WEATHER-INFORMED IMPUTATION OF BLOCK-MISSING HIGHWAY SPEED DATA

Xiaofeng Chen, Corresponding Author
Department of Computer Science and Engineering
University of Buffalo, State University of New York, Buffalo, NY 14260
xchen326@buffalo.edu

Wen Zhang
Department of Computer Science and Engineering
University of Buffalo, State University of New York, Buffalo, NY 14260

Adel W. Sadek, Ph.D.
Department of Civil, Structural and Environmental Engineering
University of Buffalo, State University of New York, Buffalo, NY 14260

Chunming Qiao, Ph.D.
Department of Computer Science and Engineering
University of Buffalo, State University of New York, Buffalo, NY 14260

Word Count: 4940 words + 4 table(s) $\times$ 250 = 5940 words

Submission Date: December 11, 2025

ABSTRACT

Reliable highway speed data are critical for traffic safety analysis, route planning, and the development of intelligent transportation systems (ITS). However, large-scale highway datasets often suffer from extensive missing values for a variety of reasons. This study focuses on cases where a significant portion (above 50%) of aggregated segment-level average speed data is missing for an extended period of time, such as weeks or even months, and aims to reconstruct the missing data using (AI-based) data imputation methods. We evaluate 16 models across three categories: classical regression models, ensemble tree-based methods and advanced deep generative models such as Frequency-aware Generative Time-series Imputation (FGTI). The models are applied to a dataset from Buffalo (NY) freeway segments, augmented with high-resolution Road Weather Information System (RWIS) data—including both mild (Dec 12–18) and harsh (Dec 19–25) winter conditions. We reveal a fundamental trade-off between model complexity and robustness. FGTI consistently achieves superior performance (MAE: 0.7–1.6), capturing both overall trends and fine-grained residuals variations under different seasonal conditions.

Keywords: Traffic Speed Imputation, Block Missing Data, Deep Generative Models, Multivariate Time Series, Weather-Traffic Interaction

INTRODUCTION

Accurate and complete highway speed data are essential for traffic monitoring, travel time estimation, and the development of Intelligent Transportation Systems (ITS) applications (*1–3*). However, such data are often missing due to sensor failures, communication interruptions, or inconsistent infrastructure coverage (*4, 5*). This issue is especially severe in large urban areas, where entire months of data may be unavailable, thereby limiting their utility in mobility analysis and performance evaluation. Desai et al. (*6*) noted that prolonged data gaps, especially during extreme weather events, can significantly affect transportation system performance assessments and hinder timely decision-making. In fact, data completeness in transportation management systems (TMSs) often falls below 70%, prompting early research efforts such as Smith et al. (2003) to explore imputation techniques as a practical remedy for restoring large-scale datasets (*7*).

To address missing data, various imputation methods have been developed - including interpolation, regression, and advanced machine learning or deep generative approaches (*5, 7–9*). However, block-missing scenarios - where long continuous intervals of data are absent - remain particularly challenging, especially when the missing segment spans weeks or months (*10*). Longer missing durations can significantly increase the imputation error and reduce the reliability of the model. This underscores the limitations of traditional gap-filling approaches. In response, many studies now incorporate auxiliary features - namely weather conditions (e.g., temperature, precipitation, wind, visibility) and temporal predictors (e.g., hour-of-day, day-of-week) - to compensate for information that is lost in prolonged, block-missing periods. Empirical evidence shows that models endowed with these environmental and temporal predictors consistently outperform those that rely solely on past traffic metrics, especially in challenging missing scenarios (*11, 12*).

In this study, we examine highway speed data collected via a fine-grained Connected Vehicle (CV) dataset for Buffalo, New York segments with a posted speed limit below 55 mph. Speed records were available to the study for the months of October, March, and part of December; however, data was missing for the other months, which thus exhibit complete block-missing speed records. Meanwhile, auxiliary features - including high-resolution (10-minute) weather data from nearby weather stations and time variables (hour-of-day, weekday/weekend indicator) - are entirely observed. Given this setting, we propose a weather-informed imputation framework designed specifically for block-missing highway speed data. Our approach takes advantage of observed environmental and temporal covariates to reconstruct missing speed values. Our main contributions are summarized as follows:

1. We study the correlation between weather and high-speed data, and evaluate classical and deep generative models for imputing missing speed data under realistic block-missing scenarios.
2. For each of the 16 models, we conduct three experiments that systematically test model generalization under 80% data missingness, by strategically splitting mild (Dec 12-18) and harsh (Dec 19-25) winter periods, revealing fundamental differences in model robustness to domain shift.
3. Through controlled train-test splits, we compare the performance of the 16 models using the testing data when available and the correlated weather data when no test data are available; and we discover that models handle extreme-to-mild extrapolation more effectively than mild-to-extreme interpolation.
4. Our results reveal that while one deep generative AI model, namely Frequency-aware Generative Time-series Imputation (FGTI) shows consistent superior performacnce, an-

other deep generative model, namely Conditional Score-based Diffusion Models, performs worse in most cases than simple models. This highlights the importance of choosing the appropriate model for the application at hand, rather than the direct trend-driven adoption of deep generative models. In fact, we demonstrate that under extreme winter conditions with 80% data loss, simple models can provide reliable alternatives to complex deep learning approaches.

RELATED WORK

Missing Data Patterns & Block-Missing

Understanding how data are missing is fundamental to choosing the right imputation method. Donald Rubin's classic framework (1976) remains foundational in that regard, defining three categories for the pattern of how data are missing. These are, 1) Missing Completely At Random (MCAR), where data points are absent purely at random; 2) Missing At Random (MAR), where missingness depends only on observed data; and 3) Missing Not At Random (MNAR), where missingness depends on unobserved or missing values (*13*). Previous studies show that the accuracy of the imputation depends not just on the imputation method itself, but also on the underlying missingness patterns (*14–16*). However, it should be noted that Rubin's classification focuses primarily on the cause of missingness and does not consider how missing values are arranged across time and space.

To bridge this gap in traffic data imputation, Chan et al. introduced a structure-based taxonomy of missing-pattern types commonly observed in transportation datasets: random missing, fiber missing, and block missing (*9*). In real-world transportation datasets, block missing is a prominent and challenging category, referring to large, continuous segments of missing data across time, space, or both. This type of missingness is usually caused by sensor malfunctions, detector removal, or communication failures, and lacks adjacent temporal or spatial reference points, which complicate the data imputation process (*17, 18*).

Beyond the missingness structure, the dimensionality of missingness (i.e. how many features are absent) also substantially affects the performance of the imputation methods (*16, 19, 20*). In transportation datasets, univariate block-missing is frequently observed: for instance, speed data are systematically absent for a given roadway segment, while weather records and temporal metadata (hour-, weekday-based features) remain fully available. This type of loss aligns well with conditional imputation frameworks, since the missing variable can be inferred from auxiliary contextual features without modeling the full joint distribution of multivariate inputs (*21–23*).

Imputation Techniques for Block Missing Data

Traditional imputation techniques in transportation research have mainly relied on simple statistical heuristics such as mean substitution, linear interpolation, or spline fitting. These methods are straightforward to implement but quickly lose accuracy when missing segments exceed a few consecutive observations—such as in block-missing scenarios that span hours or days. To address these limitations, more principled statistical approaches have been proposed. An example of these statistical approaches applies predictive mean matching within a chained equation framework for multiple imputation. This approach leverages spatial correlation among loop detectors to estimate missing values and demonstrates robustness in scenarios involving high missingness rates and fine-grained temporal resolutions. By imputing data prior to aggregation, the method preserves the variance structure and improves the reliability of the estimation, outperforming baseline techniques

such as pairwise regression, especially under block-missing conditions (*24*).

In addition, machine learning models have been widely used in traffic data imputation tasks due to their ability to model nonlinear relationships and leverage external features. Regression-based techniques, including ridge regression and expectation-maximization algorithms, offer improved flexibility by incorporating auxiliary variables. However, they typically rely on assumptions such as linearity or Gaussian error distributions and often perform inadequately when faced with prolonged or structured missing patterns, such as those spanning entire days or months. Tree-based models, such as Random Forest and XGBoost, are particularly popular for their robustness and interpretability in static settings. When applied to short gaps or randomly missing values, these models can achieve high point-wise accuracy by capturing correlations between traffic speed and auxiliary variables such as weather or time-of-day features (*25*). However, they typically lack mechanisms to capture temporal dependencies across sequences and are not explicitly designed to handle block missing patterns, where entire periods of speed data are absent. As a result, their performance may deteriorate when faced with temporally structured missingness.

Recent advancements in deep learning have led to the development of powerful generative models capable of capturing complex temporal dependencies for time-series imputation. Among these, Variational Autoencoders (VAEs) and Transformer-based architectures have shown promise in recovering missing traffic data under irregular conditions. These models often utilize temporal attention mechanisms or latent variable structures to identify global trends, but their performance may degrade when applied to large blocks of missing data due to the challenge of maintaining long-term consistency (*26*).

To directly address these limitations, several deep generative frameworks have been proposed for block-missing scenarios. One representative approach is the Conditional Score-based Diffusion Imputation (CSDI), which performs probabilistic sampling conditioned on observed variables to reconstruct missing segments with high fidelity. The score-based diffusion process explicitly models uncertainty while preserving temporal continuity and has been shown to be effective even in the case of high missingness rates (*27*). Similarly, the Flexible Generative Time-series Imputation (FGTI) framework adopts a two-stage architecture to decompose time series into low-frequency trends and high-frequency residuals. This design enables it to recover both the coarse and the fine-grained structures. FGTI has demonstrated strong performance in imputing missing speed data using contextual weather features (*28*).

In this work, we evaluated 16 imputation models - ranging from classical machine learning (Ridge, Random Forest, XGBoost, etc.) to state-of-the-art deep generative architectures (CSDI and FGTI). The evaluation was performed on a seasonal block-missing traffic speed dataset, compiled from fine-grained, CV data from Buffalo, NY. More details about the dataset, the models configuration, and the design of a three-round seasonal evaluation protocol, are described in the next section.

DATA & MODELS

Data Description

Speed data

CV speed data were compiled for the period from September 27, 2022, to March 31, 2023, for the Buffalo, NY metropolitan region. The CV dataset contained high-resolution GPS traces, from anonymized passenger vehicles, including instantaneous speed measurements, geographic coordinates, and timestamps. To identify the road segments upon which the vehicles were travel-

ing, the CV dataset was matched to OpenStreetMap (OSM) identifiers. After the matching, each observation in the resulting dataset included vehicle trajectory information, speed measurements in miles per hour (mph), and road characteristics such as posted speed limits and highway classifications.

For this analysis, we extracted speed observations exclusively for freeway segments (with posted speed limit of 55 mph) to focus on freeway travel conditions. Raw CV speed data were temporally averaged into 10-minute intervals, to match the temporal resolution of the auxiliary weather dataset. The aggregated speed data are shown in Figure 1. As illustrated in the figure, there are two substantial block-missing intervals: one spans from October 28, 2022 to December 11, 2022 and the other extends from December 23, 2022 through March 1, 2023. These two extended block-missing intervals pose serious problems for forecasting-based inference.

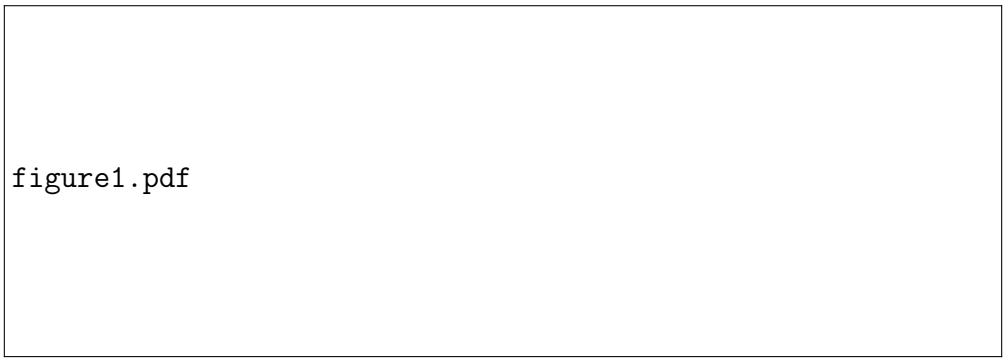

FIGURE 1 Observed 10-minute average freeway speed data with block-missing gaps.

To further understand the temporal structure of the speed data, we plotted the average speed trend by time of day for each two-week segment, as shown in Figure 2. On average, the December curve for the period from December 12 to December 25 (green) lies approximately 6-8 mph lower than those in October and March, which suggests a pervasive speed reduction pattern. Such reductions are likely caused by the historic Buffalo Blizzard of December 23-26, 2022. Moreover, speeds along these segments exhibit highly repeatable bimodal curves, highlighting consistent morning (7-9 AM) and evening (4-7 PM) rush-hour dips.

To investigate weekday-specific intraday variation, we averaged speed data at 10-minute intervals for each weekday. These weekday-averaged 10-minute profiles are plotted in Figure 3.Averaging preserves the original temporal resolution while smoothing random fluctuations and revealing stable diurnal trends: notably, distinct morning (7-9 AM) and evening rush hours (4–7 PM) speed dips characteristic of freeway traffic cycles.

The above observations imply that "hour-of-day" and "day-of-week" are informative predictors for freeway speed imputation, and their effects are cyclical. To enable models to capture these cyclical patterns smoothly—especially around midnight transitions and weekday/weekend boundaries — we encode the time features using sine and cosine transformations.

Weather data

Weather data were obtained from the Road Weather Information System (RWIS), installed at the University at Buffalo north campus. The RWIS dataset provides 20 road surface and atmospheric variables collected at 10-minute intervals, synchronized with the speed data temporal

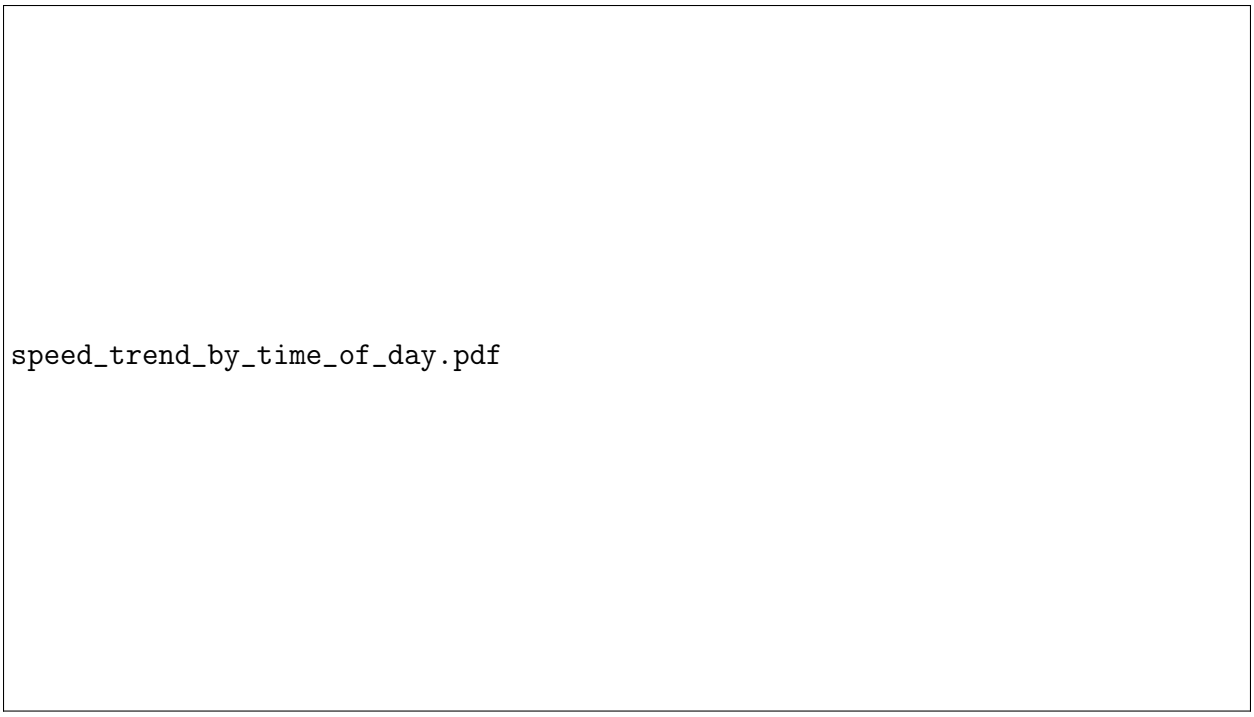

FIGURE 2 Average freeway speed profiles across different 2-week periods, plotted by time of day (10-minute bins). Each curve represents a distinct temporal window.

FIGURE 3 Weekday-averaged speed profiles by time of day.

resolution. Key variables include surface temperature, pavement conditions (dry, wet, icy), precipitation intensity, wind speed, air temperature, etc. All features were converted into numerical form for model input: continuous variables (e.g., wind speed, rainfall rate) were used directly; categorical pavement labels (Dry/Wet/Icy) were encoded ordinally (Dry = 1 $\rightarrow$ Wet = 2 $\rightarrow$ Ice or Snow = 3), reflecting increasing slipperiness; we also tested one-hot encoding in ablations. This encoding facilitates tree-based and neural models while preserving the ordinal nature of pavement condition in safety-sensitive settings.

Figure 4 shows Pearson correlations between observed vehicle speed and each of the 20 weather variables. As expected, adverse pavement conditions — such as ice and snow layers — are strongly negatively correlated with traffic speed, with coefficients nearing –0.8. Wind speed and

rolling precipitation averages over various time windows also show moderate negative associations. In contrast, air and surface temperatures display a positive correlation with speed. These results confirm that weather dynamics provide meaningful signals for reconstructing missing speed data, particularly under block-missing scenarios where temporal continuity is lost.

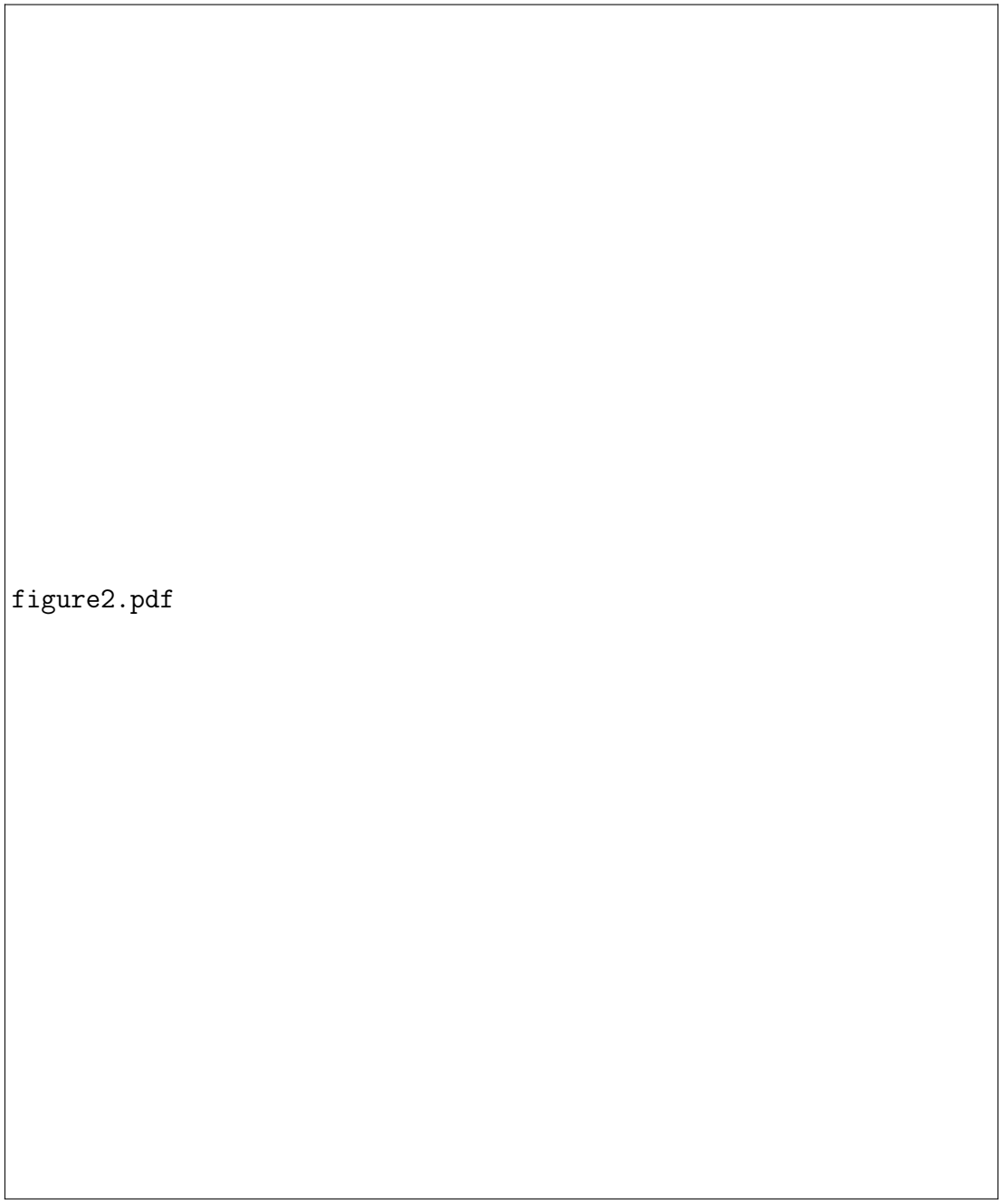

FIGURE 4 Pearson correlation between observed vehicle speed and weather variables.

Models

In this work, we compare 16 different models across three major methodological categories: linear regression, ensemble decision trees, and deep generative models(*5*).

Linear Regression Models

Linear regression models provide transparent and interpretable baselines. We evaluate Ridge, Lasso, and ElasticNet regression, each modeling traffic speed as a linear combination of predictors (e.g., weather features and encoded temporal variables). Ridge applies L2 regularization to mitigate multicollinearity, Lasso uses L1 regularization to promote sparsity, and ElasticNet combines both. These models are trained on observed samples and deployed on masked periods without temporal dynamics.

Ensemble Tree-Based Models

To capture nonlinear features between speed and auxiliary variables, we include ensemble decision tree models: Random Forest (RF), Extra Trees, Gradient Boosting Decision Trees (GBDT), AdaBoost, Histogram-Based Gradient Boosting (HistGB), and XGBoost(*5*). Bagging-based methods can (RF, Extra Trees) improve stability by averaging over multiple random tree constructions, while boosting methods iteratively reduce residuals to enhance fit. These models have shown strong empirical performance for structured tabular prediction tasks.

Deep Generative Models

We further evaluate two representative diffusion-based deep generative models for time series imputation: Conditional Score-based Diffusion Imputation (CSDI) (*27*) and Frequency-aware Generative Time-series Imputation (FGTI) (*28*). These models were originally designed for multivariate time series with randomly missing values, and both employ distinct masking strategies aligned with different learning paradigms.

CSDI follows an in-mini-batch masking strategy during training, in which random missing patterns are dynamically introduced in each batch. This design enhances generalization by exposing the model to a wide variety of missingness configurations across features and time steps (*29*). The model is trained to condition on partially observed features and infer the missing ones through a reverse diffusion process that integrates temporal information and observed conditions (*27*).In our study, we adapt CSDI for univariate block-missing scenarios by modifying the masking mechanism to target only the speed feature in long contiguous blocks (e.g., full days), while retaining weather and calendar variables as fully observed conditional inputs. This allows the model to focus on reconstructing temporally structured gaps in a single target variable while preserving its original dynamic masking framework.

FGTI, on the other hand, is trained using a pre-masking approach, where missing values are defined prior to training and kept fixed across epochs (*29*). This allows for explicit modeling of structured gaps, such as block-missing segments, and supports finer control over evaluation consistency. The model fuses frequency-domain (via Fourier transforms) and time-domain (via Transformer encoders) representations, enabling it to capture both long-term trends and local temporal patterns.While the original FGTI is designed for multivariate imputation, we modify it for our task by enforcing univariate prediction of speed using multivariate conditioning on weather and temporal features. This univariate adaptation ensures comparability with other models and aligns with the practical demands of highway speed recovery.

RESULTS AND DISCUSSION

This section evaluates the robustness of the imputation models under diverse seasonal conditions using a three–round experimental protocol that simulates winter-specific block–missing scenarios. We first summarize the experimental setup and evaluation metrics, and then present the round–wise quantitative and structural results.

Experimental Setup and Evaluation Metrics

Three–round evaluation protocol

To stress–test the models under different winter regimes, we designed a three–round evaluation protocol in which each round withholds two or more non–overlapping periods, as summarized in Table 1. The masked periods are strategically selected from different stages of the cold season, including late October, early March, and two December segments. December is further divided into two evaluation windows—Dec 12–18 (early winter) and Dec 19–25 (peak winter)—to assess how models generalize across varying levels of winter severity.

In each round, all models are independently trained on the corresponding training periods and evaluated on the withheld periods. This design allows us to compare model performance within each round (i.e., under a specific train–test split) and to examine how their behavior changes when the masking pattern and winter conditions vary across the three rounds.

TABLE 1 Summary of the training and evaluation periods for each round.

Round	Training Period	Evaluation Period
1	Sep 28–Oct 13, Dec 12–25, Mar 16–30	Oct 14–28, Mar 1–15
2	Sep 28–Oct 13, Dec 19–25, Mar 16–30	Oct 14–28, Dec 12–18, Mar 1–15
3	Sep 28–Oct 13, Dec 12–18, Mar 16–30	Oct 14–28, Dec 19–25, Mar 1–15

Error–based performance metrics

When ground–truth speed observations are available for the evaluation periods, we quantify pointwise reconstruction accuracy using four standard metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). Let y_t and $\hat{y}_t$ denote the ground–truth and imputed speeds at time step t, respectively, and let T be the number of evaluated time steps. The metrics are defined as

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{T}\left|y_t - \hat{y}_t\right|, \tag{1}$$

$$\mathrm{MSE} = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{y}_t\right)^2, \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}, \tag{3}$$

$$\mathrm{MAPE}(\%) = \frac{100}{T}\sum_{t=1}^{T}\left|\frac{y_t - \hat{y}_t}{y_t}\right|. \tag{4}$$

These metrics are reported for all three rounds and for all models in Table 2. They enable a direct comparison of numerical accuracy across methods and seasonal scenarios.

Structural consistency via weather–speed correlations

Beyond pointwise errors, we also assess whether the imputed speeds preserve the physical relationships between traffic speed and weather. For each model, we compute the Pearson correlation between the imputed speed series and each weather feature (F0–F19) over the evaluation periods. This yields a correlation vector $\rho^{(m)}$ for model m, which we compare to the ground–truth correlation vector $\rho^{(\mathrm{gt})}$. Structural consistency is quantified by the Euclidean distance $\|\rho^{(m)} - \rho^{(\mathrm{gt})}\|_2$; smaller values indicate better preservation of the underlying weather–speed dependencies. Figures 6, 8, and 10 visualize these correlation profiles and list models in ascending order of distance from the ground truth.

TABLE 2 Comparison of models based on MAE, MSE, RMSE, and MAPE (%).

Round	**Metric**	**FGTI**	GBDT	HISTGB	RF	XGB	EXTRA	RIDGE	BAYES	LR	LASSO	SVR	ELASTIC	MLP	KNN	ADA	**CSDI**
Round 1	MAE	**0.703**	1.877	1.900	1.961	2.056	2.057	2.058	2.065	2.066	2.082	2.179	2.207	2.348	2.623	3.078	**4.142**
	MSE	**2.610**	10.914	12.239	12.284	13.224	12.474	11.941	12.085	12.135	11.203	19.158	10.594	13.188	21.872	17.778	31.023
	RMSE	**1.616**	3.304	3.498	3.505	3.636	3.532	3.456	3.476	3.483	3.347	4.377	3.255	3.631	4.677	4.216	5.568
	MAPE (%)	**1.28**	3.424	3.469	3.557	3.737	3.717	3.735	3.747	3.747	3.772	4.313	3.966	4.202	4.952	5.363	7.207
Model	**Metric**	**FGTI**	EXTRA	HISTGB	XGB	GBDT	RF	BAYES	LR	LASSO	RIDGE	SVR	ELASTIC	KNN	ADA	MLP	**CSDI**
Round 2	MAE	**0.917**	1.559	1.578	1.609	1.629	1.648	1.925	1.936	1.938	1.949	2.034	2.047	2.286	2.327	2.355	**6.353**
	MSE	**4.535**	8.329	9.251	8.375	9.544	9.194	11.981	12.251	11.992	12.214	13.552	10.170	18.565	11.169	18.348	68.801
	RMSE	**2.129**	2.886	3.042	2.894	3.089	3.032	3.461	3.500	3.463	3.495	3.681	3.189	4.309	3.342	4.283	8.294
	MAPE (%)	**1.695**	2.882	2.957	2.960	3.040	3.057	3.566	3.588	3.584	3.609	3.781	3.683	4.221	4.132	4.300	10.782
Model	**Metric**	**FGTI**	LR	BAYES	RIDGE	RF	HISTGB	EXTRA	GBDT	XGB	ELASTIC	LASSO	**CSDI**	SVR	MLP	KNN	ADA
Round 3	MAE	**1.575**	2.967	3.054	3.056	3.121	3.204	3.231	3.269	3.393	3.497	3.590	**3.765**	3.778	3.783	3.979	4.024
	MSE	**23.623**	32.849	34.602	34.627	43.628	49.814	45.347	53.202	54.100	51.590	55.703	64.237	68.745	67.850	60.744	52.799
	RMSE	**4.860**	5.731	5.882	5.885	6.605	7.058	6.734	7.294	7.355	7.183	7.463	8.014	8.291	8.237	7.794	7.266
	MAPE (%)	**4.121**	7.487	7.733	7.738	8.249	8.623	8.506	8.840	9.082	9.171	9.467	10.135	10.273	10.091	10.143	9.915

Round 1

In Round 1, models are trained on speed and weather data from early autumn (Sep 28–Oct 13), peak winter (Dec 12–25), and early spring (Mar 16–30), and evaluated on mid–autumn (Oct 14–28) and early March (Mar 1–15). As shown in Table 2, FGTI achieves the best overall accuracy, with the lowest MAE (0.703) and RMSE (1.616) among all models. Ensemble tree–based methods, including GBDT (MAE: 1.877), HistGB (MAE: 1.900), and Random Forest (MAE: 1.961), also yield competitive performance.

Figure 5 presents the daily imputation results in the evaluation dataset. FGTI captures day–to–day fluctuations with remarkable fidelity, producing smooth and visually coherent reconstructions. GBDT and ElasticNet provide reasonable approximations, though with some over–smoothing of peak transitions. Although CSDI captures the overall trend, it exhibits greater instability and consistently underestimates speeds, often missing key temporal shifts in speed dynamics.

The KNN imputation results (gold line) are characterized by substantial fluctuations and inconsistent trends, especially during nighttime periods on March 10. This behavior illustrates a fundamental limitation of traditional statistical imputation techniques such as KNN, which rely on local similarity and perform poorly when extended blocks of data are missing. Without access to recent contextual information, KNN lacks the capacity to interpolate coherent temporal dynamics, rendering it unsuitable for block–missing highway speed recovery.

The structural consistency analysis in Figure 6 further highlights FGTI's advantage in preserving speed–weather correlations, ranking first with the smallest Euclidean distance to the

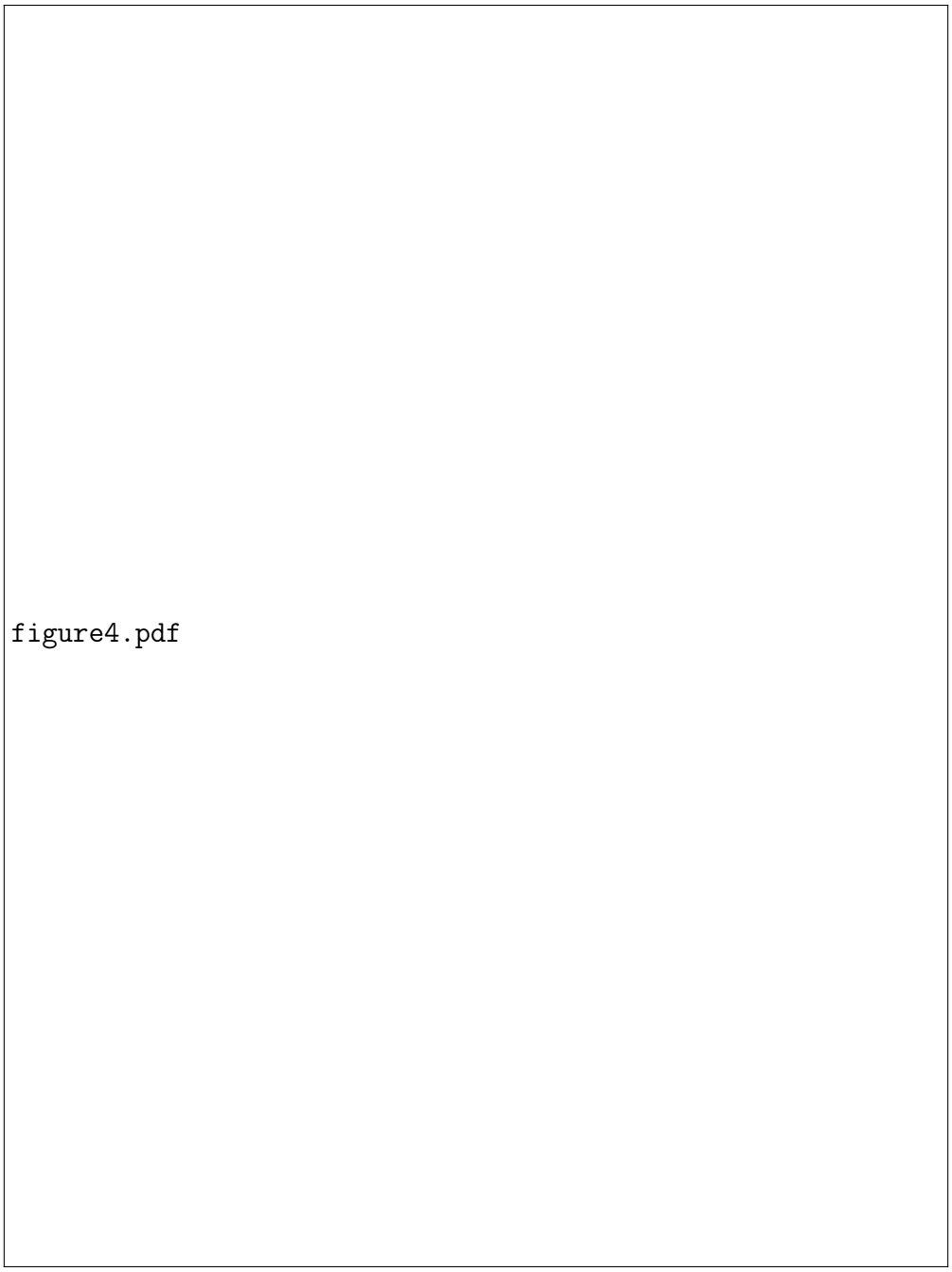

FIGURE 5 Daily imputation results for each evaluation day in Round 1. Each subplot shows the observed ground truth (orange), alongside imputed speed values from six models: CSDI (dark green), ElasticNet (purple), GBDT (blue), FGTI (red dots), KNN (gold), and Random Forest (grey). The horizontal axis represents time of day (00:00–24:00), and the vertical axis denotes vehicle speed (mph).

ground truth. ElasticNet and Lasso follow as second and third, demonstrating that linear models can effectively capture these relationships despite their simplicity. In contrast, SVR, KNN, and CSDI fail to maintain the physical relationships between speed and weather, exhibiting notable

discrepancies and suggesting limited capacity to model latent weather–speed dependencies.

FIGURE 6 Pearson correlation between weather features and imputed speed for each model in Round 1. The x–axis denotes weather feature indices (F0–F19), and the y–axis indicates the correlation coefficient. Models are listed in the legend in order of increasing Euclidean distance to the ground–truth correlation profile.

Round 2

Round 2 focuses on testing model performance for mild–winter conditions by withholding early December (Dec 12–18) in addition to late October and early March. The training window now includes only limited exposure to mild winter. Models are trained on early autumn (Sep 28–Oct 13), late winter (Dec 19–25), and early spring (Mar 16–30), and evaluated on mid–autumn (Oct 14–28), early March (Mar 1–15), and mild winter (Dec 12–18). The performance impact varies dramatically across models. FGTI maintains stable performance (MAE: 0.917), demonstrating robustness when the target conditions fall within the range of its training data. Conversely, CSDI's MAE degrades to 6.353—a 53% increase relative to Round 1. Interestingly, ensemble methods show improved performance, with Extra Trees achieving its best MAE of 1.559 and HistGB reaching 1.578.

Figure 7 reveals specific temporal patterns in model behavior. KNN continues to perform poorly with erratic predictions. Extra Trees, despite good overall metrics, exhibits systematic underestimation during nighttime hours (after 18:00), particularly visible on October 17–18, December 17, and March 10. This suggests that the model struggles to capture reduced nighttime traffic patterns when trained without representative mild–winter examples. ElasticNet produces more stable predictions across all hours, while FGTI maintains superior temporal smoothness and accurately captures slope transitions during peak hours.

The correlation–structure analysis in Figure 8 shows FGTI maintaining its top ranking, followed by ElasticNet and Lasso. The stability of linear models in preserving correlations, even under challenging training conditions, highlights their robustness. CSDI's continued poor ranking (14th), combined with its degraded MAE, suggests fundamental difficulties with domain shift.

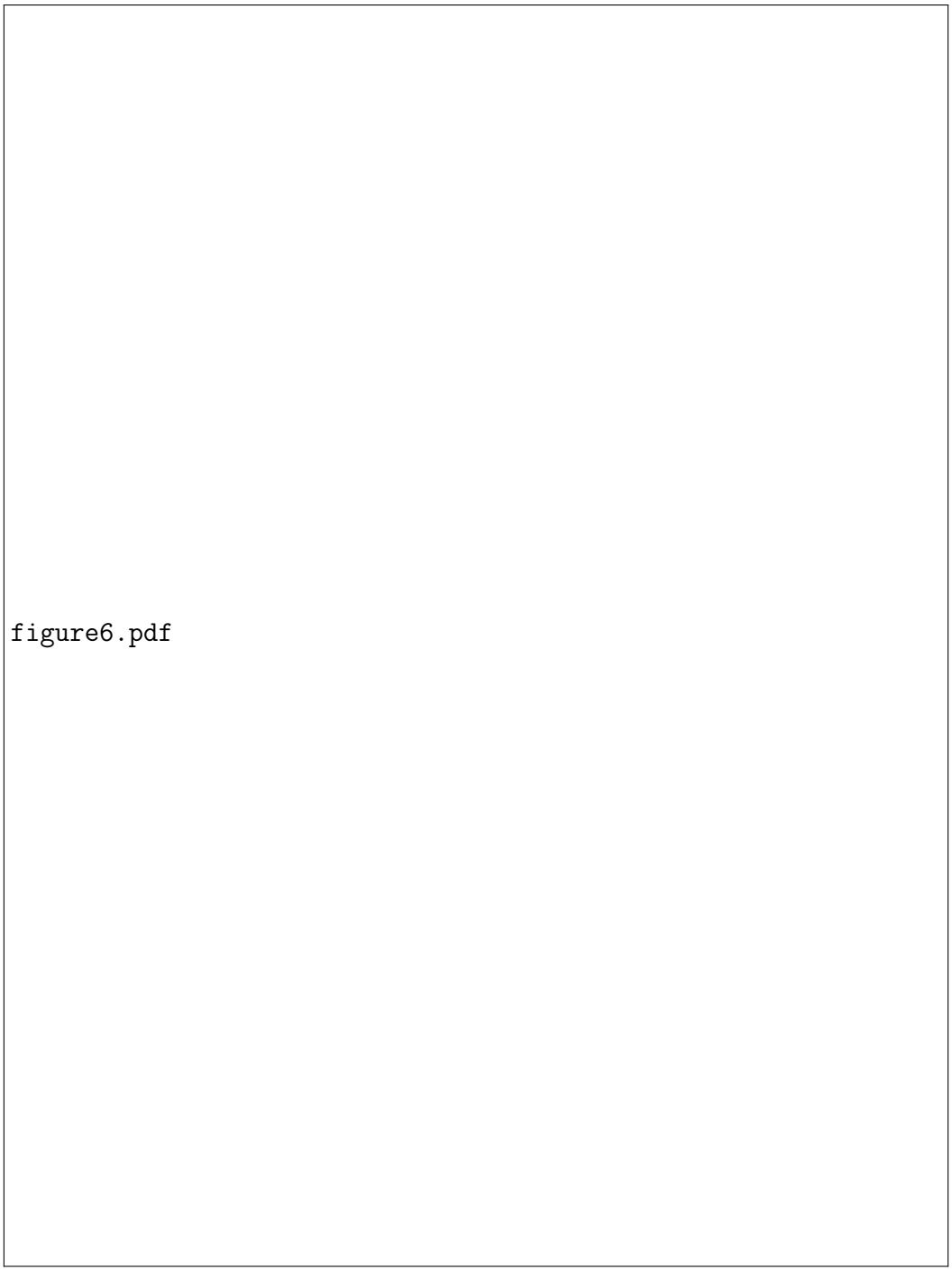

FIGURE 7 Daily imputation results for each evaluation day in Round 2. Each subplot shows the observed ground truth (orange), alongside imputed speed values from four models: CSDI (dark green), Extra Trees (blue), ElasticNet (purple), and FGTI (red dots). The horizontal axis represents time of day (00:00–24:00), and the vertical axis denotes vehicle speed (mph).

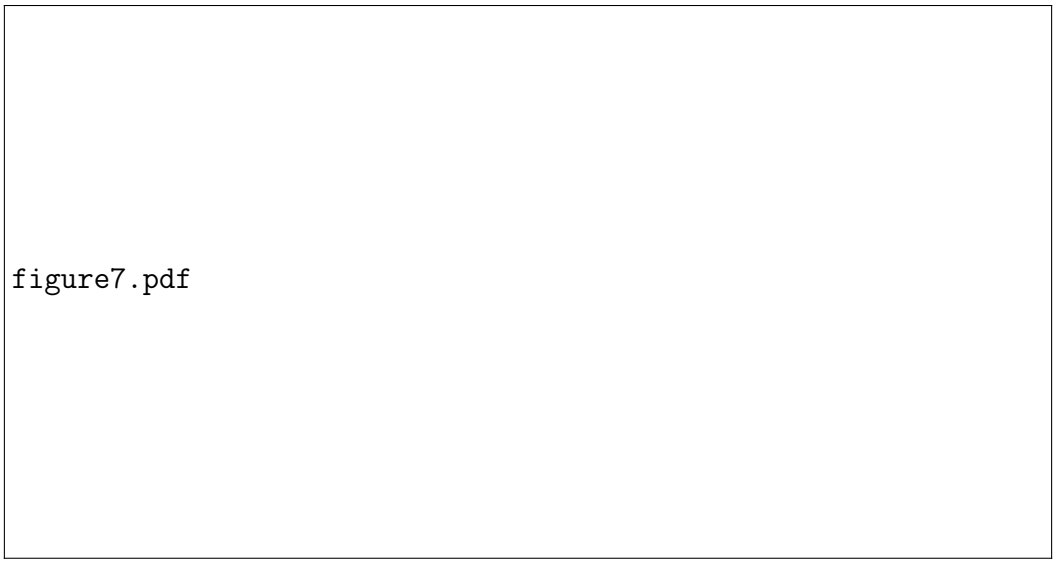

FIGURE 8 Pearson correlation between weather features and imputed speed for each model in Round 2. The x–axis denotes weather feature indices (F0–F19), and the y–axis indicates the correlation coefficient. The black line shows the ground truth, and other lines represent different imputation models. Models are listed in the legend in order of increasing Euclidean distance to the ground–truth correlation profile.

Round 3

Round 3 reverses the scenario by training on milder conditions (including Dec 12–18) and testing on extreme winter (Dec 19–25). This simulates the critical real–world challenge of predicting severe–weather impacts when historical data contain only moderate conditions. As shown in Table 2, this is the most challenging scenario: all models exhibit substantial performance degradation compared with the first two rounds, except for CSDI. Most models show several–fold increases in MSE relative to Round 2. Even our top–performing model in earlier rounds, FGTI, experiences noticeably reduced performance (MAE: 1.575).

Figure 9 shows the imputation results for each day during the extreme–winter period. FGTI exhibits the poorest alignment with ground truth throughout this challenging period—a logical consequence of its frequency–based learning paradigm encountering patterns outside its learned spectrum. Bayesian Ridge and Linear Regression, despite their simplicity, successfully capture the low–speed patterns during peak storm conditions. CSDI shows improved visual coherence compared with Round 2, suggesting that its diffusion–based approach may better handle uncertainty in extreme conditions.

The correlation analysis in Figure 10 reveals a dramatic shift: CSDI rises to second place in terms of structural consistency, indicating that it better preserves the weather–speed relationships when trained on mild conditions and tested on extreme ones. This improvement, combined with its reasonable MAE (3.765), suggests that CSDI's generative approach may be more suited for mild–to–extreme extrapolation than for extreme–to–mild interpolation.

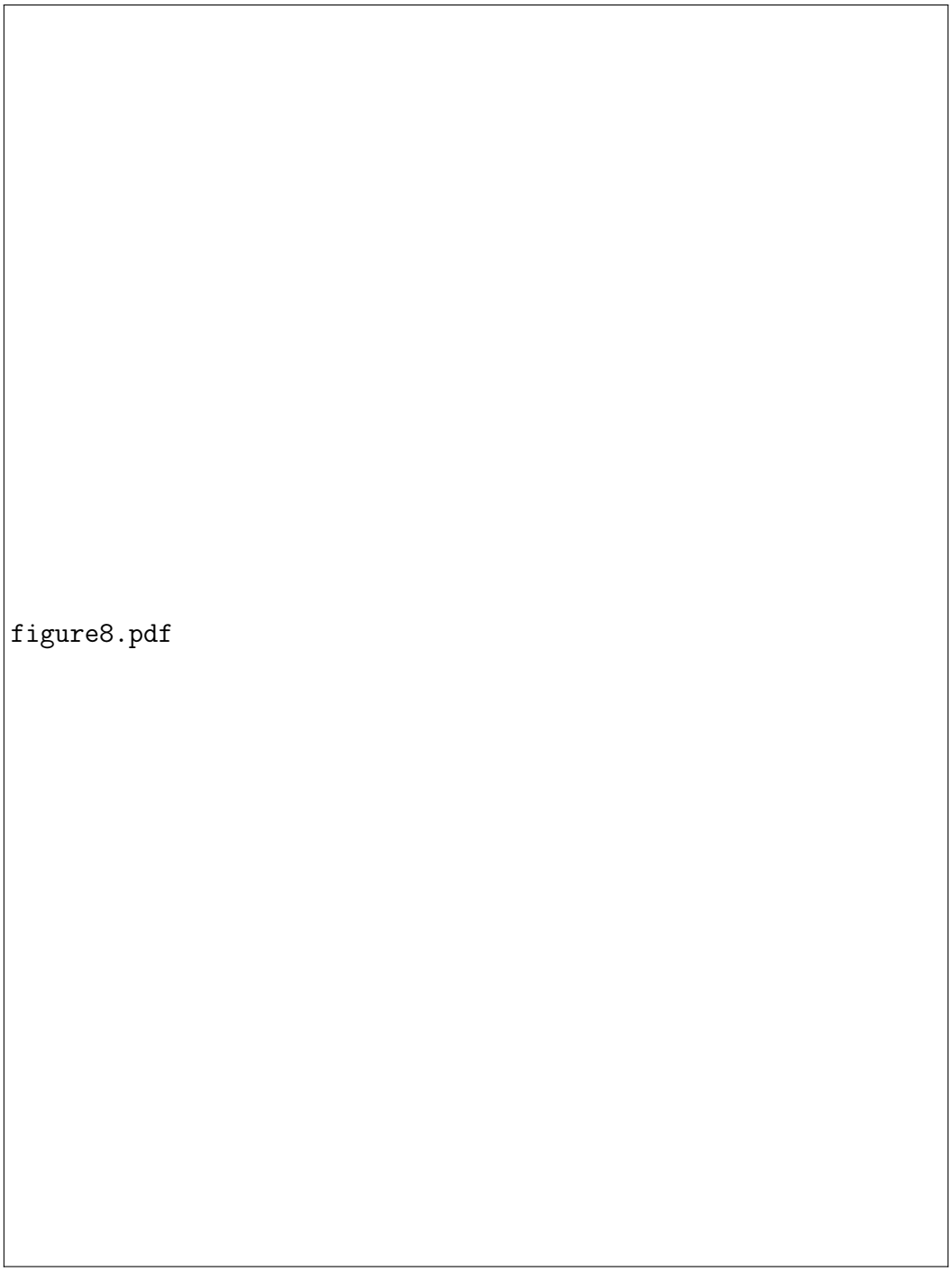

FIGURE 9 Daily imputation results for each evaluation day in Round 3. Each subplot shows the observed ground truth (orange), alongside imputed speed values from four models: CSDI (dark green), Bayesian Ridge (blue), Linear Regression (purple), and FGTI (red dots). The horizontal axis represents time of day (00:00–24:00), and the vertical axis denotes vehicle speed (mph).

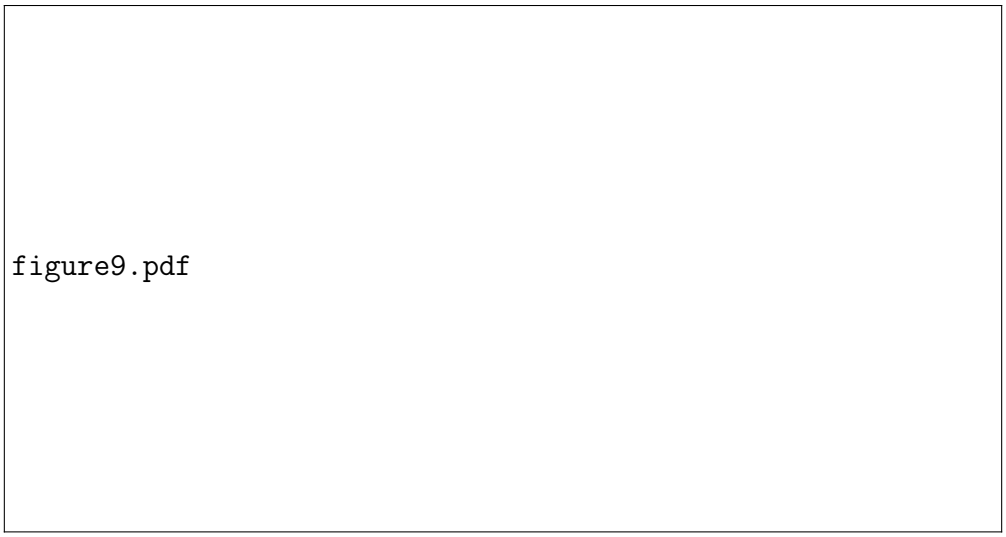

FIGURE 10 Pearson correlation between weather features and imputed speed for each model in Round 3. The x–axis denotes weather feature indices (F0–F19), and the y–axis indicates the correlation coefficient. The black line shows the ground truth, and other lines represent different imputation models. Models are listed in the legend in order of increasing Euclidean distance to the ground–truth correlation profile.

Cross-round comparison of average performance

To summarize overall robustness across seasons, we compute the average MAE, MAPE, MSE, and RMSE over all three rounds for each model (Table 3). We also report an aggregate ranking of structural consistency based on the three-round average Pearson correlations between imputed speed and weather features (Table 4)..

The cross-round averages clearly identify FGTI as the overall winner. It achieves the lowest MAE, MAPE, MSE, and RMSE among all models, indicating both high accuracy and stable behavior across different seasonal configurations. In other words, FGTI not only attains the best typical-case performance, but also maintains relatively small errors in the more challenging tail scenarios.

A second tier of methods is formed by randomized tree ensembles such as Extra Trees and Random Forest. Their average errors remain close to those of FGTI, and they exhibit strong robustness to seasonal shifts and noise. These models are attractive candidates for production deployment or as components in hybrid ensembles, especially when interpretability and training stability are priorities. By contrast, boosted trees (HistGB, XGBoost, and Gradient Boosting) are more sensitive to time-split changes and hyperparameter tuning; their performance can degrade more rapidly when the training and evaluation periods differ.

Linear models (Ridge, Lasso, ElasticNet, and ordinary Linear Regression) deliver consistent, mid-range performance. They effectively capture additive trends and first-order weather–speed relationships, which explains their relatively strong structural consistency in the correlation analyses. However, their inability to represent complex nonlinear interactions limits their accuracy under extreme weather or highly nonlinear traffic responses.

Finally, other deep and neighbor-based models underperform in this setting. The Multilayer Perceptron (MLP), SVR, KNN, and CSDI are all adversely affected by the modest data volume, the

TABLE 3 Average error-based performance of all models over the three rounds.

Model	MAE	MAPE (%)	MSE	RMSE
FGTI	**1.065**	**2.365**	**10.256**	**2.868**
HISTGB	2.227	5.016	23.768	4.533
RF	2.243	4.954	21.702	4.381
GBDT	2.258	5.101	24.553	4.562
EXTRA	2.282	5.035	22.050	4.384
LR	2.323	4.941	19.078	4.238
BAYES	2.348	5.015	19.556	4.273
XGB	2.353	5.260	25.233	4.628
RIDGE	2.354	5.027	19.594	4.279
LASSO	2.537	5.608	26.299	4.758
ELASTIC	2.584	5.607	24.118	4.542
SVR	2.664	6.122	33.818	5.450
MLP	2.829	6.198	33.129	5.384
KNN	2.963	6.439	33.727	5.593
ADA	3.143	6.470	27.249	4.941
CSDI	4.753	9.375	54.687	7.292

TABLE 4 Average Pearson correlation performance of all models over the three rounds.

Model	Rank
FGTI	1
ElasticNet	2
AdaBoost	3
Extra Trees	4
Lasso	5
Random Forest	6
Gradient Boosting	7
Ridge	8
Linear Regression	9
Bayesian Ridge	10
Multilayer Perceptron	11
Histogram GBDT	12
XGBoost	13
CSDI	14
KNN	15
Support Vector Regression	16

high dimensionality of weather features, and, in the case of CSDI, a mismatch between the original modeling objective and the block-missing imputation task considered here. Their comparatively large average errors and lower rankings underscore the challenge of applying more complex models when the training data do not sufficiently cover the full range of rare but high-impact events.

CONCLUSIONS

This study addresses the critical challenge of imputing missing vehicle speed data in highway segments, particularly under block-missing scenarios where entire months of data are absent. We curated a dataset from Buffalo freeway segments—including high-resolution RWIS weather variables and encoded cyclical time features (hour-of-day and day-of-week). We applied a combination of classical regression models, ensemble tree-based methods, and advanced deep generative models to impute the missing values.

Our results demonstrate that FGTI consistently achieves superior performance, offering the lowest MAE and maintaining structural fidelity between imputed speeds and weather correlations—even under complex block-missing patterns. Its frequency-aware architecture allows it to reconstruct both trend and residual components effectively.

Linear regression methods and ensemble tree models offered reliable and interpretable baselines. Although less precise than FGTI, tree-based learners often matched or exceeded its performance in mild conditions—especially when speed–weather dependencies were strong and the training distribution covered the test scenarios. On the other hand, CSDI—despite its success in healthcare and environmental imputation—generally underperformed compared to both linear and tree methods in our block-missing setting

All models performed poorly during December 22–24, 2022, coinciding with the Buffalo Blizzard. This underscores a fundamental limitation: when system behavior deviates beyond the range of historical variability, no data-driven imputation model can reliably reconstruct the missing state.

Future extensions of this study may include incorporating graph neural networks to capture spatial dependencies in regional weather–traffic dynamics, leveraging vehicle-level speed distributions, and using milespot-based segmentation to partition freeway corridors into finer segments, thereby enabling more precise speed imputation across different roadway sections.

REFERENCES

1. Guide, T. M., Traffic monitoring guide. *Federal Highway Administration*, 2001.
2. Wang, H., X. Tang, Y.-H. Kuo, D. Kifer, and Z. Li, A simple baseline for travel time estimation using large-scale trip data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 10, No. 2, 2019, pp. 1–22.
3. Zhu, L., F. R. Yu, Y. Wang, B. Ning, and T. Tang, Big data analytics in intelligent transportation systems: A survey. *IEEE transactions on intelligent transportation systems*, Vol. 20, No. 1, 2018, pp. 383–398.
4. Najafi, B., S. Parsaeefard, and A. Leon-Garcia, Estimation of missing data in intelligent transportation system. In *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, IEEE, 2020, pp. 1–6.
5. Zhang, Y., X. Kong, W. Zhou, J. Liu, Y. Fu, and G. Shen, A comprehensive survey on traffic missing data imputation. *IEEE Transactions on Intelligent Transportation Systems*, 2024.
6. Desai, J., J. K. Mathew, H. Li, R. S. Sakhare, D. Horton, and D. M. Bullock, Analysis of Connected Vehicle Data to Quantify National Mobility Impacts of Winter Storms for Decision Makers and Media Reports. *Future Transportation*, Vol. 3, No. 4, 2023, pp. 1292–1309.
7. Smith, B. L., W. T. Scherer, and J. H. Conklin, Exploring imputation techniques for missing data in transportation management systems. *Transportation Research Record*, Vol. 1836, No. 1, 2003, pp. 132–142.
8. Li, Y., Z. Li, and L. Li, Missing traffic data: comparison of imputation methods. *IET Intelligent Transport Systems*, Vol. 8, No. 1, 2014, pp. 51–57.
9. Chan, R. K. C., J. M.-Y. Lim, and R. Parthiban, Missing traffic data imputation for artificial intelligence in intelligent transportation systems: review of methods, limitations, and challenges. *Ieee Access*, Vol. 11, 2023, pp. 34080–34093.
10. Bae, B., H. Kim, H. Lim, Y. Liu, L. D. Han, and P. B. Freeze, Missing data imputation for traffic flow speed using spatio-temporal cokriging. *Transportation Research Part C: Emerging Technologies*, Vol. 88, 2018, pp. 124–139.
11. El Esawey, M., Using spatio-temporal data for estimating missing cycling counts: a multiple imputation approach. *Transportmetrica A: transport science*, Vol. 16, No. 1, 2020, pp. 5–22.
12. Kyrtsoglou, A., D. Asimina, D. Triantafyllidis, S. Krinidis, K. Kitsikoudis, D. Ioannidis, S. Antypas, G. Tsoukos, and D. Tzovaras, Missing data imputation and meta-analysis on correlation of spatio-temporal weather series data. In *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, 2021, pp. 0496–0502.
13. Rubin, D. B., Inference and missing data. *Biometrika*, Vol. 63, No. 3, 1976, pp. 581–592.
14. Kang, H., The prevention and handling of the missing data. *Korean journal of anesthesiology*, Vol. 64, No. 5, 2013, pp. 402–406.
15. Haji-Maghsoudi, S., A.-a. Haghdoost, A. Rastegari, and M. R. Baneshi, Influence of pattern of missing data on performance of imputation methods: an example using national data on drug injection in prisons. *International journal of health policy and management*, Vol. 1, No. 1, 2013, p. 69.

16. Aljuaid, T. and S. Sasi, Proper imputation techniques for missing values in data sets. In *2016 international conference on data science and engineering (ICDSE)*, IEEE, 2016, pp. 1–5.
17. Liang, Y., Z. Zhao, and L. Sun, Dynamic spatiotemporal graph convolutional neural networks for traffic data imputation with complex missing patterns. *arXiv preprint arXiv:2109.08357*, 2021.
18. Xing, J., W. Wu, Q. Cheng, and R. Liu, Traffic state estimation of urban road networks by multi-source data fusion: Review and new insights. *Physica A: Statistical Mechanics and its Applications*, Vol. 595, 2022, p. 127079.
19. Duruflé, H., M. Selmani, P. Ranocha, E. Jamet, C. Dunand, and S. Déjean, A powerful framework for an integrative study with heterogeneous omics data: from univariate statistics to multi-block analysis. *Briefings in Bioinformatics*, Vol. 22, No. 3, 2021, p. bbaa166.
20. Umar, N. and A. Gray, Comparing single and multiple imputation approaches for missing values in univariate and multivariate water level data. *Water*, Vol. 15, No. 8, 2023, p. 1519.
21. Nevalainen, J., M. G. Kenward, and S. M. Virtanen, Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification. *Statistics in medicine*, Vol. 28, No. 29, 2009, pp. 3657–3669.
22. Lee, K. J. and J. B. Carlin, Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *American journal of epidemiology*, Vol. 171, No. 5, 2010, pp. 624–632.
23. Chaudhry, A., W. Li, A. Basri, and F. Patenaude, A method for improving imputation and prediction accuracy of highly seasonal univariate data with large periods of missingness. *Wireless Communications and Mobile Computing*, Vol. 2019, No. 1, 2019, p. 4039758.
24. Henrickson, K., Y. Zou, and Y. Wang, Flexible and robust method for missing loop detector data imputation. *Transportation Research Record*, Vol. 2527, No. 1, 2015, pp. 29–36.
25. Tak, S., S. Woo, and H. Yeo, Data-driven imputation method for traffic data in sectional units of road links. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 6, 2016, pp. 1762–1771.
26. Kwon, D., C. Lee, H. Kang, and I. Kim, Large-scale network imputation and prediction of traffic volume based on multi-source data collection system. *Transportation research record*, Vol. 2677, No. 9, 2023, pp. 30–42.
27. Tashiro, Y., J. Song, Y. Song, and S. Ermon, Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in neural information processing systems*, Vol. 34, 2021, pp. 24804–24816.
28. Yang, X., Y. Sun, X. Chen, et al., Frequency-aware generative models for multivariate time series imputation. *Advances in Neural Information Processing Systems*, Vol. 37, 2024, pp. 52595–52623.
29. Qian, L., Z. Ibrahim, W. Du, Y. Yang, and R. Dobson, Unveiling the secrets: How masking strategies shape time series imputation. *arXiv preprint arXiv:2405.17508*, 2024.