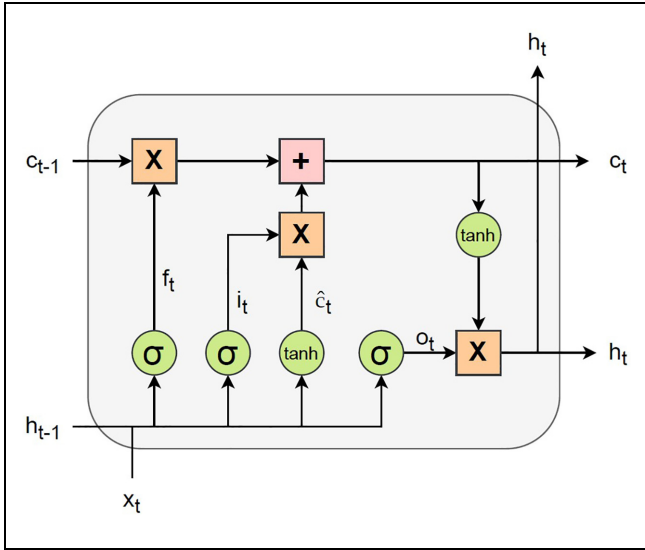
LSTM networks (24) regulate the flow of information using three gates (i.e., forget gate,  $f_t$ ; input gate,  $i_t$ ; and output gate,  $o_t$ ), and a reservoir of long-term memory known as cell state,  $c_t$ , to determine the hidden state,  $h_t$ , of the network, which corresponds to the output determined at every time step (Figure 2). The following equations indicate how the information is transmitted through the network:

$$f_t = \sigma(W_f y_t + U_f h_{t-1} + b_f) \quad (6)$$

$$i_t = \sigma(W_i y_t + U_i h_{t-1} + b_i) \quad (7)$$

$$\tilde{c}_t = \tanh(W_c y_t + U_c h_{t-1} + b_c) \quad (8)$$

$$o_t = \sigma(W_o y_t + U_o h_{t-1} + b_o) \quad (9)$$



**Figure 2.** Structure of an LSTM network.

Note: LSTM = long short-term memory;  $c_t$  = cell state;  $h_t$  = hidden state.

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (10)$$

$$h_t = o_t \odot \tanh(c_t) \quad (11)$$

Firstly, the LSTM network decides whether the information from the previous time step is discarded or maintained by means of the forget gate,  $f_t$  (Equation 6), where  $x_t$  is the input;  $h_{t-1}$  the previous hidden state;  $W_f$  and  $U_f$  are the weights for the input and previous hidden state, respectively;  $b_f$  the bias; and  $\sigma$  represents a sigmoid activation function. The next step is to renew the information contained in the cell state,  $c_t$ , based on the input and the previous hidden state,  $h_{t-1}$ . The new memory network is determined by the candidate cell state,  $\tilde{c}_t$  (Equation 8), whereas the input gate,  $i_t$  (Equation 7), acts as a filter to decide whether this new information is worth adding to the cell state,  $c_t$ , or should otherwise be filtered. In these equations,  $W_c$  and  $U_c$  are the weights for the input and previous hidden state for the candidate cell state,  $\tilde{c}_t$ ;  $b_c$  the bias of the same candidate cell state,  $\tilde{c}_t$ ;  $W_i$  and  $U_i$  the weights for the input gate; and  $b_i$  the bias of the input gate. In this case, the candidate cell state uses a hyperbolic tangent as the activation function, whereas the input gate is activated with a sigmoid activation function.

The cell state of the LSTM network is updated as shown in Equation 10, combining the elementwise product,  $\odot$ , of the forget gate and the previous cell state with the elementwise product of the input gate and the candidate cell state  $\tilde{c}_t$ . At this stage, the new hidden state  $h_t$  can be computed using the output gate (Equation 9) and the updated cell state of the network, as shown in Equation 11. The present study experimented with LSTM models under univariate and multivariate conditions.

## Quantile Regression

In this study, we implemented a QR loss function—a nonparametric approach—to estimate the PI corresponding to the lower and upper boundaries of the estimate. PI is a measure illustrating the robustness of the algorithm in relation to its ability to quote the variation within an observed dataset. The loss function is equal to

$$\rho_\tau(\epsilon) = \begin{cases} \tau\epsilon, & \text{if } \epsilon \geq 0 \\ (\tau - 1)\epsilon, & \text{otherwise} \end{cases} \quad (12)$$

Then, the error function that must be minimized is

$$E_\tau = \frac{1}{N} \sum_{i=1}^N \rho_\tau(y(i) - \hat{y}_\tau(i)) \quad (13)$$

where  $y(i)$  is the target value, and  $\hat{y}_\tau(i)$  is the forecast  $\tau$ -quantile.

## Prediction and Performance Evaluation

In this study, the accuracy of point forecasts was quantified using the mean absolute percentage error (MAPE),

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\% \quad (14)$$

where

$N$  = number of samples,

$y_i$  = observations, and

$\hat{y}_i$  = point forecasts.

However, the coverage and width of the PI must also be assessed for its evaluation. For that purpose, Prediction Interval Coverage Probability (PICP) metric was considered to measure the coverage of the PI and is defined as follows:

$$PICP = \frac{1}{N} \sum_{i=1}^N c_i \quad (15)$$

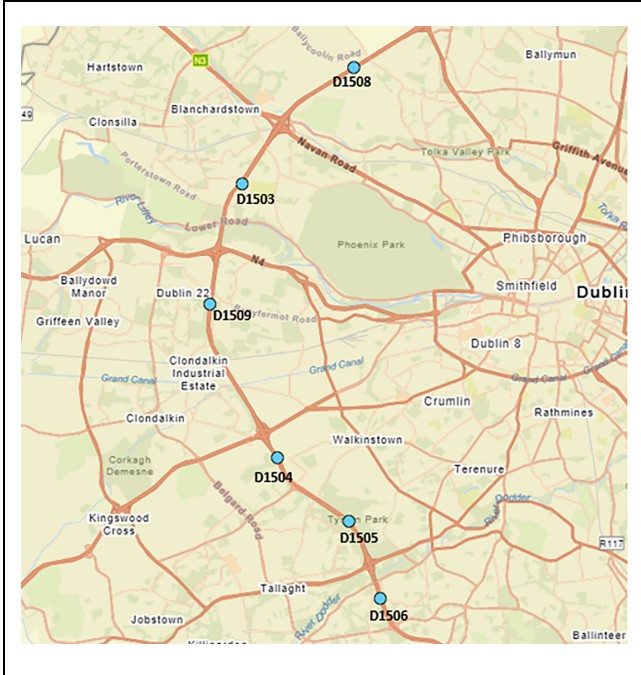
where  $N$  accounts for the number of observations, and  $c_i$  is equal to 1 if the observations fall within the PI, and 0 if not. A robust prediction algorithm would be expected to have a very high probability coverage.

The present study experimented with the aforementioned methodology in four ways (as shown in Table 1) to explore the best-performing combinations. Table 1 details the model combinations and their input variables adopted for prediction. Univariate models take past observations of speed time series as input to predict future values; multivariate models take past observations of both speed and flow time series as inputs to predict future speed values. Furthermore, VMD integrated models train dedicated LSTM models to predict values for

**Table 1.** Variations of Prediction Algorithms used for Travel Time Modeling

Serial no.	Models	Name	Inputs	No. of modes	Output
1	LSTM univariate	Uni-LSTM	Speed	NA	Travel time
2	LSTM multivariate	Multi-LSTM	Speed, Flow	NA	Travel time
3	VMD LSTM univariate	Uni-VMD-LSTM	Speed	3	Travel time
4	VMD LSTM multivariate	Multi-VMD-LSTM	Speed, Flow	3	Travel time

Note: LSTM = long short-term memory; VMD = variational mode decomposition.

**Figure 3.** Map of test bed with the chosen detectors.

each IMF, which are combined to obtain the final predicted speed signal.

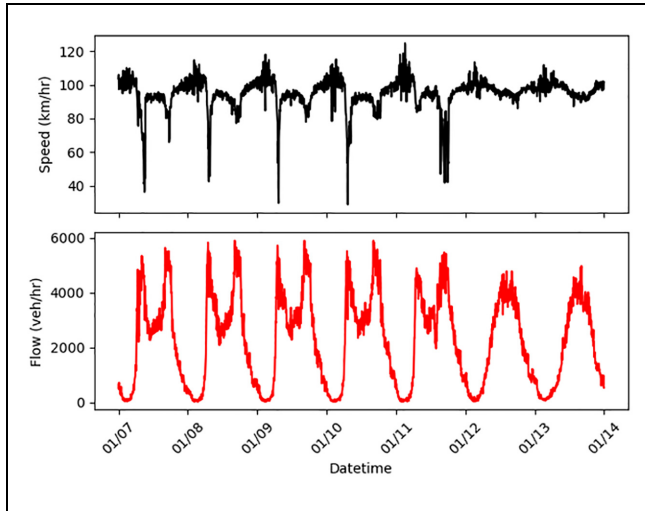
## Data Description

The data for this study were sourced from Traffic Infrastructure Ireland traffic counters (25) installed on the Irish road network. Vehicles are detected by passing over loops embedded beneath the road surface. Traffic counters provide information on the volume of traffic by time of day and by vehicle class (e.g., motorcycle, car, goods vehicles distinguished by the number of axles, etc.) with up to 12 classes being identified. In this study, we focused on six consecutive vehicle detectors located on the M50, the most prominent and busiest Irish motorway situated around the capital city, Dublin (see Figure 3). The M50 is a C-shaped, orbital, six-lane expressway corridor, with three lanes in each direction, that connects Dublin port with the M11 at Shankill, Ireland. All the other national routes radiate

outwards from Dublin, their junctions beginning at the M50. The speed limit is 120 km/h and the traffic composition consists of 79.31% passenger cars, 0.2% motorbikes, 11.74% light goods vehicles, 7.89% heavy motor vehicles, 0.34% buses, and 0.525% caravans.

The raw data obtained were vehicle transactions consisting of time of passage, speed, vehicle type, and lane identifiers. For this study, the flow and speed values from the vehicle class “passenger cars” were considered for a period of 5 months (January to May 2019). Reserving the last month for testing (80:20 ratio), the remaining data were utilized for training and validation. The sourced data were processed in four stages: data cleaning, outlier removal, time series formation, and data imputation. Data cleaning involves extraction of the necessary information from the raw data, which consists of location-related details, lane identifiers, and vehicle identities such as tag-IDs and length, which were removed from the database to prepare the necessary inputs for the developed methodology. In the outlier removal stage, unreasonable data points that did not reflect the characteristics of the study sites were removed. Vehicle transactions with zero speed values, extremely high speed values of more than 200 km/h, and negative speed values were identified as outliers and removed from the database. Such values may have been incorrectly reported owing to sensor or communication errors.

In the next stage, the cleaned flow and speed values were processed to set up the time series. In the present study, the traffic flow and speed values observed at different times of the day were viewed as sequential data or a time series. The entire 24-h time window was divided into 5-min slots, such that we had twelve 5-min slots in an hour totaling 288 slots in a 24-h window. Further, the data were preprocessed such that at each time slot there was only one observation. In this regard, the traffic flow observation for any slot was the cumulative number of vehicles passing over the counter during a particular 5-min interval. The speed values were obtained by averaging the speeds of all the vehicles that passed over the counter during the 5-min period. Missing speed values resulting from there being no vehicles during a 5-min period were imputed by temporal substitution, in which temporally lagged observations were used for data imputation. Substitutions were designed based on the



**Figure 4.** Sample plot of speed and flow time series.

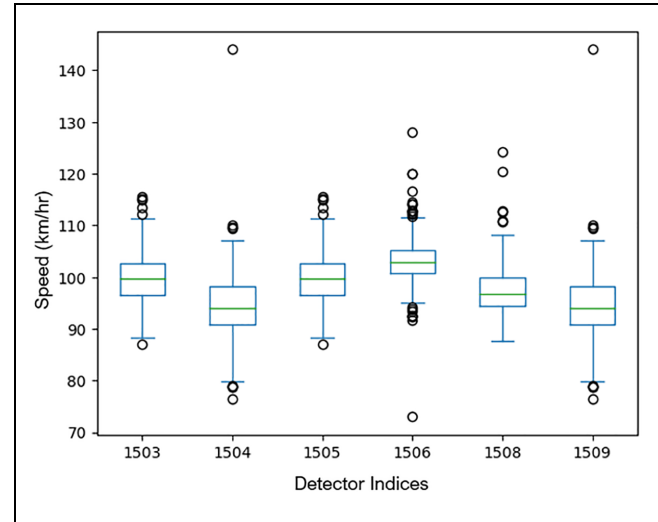
availability of data checked at different levels, such as an immediate past observation in time, and a week past observation, by taking advantage of the daily and weekly seasonality in the traffic data. This process of handling missing values is generally termed data imputation. The percentage of missing values was found to be less than 0.2% for the chosen dataset. The processed database comprised 43,488 observations in the continuous time series format with a 5-min resolution (frequency). A sample plot of processed speed and flow time series is shown in Figure 4. The descriptive statistics of the speed time series are clearly illustrated by boxplots presented in Figure 5.

From Figures 4 and 5 it can be observed that the statistical characteristics of the speeds identified by each of the detectors were significantly different, despite being situated consecutively on the same motorway. On that note, Figures 4 and 5 collectively reflect the spatiotemporal variation in speed values observed on the M50. The processed speed time series was given as input to the developed variable mode decomposition algorithm, and three different band-limited IMFs (modes) were generated. A sample plot of the original speed signal and decomposed modes is shown in Figure 6.

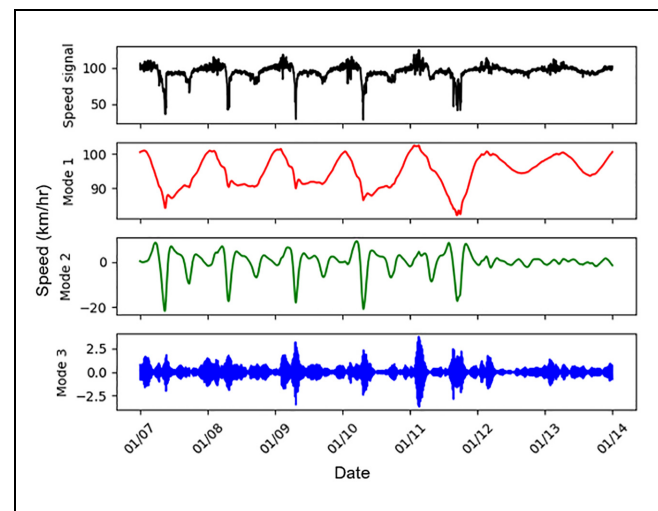
Further, each IMF was trained using dedicated LSTM models along with flow time series, and speed values were predicted. The present study considered 24 time-lagged observations to predict future travel time values with a 5-min horizon. In the subsequent stage, the travel time values were estimated from the predicted speed values.

## Results

To explore the efficiency and performance of the developed model, the results were evaluated and compared



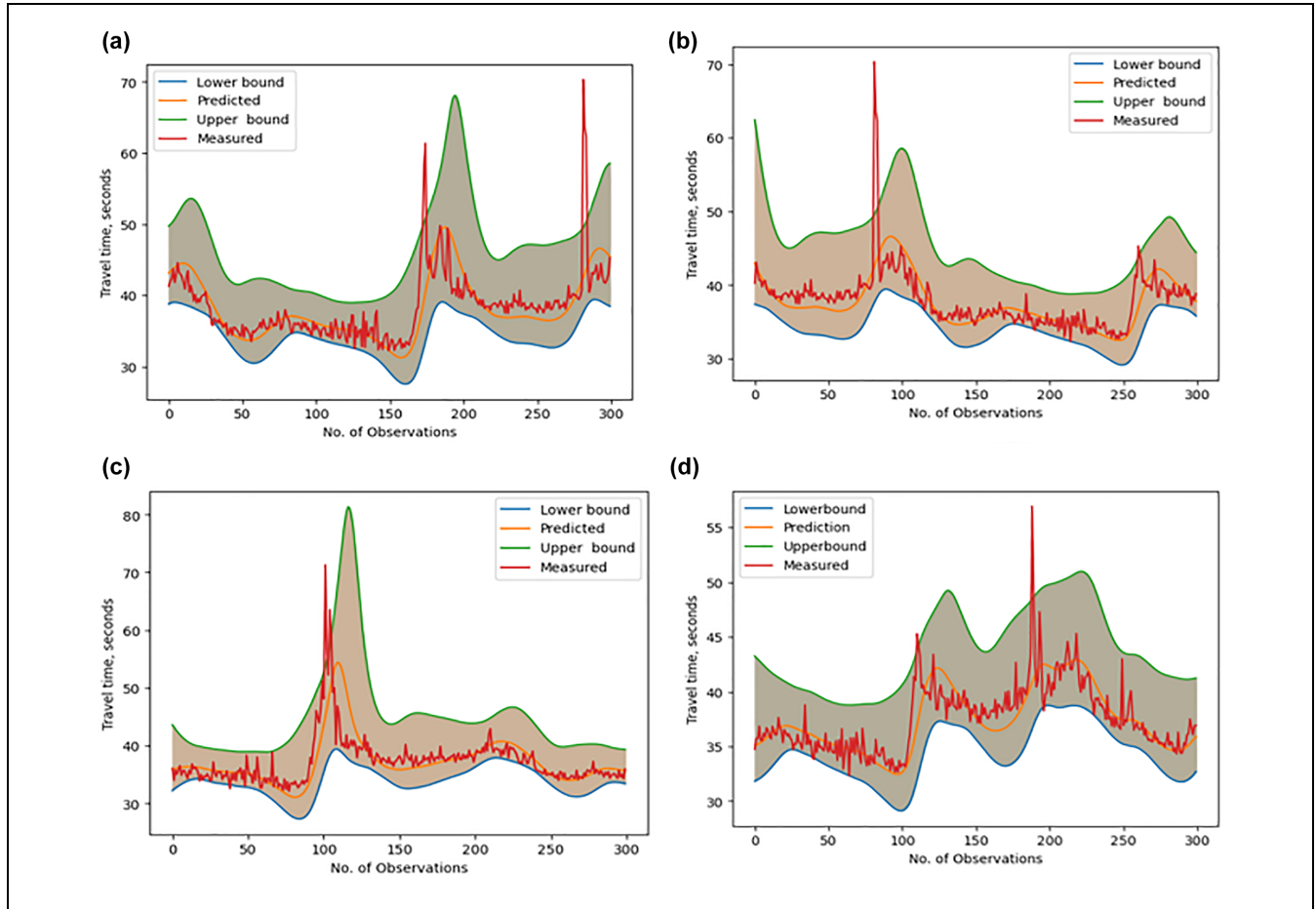
**Figure 5.** Box plot of speed sample across the considered detectors.



**Figure 6.** Mode decomposition of speed signal using VMD.  
Note: VMD = variational mode decomposition.

against the benchmark models. To check the importance of the mode decomposition step during prediction, the performance of the VMD LSTM model was compared with a simple LSTM model, which takes the input without any preprocessing. To identify the advantages of considering traffic flow in travel time prediction, performances were compared between multivariate and univariate versions of the deep learning models. Overall, the four test cases (shown in Table 1) Multi-VMD-LSTM, Uni-VMD-LSTM, Multi-LSTM, and Uni-LSTM were considered and the prediction performances of all models compared.

Figure 7 shows the predicted travel time intervals of all the explored model combinations and measured travel



**Figure 7.** Prediction intervals across developed methods.

times. It can be seen that intervals predicted by the multivariate models included all or most of the observed data points within the PIs, unlike the univariate models. It was also observed that the performance of the Multi-VMD-LSTM model was better than the other design variations, illustrating the advantage of adopting a signal processing tool like VMD when considering multiple variables. The VMD LSTM model presented a good adaptation to the data, even if some of the observations fall outside the interval.

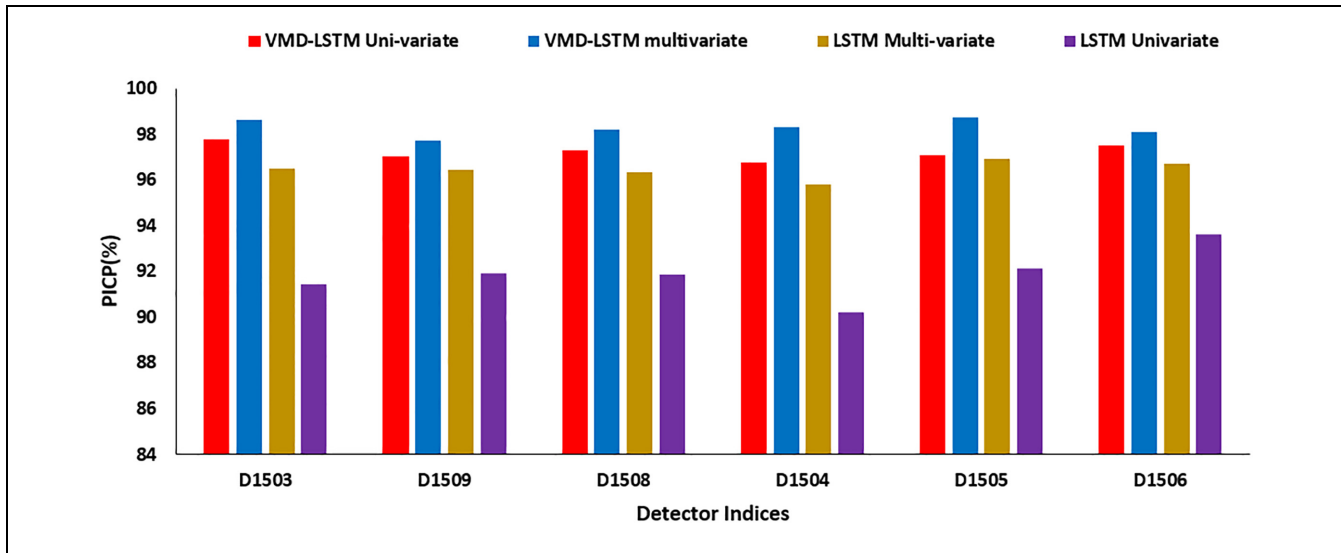
Figure 8 shows a comparison of PICP values across all the detectors among the four model variants. It was observed that the VMD LSTM model provided better coverage probability when compared with the LSTM models.

Further, the performance of all the modeled datasets considered in this study was compared to illustrate the consistency and effectiveness of the proposed methodology (Table 2). From the data presented in the table, it can be observed that the MAPE values of all the tested cases were between 3% and 6%. In the case of the six studied loop detectors, the VMD LSTM multivariate

version of the proposed model outperformed the other models tested in this study. Preprocessing using the VMD proved to be the most useful addition to a conventional deep learning model such as LSTM. The use of both speed and flow in traffic prediction proved effective in the case of four detectors, whereas the other two did not show any effective improvement. This outcome was similar to model performance without the preprocessing step. Furthermore, preprocessing seemed to improve the impact of multivariate inputs.

## Conclusion

Travel time prediction is essential to the developing and implementation of the majority of ITS applications in real time. The present study formulated a travel time prediction methodology by decomposing the input time series into multiple modes (IMFs) using VMD, and exclusive multivariate LSTM models were built for each of the IMFs, integrating QR to obtain the probabilistic intervals for the predicted travel time.



**Figure 8.** Comparison of PICP values across test detectors.

Note: PICP = prediction interval coverage probability.

**Table 2.** Table of MAPE Values for all Test Cases

Detector ID	MAPE (%)			
	VMD LSTM multi-var	VMD LSTM uni-var	LSTM multi-var	LSTM uni-var
D1503	3.41	3.31	4.35	5.10
D1508	4.52	5.13	6.1	5.98
D1509	5.21	5.25	6.4	6.72
D1504	2.41	4.41	4.95	5.60
D1505	3.82	5.78	6.1	6.98
D1506	4.72	5.62	5.9	6.32

Note: MAPE = mean absolute percentage error; VMD = variational mode decomposition; LSTM = long short-term memory; multi-var = multivariate; uni-var = univariate.

The probabilistic intervals produced the upper and lower bounds of the predicted travel time, providing a measure of uncertainty. Performance of the developed methodology was found to be efficient for both point forecasts, in which MAPE scores varied between 3% and 5%, and prediction intervals with PICP values varying between 97% and 99%. This performance was compared with simple LSTM models under univariate and multivariate cases to explore the advantages of a VMD–LSTM model combination. The results showed that the proposed method outperformed the benchmark methods in all cases, consistently showing the superiority of the developed methodology. Overall, the results showed that the VMD–LSTM–QR-based method was efficient and reliable for the purpose of travel time prediction.

Furthermore, the probabilistic estimates around the point predictions (i.e., probabilistic prediction interval) acted as a measure of the robustness of the prediction algorithms and are essential for real-time implementations. Under unexpected traffic conditions during incidents, pandemics, and extreme weather events, PIs would be expected to provide meaningful bounds with which to understand the expected variations of travel time in the near future. The developed methodology would be completely transferable to any location with the availability of the aforementioned data source and initial training to learn about the model parameters. Further, the multivariate LSTM could be extended by adding suitable weather factors to develop a weather-adaptive travel time prediction system—a possible future extension to this study.



## Author Contributions

The authors confirm their contribution to the paper as follows: study conception and design: D. Bharathi, J. M. González-Sopeña; data collection: D. Bharathi; analysis and interpretation of results: D. Bharathi; draft manuscript preparation: D. Bharathi, J. M. González-Sopeña, B. Ghosh, S. Clarke. All authors reviewed the results and approved the final version of the manuscript.


## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This publication emanated from research supported in part by a grant from Science Foundation Ireland (grant no. 13/RC/2077P2).

## ORCID iDs

Dhivya Bharathi  <https://orcid.org/0000-0002-3101-1752>

Juan Manuel González-Sopeña  <https://orcid.org/0000-0003-3282-7224>

Bidisha Ghosh  <https://orcid.org/0000-0002-1924-3040>

## References

- Williams, B., and L. Hoel. Modeling and Forecasting Vehicle Traffic Flow as a Seasonal Arima Process: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering*, Vol. 129, No. 6, 2003, pp. 664–672.
- Bharathi, B. D., B. A. Kumar, A. Achar, and L. Vanajakshi. Bus Travel Time Prediction: A Log-Normal Autoregressive (AR) Modelling Approach. *Transportmetrica A: Transport Science*, Vol. 16, No. 3, 2020, pp. 807–839.
- Ghosh, B., B. Basu, and M. Mahony. Multivariate Short-Term Traffic Flow Forecasting Using Time-Series Analysis. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, No. 2, 2009, pp. 246–254.
- Tirachini, A. Estimation of Travel Time and the Benefits of Upgrading the Fare Payment Technology in Urban Bus Services. *Transportation Research Part C: Emerging Technologies*, Vol. 30, 2017, pp. 239–256.
- Yu, Z., J. Wood, and V. Gayah. Using Survival Models to Estimate Bus Travel Times and Associated Uncertainties. *Transportation Research Part C: Emerging Technologies*, Vol. 74, 2017, pp. 366–382. <https://doi.org/10.1016/j.trc.2016.11.013>.
- Yin, T., G. Zhong, J. Zhang, S. He, and B. Ran. A Prediction Model of Bus Arrival Time at Stops with Multi-Routes. *Transportation Research Procedia*, Vol. 25, 2017, pp. 4623–4636.
- Chien, S. J., Y. Ding, and C. Wei. Dynamic Bus Arrival Time Prediction with Artificial Neural Networks. *ASCE Journal of Transportation Engineering*, Vol. 128, No. 5, 2002, pp. 429–438.
- Bin, Y., Y. Zhinzhen, and Y. Baozhen. Bus Arrival Time Prediction Using Support Vector Machines. *Journal of Intelligent Transportation Systems*, Vol. 10, No. 4, 2006, pp. 151–158.
- Vanajakshi, L., and L. Rilett. Support Vector Machine Technique for the Short Term Prediction of Travel Time. *Proc., IEEE Intelligent Vehicles Symposium*, Istanbul, Turkey, 2007, IEEE, New York, pp. 600–605.
- Agafonov, A. A., and A. Yumaganov. Bus Arrival Time Prediction Using Recurrent Neural Network with LSTM Architecture. *Optical Memory Neural Networks*, Vol. 28, 2019, pp. 222–230.
- Petersen, N., F. Rodrigues, and F. Pereira. Multi-Output Bus Travel Time Prediction with Convolutional LSTM Neural Network. *Expert Systems with Applications*, Vol. 120, 2019, pp. 426–435. <https://doi.org/10.1016/j.eswa.2018.11.028>.
- Gu, Y., W. Lu, X. Xu, L. Qin, Z. Shao, and H. Zhang. An Improved Bayesian Combination Model for Short-Term Traffic Prediction with Deep Learning. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 21, No. 3, 2020, pp. 1332–1342. <https://doi.org/10.1109/TITS.2019.2939290>.
- Aloysius, N., and M. Geetha. A Review on Deep Convolutional Neural Networks. *Proc., International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, IEEE, New York, 2017. <https://doi.org/10.1109/iccsp.2017.8286426>.
- Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, Vol. 454, No. 1971, 1998, pp. 903–995.
- Wu, Z., and N. E. Huang. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. *Advances in Adaptive Data Analysis*, Vol. 1, No. 1, 2009, pp. 1–41.
- Zheng, Z., L. Pan, and K. Pholsena. Mode Decomposition Based Hybrid Model for Traffic Flow Prediction. *Proc., IEEE Third International Conference on Data Science in Cyberspace (DSC)*, Guangzhou, China, 2018, New York, IEEE, pp. 521–526.
- Tian, Z. Approach for Short-Term Traffic Flow Prediction Based on Empirical Mode Decomposition and Combination Model Fusion. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 22, No. 9, 2021, pp. 5566–5576. <https://doi.org/10.1109/TITS.2020.2987909>.
- Xiu, C., Y. Sun, Q. Peng, C. Chen, and X. Yu. Learn Traffic as a Signal: Using Ensemble Empirical Mode Decomposition to Enhance Short-Term Passenger Flow Prediction in Metro Systems. *Journal of Rail Transport Planning Management*, Vol. 22, 2022, p. 100311.
- Sopeña, J. M. G., V. Pakrashi, and B. Ghosh. Interval Prediction for Short-Term Traffic Forecasting Using Hybrid Mode Decomposition Models. *Proc., IEEE International Intelligent Transportation Systems Conference (ITSC)*, Indianapolis, IN, 2021, New York, IEEE, pp. 3246–3251.

20. Sopena, J. M. G., V. Pakrashi, and B. Ghosh. Decomposition-Based Hybrid Models for Very Short-Term Wind Power Forecasting. *Proc., 7th International Conference on Time Series and Forecasting*, Gran Canaria, Spain, 2021.
21. Dragomiretskiy, K., and D. Zosso. Variational Mode Decomposition. *IEEE Transactions on Signal Processing*, Vol. 62, No. 3, 2013, pp. 531–544.
22. Hestenes, M. R. Multiplier and Gradient Methods. *Journal of Optimization Theory and Applications*, Vol. 4, No. 5, 1969, pp. 303–320.
23. Boyd, S., N. Parikh, and E. Chu. *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. Now Publishers Inc., Delft, the Netherlands, 2011.
24. Hochreiter, S., and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, Vol. 9, No. 8, 1997, pp. 1735–1780.
25. Traffic Counts/Data Map. 2013. <https://trafficdata.tii.ie/publicmultinodemap.asp>. Accessed March 15, 2022.