

1 **Freeway Traffic Flow Forecasting Using Physics-Guided LSTM with Flawed Data**

2
3
4 **Zhao Zhang**

5 Ph.D., Postdoctoral Researcher

6 Department of Civil and Environmental Engineering, University of Utah

7 100 Central Campus Dr., Suite 2000, Salt Lake City, UT, U.S., 84112

8 Email: zhao.zhang@utah.edu

9
10 **Qinzheng Wang**

11 Ph.D., Postdoctoral Researcher

12 Department of Civil and Environmental Engineering, University of Utah

13 100 Central Campus Dr., Suite 2000, Salt Lake City, UT, U.S., 84112

14 Email: qinzheng.wang@utah.edu

15
16 **Hao Yang**

17 Ph.D., Assistant Professor

18 Department of Civil Engineering, McMaster University, Hamilton, ON, Canada, L8S 4L8

19 Email: haoyang@mcmaster.ca

20
21 **Xianfeng (Terry) Yang**

22 Ph.D., Assistant Professor

23 Department of Civil and Environmental Engineering, University of Maryland

24 1173 Glenn L. Martin Hall, College Park, MD, U.S., 20742

25 Email: xyang@umd.edu

26
27
28 Word Count: 6500 words + 4 Tables * 250 = 7500 words

29
30
31 Submitted to 102nd Transportation Research Board Annual Meeting for Presentation

32
33
34
35 Submission Date: August 1, 2022

1 **Abstract**

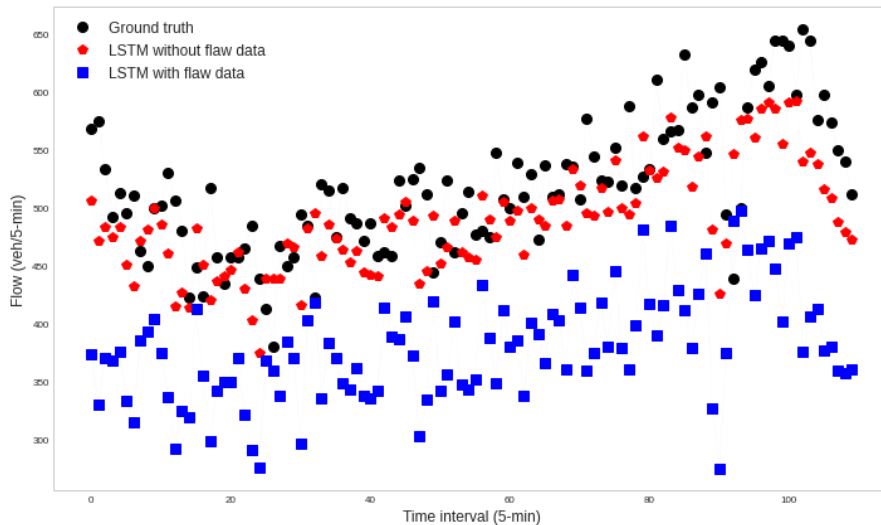
2 Having an accurate traffic flow predictions play an important role in intelligent transportation
3 systems. Recently, long short-term memory (LSTM) model is prevailing in utilizing big data for
4 traffic state prediction. The performance of LSTM model heavily depends on high-quality data
5 due to their data-driven nature. However, historical traffic data usually contains flaws (e.g.,
6 incomplete, or incorrect information) since the malfunctioning of traffic sensors. To tackle this
7 issue, this study aims to propose and evaluate a new advanced model, named as physics-guided
8 LSTM, that adopt traffic physical knowledge into a modified loss function to guide the training
9 process of LSTM. More specifically, PG-LSTM network could bring the physical knowledge
10 from upstream traffic flow to overcome the data flaw problem downstream for traffic flow
11 forecasting. To illustrate the effectiveness of the PG-LSTM, this study implements empirical
12 studies with a real-world dataset collected from a stretch of I-15 freeway in Utah. Experimental
13 study results show that the proposed PG-LSTM model could outperform the other compatible
14 methods.

15 **Keywords:** Traffic flow forecasting, LSTM, Physics-guided LSTM, Macroscopic traffic flow
16 modeling

1 **1. INTRODUCTION**

2 Accurate prediction of future traffic state is an essential component of intelligent transportation
 3 systems (ITS), which helps travelers plan their trips, allows transportation agencies to take
 4 actions to mitigate traffic congestions, and therefore promote a more efficient and safer driving
 5 environment. The acquisition of accurate future traffic information has always been complicated
 6 due to the stochastic nature of traffic patterns. Traffic state prediction (TSP) is a method that can
 7 predict future traffic information based on historical traffic information (1–6), which is an
 8 effective way to obtain the future traffic information.

9 Data collection and modeling approach selection are two key tasks in TSP. Most of the
 10 existing studies on traffic state predictions are dependent on historical traffic information
 11 collected by various stationary traffic sensors (1–5). In practice, stationary data can be easily
 12 retrieved because it is collected by fixed traffic detectors (e.g., inductive loops and radar
 13 detectors) on freeways. Each stationary detector counts the number of vehicles that pass every
 14 minute and detects the speed of each vehicle. However, those collected historical data usually
 15 contains missing and erroneous value. In recent decades, ML models are prevailing in utilizing
 16 big data for TSP, such as pure LSTM model (2, 3, 7–13) and hybrid LSTM model (14–16).
 17 However, accurate future traffic information tends to be hard to obtain because historical data
 18 usually contains missing and erroneous data. The flawed data could significantly downgrade the
 19 performance of traditional prediction models. Figure 1 illustrates an example of biased prediction
 20 (marked as blue square) of an LSTM network with a dataset that contains flawed data diverges
 21 from LSTM network on accurate data (marked as red pentagon). Hence, there is an urgent need
 22 for exploring an efficient TSP method that has the capability to address those data issues.



23 **Figure 1 Comparison between LSTM with and without flawed data**

24
 25 During the past decades, classical traffic physics models have been reported in the
 26 literature, which can be generally classified into two categories: (a) continuous models, such as
 27 the first-order Lighthill-Whitham-Richards (LWR) model (17, 18), the second-order Payne-
 28 Whitham (PW) model (19, 20), and the second-order Aw-Rascole-Zhang (ARZ) model (21, 22);
 29 and (b) discretized models, such as a discrete PW-like TSE model, named METANET (23),
 30 which were presented to simulate traffic states of subsegments and time intervals because of their
 31 tremendous computational efficiency. Grounded on the physics laws, those models often require

1 great efforts for parameter calibrations and are difficult to work with data fluctuations and
 2 uncertainties. In view of the increasing data availability in transportation systems, ML models
 3 have been rapidly developed and widely implemented in recent years. Compared with the
 4 classical traffic physics models, ML models does not require the compliance of physics laws. So
 5 far, the fundamental theories behind ML models in traffic flow modeling. Recently, researchers
 6 started to utilize hybrid machine learning for traffic flow modeling, which incorporates traffic
 7 physics knowledge into machine learning model to concurrently take the benefits and overcome
 8 the limitations of both. The modeling performance can be improved from two perspectives: (1)
 9 data availability (24), the traffic physics model increases the data availability for ML model
 10 training to improve the modeling performance; (2) model structure (25), traffic physics model is
 11 used for building up the physics laws in the whole training process of ML models, which makes
 12 the ML models comply with the physics laws to further improve the modeling performance.

13 Therefore, this research aims to develop an innovative framework, named as physics-
 14 guided LSTM (PG-LSTM), to solve the flawed data problem in TSP tasks. More specifically,
 15 PG-LSTM network could adopt traffic physics knowledge into a modified loss function to guide
 16 the training process of LSTM to tackle the flaw data limitation and obtain accurate traffic flow
 17 predictions. Hence, it could bring the traffic physics knowledge from traffic flows at the
 18 upstream segment to overcome the data flow problem in traffic flow forecasting at the
 19 downstream section. This study makes significant contributions to the literature from the
 20 following perspectives: (1) an innovative PG-LSTM model is proposed for TSP problem; (2) a
 21 modified loss function is developed to integrate traffic physics knowledge to participate in the
 22 whole training process; (3) an advanced method is provided to solve the flawed data problem in
 23 TSP tasks. This research is expected to bridge the gap between the research of fundamental
 24 transportation theory and data-driven approaches proposed by an innovative PG-LSTM
 25 framework.

26 2. LITERATURE REVIEW

27 *Traditional traffic flow model*

28 In early stages, macroscopic traffic flow was found to be similar to hydrodynamic theory
 29 (26). Based on that finding, the fundamental diagram was defined as the relationship between
 30 traffic speed, flow, and density. The fundamental diagram is one of the most basic concepts in
 31 traffic flow theory, which is described in **Equations 1-2**.

$$32 \quad v = V(\rho) \quad (1)$$

$$33 \quad q = \rho V(\rho) \quad (2)$$

34 where V represents the speed-density fundamental diagram.

35 According to the fundamental diagram, macroscopic traffic flow models were developed
 36 using partial differential equations (PDE) to represent the aggregated traffic behavior. The traffic
 37 flow models can be generally classified into continuous models and discretized models. The
 38 Lighthill-Whitham-Richards (LWR) model (17, 18) is a continuous first-order model and can be
 39 formulated in **Equations 3-4**.

$$40 \quad \partial_t \rho + \partial_x(\rho v) = 0 \quad (3)$$

$$41 \quad v = V(\rho) \quad (4)$$

1 The LWR model succeeds in mimicking simple traffic conditions (e.g., traffic jam and
2 shockwave) but it cannot reproduce more complicated traffic phenomena well.

3 To tackle such limitations, the well-known second-order PW model (19, 20) was
4 developed by adding the momentum equation to capture complex traffic behavior. The PW
5 model is formulated as **Equations 5-6**, where **Equation 6** is the momentum equation.

$$6 \quad \partial_t \rho + \partial_x(\rho v) = 0 \quad (5)$$

$$7 \quad \partial_t v + v \partial_x v = -\frac{v-v(\rho)}{\tau} - \frac{c_0^2}{\rho} \partial_x(\rho) \quad (6)$$

8 where τ is the relaxation time and c_0^2 is a parameter related to driver anticipation. Papageorgiou
9 et al. (23) proposed a discrete PW-like TSE model, named METANET, which is an extension of
10 PW model. It can reproduce complex traffic phenomena but does not require tremendous
11 computation efforts at a certain level.

12 *Traffic state prediction*

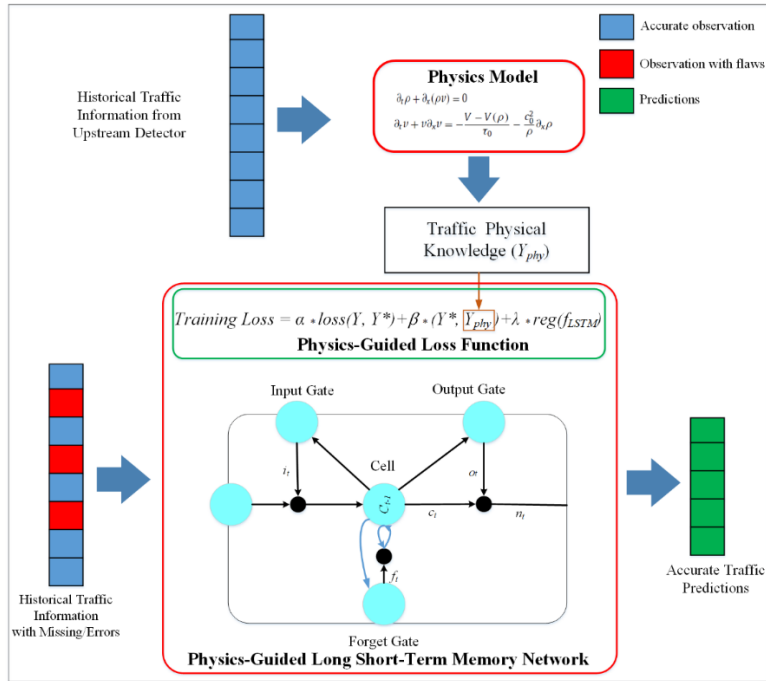
13 In recent decades, many methods have been developed to forecast short-term traffic state.
14 The prediction methods can be generally classified into parametric methods and Nonparametric
15 methods (15). Parametric methods mainly include ARIMA model (27–29) and Kalman filter (30,
16 31). The prediction performance of parametric methods tend to cannot obtain satisfying result
17 under irregular traffic variations. To tackle this issue, nonparametric methods are developed to
18 obtain the acquisition of nonlinear laws from historical data. It mainly includes ANN (32, 33), k-
19 Nearest Neighbors (34–37), SVM (35, 38), and Bayesian model (39, 40). However, parametric
20 methods heavily relied on data quality and quantity of training data. In recent years, deep
21 learning gained prominence in TSP tasks regarding performance improvements. Compared with
22 other deep learning networks, the recurrent neural network (RNN) could better capture the
23 temporal evolution of traffic flow by self-loops and chain-like structures (16). The RNNs have
24 been broadly implemented for predicting freeway traffic (41–43). Traditional RNN models have
25 two constraints (2, 44): (1) Traditional RNNs cannot train time series with long time lags, and (2)
26 Traditional RNNs rely on the predetermined time lags to learn the temporal sequence processing,
27 but it is challenging to find the optimal time window size in an automatic way. To overcome
28 those limitations, Long Short-term Memory (LSTM) Network is a special Recurrent Neural
29 Network (RNN) architecture, was initially proposed by Hochreiter and Schmidhuber (45). The
30 LSTM network could overcome those limitations because it can learn information with long time
31 spans and determine the optimal time lags for prediction in an automated manner (2, 15). These
32 attributes made the LSTM network widely implemented in the transportation domain, especially
33 for traffic state prediction (2, 3, 7–13). Recently, many research started to develop hybrid LSTM
34 structures to further improve the performance of pure LSTM, such as convolutional LSTM
35 (ConvLSTM) (14), self-attention LSTM (SA-LSTM) (15), end-to-end hybrid deep learning
36 network (M-B-LSTM) (16). However, the better prediction performance obtained by all pure
37 LSTM and hybrid LSTM depends on the high-quality historical information. The performance of
38 these models is significantly downgraded if historical training data contains flaws.

39 In summary, a hybrid TSP framework that integrates the physics knowledge and the data-
40 driven methods to overcome flawed data problem is still lacking. This paper focuses on filling
41 the gap by proposing a physics-guided LSTM model for TSP problem.

1 **3. METHODOLOGY**

2 ***Physics-Guided Long Short-Term Memory (PG-LSTM) Network Structure***

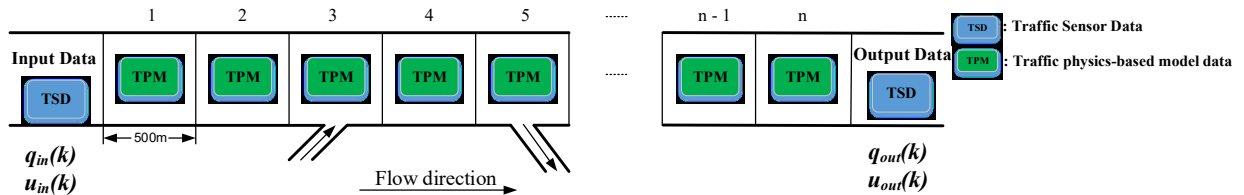
3 To deal with the flawed data problems in traffic flow forecasting, the PG-LSTM network
 4 is constructed. As shown in Figure 2, the proposed PG-LSTM includes four key steps. First,
 5 traffic physics knowledge is built up for the target spot by the traffic flow model. Second, the
 6 physics inconsistency is obtained for the target segment. Third, a new physics-guided loss
 7 function is constructed for the LSTM network. Then, the model training process would be guided
 8 by the new loss function. The new loss function could bring the traffic physics knowledge to the
 9 training process, which will help the LSTM network to overcome the data missing problem. Last,
 10 the trained PG-LSTM is used to forecast the future traffic flow. The LSTM network and physics-
 11 guided loss function are described in the following sections.



12 **Figure 2 The diagram of PG-LSTM network**

14 ***Second-order traffic flow model***

15 In this study, the second-order macroscopic traffic flow model is implemented as a traffic
 16 physics model to construct traffic physics knowledge, which is developed by Papageorgiou et al.
 17 (46).



18 **Figure 3 Freeway stretch example**

1 In Figure 3, the target freeway segment is conceptually divided into n subsegments with a
 2 unit length of ΔL (500m). For each subsegment i , the mean density, $d_j(k)$, can be determined by
 3 the difference between the input and output flows as follows:

$$4 \quad d_i(k + 1) = d_i(k) + \frac{\Delta T}{\lambda_i \Delta L} [q_{i-1}(k) - q_i(k) + r_i(k) - s_i(k)] \quad (7)$$

5 The departure flow is assumed to be a portion of the flow at the segment in **Equation 8**. The
 6 ramp flow is captured by a random walk through assuming any unmeasured on-ramp and off-
 7 ramp are constant.

$$8 \quad s_i(k) = \beta_i(k) * q_{i-1}(k) \quad (8)$$

9 For dynamically updating the average speed, $u_i(k)$, a well-developed equation proposed by
 10 the METANET model (46) is adopted and shown as follows:

$$u_i(k + 1) = u_i(k) + \frac{\Delta T}{\tau_i} [V_i\{d_i(k)\} - u_i(k)] + \frac{\Delta T}{L_i} u_i(k) [u_{i-1}(k) - u_i(k)] \\ - \frac{\gamma_i \Delta T [d_{i+1}(k) - d_i(k)]}{\tau_i \Delta L [d_i(k) + \kappa]} \quad (9)$$

11 where $V[d_i(k)]$ is the static speed for segment i at time k with respect to the density $d_i(k)$:

$$12 \quad V[d_i(k)] = u_f \exp \left[-\frac{1}{a} \left(\frac{d_i(k)}{d_{cr}} \right)^a \right] \quad (10)$$

13 Also, the relationship between flow, density, and speed is given by the following:

$$14 \quad q_i(k) = d_i(k) u_i(k) \lambda_i \quad (11)$$

15 where **Equations 8-11** are the conservation equation, dynamic speed equation, stationary speed
 16 equation, and flow equation, respectively; $\tau, \gamma, \tau, \kappa, d_{cr}, u_f, a$ are positive model parameters
 17 which are given the same values for all segments. Using the traffic flow and speed from traffic
 18 sensors at upstream, the traffic physics knowledge can be directly estimated by **Equations 8-11**
 19 on the downstream sections.

20 **Long Short-Term Memory (LSTM) Network**

21 To overcome the vanishing gradient problem of traditional RNN, a special Recurrent
 22 Neural Network (RNN) architecture – LSTM network was proposed by Hochreiter and
 23 Schmidhuber (45). The LSTM network can determine the optimal time lags for the time series
 24 problem by giving the memory cell ability to determine when to forget information. This makes
 25 the LSTM network particularly desirable for short-term TSP. A typical LSTM network is made
 26 up of one input layer, one recurrent hidden layer (memory block), and one output layer. The
 27 memory block contains memory cells with self-connections memorizing the temporal state, and
 28 pair of adaptive, multiplicative gating units to control information flow in the block. The LSTM
 29 network can solve the vanishing gradient problem because multiplicative gates can learn to open
 30 and close, and thus LSTM network cells can store and access information for a long time. The
 31 structure of the LSTM network is illustrated in Figure 4.

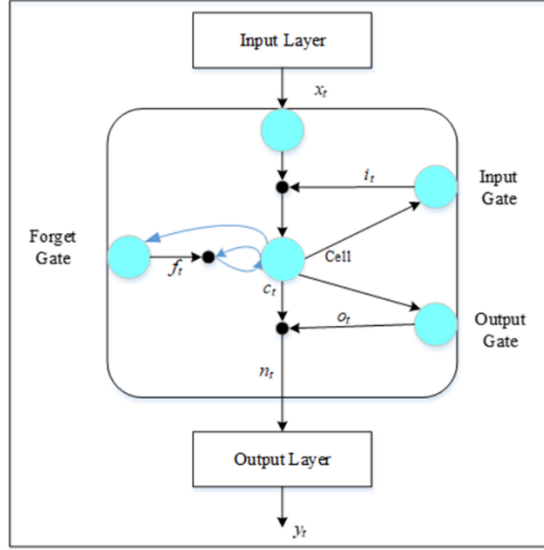


Figure 4 The structure of LSTM network

The input of the historical traffic flow sequence is denoted as $x = (x_1, x_2, \dots, x_T)$ (T is the prediction period), and the output sequence is $y = (y_1, y_2, \dots, y_T)$ can be iteratively calculated as follows:

$$i_t = \sigma(W_{ix}x_t + W_{in}n_{t-1} + b_i) \quad (12)$$

$$f_t = \sigma(W_{fx}x_t + W_{fn}n_{t-1} + b_f) \quad (13)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cn}n_{t-1} + b_c) \quad (14)$$

$$o_t = \sigma(W_{ox}x_t + W_{on}n_{t-1} + W_{oc}c_t + b_o) \quad (15)$$

$$n_t = o_t \odot h(c_t) \quad (16)$$

where i_t , f_t , and o_t denote the output of three gates; c_t and n_t denote the activation vectors for each cell and memory block; W denotes weight matrices; b denotes bias vectors; \odot denotes the scalar product of two vectors; and $\sigma(\cdot)$ denotes the standard logistics sigmoid function:

$$\sigma(x) = \frac{1}{1 + e^x} \quad (17)$$

$g(\cdot)$ is a centered logistic sigmoid function with range $[-2, 2]$:

$$g(x) = \frac{4}{1 + e^x} - 2 \quad (18)$$

$h(\cdot)$ is a centered logistic sigmoid function with range $[-1, 1]$:

$$h(x) = \frac{2}{1 + e^x} - 1 \quad (19)$$

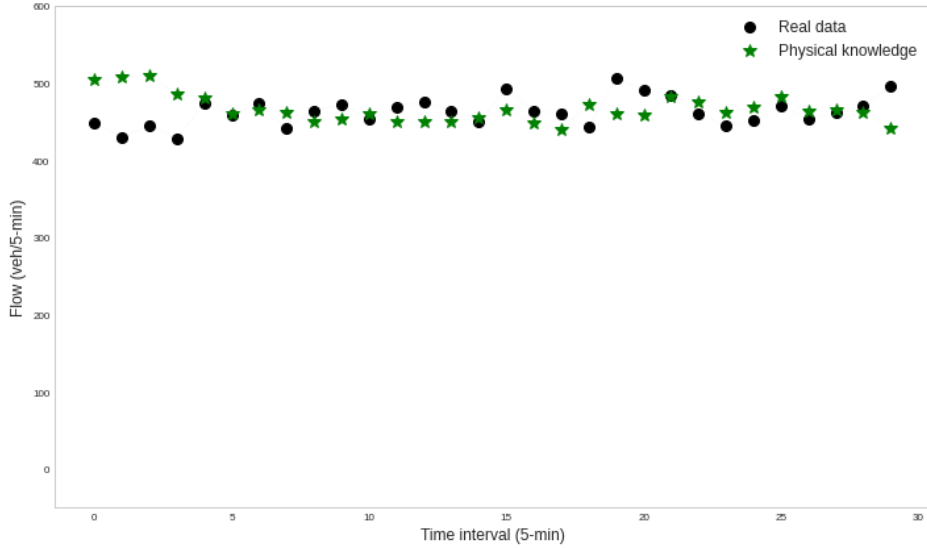
Problem Introduction

The time series data on studied spots are defined as S . Let S_T denote the data on target spot and S_u denote the data on upstream segment. The traffic information from the target detector is denoted by $S_t = \{x_1^T, x_2^T, \dots, x_n^T\}$ (n denotes the number of data points), which is used for model training. The traffic information from the upstream detector is denoted by $S_u = \{x_1^u, x_2^u, \dots, x_n^u\}$, which is used for constructing the traffic physics knowledge for target segment.

1 In this research, the TFM is utilized to construct traffic physics knowledge for the target spot
 2 based on S_u , named physics data Y_{phy} . The input of TFM includes speed and flow from the
 3 upstream detector. Then, unobserved historical traffic physics knowledge on the target spot could
 4 be estimated by TFM using **Equation 20**.

$$5 \quad S_u \xrightarrow{TFM} Y_{phy} \quad (20)$$

6 Figure 5 illustrates the comparison between constructed physics knowledge and real
 7 observed data. It can be clearly seen that the pattern of constructed physics knowledge is similar
 8 to the real observed data. This accurate physics knowledge established at the target spot could
 9 provide the cherish information for flawed data inherent, especially for ML models to find the
 10 correct pattern with flawed data.



11 **Figure 5 The comparison between constructed physics knowledge and real observed data**

12 Hence, the purpose of this study is to integrate physics knowledge into the LSTM model
 13 to improve the prediction performance with flawed data. The goal of traffic flow forecasting is to
 14 predict the traffic flow in the future time series length based on the current historical traffic
 15 information. Let F denote the PG-LSTM transform function, then the prediction process can be
 16 described as:
 17

$$18 \quad \hat{x}_{t+1}^T = [f_0, f_1, \dots, f_n] \begin{bmatrix} x_{t-n}^T \\ \vdots \\ x_t^T \end{bmatrix} \quad (21)$$

19 where $[x_{t-n}^T, x_{t-n+1}^T, \dots, x_{t-1}^T, x_t^T]^T$ is a sequence of traffic flow observations with n historical
 20 time steps as the network input and \hat{x}_{t+1}^T is the predicted value at the next state. After obtaining
 21 the x_{t+1}^T , we can add the x_{t+1}^T to the training set to get \hat{x}_{t+2}^T . By repeating this process, the traffic
 22 flow and prediction of multiple time series can be forecasted.

23 **Physics-Guided Loss Function**

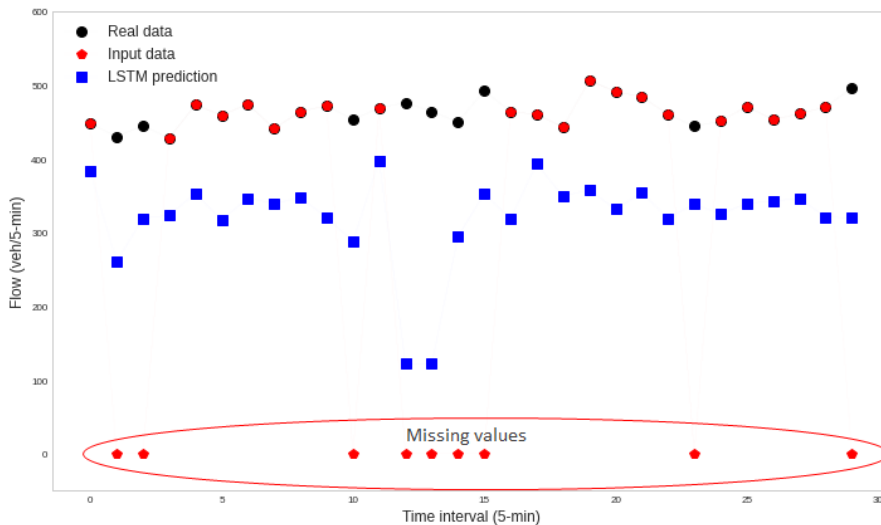
24 The purpose of this research is to predict traffic flow at time $T + 1$ using historical data.
 25 If the data is complete and accurate, we can easily perform this prediction using recurrent
 26 dynamics methods like LSTM. However, flawed data are unpredictable and beyond our control

1 due to a variety of factors such as sensor malfunctioning and transmission failure. Flawed data
 2 frequently leads to unreliable predictions, which has a negative impact on traffic management
 3 system.

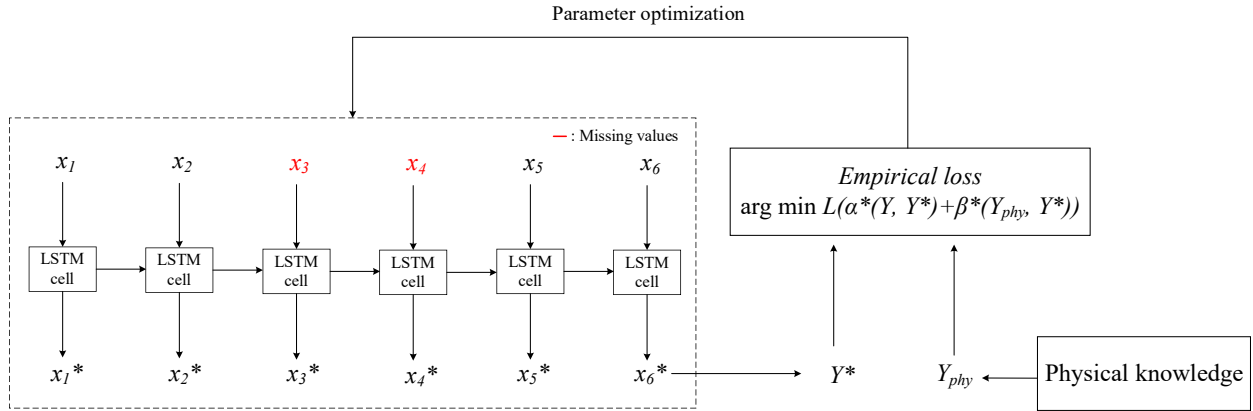
4 It should be noted that the aforementioned strategy may result in poor prediction
 5 performance because the performance of the LSTM model is strongly dependent on the input
 6 data. This is because that machine learning methods leverage loss function to evaluate how well
 7 a specific algorithm models the given data. If predictions deviate too much from the ground truth
 8 data, the loss function will return a very large prediction error. Gradually, the loss function learns
 9 to reduce the prediction error (i.e., close to real data) with the assistance of some optimization
 10 functions. When a random value is used to impute missing data, the loss function will activate to
 11 learn the data distribution including missing data. Since the quality of imputed data cannot be
 12 guaranteed, thus the prediction performance may be low. Figure 6 is an example that presents the
 13 comparison between LSTM prediction with input data including missing values and real
 14 observed data. As shown in the figure, the performance of LSTM is significantly downgraded
 15 because the input training data contains missing values. Therefore, a new loss function that can
 16 bring in the physics knowledge to guides the whole training process of LSTM is developed.
 17 Figure 7 shows the detailed procedure that PG-LSTM integrates the physics knowledge into loss
 18 function to overcome the flawed data. The physics knowledge can guide the loss function to get
 19 closer to the ground truth data because the physics model can well capture the real data
 20 distribution, as shown in Figure 5. Figure 8 is an example that presents the comparison between
 21 LSTM prediction and real data with the guidance of physics knowledge. Note that the same
 22 method is applied to conduct prediction for the condition of flawed data. The physics-guided loss
 23 function is described as follows:

$$24 \quad \text{PGLoss} = \alpha * \text{Loss}(Y, \hat{Y}) + \beta * \text{Loss}(\hat{Y}) + \lambda * R(f) \quad (22)$$

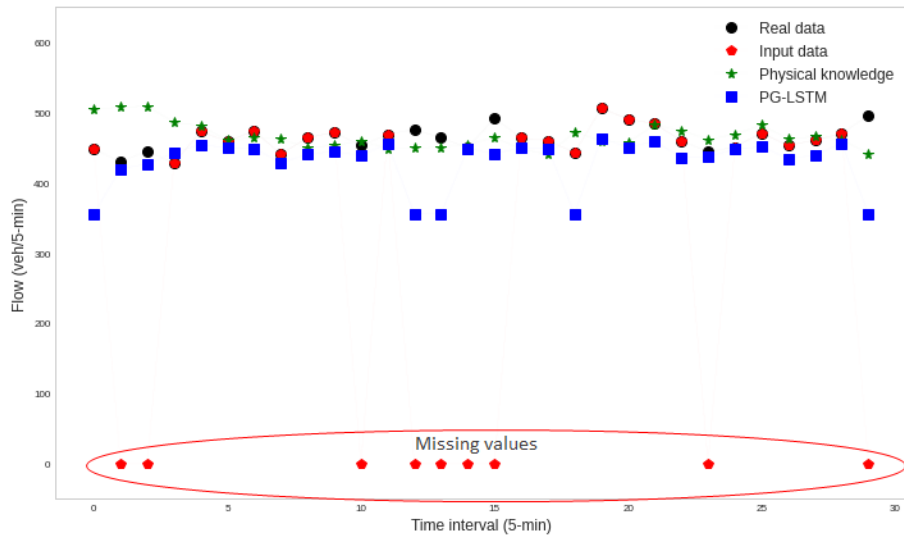
25 where the training $\text{Loss}(Y, \hat{Y})$ measures the empirical error (e.g., MSE) between labels Y and
 26 predictions \hat{Y} ; $\text{Loss}(\hat{Y})$ denotes the physics inconsistency (also termed as physics-based loss)
 27 that aims to keep the consistency between predictions and physics knowledge. $R(f)$ denotes the
 28 model structural error that measures the model complexity; α , β , and λ represent the trade-off
 29 hyper-parameters of empirical error, structural error, and physics inconsistency, respectively.



30
 31 **Figure 6 The comparison between LSTM prediction, input data and real observed data**



1
2 **Figure 7 Illustration of the training process of PG-LSTM network with missing data**



3
4 **Figure 8 The comparison between PG-LSTM prediction, input data, physics knowledge,**
5 **and real observed data**
6

7 Based on previous successful experience with PGML for TSE problems, the physics-
8 guided loss function is developed for traffic flow forecasting tasks, especially for the case of
9 training data with missing values. The detailed description of how to establish the physics-based
10 loss for the traffic flow forecasting problem is as follows.

11 Physics value for all target segments could be computed by **Equation 23**. To ensure
12 those physics values comply with traffic flow physics knowledge, Y_{phy} , this research first
13 calculates the difference between physics knowledge and model predictions on the training set
14 for traffic flow forecasting problem:

$$15 \quad \Delta = |Y_{phy} - \hat{Y}| \quad (23)$$

16 A positive value of Δ can be viewed as a violation of physics laws, which can be further
17 evaluated as the square of Δ . Hence, the mean of physics violations across all observations can
18 be considered as an additional term in the physics-based loss function of the training process:

$$19 \quad PHY.Loss(\hat{Y}) = \frac{1}{N} \sum_{i=1}^n \Delta^2 \quad (24)$$

1 Adding the term to the original LSTM loss function, the physics-guided loss function of
 2 training and testing process can be expressed as follows:

$$3 \quad \text{PGLoss} = \alpha * \frac{1}{N} \sum_{i=1}^n (Y - \hat{Y})^2 + \beta * \frac{1}{N} \sum_{i=1}^n |Y_{phy} - \hat{Y}|^2 + \lambda * R(f) \quad (25)$$

4 To the best of my knowledge, this is the first time to incorporate physics knowledge into
 5 sequential models to improve the traffic flow forecasting performance. The physics knowledge
 6 could provide valuable information to overcome the limitation of missing data, which can steer
 7 the whole training process in the right direction close to the ground truth.

8 4. EXPERIMENTAL STUDY

9 *Experimental Setup*

10 To evaluate the effectiveness of the proposed PG-LSTM network, Performance
 11 Measurement System (PeMS) data collected from interstate freeway I-15 is utilized for
 12 experiments. It is the most extensively used data type in traffic flow forecasting tasks (9, 15, 16,
 13 47). The studied freeway stretch is illustrated in Figure 9. In the studied scenario, the separate
 14 freeway segment in I-15 has 4 detectors (e.g., 407, 409, 411, and 412). 407 marked as blue,
 15 which is located on upstream normal freeway segment and used for constructing physics
 16 knowledge for downstream spot. 409, 411, and 412 are marked in yellow and are located on the
 17 downstream off-ramp, on-ramp, and normal freeway segments respectively for testing the
 18 proposed PG-LSTM network. The data was collected from January 4, 2021, to January 8, 2021.
 19 Since the data is collected every 5 minutes, there are 288 observations per detector per day.



-N: Normal Segment; -Off: Off-ramp segment; -On: On-ramp segment

20 **Figure 9 The stretch of the studied freeway segment**

21 To evaluate the performance of the proposed PG-LSTM network with flawed data, the
 22 testing cases are constructed regarding the basic TSP problem with the scarceness with random
 23 missing data. We investigate the biased data scenarios by artificially removing the traffic flow in
 24 the training data to mimic the common device malfunctioning situations. The robustness analysis
 25 is conducted to show the capability of dealing with the unpredictable missing inputs in the
 26 training phase. Theoretically, the proposed PG-LSTM is more robust than LSTM on a noisy
 27 dataset. To justify this feature, a certain portion of the training dataset is removed. The test set is
 28 the complete dataset. In the robustness study, 10%, 30%, and 50% of the training data are
 29 removed, and the testing data keep unchanged. In the setup of the experiments, the prefixed
 30 parameters of the proposed method are summarized in Table 1.
 31

Table 1 The prefixed parameters of LSTM and PG-LSTM models

Parameter	Value
Training set size	864
Testing set size	576
The learning rate	0.001
Optimizer	Adam
Activation fuction	Relu
Hidden layer	3
Epoch	1000

Performance Index

To quantify the accuracy of predictions, this research selects three common prediction evaluation indexes RMSE, MAPE, and MAE of each dimension are used as the performance metric, which are defined in Equations 26-28. The smaller the value of these three evaluation methods represent the smaller the prediction error.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2} \quad (26)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i^* - y_i|}{y_i} * 100\% \quad (27)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^*| \quad (28)$$

where, y_i is the observed traffic speed and flow and y_i^* is the estimated traffic speed and flow.

Results Analysis

Table 2 summarizes the traffic flow prediction results from PG-LSTM and other benchmark methods (e.g., SVM, RF, ANN, and LSTM) of the spot of the normal segment with missing rates 0.1, 0.3, and 0.5). For the missing rate of 0.1, most comparable ML models could achieve relatively accurate prediction. For example, the LSTM obtained the best performance, which can yield a 48.14 vehicles/5-minutes of RMSE, an 11.28% of MAPE, and 36.06 vehicles/5-minutes of MAE. In comparison to these benchmark methods, the PG-LSTM generates a 40.03 vehicles/5-minutes of RMSE, a 9.49% of MAPE, and 29.00 vehicles/5-minutes of MAE. Further comparison between comparable methods and PG-LSTM reveals that PG-LSTM can improve the traffic prediction accuracy under the missing rate of 0.1. For the missing rata of 0.3, the prediction performance of all comparable ML models is significantly dropped, especially for LSTM. However, the PG-LSTM still could obtain accurate predictions, which yields a 41.12 vehicles/5-minutes of RMSE, a 11.79% of MAPE, and 32.10 vehicles/5-minutes of MAE. For the missing rate of 0.5, the prediction results obtained by all comparable methods are far from the ground truth. For example, the LSTM yields a 187.03 vehicles/5-minutes of RMSE, an 83.29% of MAPE, and 162.10 vehicles/5-minutes of MAE. However, PG-LSTM can still achieve accurate prediction results (e.g., a 44.80 vehicles/5-minutes of RMSE, a 12.03% of MAPE, and 33.18 vehicles/5-minutes of MAE). All results proved that the proposed PG-LSTM

1 could perform well for traffic flow forecasting with missing data, while the performance of other
 2 models has been downgraded significantly. Figure 10 (a)-(c) show the comparison of prediction
 3 results of SVM, RF, ANN, LSTM and PG-LSTM with ground truth on the spot of the normal
 4 segment. It can be clearly seen that the line of PG-LSTM could better fit the ground truth under
 5 different missing rate. The patterns of lines of other comparable methods are different from the
 6 ground truth, especially for the cases with a large missing rate. This demonstrates the excellent
 7 prediction performance of PG-LSTM under different missing rates.

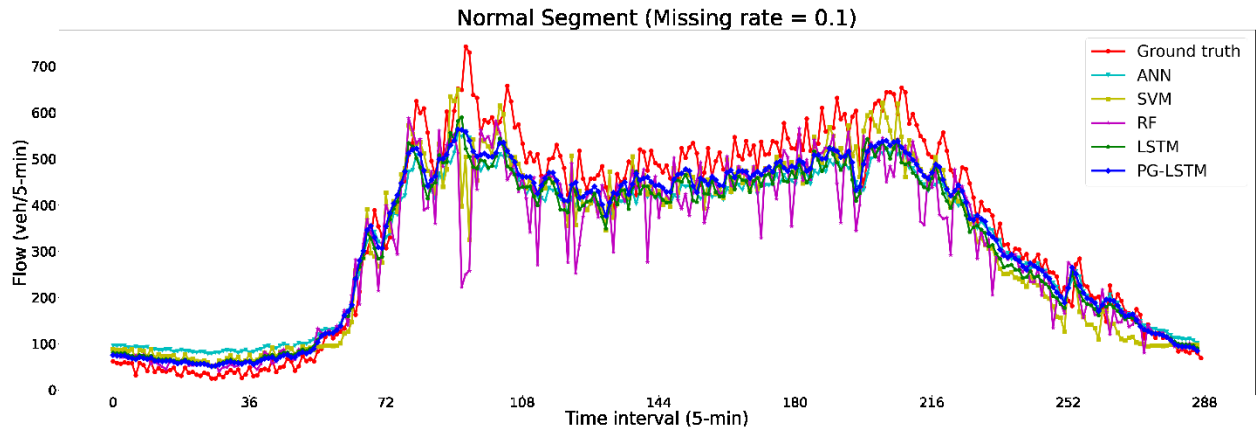
8 To further confirm this finding, the prediction results obtained by LSTM and PG-LSTM
 9 are compared to the observed data. In Figure 11 (a)-(f), the prediction results can be seen as
 10 fitting the ground truth well if the coefficient of the trend line is close to one and the intercept is
 11 close to zero. For the missing rate of 0.1, the coefficient is 0.87 and the intercept is 12.21 for
 12 LSTM, and the coefficient is 0.91 and the intercept is 11.33 for PG-LSTM. For the missing rate
 13 of 0.3, the coefficient is 0.74 and the intercept is 10.83 for LSTM, and the coefficient is 0.94 and
 14 the intercept is 5.70 for PG-LSTM. For the missing rate of 0.5, the coefficient is 0.53 and the
 15 intercept is 3.94 for LSTM, and the coefficient is 0.90 and the intercept is 13.40 for PG-LSTM.
 16 These results indicate that the prediction performance of LSTM is significantly downgraded
 17 when the missing rate increases. At the same time, the PG-LSTM still could achieve stable
 18 prediction performance under different missing rates.

19 To further illustrate the variation of MAPEs in different models, Figure 12 shows the
 20 violin plots of MAPEs on the test set by LSTM and PG-LSTM models with different missing
 21 rates. In each "violin", its margin shows the Gaussian distribution of the dataset, and a box plot is
 22 drawn inside. In Figure 12, it is noted that the MAPE values of LSTM is dramatically upgraded
 23 with the missing rate increasing. In contrast, the PG-LSTM could always generate lower MAPEs
 24 value under different missing rates. All results proved that PG-LSTM could perform well for
 25 traffic flow forecasting.

26 **Table 2 Prediction results of spots on normal segment with missing data**

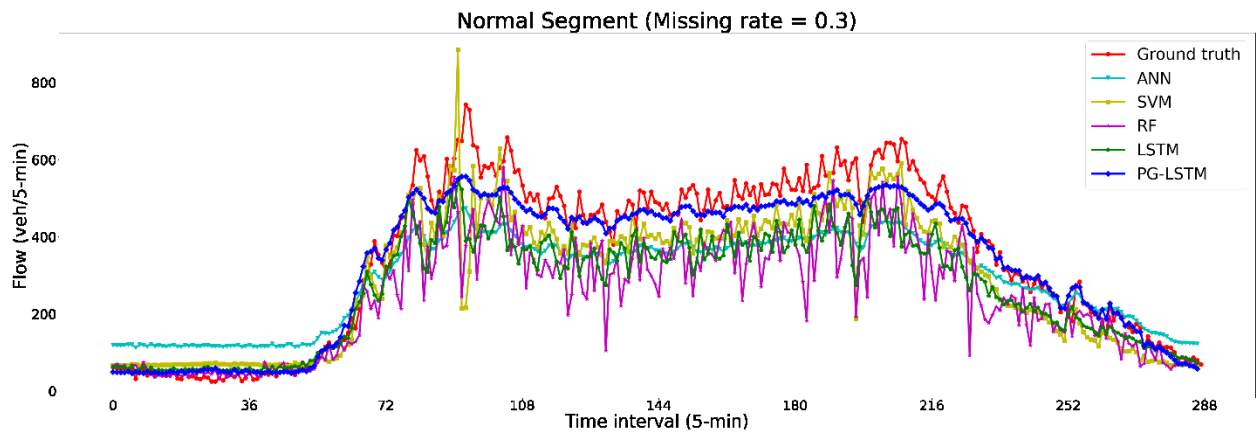
Method	Missing rate	RMSE	MAPE	MAE
SVM	0.1	50.88	15.52	36.62
RF	0.1	74.88	17.61	48.72
ANN	0.1	61.87	15.27	48.26
LSTM	0.1	48.14	11.28	36.06
PG-LSTM	0.1	40.03	9.49	29.00
SVM	0.3	77.38	25.25	58.55
RF	0.3	122.57	44.90	95.72
ANN	0.3	98.34	27.86	82.01
LSTM	0.3	98.19	28.07	80.20
PG-LSTM	0.3	41.12	11.79	32.10
SVM	0.5	237.94	190.57	172.91
RF	0.5	193.32	100.31	160.25
ANN	0.5	173.98	57.65	146.34
LSTM	0.5	187.03	83.29	162.10
PG-LSTM	0.5	44.80	12.03	33.18

27



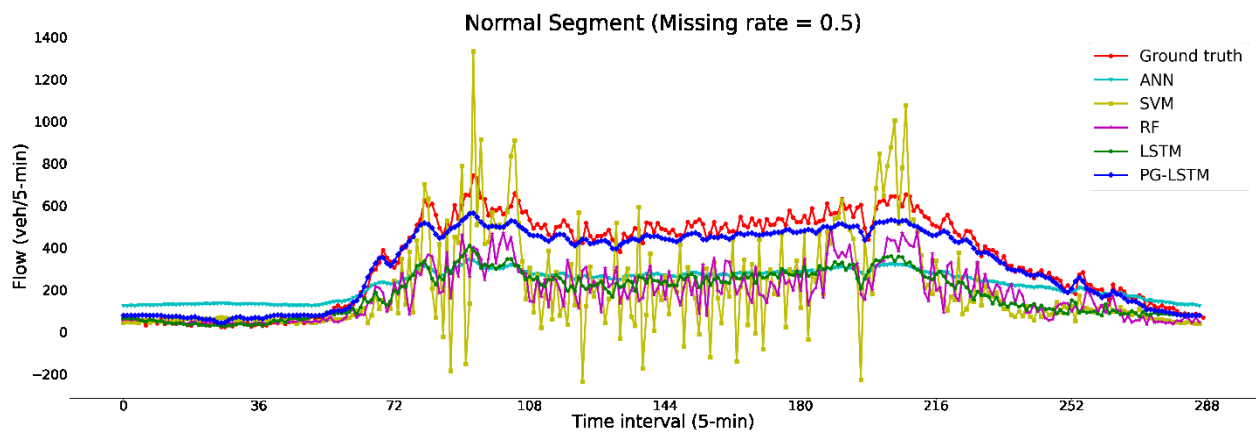
1
2

(a)



3
4

(b)



5
6
7

(c)

Figure 10 Prediction estimates v.s. ground truth on normal spot

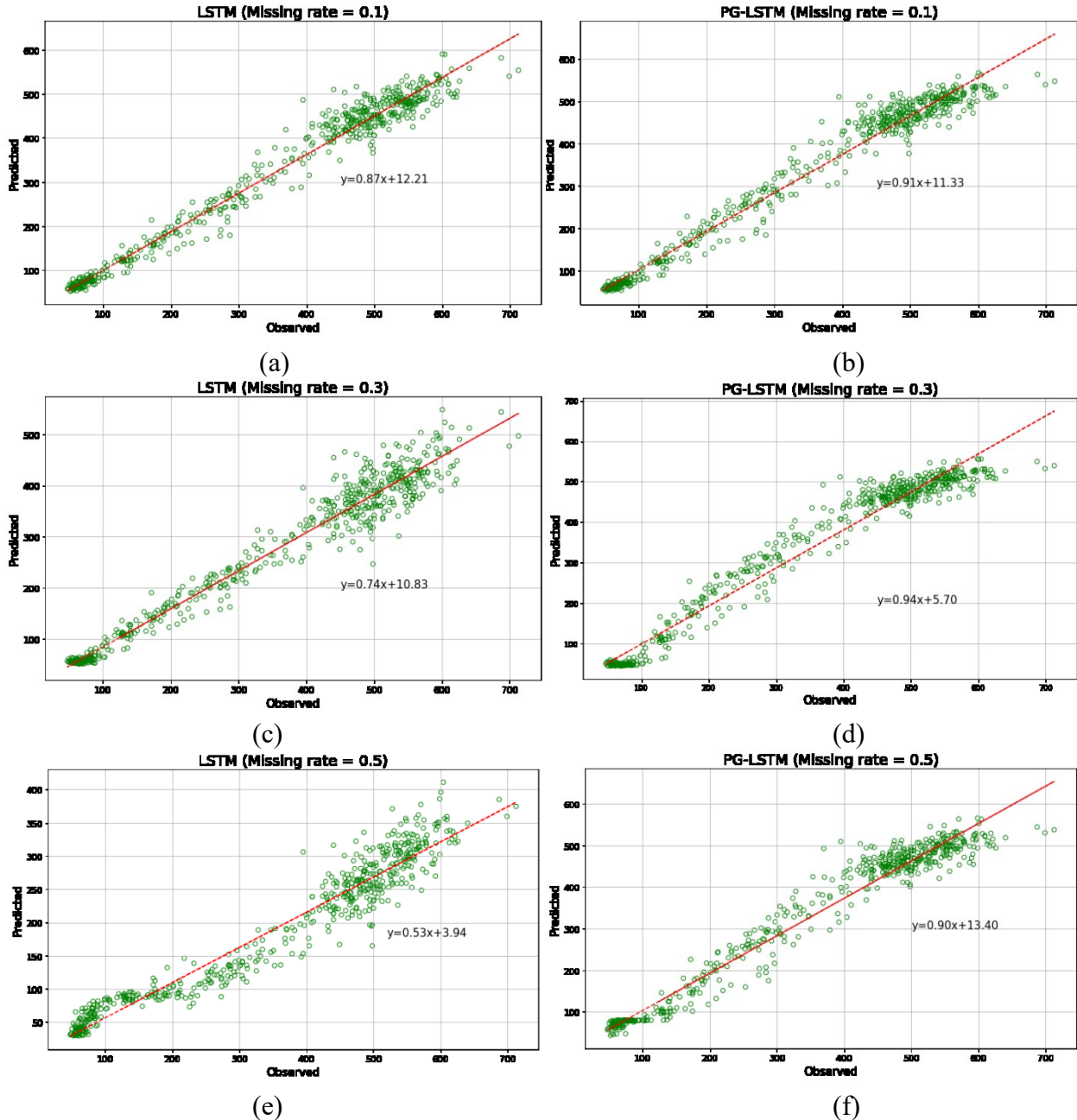
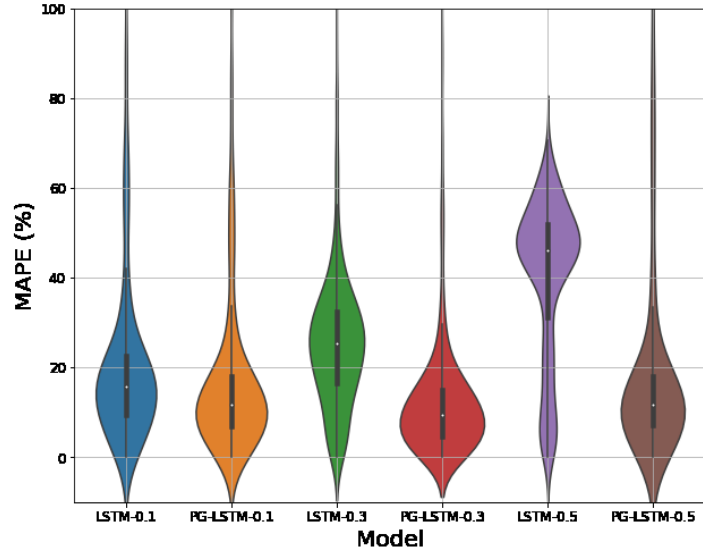


Figure 11 Predicted flow by LSTM and PG-LSTM v.s. ground truth on normal spot

The proposed PG-LSTM and comparable methods were also implemented on spots on the on-ramp and off-ramp segments. The prediction results are shown in Tables 3-4. Compared with Table 2, the conclusion can be reached that similar results pattern were found on spots of these three types of segments. The performance of comparable methods is significantly downgraded when the missing rate increases on spots of the normal, on-ramp, and off-ramp segments. At the same time, the PG-LSTM still could achieve stable prediction performance under different missing rate. Hence, the proposed PG-LSTM network could obtain accurate traffic flow prediction under different data missing rates.



1
2 **Figure 12 MAPE of predicted flow with different missing rates on normal spot**

3 **Table 3 Prediction results of on-ramp spot segment with missing data**

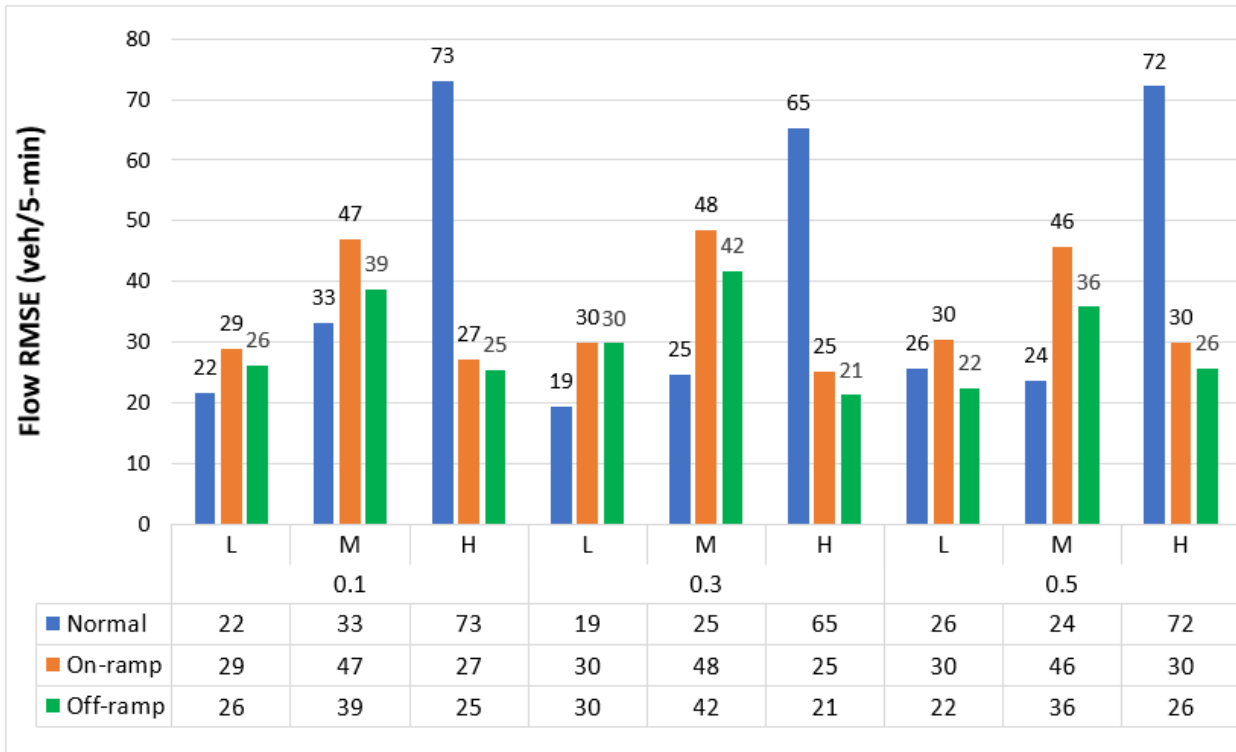
Method	Missing rate	RMSE	MAPE	MAE
SVM	0.1	47.77	15.31	34.61
RF	0.1	65.38	17.81	44.76
ANN	0.1	49.02	14.16	37.98
LSTM	0.1	47.34	12.03	35.38
PG-LSTM	0.1	36.84	9.54	26.67
SVM	0.3	82.47	32.94	60.81
RF	0.3	130.28	48.94	97.27
ANN	0.3	112.52	34.03	94.57
LSTM	0.3	105.62	31.30	86.32
PG-LSTM	0.3	41.72	11.50	30.88
SVM	0.5	231.93	400.08	168.07
RF	0.5	195.07	107.10	157.31
ANN	0.5	198.36	75.46	167.68
LSTM	0.5	193.47	97.72	167.24
PG-LSTM	0.5	42.67	15.22	31.93

4 To further verify the robustness of the proposed PG-LSTM, Figure 13 illustrates the
 5 prediction error of PG-LSTM with different missing rates under different traffic dynamic orders.
 6 The traffic conditions are divided into three categories according to traffic volume: low traffic
 7 volume (0-200 vehicles/5-minutes), medium traffic volume (200-450 vehicles/5-minutes), and
 8 high traffic volume (>450 vehicles/5-minutes). As shown in the figure, the performance of PG-
 9 LSTM is almost the same on normal, on-ramp, and off-ramp spots with different missing rates
 10 under low traffic volume. For medium traffic volume, PG-LSTM performs better on the normal
 11 spot than on on-ramp and off-ramp spots with different missing rates. PG-LSTM obtained the
 12 largest error on normal spots with different missing rates under high traffic volume, this may be
 13 caused by the traffic volume on the normal spot being higher than on-ramp and off-ramp spots.

1

Table 4 Prediction results of off-ramp spot with missing data

Method	Missing rate	RMSE	MAPE	MAE
SVM	0.1	49.18	16.39	35.84
RF	0.1	66.54	17.11	44.09
ANN	0.1	55.18	15.29	42.73
LSTM	0.1	48.25	11.92	36.26
PG-LSTM	0.1	35.68	9.36	26.06
SVM	0.3	80.99	26.10	58.38
RF	0.3	123.47	45.44	94.06
ANN	0.3	109.66	32.12	90.46
LSTM	0.3	108.82	37.50	91.72
PG-LSTM	0.3	36.98	10.89	27.81
SVM	0.5	259.72	264.24	183.30
RF	0.5	184.06	98.95	152.30
ANN	0.5	203.70	78.53	173.36
LSTM	0.5	195.86	90.8	167.11
PG-LSTM	0.5	37.73	11.52	28.07



L-low traffic volume (0-200 veh/5-min); M-Medium traffic volume (200-450 veh/5-min); H-High traffic volume (>450 veh/5-min).

Figure 13 Prediction performance of PG-LSTM with different missing rates in different traffic volume and different spots

2
3
4
5

1 **5. CONCLUSIONS**

2 Accurate traffic flow prediction is important for the successful operation of ITS on freeways.
3 However, traffic detectors may provide incomplete and incorrect traffic information, which is a
4 major constraint for existing methods to get accurate predictions. To tackle this issue, this
5 chapter introduces an advanced PG-LSTM network for traffic flow forecasting. The PG-LSTM
6 framework could incorporate traffic physics knowledge into loss functions to help LSTM models
7 capture generalizable dynamic patterns that are consistent with established traffic physics
8 knowledge. More specifically, the application of physics-guided loss functions in the learning
9 objective of our PG-LSTM framework ensures that the model predictions will not only show
10 lower errors on the training set but also have scientific consistency with the known physics on
11 the unlabeled set. This attribute makes the PG-LSTM network could overcome the data flow
12 limitation. To test the effectiveness of the proposed PG-LSTM network to cope with the
13 problems of freeway traffic flow prediction with flawed data, this study conducted empirical
14 studies on a real-world dataset collected from a stretch of interstate freeway I-15 in Utah.
15 Experimental study results indicate that the proposed PG-LSTM network has been proved to
16 perform better than the previous ML methods, including SVM, RF, ANN, and LSTM, especially
17 in terms of prediction accuracy. The proposed PG-LSTM network combined machine learning
18 models with a traffic physics model, potentially leading to a revolution in ITS development.

19 **ACKNOWLEDGMENT**

20 This research is supported by the project “CMMI #2047268 CAREER: Physics Regularized
21 Machine Learning Theory: Modeling Stochastic Traffic Flow Patterns for Smart Mobility
22 Systems” funded by the National Science Foundation (NSF).

23 **AUTHOR CONTRIBUTIONS**

24 The authors confirm contribution to the paper as follows: study conception and design: Z. Zhang,
25 X. Yang; data collection: Z. Zhang; analysis and interpretation of results: Z. Zhang, Q. Wang,
26 and X. Yang; draft manuscript preparation: Z. Zhang, Q. Wang, H. Yang, and X. Yang. All
27 authors reviewed the results and approved the final version of the manuscript.

1 **REFERENCE**

- 2 1. Lv, Y., Y. Duan, W. Kang, Z. Li, and F. Y. Wang. Traffic Flow Prediction with Big Data:
3 A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*,
4 Vol. 16, No. 2, 2015, pp. 865–873. <https://doi.org/10.1109/TITS.2014.2345663>.
- 5 2. Ma, X., Z. Tao, Y. Wang, H. Yu, and Y. Wang. Long Short-Term Memory Neural
6 Network for Traffic Speed Prediction Using Remote Microwave Sensor Data.
7 *Transportation Research Part C: Emerging Technologies*, Vol. 54, 2015, pp. 187–197.
8 <https://doi.org/10.1016/j.trc.2015.03.014>.
- 9 3. Tian, Y., K. Zhang, J. Li, X. Lin, and B. Yang. LSTM-Based Traffic Flow Prediction with
10 Missing Data. *Neurocomputing*, Vol. 318, 2018, pp. 297–305.
11 <https://doi.org/10.1016/j.neucom.2018.08.067>.
- 12 4. Zou, Y., X. Zhu, Y. Zhang, and X. Zeng. A Space-Time Diurnal Method for Short-Term
13 Freeway Travel Time Prediction. *Transportation Research Part C: Emerging*
14 *Technologies*, Vol. 43, 2014, pp. 33–49. <https://doi.org/10.1016/j.trc.2013.10.007>.
- 15 5. Zhang, Y., and H. Ge. Freeway Travel Time Prediction Using Takagi-Sugeno-Kang Fuzzy
16 Neural Network. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 28, No. 8,
17 2013, pp. 594–603. <https://doi.org/10.1111/mice.12014>.
- 18 6. Liu, Y., J. J. Q. Yu, J. Kang, D. Niyato, and S. Zhang. Privacy-Preserving Traffic Flow
19 Prediction: A Federated Learning Approach. *IEEE Internet of Things Journal*, Vol. 7, No.
20 8, 2020, pp. 7751–7763. <https://doi.org/10.1109/JIOT.2020.2991401>.
- 21 7. Yang, B., S. Sun, J. Li, X. Lin, and Y. Tian. Traffic Flow Prediction Using LSTM with
22 Feature Enhancement. *Neurocomputing*, Vol. 332, 2019, pp. 320–327.
23 <https://doi.org/10.1016/j.neucom.2018.12.016>.
- 24 8. Tian, Y., K. Zhang, J. Li, X. Lin, and B. Yang. Neurocomputing LSTM-Based Traffic
25 Flow Prediction with Missing Data. Vol. 318, 2018, pp. 297–305.
26 <https://doi.org/10.1016/j.neucom.2018.08.067>.
- 27 9. Fu, R., Z. Zhang, and L. Li. Using LSTM and GRU Neural Network Methods for Traffic
28 Flow Prediction. In *2016 31st Youth Academic Annual Conference of Chinese Association*
29 *of Automation (YAC)*. *IEEE*, 2016, pp. 324–328.
- 30 10. Luo, X., D. Li, Y. Yang, and S. Zhang. Spatiotemporal Traffic Flow Prediction with KNN
31 and LSTM. *Journal of Advanced Transportation*, Vol. 2019, 2019.
32 <https://doi.org/10.1155/2019/4145353>.
- 33 11. Mackenzie, J., J. F. Roddick, and R. Zito. An Evaluation of HTM and LSTM for Short-
34 Term Arterial Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation*
35 *Systems*, Vol. 20, No. 5, 2019, pp. 1847–1857.
36 <https://doi.org/10.1109/TITS.2018.2843349>.
- 37 12. Kang, D., Y. Lv, and Y. Chen. Short-Term Traffic Flow Prediction with LSTM Recurrent
38 Neural Network. *IEEE 20th International Conference on Intelligent Transportation*

- 1 *Systems*, 2017, pp. 1–6.
- 2 13. Do, L. N. N., H. L. Vu, B. Q. Vo, Z. Liu, and D. Phung. An Effective Spatial-Temporal
3 Attention Based Neural Network for Traffic Flow Prediction. *Transportation Research*
4 *Part C: Emerging Technologies*, Vol. 108, No. August, 2019, pp. 12–28.
5 <https://doi.org/10.1016/j.trc.2019.09.008>.
- 6 14. Chen, X., X. Xie, and D. Teng. Short-Term Traffic Flow Prediction Based on ConvLSTM
7 Model. *2020 IEEE 5th Information Technology and Mechatronics Engineering*
8 *Conference (ITOEC)*, 2020, pp. 846–850. [https://doi.org/10.3963/j.issn.1671-](https://doi.org/10.3963/j.issn.1671-4431.2020.08.014)
9 [4431.2020.08.014](https://doi.org/10.3963/j.issn.1671-4431.2020.08.014).
- 10 15. Yu, J., H. Wei, H. Guo, and Y. Cai. Urban Traffic State Prediction Based on SA-LSTM.
11 *IOP Conference Series: Earth and Environmental Science*, Vol. 783, No. 1, 2021.
12 <https://doi.org/10.1088/1755-1315/783/1/012153>.
- 13 16. Qu, Z., H. Li, Z. Li, and T. Zhong. Short-Term Traffic Flow Forecasting Method with M-
14 B-LSTM Hybrid Network. *IEEE Transactions on Intelligent Transportation Systems*, Vol.
15 23, No. 1, 2022, pp. 225–235. <https://doi.org/10.1109/TITS.2020.3009725>.
- 16 17. Richards, P. I. . Shock Waves on the Highway. *INFORMS*, Vol. 4, No. 1, 1956, pp. 42–51.
- 17 18. Lighthill, M. . J. ., and G. . B. . Whitham. On Kinematic Waves . II . A Theory of Traffic
18 Flow on Long Crowded Roads. *Proceedings of the Royal Society of London. Series A ,*
19 *Mathematical and Physical Science*, Vol. 229, No. 1178, 1955, pp. 317–345.
- 20 19. Payne, H. J. Models of Freeway Traffic and Control. *Mathematical Models of Public*
21 *Systems*. Simulation Councils. 1971.
- 22 20. Whitham, G. B. *Linear and Nonlinear Waves*. Modern Book Incorporated, 1975.
- 23 21. Aw, A., and M. Rascle. Resurrection of “Second Order” Models of Traffic Flow. Vol. 60,
24 No. 3, 2000, pp. 916–938.
- 25 22. Zhang, H. M. A Non-Equilibrium Traffic Model Devoid of Gas-like Behavior.
26 *Transportation Research Part B*, Vol. 36, 2002, pp. 275–290.
- 27 23. Papageorgiou, M., J. M. Blosseville, and H. Hadj-Salem. Macroscopic Modelling of Tra c
28 Ow on the Boulevard Peripherique in Paris. *Transportation Research Part B:*
29 *Methodological*, Vol. 23, 1989, pp. 29–47.
- 30 24. Zhang, Z., Y. Yuan, and X. T. Yang. A Hybrid Machine Learning Approach for Freeway
31 Traffic Speed Estimation. 2020. <https://doi.org/10.1177/0361198120935875>.
- 32 25. Yuan, Y., Z. Zhang, X. T. Yang, and S. Zhe. Macroscopic Traffic Flow Modeling with
33 Physics Regularized Gaussian Process: A New Insight into Machine Learning
34 Applications in Transportation. *Transportation Research Part B: Methodological*, Vol.
35 146, 2021, pp. 88–110. <https://doi.org/10.1016/j.trb.2021.02.007>.
- 36 26. Seo, T., A. M. Bayen, T. Kusakabe, and Y. Asakura. Traffic State Estimation on Highway:

- 1 A Comprehensive Survey. *Annual Reviews in Control*, Vol. 43, 2017, pp. 128–151.
2 <https://doi.org/10.1016/j.arcontrol.2017.03.005>.
- 3 27. Van Der Voort, M., M. Dougherty, and S. Watson. Combining Kohonen Maps with
4 ARIMA Time Series Models to Forecast Traffic Flow. *Transportation Research Part C:
5 Emerging Technologies*, Vol. 4, No. 5, 1996, pp. 307–318. [https://doi.org/10.1016/S0968-
6 090X\(97\)82903-8](https://doi.org/10.1016/S0968-090X(97)82903-8).
- 7 28. Williams, B. M., and L. A. Hoel. Modeling and Forecasting Vehicular Traffic Flow as a
8 Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *Journal of
9 Transportation Engineering*, Vol. 129, No. 6, 2003, pp. 664–672.
10 [https://doi.org/10.1061/\(ASCE\)0733-947X\(2003\)129:6\(664\)](https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664)).
- 11 29. Shi, G., J. Guo, W. Huang, and B. M. Williams. Modeling Seasonal Heteroscedasticity in
12 Vehicular Traffic Condition Series Using a Seasonal Adjustment Approach. *Journal of
13 Transportation Engineering*, Vol. 140, No. 5, 2014, pp. 1–11.
14 [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000656](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000656).
- 15 30. Ojeda, L. L., A. Y. Kibangou, and C. C. De Wit. Adaptive Kalman Filtering for Multi-
16 Step Ahead Traffic Flow Prediction. *Proceedings of the American Control Conference*,
17 2013, pp. 4724–4729. <https://doi.org/10.1109/acc.2013.6580568>.
- 18 31. Guo, J., W. Huang, and B. M. Williams. Adaptive Kalman Filter Approach for Stochastic
19 Short-Term Traffic Flow Rate Prediction and Uncertainty Quantification. *Transportation
20 Research Part C: Emerging Technologies*, Vol. 43, 2014, pp. 50–64.
21 <https://doi.org/10.1016/j.trc.2014.02.006>.
- 22 32. Smith, B., and M. J. Demetsky. SHORT-TERM TRAFFIC FLOW PREDICTION
23 MODELS - A COMPARISON OF NEURAL NETWORK AND NONPARAMETRIC
24 REGRESSION APPROACHES. *Sensors (Peterborough, NH)*, 1994, pp. 1–4.
- 25 33. Chen, D. Research on Traffic Flow Prediction Method Based on BP Neural Network. *IEEE
26 Transactions on Industrial Informatics*, Vol. 13, No. 4, 2017, pp. 2000–2008.
- 27 34. Dell’acqua, P., F. Bellotti, R. Berta, and A. De Gloria. Time-Aware Multivariate Nearest
28 Neighbor Regression Methods for Traffic Flow Prediction. *IEEE Transactions on
29 Intelligent Transportation Systems*, Vol. 16, No. 6, 2015, pp. 3393–3402.
30 <https://doi.org/10.1109/TITS.2015.2453116>.
- 31 35. Cai, P., Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun. A Spatiotemporal Correlative K-
32 Nearest Neighbor Model for Short-Term Traffic Multistep Forecasting. *Transportation
33 Research Part C: Emerging Technologies*, Vol. 62, 2016, pp. 21–34.
34 <https://doi.org/10.1016/j.trc.2015.11.002>.
- 35 36. Wu, S., Z. Yang, X. Zhu, and B. Yu. Improved K-Nn for Short-Term Traffic Forecasting
36 Using Temporal and Spatial Information. *Journal of Transportation Engineering*, Vol.
37 140, No. 7, 2014, pp. 1–9. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000672](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000672).
- 38 37. Sun, B., W. Cheng, P. Goswami, and G. Bai. Short-Term Traffic Forecasting Using Self-

- 1 Adjusting k-Nearest Neighbours. *IET Intelligent Transport Systems*, Vol. 12, No. 1, 2018,
2 pp. 41–48. <https://doi.org/10.1049/iet-its.2016.0263>.
- 3 38. Castro-Neto, M., Y. S. Jeong, M. K. Jeong, and L. D. Han. Online-SVR for Short-Term
4 Traffic Flow Prediction under Typical and Atypical Traffic Conditions. *Expert Systems*
5 *with Applications*, Vol. 36, No. 3, 2009, pp. 6164–6173.
6 <https://doi.org/10.1016/j.eswa.2008.07.069>.
- 7 39. Wang, J., W. Deng, and Y. Guo. New Bayesian Combination Method for Short-Term
8 Traffic Flow Forecasting. *Transportation Research Part C: Emerging Technologies*, Vol.
9 43, 2014, pp. 79–94. <https://doi.org/10.1016/j.trc.2014.02.005>.
- 10 40. Xu, Y., Q. J. Kong, R. Klette, and Y. Liu. Accurate and Interpretable Bayesian MARS for
11 Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol.
12 15, No. 6, 2014, pp. 2457–2469. <https://doi.org/10.1109/TITS.2014.2315794>.
- 13 41. Park, D., L. R. Rilett, and G. Han. Spectral Basis Networks for Real-Time Travel Time
14 Forecasting. *Journal of Transportation Engineering*, Vol. 3, No. December, 1999, pp.
15 515–523.
- 16 42. Van Lint, J. W. C., S. P. Hooqendoorn, and H. J. Van Zuvlen. Freeway Travel Time
17 Prediction with State-Space Neural Networks Modeling State-Space Dynamics with
18 Recurrent Neural Networks. *Transportation Research Record*, No. 1811, 2002, pp. 30–39.
19 <https://doi.org/10.3141/1811-04>.
- 20 43. Lingras, P., S. Sharma, and M. Zhong. Prediction of Recreational Travel Using
21 Genetically Designed Regression and Time-Delay Neural Network Models.
22 *Transportation Research Record*, No. 1805, 2002, pp. 16–24.
23 <https://doi.org/10.3141/1805-03>.
- 24 44. Gers, F. A., J. Schmidhuber, and F. Cummins. *Lerning to Forget: Continual Prediction*
25 *with LSTM*. 1999.
- 26 45. Hochreiter, S., and J. Urgan Schmidhuber. Long Shortterm Memory. *Neural Computation*,
27 Vol. 9, No. 8, 1997, p. 17351780.
- 28 46. Papageorgiou, M., J.-M. Blosseville, and H. Haj-Salem. Modeling and Real-Time Control
29 of Traffic Flow on the Southern Part of Boulevard Peripherique in Paris—Part I: Modeling.
30 *Transportation Research A*, Vol. 24, 1990, pp. 345–359.
- 31 47. Wu, Y., H. Tan, L. Qin, B. Ran, and Z. Jiang. A Hybrid Deep Learning Based Traffic Flow
32 Prediction Method and Its Understanding. Vol. 90, No. April 2017, 2018, pp. 166–180.
33 <https://doi.org/10.1016/j.trc.2018.03.001>.

34