sustainability

MDPI

Article

Impact of High Resolution Radar-Obtained Weather Data on Spatio-Temporal Prediction of Freeway Speed

Mustafa Attallah [1], Jalil Kianfar [1,*] and Yadong Wang [2]

[1] Department of Civil, Computer and Electrical Engineering, Saint Louis University, St. Louis, MO 63103, USA
[2] Electrical and Computer Engineering Department, Southern Illinois University Edwardsville, Edwardsville, IL 62026, USA
* Correspondence: jalil.kianfar@slu.edu

Abstract: Inclement weather and environmental factors impact traffic operations resulting in travel delays and a reduction in travel time reliability. Precipitation is an example of an environmental factor that affects travel conditions, including traffic speed. While Intelligent Transportation Systems services aim to proactively mitigate congestion on roadways, these services are often not sensitive to weather conditions. This paper investigates the application of high-resolution weather data in improving the performance of proactive transportation management models and proposes short-term speed prediction models that fuse real-time high-resolution weather surveillance radar data with traffic stream data to conduct spatial and temporal prediction of the speed of roadway segments. Extreme gradient boosting weather-aware speed prediction models were developed for a 7-km segment of Interstate 270 in St. Louis, MO, USA. The performance of the weather-aware models was compared with the performance of weather-insensitive speed prediction models that did not take precipitation into account. The results indicated that in the majority of instances, the weather-aware models outperformed the weather-insensitive models. The extreme gradient boosting models were compared with the K-nearest neighbors algorithm and feed-forward neural network models. The extreme gradient boosting model consistently outperformed the other two methods. In addition to speed prediction models, van Aerde speed-flow traffic stream models were developed for rain and no-rain conditions to study the impact of precipitation on the traffic stream across the corridor. Results indicated that the impact of precipitation is not identical across the corridor, which was mirrored in the results obtained from weather-aware speed prediction models.

Keywords: Intelligent Transportation Systems; precipitation rate; speed prediction; weather radar

check for updates

Citation: Attallah, M.; Kianfar, J.; Wang, Y. Impact of High Resolution Radar-Obtained Weather Data on Spatio-Temporal Prediction of Freeway Speed. *Sustainability* **2022**, *14*, 14932. https://doi.org/10.3390/su142214932

Academic Editor: Shuai Su

Received: 30 September 2022
Accepted: 2 November 2022
Published: 11 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Proactive traffic management systems are a subset of Intelligent Transportation Systems (ITS) that allow the transportation operators to anticipate short-term traffic conditions (e.g., traffic speed in the next five minutes) and enable them to take preventive actions to mitigate traffic congestion and improve the efficiency of the transportation system. Such actions include proactively harmonizing the traffic through variable speed limits or informing the public through advanced traveler information systems [1–4]. Although great progress has been made in developing ITS subsystems that fuse diverse data sources and combining historical and real-time data to predict the state of transportation networks [5–9], very limited studies have focused on incorporating the impact of weather in proactive transportation management systems [10,11]. Even though every commuter is intuitively aware of the impact of inclement weather on the quality of travel, transportation management systems are often insensitive to weather conditions. Several factors may have contributed to the limited research on the implementation of weather-aware transportation management strategies. First, the primary focus of the researchers and practitioners has been improving traffic management strategies for normal weather conditions because the majority of trips

occur during normal weather conditions. Moreover, inclement weather is less frequent and has not historically gained the attention of transportation operators. Second, the absence or lack of access to real-time weather data may have hindered the development of weather-aware systems. Even though Road Weather Information Systems (RWIS) are used by some transportation agencies to collect environmental data, these systems are installed sparsely on the roadway network and do not have the same density as traffic detectors. Thus, they do not provide real-time and full coverage of the roadway network.

Recently, the weather has been recognized as an important factor that could create challenges to the operation of the transportation system [12,13]. This paper contributes to the very limited body of research on weather-aware transportation management models by proposing a weather-aware model for short-term traffic speed prediction. High-resolution weather surveillance radar data were processed to obtain real-time and localized precipitation rates at each freeway traffic sensor location and are fused with traffic stream data, including flow, speed, and occupancy obtained from roadway detectors. The eXtreme Gradient Boosting (XGBoost) method is used to develop two categories of models: (1) weather-aware models and (2) weather-insensitive models. The results obtained from XGBoost are compared with the K-Nearest Neighbor (KNN) and Neural Network (NNET) methods. Even though the cause-and-effect relationship between traffic congestion and inclement weather is well-known to any commuter, modeling the complex relationship between precipitation and other weather events and traffic congestion is not a trivial task. To better understand the impact of precipitation on traffic flow, van Aerde traffic stream models were developed for no-rain and with-rain conditions. These models also help to explain the impact of precipitation on speed changes across the case study roadway.

The outline of this paper is as follows: The next section reviews the relevant literature on short-term traffic flow and speed prediction. Then information on the case study location is provided. Next, the framework for the development of the short-term speed prediction models is described. Following that, the results and findings obtained from the prediction models and van Aerde models are discussed. Lastly, the conclusion summarizes the findings from this study and offers suggestions for future research.

2. Literature Review

Traffic prediction models are broadly classified as long-range and short-term prediction models. Long-range models consider factors such as land use, population, and socio-economic characteristics to predict travel in future years. These models, which are more commonly known as travel demand models or long-range transportation planning models, are often developed for an entire region rather than one corridor. Input data for these models is costly to collect and do not significantly change on a daily or hourly basis. For the aforementioned reasons, these models cannot be practically used for short-term traffic prediction. To address this shortcoming, machine learning methods are employed for short-term traffic prediction. These short-term traffic prediction models utilize data that can be regularly, frequently, reliably, and consistently collected to predict the state of roadway systems on a short-term basis. For this reason, data that can be automatically collected by detectors are utilized as predictor variables in these models. Even though events such as marches, protests, and sports events could influence traffic, they are not used as a predictor variable because they are difficult to automatically obtain and are hard to quantify and classify to be used in the models. In this section, the state-of-the-art in short-term speed prediction is reviewed with a particular focus on studies that have utilized weather data in the prediction of traffic speed or volume. Long-range transportation planning models are out of the scope of this paper and are not reviewed.

Asif et al. [14] suggested that unsupervised data mining methods, K-means clustering, and self-organizing maps could be used to discover the spatial and temporal trends in traffic patterns of local roads and major networks on 5024 road segments in Singapore. The roadway segments were grouped into four clusters, and the support vector regression (SVR) algorithm was then used to develop a speed prediction model for each cluster. Two

months of traffic data were used to train and test the models and predict speed in 5-min to 60-min prediction horizons. The reported mean absolute percentage error (MAPE) values ranged between 2.69% and 24.58% for 5-min and 60-min prediction horizons.

Bouktif et al. [15] developed weather-sensitive models that forecasted speed for 6-, 24- and 36-h horizons in Manhattan, New York, United States. Three distinct models were developed to predict speed on the basis of: (1) only traffic data, (2) traffic and weather data, and (3) Bayesian-based optimized traffic and weather data. The traffic data were collected by detectors located in uneven locations by the New York City department of transportation. The weather data included hourly readings of temperature, total snow, sun hours, wind speed, wind direction, and precipitation.

Ma et al. [16] proposed Convolutional Neural Networks (CNN) for network speed prediction. CNN models are often used for image processing; however, Ma et al. [16] proposed an approach similar to image processing for predicting the roadway network speed. The traffic data were collected with a 2-min resolution for 37 days in 2015. The results obtained from the CNN model were compared with results obtained from K-Nearest Neighbors, Ordinary Least Squares, Artificial Neural Networks, Random Forests, Recurrent Neural, Stacked Autoencoder, and Long-Short-Term Memory Networks. The model was developed to predict speed in the next 10 and 20 min based on the traffic condition in the previous 30 and 40 min. The mean squared error (MSE) for the 10-min prediction horizon was 22.825 km^2/h^2 and was less than the error reported for the 20-min speed prediction horizon.

Wanli and Laura [17] proposed models to predict speed and volume in 5-min intervals for up to one hour in advance. The model was developed for a network that consisted of 502 links. Loop detector data were aggregated to 5-min intervals for each link and were supplemented with data collected by taxis to infer the traffic conditions between 7 a.m. and 8 p.m. The models were evaluated based on accuracy, where the accuracy was defined as one minus the average absolute relative error of the predicted traffic parameter for all links and for all prediction intervals. The volume prediction accuracy ranged between 0.82 and 0.89, whereas the speed prediction accuracy ranged between 0.82 and 0.941. These results indicated that the accuracy of speed prediction was better than volume prediction.

Wang et al. [18] utilized Long Short-Term Memory Neural Networks (LSTM NN) to develop speed prediction models and compared the results with K-nearest neighbors (KNN), Convolutional Neural Networks (CNN), and Artificial Neural Networks (ANN) models. Traffic data collected over 90 days on a roadway network that consisted of 112 road segments were used to train and validate the models. Three sets of predictor variables were used in this study to predict the next 5-, 15-, and 30-min speeds, namely a 15-min span, 30-min span, and 45-min span. The best MSE values of the LSTM NN were approximately 10.14 km^2/h^2, 14.12 km^2/h^2, and 15.65 km^2/h^2 for each prediction span, respectively. The weather data were not used to predict speed in this study.

Yu et al. [19] proposed the Spatio-Temporal Graph Convolutional Networks (STGCN) method to predict traffic speed in five-minute intervals for a highway in Beijing City, China. Traffic data collected over two months were used to train and test the model. The mean absolute error obtained from the model was 3.78 km/h, the mean average percentage error was 9.11%, and the root-mean-squared was 5.2 km/h. This study did not utilize weather data in the prediction process.

Huang and Ran [20] used multi-layer perceptron (MLP) networks to predict traffic speed with the effect of the weather being considered. Historical data for the Chicago metropolitan area were retrieved from an internet archive with a 5-min resolution for traffic data and a 1-h resolution for weather data. The results obtained from the MLP model were compared with a time-series-based model. It was reported that the MLP model slightly outperformed the time-series model. The mean absolute error for the MLP model was 12.4 km/h.

Leong et al. [21] investigated the impact of including rainfall intensity data on improving the accuracy of low-resolution traffic speed band prediction models. The authors

developed Support Vector Machine (SVM) models using three months of traffic speed data collected from 616 road segments in Singapore. The traffic data were collected in 5 min intervals, and the weather data were collected from radar at intervals ranging between 5 and 10 min. The rainfall intensity data were converted to a scale of 0 to 10 to be used in the model, where 0 was no rain, and 10 indicated very heavy rain. A base model that predicted speed in the next 5-min on the basis of current speed and speed at the previous 5-min interval was developed and compared with the weather-aware model that predicted speed on the basis of speed and rainfall data. The results demonstrated that including rainfall data in the models reduced the speed prediction error on some links, while other links did not show improvements when using rainfall.

Jia et al. [22] combined hourly weather data with high-resolution traffic data to develop traffic flow prediction models on the basis of the LSTM method to investigate the impact of weather data on a traffic variable. The results obtained from the LSTM model were compared with a deep belief network (DBN) model. Six months of data collected in Beijing, China, were used to train and test the model. The results indicated that LSTM outperformed the DBN model, and the prediction errors decreased when weather data was included in the model. Table 1 summarizes the methods, data, and results of all the research mentioned above.

Table 1. Summary of Literature Review.

Study	Method	Weather Sensitive	Temporal Resolution of Weather Data	Case Study	Prediction Horizon	Performance Measure
Asif et al. [14]	Support Vector Regression	No	-	60 days/ Singapore	5-min to 60-min	5.69–24.58% (MAPE) *
Bouktif et al. [15]	Bayesian Optimized XGBoost	Yes	1 h	834 days/ Manhattan	6, 24, and 36 h	2.8, 4.82, 4.26 (RMSE) **
Ma et al. [16]	Convolutional Neural Networks	No	-	37 days/Beijing	10-min to 20-min	22.825 (MSE) ***
Wanli and Laura [17]	Multivariate Spatial-Temporal Autoregressive	No	-	60 days/business district in an urban area	5-min to 60-min	82% to 94.1% Accuracy
Wang et al. [18]	Long Short-Term Memory Neural Networks	No	-	90 days/China	5-min, 15-min, 30-min	10.14, 14.12, 15.65 (RMSE)
Yu et al. [19]	Spatio-Temporal Graph Convolutional Networks	No	-	88 days/Beijing and California	5-min	5.2 (RMSE)
Huang and Ran [20]	Multi-Layer Perceptron (MLP) networks	Yes	1 h	180 days/ Chicago, Seattle, Minneapolis	5-min	12.4 (MAE) ****
Leong et al. [21]	Support Vector Machines	Yes	5-min to 10-min	90 days/ Singapore	5-min	None to significant improvement using rain
Jia et al. [22]	Long Short-Term Memory	Yes	1 h	180 days/Beijing	10-min and 30-min	255.04 and 346.83 (RMSE (veh/h))

* Mean Average Percentage Error; ** Root Mean Squared Error (km/h); *** Mean Squared Error (km^2/h^2); **** Mean Absolute Error (km/h).

In summary, several patterns were observed in the literature review. First, while various prediction methods have been utilized for roadway speed prediction, there are very few studies that have incorporated weather and precipitation data in the speed prediction process. Bouktif et al. [15] and Huang and Ran [20] utilized weather data with 1-h resolution and Leong et al. [21] utilized high-resolution precipitation data to predict speed. However, Leong et al. [21] did not directly use the precipitation intensity to predict the speed; instead,

the intensity of rainfall was classified on a scale of one to ten, and speed prediction was conducted based on the scale of rainfall instead of the direct measurement of rainfall.

Second, previous studies have utilized weather data with low-spatial and low-temporal resolutions to predict traffic stream parameters. Low-spatial resolution indicates that the data obtained from a weather station was generalized to several traffic detectors in the vicinity of the weather station, and it was assumed that weather-related parameters are all the same across the traffic detectors in that vicinity. For example, Bouktif et al. [15] assumed that the data obtained from the weather station is even across all the traffic detectors in a study area of about a 3.27 km radius. Low-temporal resolution indicates that the weather records were aggregated to intervals such as 1 h, and it was assumed that the weather did not change during the aggregation period.

Third, factors such as marches, sports events, and accidents are not considered in short-term traffic prediction. However, as data collection technologies improve in the future, these factors could be included in the short-term prediction models to account for their impact on traffic flow.

To the best of the authors' knowledge, this is the first paper that investigates the application of high-spatial and high-temporal resolution rainfall data for short-term speed prediction. Furthermore, this paper contributes to the limited body of knowledge related to the impact of weather on traffic stream models by developing speed-flow traffic stream models for rain and no-rain conditions. While short-term traffic speed prediction models could be directly implemented into ITS and are practice-ready, traffic stream models have more theoretical value and can provide the foundation for the development of weather-aware traffic simulation models and other engineering design criteria. Thus, this paper contributes to both applied research and fundamental research.

3. Data Description and Study Area

Short-term speed prediction models were developed for an approximately 7.3 km-long section of Interstate 270 Northbound (I-270 NB) in St. Louis, MO, USA, as shown in Figure 1. Traffic data were obtained from six microwave radar traffic detectors. These detectors are referred to as locations A, B, C, and D, in addition to two other traffic detectors, A-U and D-D. Data from locations A-U and D-D were used to inform the models on the upstream and downstream traffic conditions and the speed was not predicted for these two locations. Figure 2 shows an example of a microwave radar traffic detector installed on a freeway location in the study corridor. The radar field of view coverage area includes 50 degrees of the elevation angle, 12 degrees azimuth, 0 m to 76 m range, 24 GHz frequency band, −40° to +74 °C temperature range, and up to 190 km/hr maximum wind speed. The detectors at the corridor were permanently installed on the side of the road by the department of transportation to monitor all the travel lanes in each direction. Flow rate, speed, and occupancy were collected at these locations in five-minute intervals and reported to the Traffic Management Center (TMC). Data collected over 43 days between 6:00 a.m. to 7:00 p.m. were used to train and test the models. A total of 38,346 detector readings collected at six detector locations was fused to create 6391 data points for training and testing the models.

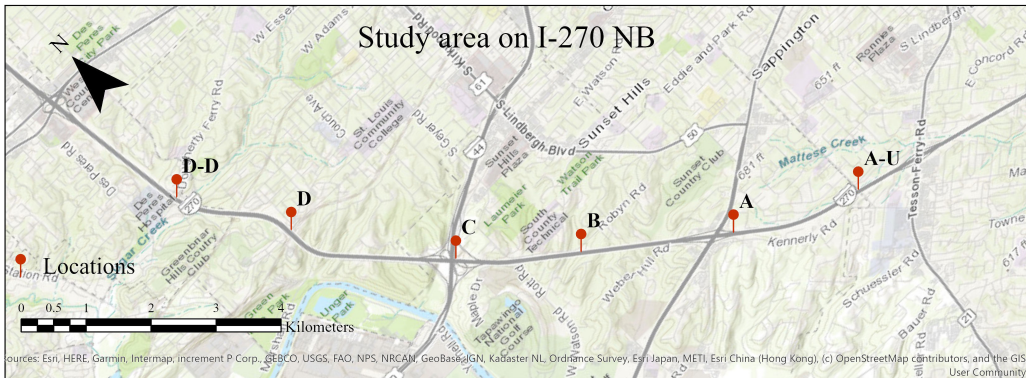

Figure 1. Study area on I-270 NB, St. Louis, MO, USA.

Figure 2. Side-fired microwave radar traffic detector [23].

Traffic flow parameters fluctuated on a daily basis and varied over the 43 days. The average traffic speed during the rain was between 86.32 km/h and 99.21 km/h across the six detectors locations and was between 94.81 km/h and 104.46 km/h during no-rain weather conditions. The average traffic volume in rainy conditions varied between 64.18 veh/5 min/ln and 84.35 veh/5 min/ln across the corridor while traffic volume varied between 70.30 veh/5 min/ln and 85.9 veh/5 min/ln during no-rain conditions. A similar trend was observed for occupancy. During rainy periods, an average occupancy between 6.95% and 9.94% was reported, while the average occupancy was measured to be between 6.14% and 7.90% during no-rain conditions.

In addition to traffic data, precipitation data were considered in the prediction of traffic speed. The precipitation rate was estimated using radar data collected by a Weather Surveillance Radar 1988 Doppler (WSR-88D) located in St. Louis (KLSX). WSR-88Ds utilize a dual-polarization mode, which significantly enhances radar capabilities in quantitative precipitation estimation (QPE), severe weather detection and forecasting, and cloud microphysics studies. In this work, radar data were first processed with a physically based quality control approach [24], which removes those radar echoes contaminated by clutters (such as ground clutter, insects, birds, etc.). The precipitation rate was then estimated using a novel developed radar QPE approach utilizing radar-specific attenuation and a specific differential phase [25]. Compared to conventional radar QPE approaches using reflectivity and/or differential reflectivity, the newly developed approach can significantly enhance QPE in robustness and accuracy [25]. The spatial and temporal resolutions of QPE are 1 km by 1 km, and 5 min, respectively, which enables the models to capture higher precipitation resolutions spatially and temporally. Precipitation was reported on 43 days of the study period. The precipitation rate varied from one location to another from zero to a maximum of 160 mm/h at location A-U, 116 mm/h at location A, 130 mm/h at location B, 150 mm/h at location C, 200 mm/h at location D, and 170 mm/h at location D-D. The averages of the precipitation rate at each location during rainy conditions were 4.06 mm/h at location A-U, 4.68 mm/h at locations A and B, 4.86 mm/h at location C, 5.27 and 5.01 mm/h at locations D and E, respectively.

4. Methods

This section provides an overview of the proposed methodology for short-term speed prediction. First, the structure of the models is discussed. Second, the eXtreme Gradient Boosting (XGB) algorithm is explained. Third, the k-folds cross-validation technique is described. Lastly, the performance metrics used to evaluate the models are defined.

The objective of this paper is to develop a model that predicts speed in short-term intervals (e.g., next 5 min) at a location on the freeway on the basis of the prevailing freeway traffic and weather conditions at the current time (t) as Equation (1) shows.

$$y_{(l,t+i)} = f_{l,p}(\mathbf{X}_t) \tag{1}$$

where $y_{(l,t+i)}$ is the predicted speed at location l at time $t+i$ and $i \in \{5, 10, 15\}$. $f_{l,p}(\mathbf{X}_t)$ is the weather-aware model that predicts the speed at location l and time $t+i$. The predictor variables for the model, traffic and weather conditions on the corridor at time t, are shown as $\mathbf{X}_t$ in Equation (2).

$$\mathbf{X}_t = \begin{pmatrix} u_{(A-U)}, & u_{(A)}, & \cdots & , u_{(D-D)} \\ q_{(A-U)}, & q_{(A)}, & \cdots & , q_{(D-D)} \\ o_{(A-U)}, & o_{(A)}, & \cdots & , o_{(D-D)} \\ p_{(A-U)}, & p_{(A)}, & \cdots & , p_{(D-D)} \end{pmatrix} \tag{2}$$

where,

$\mathbf{X}_t$ is traffic and weather conditions at time t.
u_m is speed at location m in km/h at time t.
q_m is traffic volume at location m in veh/5 min/ln at time t.

o_m is occupancy at location m in % at time t.
p_m is precipitation at location m in mm/h at time t.
m is the location of the traffic detector.

The proposed model in Equation (1) was compared with the weather-insensitive model in Equation (3), where precipitation is not considered in the prediction process.

$$y_{(l,t+i)} = f_l(\mathbf{X}_t) \tag{3}$$

where $f_l(\mathbf{X}_t)$ is the weather-insensitive model that predicts the speed at location l and time $t+i$.

$$\mathbf{X}_t = \begin{pmatrix} u_{(A-U)}, & u_{(A)}, & \cdots & , u_{(D-D)} \\ q_{(A-U)}, & q_{(A)}, & \cdots & , q_{(D-D)} \\ o_{(A-U)}, & o_{(A)}, & \cdots & , o_{(D-D)} \end{pmatrix} \tag{4}$$

where,

$\mathbf{X}_t$ is traffic conditions at time t. The other parameters were defined previously.

The proposed framework for developing the speed prediction models includes the following steps: (1) The radar and traffic data observations are fused on the basis of their spatio-temporal identities to ground truth data; (2) Training and testing data sets are created by assigning 70% of observations to the training set and the remaining 30% to the holdout testing set; (3) A prediction algorithm (e.g., XGBoost), in conjunction with 20-fold cross-validation, is utilized to train the models and choose the model with lowest error function value; (4) The 30% holdout testing set is utilized to evaluate the chosen model and to calculate the performance metrics. The remaining subsections provide additional information about the models' development process.

4.1. Extreme Gradient Boosting

eXtreme Gradient Boosting (XGB) is an ensemble learning method that has been successfully applied to the regression and classification problems. Ensemble learning methods consist of a group of weak learners that are iteratively trained to construct a concrete predictive model. In the XGB method, decision trees are successively built using the boosting technique [26]. Boosting is an iterative process in which trees are consecutively added to the model, such that each new tree reduces the residual error of the previous tree. The boosting method consists of three steps:

First, the initial model, $F_0(x)$, is developed to predict the response variable (y). The residual of this model is denoted by $(y - F_0)$, with which the model is associated. Second, the model $h_1(x)$ is fitted to the residual obtained in the first step. In the third step, the initial model $F_0(x)$ is combined with the model $h_1(x)$ to generate $F_1(x)$, the boosted version of F_0 and with a lower mean squared error as shown in Equation (5).

$$F_1(x) = F_0(x) + h_1(x) \tag{5}$$

To improve the performance of F_1, F_2 is generated by modeling the residuals from F_1, as follows:

$$F_2(x) = F_1(x) + h_2(x) \tag{6}$$

The process of boosting the model is repeated for m iterations until an acceptable mean squared error is achieved. Each new learner contributes to the models by reducing the error of the learner that was added to it in a previous step, resulting in a final model $F_m(x)$ as shown in Equation (7).

$$F_m(x) = F_{m-1}(x) + h_m(x) \tag{7}$$

The model is initialized with a function $F_0(x)$ that minimizes the loss function (8) and (9).

$$F_0(x) = argmin_\gamma \sum_{i=1}^{n} (y_i, \gamma) \tag{8}$$

$$argmin_{\gamma} \sum_{i=1}^{n} (y_i, \gamma) = argmin_{\gamma} \sum_{i=1}^{n} (y_i - \gamma)^2 \quad (9)$$

where, γ is the constant value of function $F_0(x)$ that serves as a threshold for the splitting process of the trees. Taking the first differential with respect to γ of the above equation, the boosted model is initialized with:

$$F_0(x) = \frac{\sum_{i=1}^{n} (y_i)}{n} \quad (10)$$

The gradient boosting involves the following steps: First, Defining $F_0(x)$ with which the boosting algorithm is initialized as Equation (8) shows. Second, the gradient of the loss function in the ith observations is calculated for each iteration m, as shown Equation (11):

$$r_{im} = -\alpha[\frac{\partial(L(y_i, F(x_i))}{\partial F(x_i)}]F(x) = F_{(m-1)}(x) \quad (11)$$

where, α is the learning rate of the model. $L(y_i, F(x_i))$ is the loss function which is calculated as the squared loss as shown in Equation (12).

$$L(\theta) = \sum_{i} (\bar{y}_i - y_i)^2 \quad (12)$$

Each $h_m(x)$, obtained from step two, is fit on the obtained gradient at each step. Finally, the multiplicative γ_m is derived for each terminal node and the boosted model $F_m(x)$ is defined as Equation (13):

$$F_m(x) = F_{m-1}(x) + \gamma_m * h_m(x) \quad (13)$$

A grid of values is used to investigate the best set of algorithmic parameters to train XGB for each location, prediction horizon, and weather sensitivity.

4.2. K-Folds Cross-Validation

Machine learning methods, including the Extreme Gradient Boosting method, have been successfully applied to model complex problems in various fields [27]; often producing error indexes smaller than the common statistical models. However, overfitting and memorizing the training data are common concerns with the use of machine learning methods. Thus, machine learning models are often validated. Traditionally, a data set that was not previously used for training is presented to the model and the prediction error for this previously unseen data set is then calculated. The model is considered acceptable when the error indexes for the training and validation data sets are comparable.

In the k-fold cross-validation method, the available data are randomly partitioned into k groups or folds. The number of observations in each fold is approximately the same. The model is trained and validated k times using various combinations of folds. In the first iteration, the first fold is used for validation, and the remaining k−1 folds are used for training the model. In the second iteration, the second fold is used for validation and the model is trained using the remaining folds. This process is repeated until every fold is treated as a validation set. The average training error for the models developed in k iterations is reported as the overall training error; the validation error is calculated similarly. In this paper, the case study data were split into 70% and 30% training and testing data sets. The 70% training data were split into 20 folds in the training process.

4.3. Performance Metrics

Root Mean Square Error (RMSE), Average Absolute Error (AAE), and Mean Absolute Percentage Error (MAPE) were used to evaluate and compare the models. These performance measures are defined in Equations (14)–(16):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{14}$$

$$AAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{15}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{|y_i|}\right) * 100 \tag{16}$$

where,

y_i: observed speed for interval i.
$\hat{y}_i$: predicted speed for interval i.
n: number of observations (i.e., prediction intervals).

5. Results

One of the objectives of this paper was to investigate whether including the precipitation rate in the prediction models improves the predictive performance of the speed prediction models. Figure 3 provides an example of the potential impacts of precipitation on speed on the I-270 NB corridor during a day that was not used for training or testing the models. The precipitation rate is reported in millimeters per hour (mm/h), and speed is reported in kilometers per hour (km/h). Light rain with an intensity of less than 15 mm/h is noticed after 6:00 a.m. and rain with an intensity of less than 15 mm/h is also noticed around 9:00 a.m. During these rainfall periods, speed dropped from about 105 km/h to between 40 km/h to 60 km/h. After 4:00 pm, the corridor witnessed very heavy rain with varying intensity between 50 mm/h to 200 mm/h, especially at location D with an intensity of 200 mm/h for about 5 min. As illustrated in Figure 3, the heavy rainfall at location D is potentially contributing to a speed drop with a magnitude of 55 km/h. However, the most significant reductions in speed are not necessarily observed during the most intense periods of rainfall. For example, between 16:35 (4:35 p.m.) and 17:50 (5:50 p.m.), location C witnessed intermittent rainfall with varying intensity of 15 mm/h to 50 mm/h, which resulted in speed drop with magnitudes ranging from 35 km/h to 100 km/h. During this period, the speed was on average 60 km/h less than the free flow speed.

Comparing the precipitation rate at locations C and D between 17:00 (5 p.m.) and 18:00 (6 p.m.) in Figure 3 illustrates that the intensity and duration of rain at location C were lighter and shorter than the intensity and duration of rain at station D. In spite of the lighter rain at location C, the magnitude of speed dropped at location C was more significant than the speed drop at station D. The magnitude of the reduction in speed at location D does not appear to correlate with the intensity of precipitation at location D. It appears that location C became an active bottleneck and reduced the number of vehicles that were arriving at location D, and thus location D did not witness a significant speed drop during the precipitation period. The active bottleneck at location C mitigated the effect of precipitation at location D. This hypothesis was further investigated by developing traffic stream models for locations A to D.

To investigate the impact of precipitation on the speed at the corridor, speed-flow traffic stream models for rain and no-rain conditions were developed for locations A to D. The van Aerde traffic-stream model was used to describe the relationship between the speed and gap between vehicles during rain and no-rain conditions. Equation (17) shows the macroscopic speed–flow relationship of the van Aerde models.

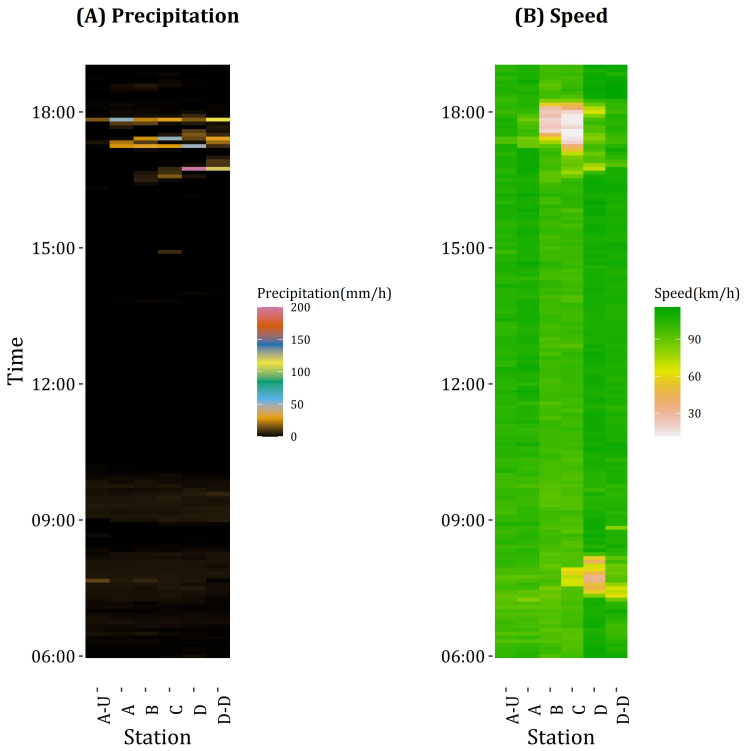

Figure 3. (**A**) Precipitation on the I-270 NB corridor during the 27 May 2016 reporting period; (**B**) Speed on the I-270 NB corridor during the 27 May 2016 reporting period (mm/h) and speed (km/h).

In Equation (17), q represents the flow rate (veh/h/ln), u indicates speed (km/h), u_f shows the free-flow speed (km/h), and c_1, c_2, and c_3 are constants [28]. A computer program developed by Rakha and Arafeh [29] was used to calibrate the van Aerde models for locations A to D, which is shown in Figure 4. Table 2 reports the parameters of the traffic stream models.

$$q = \frac{u}{c_1 + \frac{c_2}{u_f - u} + c_3 u} \tag{17}$$

The van Aerde models for locations A to D indicate that precipitation negatively affects the traffic stream. Specifically, the models indicate that precipitation consistently reduces free flow speed and capacity at all four locations. However, the magnitude of precipitation influence on the four locations is not the same. The difference in the impact of rain, as well as the traffic flow characteristics of these locations, is partially related to the geometric design of these locations. Location A is a diamond interchange; location B is on a tangent section of I-270; location C is at a directional interchange, and location D is a partial diamond interchange. As demonstrated in Figure 4, the capacity at location C is lower than the capacity at location D, and location C limits the number of vehicles arriving at location D during congestion periods. This potentially explains the reason for the minimal impact of heavy precipitation periods on location D speed.

Table 2. Traffic stream characteristics of I-270 NB corridor.

Location	Model	Free Flow Speed (km/h)	Speed at Capacity (km/h)	Capacity (veh/h/ln)	Jam Density (veh/km/ln)	c_1 (km)	c_2 (km^2/h)	c_3 (h)
A	No-Rain	107.5	100	2147.6	77.4	0.01285	0.00781	0.00033
	With Rain	96.3	86.9	1882.8	70.1	0.01410	0.01607	0.00035
B	No-Rain	96.2	85	1875	51.1	0.01923	0.03269	0.00027
	With Rain	94.6	82.5	1575	58.8	0.01664	0.03461	0.00040
C	No-Rain	98.7	85	1737.8	123.8	0.00787	0.02071	0.00047
	With Rain	94.3	85	1692	109.4	0.00903	0.01032	0.00047
D	No-Rain	111.7	100	1855.5	75	0.01315	0.02039	0.00039
	With Rain	107.5	100	1649.5	69.2	0.01437	0.00874	0.00045

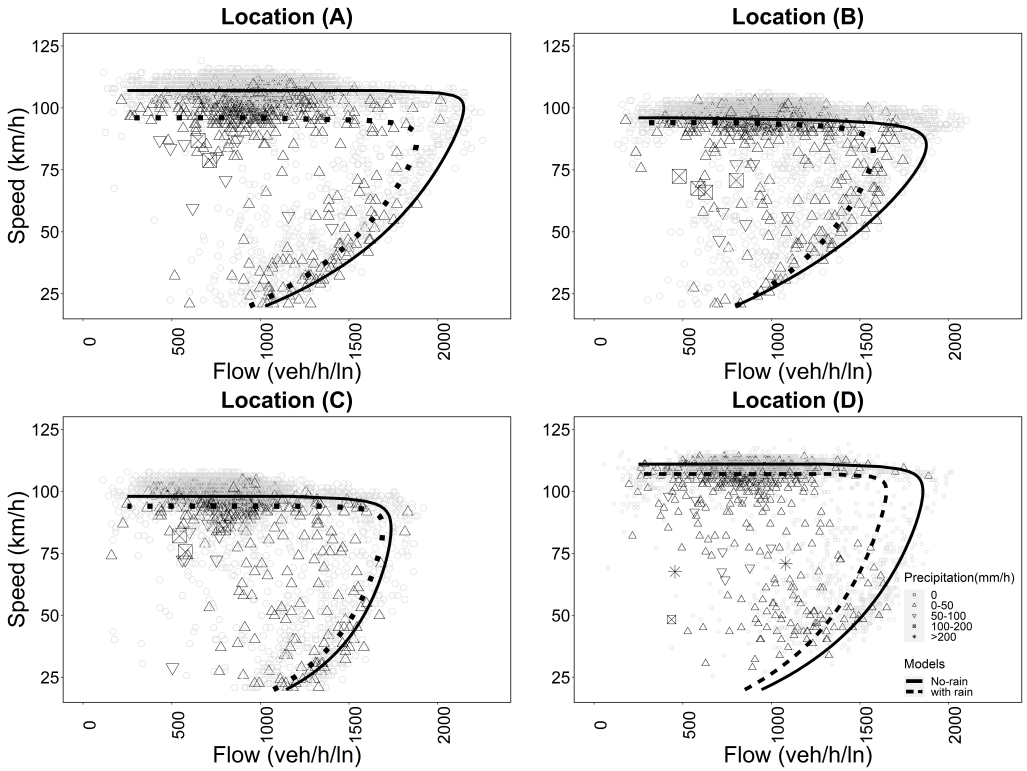

Figure 4. Traffic stream models for I-270 NB corridor.

In the next step, two categories of speed prediction models were developed for the I-270 NB corridor: (1) weather-insensitive models and (2) weather-aware models. Weather-insensitive models did not consider the precipitation rate in the prediction of speed and serve as base models to evaluate the hypothesis that the precipitation rate is an important factor for short-term speed prediction. In contrast, weather-aware models took into account the precipitation rate in predicting the I-270 NB corridor speed. Speed was predicted on the basis of flow rate, speed, occupancy, and precipitation rate that were observed at the six locations on the I-270 NB corridor. Speed was predicted on the basis of the current roadway conditions (i.e., conditions at time t) for the next three consecutive five-minute intervals, which are referred to as t + 5, t + 10, and t + 15 prediction intervals. Table 3 provides the performance metrics for weather-insensitive and weather-aware models obtained from 20-fold cross-validation.

Several trends are observed in the results. First, the predictions are most accurate in the 5-min interval immediately after the current traffic status. Second, in most cases, including the precipitation data in the prediction process reduces the prediction error. In two cases out of 36 cases, including the weather in the prediction process did not change the results obtained from the models. In eight out of 36 cases, the performance metrics slightly increased by 0.20% to 2.68% when the weather was included in the models. However,

including the weather in the prediction process improved the results in 26 cases out of 36 cases. The improvements gained from utilizing the weather data were on average 0.20 km/h for RMSE, 1.87 km/h for AAE and 3.07% for MAPE.

Table 3. Performance metrics for weather-insensitive and weather-aware models.

Location	Prediction Interval	RMSE (km/h)		AAE (km/h)		MAPE (%)	
		Weather-Insensitive	Weather-Aware	Weather-Insensitive	Weather-Aware	Weather-Insensitive	Weather-Aware
A	t + 5	4.98	4.99	2.94	2.94	4.11	4.22
	t + 10	6.7	5.85	3.49	3.33	5.51	4.71
	t + 15	7.84	7.97	4.07	4.10	6.74	6.83
B	t + 5	4.23	4.21	2.56	2.57	3.47	3.47
	t + 10	5.80	5.53	3.24	3.14	4.50	4.33
	t + 15	6.27	6.12	3.45	3.40	4.96	4.90
C	t + 5	5.32	5.27	2.97	3.00	4.38	4.32
	t + 10	6.03	6.07	3.26	3.24	5.10	5.07
	t + 15	6.66	6.64	3.51	3.49	5.53	5.52
D	t + 5	5.00	4.90	2.65	2.61	3.28	3.21
	t+10	6.75	6.52	3.25	3.19	4.15	4.10
	t + 15	7.20	7.18	3.65	3.60	4.77	4.64

A comparison of the performance of weather-insensitive and weather-aware models revealed that in 26 out of 36 cases, when the weather was included in the prediction process, the prediction error was decreased. These prediction error measures were calculated regardless of the precipitation event. In other words, the models were evaluated on the basis of the entire observations from 6:00 a.m. to 7:00 p.m. on study days. The performance of the models was further investigated during the precipitation events. Figure 5 reports the precipitation rate and the ground truth speed and compares the predicted speed for t + 5, t + 10, and t + 15 prediction intervals from weather-insensitive and weather-aware models at location D. These predictions were obtained based on data that was not previously utilized for training the model. During off-peak periods, the performance of weather-insensitive and weather-aware models is comparable. However, during precipitation events, the weather-aware model provides speed estimates that follow a similar trend to the ground truth speed. For example, during light precipitation in the morning peak at 7:30 a.m., and during heavy precipitation in the afternoon peak at around 4:45 p.m., the speed estimated by the weather-insensitive model demonstrated an average error of 12 km/h, while the speed estimated by the weather-aware model demonstrated an average error of 2 km/h. In addition, as illustrated in Figure 5, the weather-aware model successfully captured the impact of the rainfall on traffic stream speed.

The predictive performance of the XGBoost model was compared with the predictive performance of the K-nearest neighbor algorithm (KNN) and feed-forward neural network (NNET) models. KNN [30] is a non-parametric algorithm that calculates the average of a specific number (n) of the nearest neighbors to the value of interest $y_{(l,t+i)}$ for weather-aware and weather-insensitive scenarios. A grid of neighbors $N = \{1, 2, 3, \ldots, 70\}$ was investigated to determine the optimal set of neighbors for weather-aware and weather-insensitive models. For weather-insensitive models, 18 neighbors minimized the prediction error. In comparison, 15 neighbors were sufficient to produce the minimum RMSE for weather-aware models.

NNET is an artificial intelligence method that mimics the human brain by connecting artificial neurons [31]. Two sets of hyperparameters were investigated to determine the best NNET weather-aware and weather-insensitive models. The set for the number of neurons was $\{1, 2, 3, \ldots, 10\}$ and the weight decay set was $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. One hidden layer with two neurons and a weight decay of 0.5 resulted in minimum RMSE for weather-insensitive models. However, three neurons and a weight decay of 0.4 produced the best performance of weather-aware models. Figure 6 compares AAE, MAPE, and RMSE

for t + 5, t + 10, and t + 15 prediction horizons obtained from XGBoost models with the performance metrics of KNN and NNET models. XGBoost models consistently outperform NNET models and provide performance measures that are slightly more desirable than KNN models. This pattern is observed for predictions from both the weather-aware and weather-insensitive models. It is also observed that for KNN, NNET, and XGBoost models, the weather-aware models consistently provide lower error measures in comparison to the weather-insensitive models.

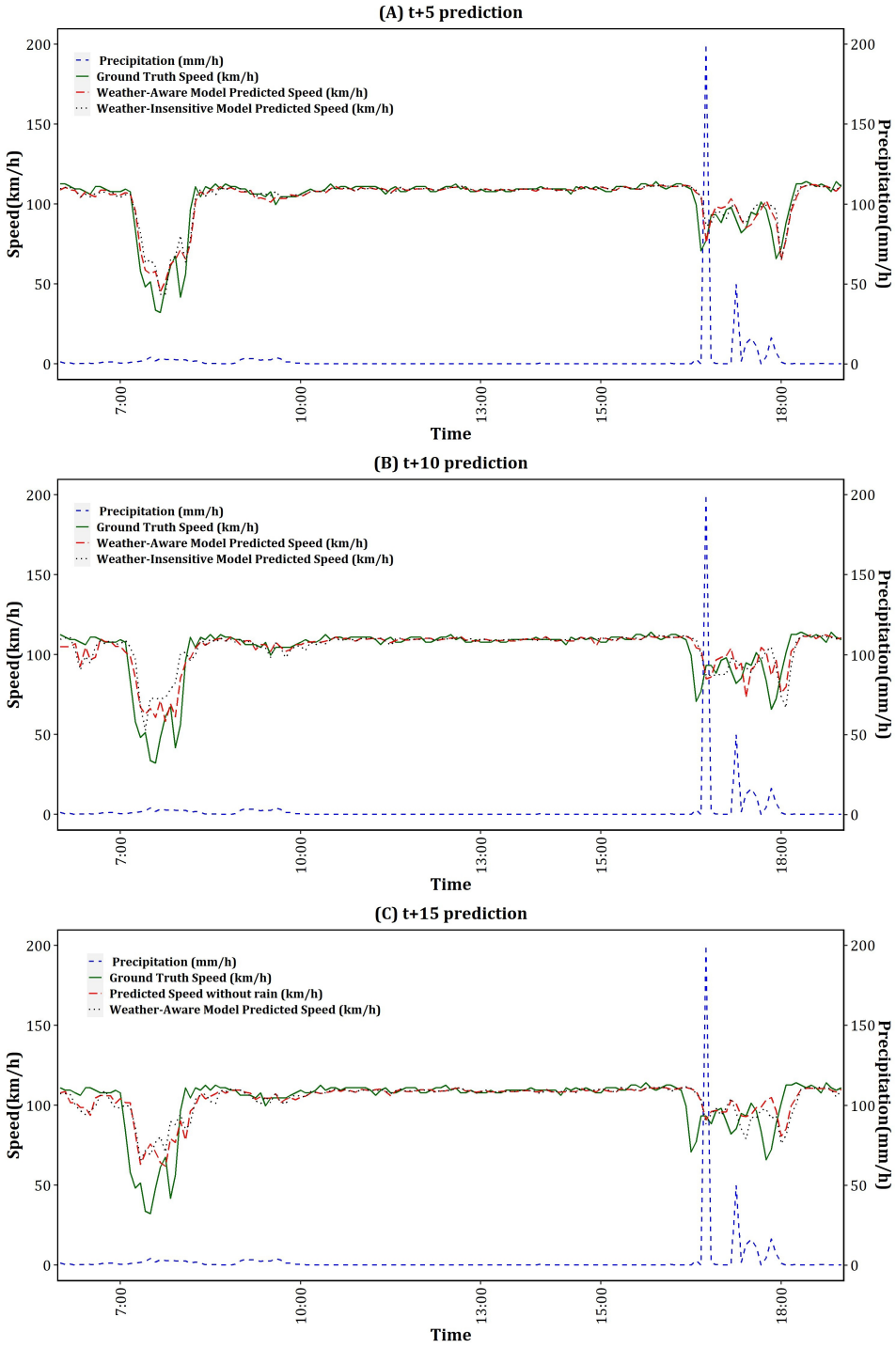

Figure 5. Ground truth speed and predicted speed obtained from weather-insensitive and weather-aware models: (**A**) t + 5 Prediction interval, (**B**) t + 10 prediction interval, (**C**) t + 15 prediction interval.

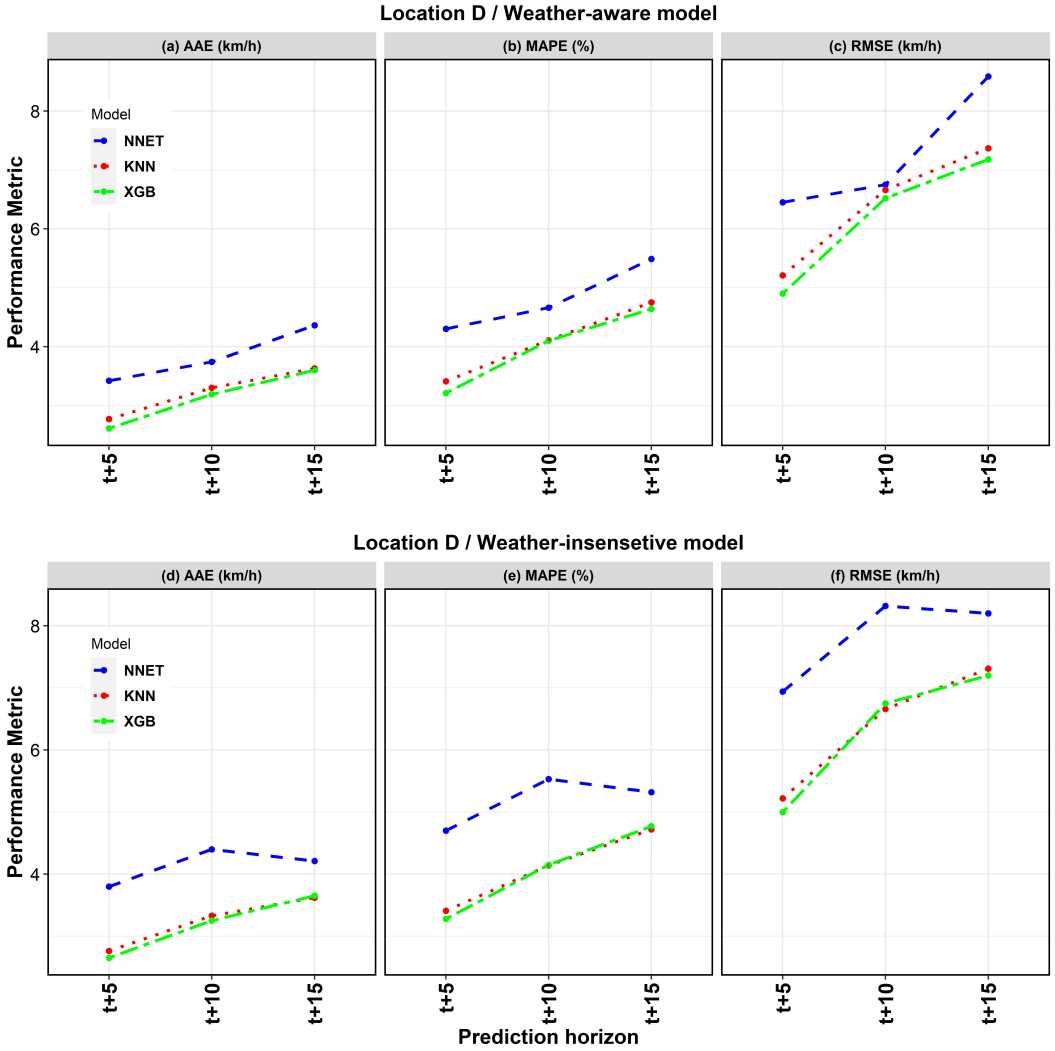

Figure 6. Performance measures for KNN, NNET, and XGBoost Models for t + 5, t + 10, and t + 15 prediction horizons at location D.

As discussed in the literature review, very few studies investigated the impact of weather on short-term speed prediction. Among those, Huang and Ran [20] are the only researchers who reported AAE as the performance metric. Their study reported a 12.4 km/h AAE for a 5-min speed prediction horizon, which is higher in comparison to the metrics reported for I-270 NB in St. Louis, MO, USA. The AAE values obtained for four locations on I-270 NB were between 2.57 km/h to 3.00 km/h. Leong et al. [21] did not report metrics such as RMSE, AAE, or MAPE. Instead, they only reported the results of the comparison between the weather-aware and weather-insensitive models. Their findings were consistent with the result obtained in this paper that, in some cases, the weather data improved the performance of the speed prediction models, and in some other cases, the rainfall data did not influence the models' performance.

6. Conclusions

Short-term speed prediction models are examples of ITS subsystems that enable traffic management center operators to communicate anticipated traffic conditions to the public and take steps to harmonize the traffic on the roadway. While traffic management centers often issue weather-related advisories to the public, high-resolution weather data are usually not quantitatively included in the transportation management tools.

In this paper, weather-aware models for short-term speed prediction were proposed. The high-resolution data obtained from weather surveillance radar were processed to obtain

precipitation rates at roadway sections. Real-time precipitation rates were fused with data obtained from traffic detectors to predict speed in five-minute intervals for the next 5-, 10-, and 15-min periods. The proposed approach for short-term speed prediction was applied to a segment of Interstate 270 NB in St. Louis, MO, USA. Weather-aware and weather-insensitive models were developed utilizing the XGBoost method. The performance of weather-aware models was compared with the performance of weather-insensitive speed prediction models that did not take precipitation rate into account in the speed prediction process. The results indicated that weather data improved the predictions in 72% of cases, did not have an impact on 6% of the cases, and slightly increased the error metrics in 22% of the cases. In those cases, the increase in AAE error measure was, on average, 0.02 km/h. The predictions made by the XGBoost method were compared with KNN and NNET methods. The performance metrics obtained from the XGBoost method outperformed the performance metrics obtained from KNN and NNET models.

van Aerde speed-flow traffic stream models were developed for the bottleneck locations on the corridor for no-rain and rainy conditions. The speed-flow models were distinctly different for rain and no-rain conditions at each location. However, the impact of rain on each location was not similar. The rain resulted in a reduction of the free flow speed ranging from 2% to 10% and a reduction in capacity ranging from 2% to 16%. The variations in the impact of rain on speed-flow traffic stream models are consistent with the variations in the speed predictions obtained from the weather-aware and weather-insensitive models. This indicates that in addition to the intensity of precipitation, additional factors such as the geometry of the roadway and the overall traffic demand may influence the magnitude of the impact of adverse weather on travel speed. In future research, it is recommended to expand the case study area to a larger network and to study the impact of other weather events, such as snow and drizzle, and other weather metrics, such as visibility index.

Author Contributions: M.A. and J.K. contributed to the design and implementation of the research, analysis of the results, and writing of the manuscript. Y.W. contributed to development of the methods for analyzing Weather Surveillance Radar data. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available upon request from the corresponding author.

Acknowledgments: We would like to thank the Missouri Department of Transportation for providing traffic detectors data.

Conflicts of Interest: The authors do not have any conflict of interest with other entities or researchers.

References

1. Akallouch, M.; Akallouch, O.; Fardousse, K.; Bouhoute, A.; Berrada, I. Prediction and Privacy Scheme for Traffic Flow Estimation on the Highway Road Network. *Information* **2022**, *13*, 381. [CrossRef]
2. Sihag, G.; Parida, M.; Kumar, P. Travel Time Prediction for Traveler Information System in Heterogeneous Disordered Traffic Conditions Using GPS Trajectories. *Sustainability* **2022**, *14*, 10070. [CrossRef]
3. Mohammed, O.; Kianfar, J. A machine learning approach to short-term traffic flow prediction: A case study of interstate 64 in Missouri. In Proceedings of the 2018 IEEE International Smart Cities Conference (ISC2), Kansas City, MO, USA, 16–19 September 2018 ; pp. 1–7.
4. Kianfar, J.; Sun, C. *Operational Analysis of Freeway Variable Speed Limit System: Case Study of Deployment in Missouri*; Technical Report; National Academies of Sciences, Engineering, and Medicine: Washington, DC, USA, 2013.
5. Li, T.; Ma, J.; Lee, C. Markov-based time series modeling framework for traffic-network state prediction under various external conditions. *J. Transp. Eng. Part A Syst.* **2020**, *146*, 04020042. [CrossRef]
6. Zardosht, B.; Beauchemin, S.S.; Bauer, M.A. A predictive accident-duration based decision-making module for rerouting in environments with V2V communication. *J. Traffic Transp. Eng. (Engl. Ed.)* **2017**, *4*, 535–544. [CrossRef]

7. Chavhan, S.; Venkataram, P. Prediction based traffic management in a metropolitan area. *J. Traffic Transp. Eng. (Engl. Ed.)* **2020**, *7*, 447–466. [CrossRef]
8. Gutierrez-Osorio, C.; Pedraza, C. Modern data sources and techniques for analysis and forecast of road accidents: A review. *J. Traffic Transp. Eng. (Engl. Ed.)* **2020**, *7*, 432–446. [CrossRef]
9. Angel, M.L.; Sando, T.; Chimba, D.; Kwigizile, V. Effects of rain on traffic operations on Florida freeways. *Transp. Res. Rec.* **2014**, *2440*, 51–59. [CrossRef]
10. Hammit, B.E.; James, R.; Ahmed, M.; Young, R. Toward the development of weather-dependent microsimulation models. *Transp. Res. Rec.* **2019**, *2673*, 143–156. [CrossRef]
11. Ni, X.; Huang, H.; Chen, A.; Liu, Y.; Xing, H. Effect of heavy rainstorm and rain-induced waterlogging on traffic flow on urban road sections: Integrated experiment and simulation study. *J. Transp. Eng. Part A Syst.* **2021**, *147*, 04021057. [CrossRef]
12. Otim, T.; Dörfer, L.; Ahmed, D.B.; Munoz Diaz, E. Modeling the Impact of Weather and Context Data on Transport Mode Choices: A Case Study of GPS Trajectories from Beijing. *Sustainability* **2022**, *14*, 6042. [CrossRef]
13. Guo, P.; Sun, Y.; Chen, Q.; Li, J.; Liu, Z. The Impact of Rainfall on Urban Human Mobility from Taxi GPS Data. *Sustainability* **2022**, *14*, 9355. [CrossRef]
14. Asif, M.T.; Dauwels, J.; Goh, C.Y.; Oran, A.; Fathi, E.; Xu, M.; Dhanya, M.M.; Mitrovic, N.; Jaillet, P. Spatiotemporal patterns in large-scale traffic speed prediction. *IEEE Trans. Intell. Transp. Syst.* **2013**, *15*, 794–804. [CrossRef]
15. Bouktif, S.; Fiaz, A.; Ouni, A.; Alnaqbi, B.; Alsereidi, F.S.; Alsereidi, F.A. Bayesian optimized XGBoost model for traffic speed prediction incorporating weather effects. In Proceedings of the 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, 21–23 October 2020; pp. 1–7.
16. Ma, X.; Dai, Z.; He, Z.; Ma, J.; Wang, Y.; Wang, Y. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors* **2017**, *17*, 818. [CrossRef] [PubMed]
17. Min, W.; Wynter, L. Real-time road traffic prediction with spatio-temporal correlations. *Transp. Res. Part C Emerg. Technol.* **2011**, *19*, 606–616. [CrossRef]
18. Wang, J.; Chen, R.; He, Z. Traffic speed prediction for urban transportation network: A path based deep learning approach. *Transp. Res. Part C Emerg. Technol.* **2019**, *100*, 372–385. [CrossRef]
19. Yu, B.; Yin, H.; Zhu, Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv* **2017**, arXiv:1709.04875.
20. Huang, S.H. *An Application of Neural Network on Traffic Speed Prediction under Adverse Weather Conditions*; The University of Wisconsin-Madison: Madison, WI, USA, 2003.
21. Leong, L.W.; Lee, K.; Swapnil, K.; Li, X.; Victor, H.Y.T.; Mitrovic, N.; Asif, M.T.; Dauwels, J.; Jaillet, P. Improving traffic prediction by including rainfall data. In Proceedings of the ITS Asia-Pacific Forum, Nanjing, China, 27–29 April 2015; Volume 14.
22. Jia, Y.; Wu, J.; Xu, M. Traffic flow prediction with rainfall impact using a deep learning method. *J. Adv. Transp.* **2017**, *2017*, 6575947. [CrossRef]
23. Missouri Department of Transportation. MoDOT's Gateway Guide. 2017. Available online: https://www.gatewayguide.com//speed-sensors.html (accessed on 15 October 2022).
24. Tang, L.; Zhang, J.; Langston, C.; Krause, J.; Howard, K.; Lakshmanan, V. A physically based precipitation–nonprecipitation radar echo classifier using polarimetric and environmental data in a real-time national system. *Weather Forecast.* **2014**, *29*, 1106–1119. [CrossRef]
25. Wang, Y.; Cocks, S.; Tang, L.; Ryzhkov, A.; Zhang, P.; Zhang, J.; Howard, K. A prototype quantitative precipitation estimation algorithm for operational S-band polarimetric radar utilizing specific attenuation and specific differential phase. Part I: Algorithm description. *J. Hydrometeorol.* **2019**, *20*, 985–997. [CrossRef]
26. Sundaram, R.B. An End-to-End Guide to Understand the Math behind XGBoost. 2018. Available online: https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/ (accessed on 10 June 2022).
27. Kianfar, J.; Edara, P. A data mining approach to creating fundamental traffic flow diagram. *Procedia-Soc. Behav. Sci.* **2013**, *104*, 430–439. [CrossRef]
28. Van Aerde, M.; Rakha, H. Multivariate calibration of single regime speed-flow-density relationships [road traffic management]. In Proceedings of the Pacific Rim TransTech Conference. 1995 Vehicle Navigation and Information Systems Conference Proceedings. 6th International VNIS. A Ride into the Future, Seattle, WA, USA, 30 July–2 August 1995; pp. 334–341.
29. Rakha, H.; Arafeh, M. Calibrating steady-state traffic stream and car-following models using loop detector data. *Transp. Sci.* **2010**, *44*, 151–168. [CrossRef]
30. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 112.
31. Park, H.; Haghani, A.; Zhang, X. Interpretation of Bayesian neural networks for predicting the duration of detected incidents. *J. Intell. Transp. Syst.* **2016**, *20*, 385–400. [CrossRef]