Responses to Comments for Paper T-ITS-23-12-3607

Dear Editors & Reviewers:

We would like to express our sincere gratitude to the editor for handling our manuscript "How to Accurately Forecast Highway Traffic Speed Using Limited Input Variables? A Self-Supervised Spatio-Temporal Bilateral Learning Network" and to the two anonymous reviewers for their insightful comments and valuable suggestions.

Based on all comments and suggestions, the authors conducted major revisions to the paper, which are summarized as follows:

- have modified the title of this work;
- have enhanced the abstraction, related work, approach, equations, experiments, and conclusion;
- have highlighted the novelties and contribution;
- have added more recent references to related work;
- have added the description of the "Temporal-Att" module to this paper;
- have standardized and italicized the variables and equations in the paper;
- have added a new dataset in the experiment;
- have added more baselines to the experiments;
- have updated comparisons and analysis in experiment;
- have extended the prediction steps from 12 to 48;
- have optimised the figures in the paper;
- have added the codes and parameters of the proposed model and baselines to my personal GitHub page (https://github.com/zouguojian/Traffic-speed-prediction/tree/main/3S-TBLN).

Please check the detailed responses for each comment below, which are marked in red color. All changes made to the manuscript are in yellow so that they will be easily identified. Should you have any questions, please contact us without hesitation. Once again, thank you very much for your work. We hope that the resubmitted manuscript is now suitable for publication in the IEEE Transactions on Intelligent Transportation Systems.

Sincerely,
Dr. Guojian Zou

We thank the Editor for organizing the review for our submission. We have carefully followed the reviewers' comments and prepared a major revision in this submission. The detailed comments are addressed point-to-point as follows. We hope that the Editor and the anonymous reviewers find this revision satisfactory.

Section #1: The entire editorial decision letter from Editor-in-Chief

Dear Mr. Guojian Zou,

The review of your paper, "How to Accurately Forecast Highway Traffic Speed Using Limited Input Variables? A Self-Supervised Spatio-Temporal Bilateral Learning Network," T-ITS-23-12-3607, has been completed. The reviewers' comments and those of the Associate Editor are copied below. Based on these comments and recommendation of the Associate Editor your paper is not ready for publication in its present form, A properly revised version that takes care of the concerns and drawbacks pointed out by the reviewers and Associate Editor is potentially publishable.

Therefore, I suggest that you revise your paper along the lines described by the reviewers and resubmit the paper. Please include a description on how you took into account the reviewers' comments in preparing your revision. We would then hope to determine a publication decision soon thereafter. Please note that if your revision is not submitted within the next 3 months (90 days), your paper will be treated as a new paper.

You may want to resubmit your paper as a REGULAR PAPER (suggested length: 10 Transactions pages, authors' biographies included). The Associate Editor will determine whether your paper is best suited to be considered as a Regular Paper or shortened to a Short Paper.

Thank you for submitting your manuscript to the Transactions on Intelligent Transportation Systems.

Sincerely,

Simona Sacone
Editor-in-Chief
Transactions on Intelligent Transportation Systems

Simona Sacone
Editor-in-Chief
IEEE Transactions on Intelligent Transportation Systems
Dept. of Computer Science, Bioengineering, Robotics and Systems Engineering,
University of Genova, Italy
simona.sacone@unige.it

Response: We appreciate the feedback and comments from the Reviewers, Associate Editor, and Editor-in-Chief regarding our paper, titled "How to Accurately Forecast Highway Traffic Speed Using Limited Input Variables? A Self-Supervised Spatio-Temporal Bilateral Learning Network," T-ITS-23-12-3607. Their insights and suggestions have been invaluable in guiding our revision process. We have meticulously addressed each of the concerns and drawbacks raised by the Reviewers and Editors. Specifically, we have revised the paper to enhance the clarity of technical details and ensure that the methods and experiments are described with precision. Additionally, we have updated the abstraction, related work, approach, equations, experiments, and conclusion to bolster the paper's significance and validity. We have incorporated revisions based on the Reviewers' suggestions, which are indicated in yellow in the resubmitted manuscript. Furthermore, we have provided detailed explanations in Sections #2 and #3 outlining how we have considered the Reviewers' comments in preparing the revised version of our paper. We are grateful for the invaluable feedback provided by the Reviewers and Editors. We extend special thanks to Simona Sacone (Editor-in-Chief) and Associate Editor for granting us the opportunity to revise and enhance our paper. We are dedicated to ensuring that it meets the rigorous standards of the Transactions on Intelligent Transportation Systems.

Section #2: The entire editorial decision letter from the Editor

Editor's comments:
Editor
Comments to the Author:

Some reviewers have concerns on the paper, in particular reviewer 2 with comments 7 and 8. I am willing to consider a revision that addresses all their comments.

Response: Thank you for your constructive feedback and for bringing the concerns of Reviewer 2 to our attention. We deeply value the input provided by the reviewers, and we are committed to addressing their comments thoroughly in the revised manuscript. In particular, we have carefully reviewed Comments 7 and 8 from Reviewer 2, and have already implemented significant improvements to address the raised concerns. (1) We have carefully reviewed the comments and concerns raised by Reviewers 1 and 2, and are making necessary revisions to improve clarity. We recognize the importance of addressing these issues to enhance the quality and impact of our manuscript. Incorporating the suggested revisions into our work is our top priority to ensure that the concerns raised are adequately addressed. Additionally, we have provided detailed responses to each of the reviewers' comments in the revised manuscript to demonstrate how we have addressed their concerns. We understand the significance of this major revision and are committed to ensuring that the revised manuscript meets the standards of excellence expected by the journal. Thank you again for your guidance and assistance. (2) Moreover, we kindly request the Editor to forward this response to Reviewer 2. We are grateful for the thoughtful review and constructive feedback provided by Reviewer 2. We have diligently addressed his/her concerns and made improvements to the manuscript following the first round of review. We would like to emphasize three key novelties of our work: multi-step traffic diffusion extraction, traffic pattern modeling, and reconstructing historical variables based on self-supervised learning. Additionally, our proposed model has been successfully implemented in a real-world system, providing travel services to traffic managers and travelers. Specifically, the model has been deployed in the smart highway platform in the Ningxia province, China. Over the past two years, we have received significant interest in our work, with many researchers subscribing to our GitHub repository (https://github.com/zouguojian/Traffic-speed-prediction) and contacting us via email. We hope that the reviewer will consider giving us an opportunity to share this work with a broader audience, enabling its application in more real-world highway networks.

We greatly appreciate your consideration of the revised manuscript and are hopeful that the revisions will resolve the concerns raised by the reviewers. Please let us know if further clarification or additional changes are needed. Thank you for your time and support.

Section #3: The entire editorial decision letter from reviewers 1 and 2

Reviewer: 1

Recommendation: Prepare A Major Revision For A New Review

Comments:
(There are no comments. Please check to see if comments were included as a file attachment with this e-mail or as an attachment in your Author Center.)

(1) What are the contributions of this paper?: The paper introduces a novel self-supervised spatio-temporal bilateral learning network (3S-TBLN) designed to enhance long-term traffic speed forecasting on highways. This model combines a semantic transformer, spatio-temporal blocks (ST-Blocks), and a bridge transformer (BridgeTrans) within a bilateral encoder-decoder architecture to capture complex spatial-temporal dynamics and bi-directional learning of traffic patterns.

Response: Thank you for taking the time to review our manuscript.

(2) What are some ways in which the paper could be improved? Please supply any additional important references that you feel the author omitted which should be noted in the paper.: The paper could be improved by incorporating external factors such as weather conditions, traffic incidents, and roadwork data, which significantly impact traffic flow and speed predictions. Integrating these variables could enhance the model's robustness and accuracy in real-world scenarios. Advanced models like 3S-TBLN can often become "black boxes," where the decision-making process is not transparent. Authors should talk about interpretability of the model, possibly by integrating techniques that explain the model's predictions.

Response: We sincerely appreciate your valuable time and insightful comments, which have significantly helped us refine our work. Below, we address your concerns in detail.

1. Incorporating External Factors (Weather, Traffic Incidents, and Roadwork Data)

The reviewer is correct in noting that external data, such as meteorological and accident information, can positively influence traffic prediction tasks [7]. However, as mentioned in this paper, our focus is primarily on traffic data, excluding external factors. This decision stems from practical challenges, such as the difficulty of obtaining meteorological data for highways in remote areas and the incomplete or delayed traffic accident records in certain regions. To ensure the research's feasibility and applicability, we chose to exclude external data in this study. Moreover, in future work, once we have access to sufficient external data, we intend to integrate it into the model and release an updated framework for researchers to replicate and utilize. Nonetheless, we agree that incorporating these factors could further improve model performance and generalizability. As a result, we have revised our conclusion section to acknowledge this limitation and highlight future research directions where multimodal data integration could be explored.

2. Model Interpretability and Explainability

We appreciate your concern regarding the interpretability of advanced deep learning models like the **3S-TBLN**, which, like many deep learning architectures, can sometimes function as a "black box." As is well known, deep learning models present challenges in interpreting the specific meaning of each internal parameter, particularly in prediction tasks within the transportation domain. While some researchers use heatmaps to explain spatial dependencies [7, 40], this approach fails to distinguish whether the heatmap reflects the similarity between spatial locations or the mutual influence between them. In response, this paper approaches model interpretability by examining the **impact of input variables and key components** on the model's predictions, providing a more direct understanding of how key factors contribute to the results.

- **Influence of Each Essential Variable in the Results:** Quantify feature importance and assess how different inputs influence model predictions. We have briefly discussed how this method can provide further insights by tracing the contribution of each input feature to the final prediction. W/o Position Information, remove position information from input embeddings. W/o Day of Week, delete the day information from inputs. W/o Minute of Day, 3S-TBLN excludes consideration minute information from inputs.

Table R1 Performance of the Next Forty-Eight-Time Steps Prediction for Different Variables on Ningxia-YC Dataset

Data	*Method*	*Horizon 6*			*Horizon 12*			*Horizon 24*			*Horizon 48*		
		MAE	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*
Ningxia-YC	*W/o Position Information*	*5.276*	*9.039*	*18.395%*	*5.283*	*9.052*	*18.362%*	*5.308*	*9.065*	*18.358%*	*5.318*	***9.077***	*18.338%*
	W/o Day of Week	*5.194*	*9.101*	*14.346%*	*5.199*	*9.121*	*14.340%*	*5.216*	*9.157*	*14.356%*	***5.231***	*9.137*	*14.301%*
	W/o Minute of Day	*5.223*	*9.089*	*17.054%*	*5.225*	*9.109*	*17.061%*	*5.237*	*9.134*	*17.027%*	*5.274*	*9.184*	*17.020%*
	3S-TBLN	***5.190***	***8.952***	***11.473%***	***5.195***	***8.984***	***11.439%***	***5.213***	***9.037***	***11.431%***	*5.250*	*9.078*	***11.444%***
	Gains	*0.077%*	*0.962%*	*20.026%*	*0.077%*	*0.751%*	*20.230%*	*0.058%*	*0.309%*	*20.375%*	*-0.363%*	*-0.011%*	*19.978%*

As shown in Table R1, there is no doubt that traffic speed plays a crucial role in the prediction results, with position, day, and minute information also significantly influencing the outcomes. Among these variables, position information exerts the greatest impact, followed by minute information, and then day information. The ranking of their influence is determined by the increase in MAE, RMSE, and MAPE values. For both Horizon 24 and Horizon 48, the results clearly indicate that **position information is more influential than minute information, which in turn holds more significance than day information**. The specific ranking of importance for each prediction is reflected in the prediction errors. The model parameters have been uploaded to the GitHub platform, where readers can download and run them and observe the errors. We have known the critical role of these variables (speeds, position, day of week, and minute of day) in the prediction results, therefore, we have not provided a detailed analysis of their impact in the article. However**, the revision responses, including Table R1 and all responses in this file, are available on my personal GitHub repository** (https://github.com/zouguojian/Traffic-speed-prediction/tree/main/3S-TBLN), where readers can access them online at any time.

- **Influence of Each Essential Component in the Results:** To evaluate the effectiveness of specific components within the proposed 3S-TBLN model, four variants are examined in this part. W/o MSGraph-Att, 3S-TBLN omits traffic flow diffusion and uses a spatial attention layer instead. W/o fusion gate mechanism, adopting an addition operation replaces the adaptive fusion mechanism from 3S-TBLN. W/o traffic patterns, 3S-TBLN excludes consideration of traffic patterns from the previous week during long-term prediction. W/o self-supervised learning, not considering bilateral learning architecture, thereby not reconstructing historical observations.

Table III Performance of the Next Forty-Eight-Time Steps Prediction for Different Variants on Ningxia-YC Dataset

Data	*Method*	*Horizon 6*			*Horizon 12*			*Horizon 24*			*Horizon 48*		
		MAE	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*
Ningxia-YC	*W/o MSGraph-Att*	*5.202*	***8.879***	*13.540%*	*5.206*	***8.913***	*13.576%*	*5.219*	***8.950***	*13.630%*	*5.258*	***8.983***	*13.672%*
	W/o fusion gate mechanism	***5.177***	*9.044*	*16.736%*	***5.185***	*9.098*	*16.517%*	***5.208***	*9.147*	*16.325%*	*5.261*	*9.183*	*16.316%*
	W/o traffic patterns	*5.341*	*8.966*	*16.780%*	*5.349*	*9.001*	*16.649%*	*5.431*	*9.129*	*16.744%*	*5.609*	*9.256*	*16.931%*
	W/o self-supervised learning	*5.268*	*9.101*	*19.991%*	*5.283*	*9.156*	*19.939%*	*5.302*	*9.192*	*19.850%*	*5.334*	*9.187*	*19.795%*
	3S-TBLN	*5.190*	*8.952*	***11.473%***	*5.195*	*8.984*	***11.439%***	*5.213*	*9.037*	***11.431%***	***5.250***	*9.078*	***11.444%***
	Gains	*-0.251%*	*-0.822%*	*15.266%*	*-0.193%*	*-0.797%*	*15.741%*	*-0.096%*	*-0.972%*	*16.134%*	*0.152%*	*-1.058%*	*16.296%*

W/O Multi-Step Graph Attention *Compared to 3S-TBLN, a decline in performance was observed. For the horizon 24 prediction**, w/o MSGraph-Att** resulted in increases in MAE, RMSE, and MAPE by 0.115%, -0.963%, and 19.237%, respectively, in Ningxia-YC, and by 0.152%, -1.046%, and 19.469% for forecasting the horizon 48. These results confirm the beneficial role of the multi-step graph attention (MSGraph-Att) mechanism in traffic speed prediction. Specifically, effective traffic diffusion modeling is crucial, and the design of MSGraph-Att is tailored to address this challenge.*

W/O Fusion Gate Mechanism *Spatial and temporal properties are essential elements that need integration in existing spatio-temporal methods. This paper introduces an adaptive fusion gate mechanism to achieve this integration. The performance of **w/o fusion gate mechanism** is lower than 3S-TBLN regarding the three metrics in Ningxia-YC. For instance, concerning horizon 24 in Ningxia-YC, **w/o fusion gate mechanism** exhibited increases of -0.096%, 1.217%, and 42.813% in terms of MAE, RMSE, and MAPE, respectively, compared to 3S-TBLN; similarly, it showed increases of 0.210%, 1.157%, and 42.573% for the horizon 48. The experiments underscore the significant impact of adaptively fusing spatial and temporal correlations on traffic speed prediction, highlighting its crucial role in the methodology.*

W/O Traffic Patterns *Illustrated in Fig.2, neighboring weeks exhibit similar traffic patterns, reflected in speeds and timestamps. Timestamps encompass weekly and daily periodic information, serving as indicators for traffic speeds. Subsequently, the speeds from the previous week and the current week's timestamps are embedded into the proposal. Regarding the three metrics on all datasets, the results of **w/o traffic patterns** are inferior to 3S-TBLN. For instance, in Ningxia-YC, **w/o traffic patterns** leads to increases in MAE, RMSE, and MAPE by 4.182%, 1.018%, and 46.479%, respectively, for the horizon 24, compared to 3S-TBLN; by 6.838%, 1.961%, and 47.947% for the horizon 48. In addition, compared with all other variants, **w/o traffic patterns** produces the worst forecast. The experiments validate that traffic pattern has a crucial impact on long-term speed prediction, alleviating forecasting errors and shifts.*

W/O Self-Supervised Learning *Traffic speeds exhibit continuous variations across different time steps, creating a bi-directional property in speed forecasting, where historical data influences future predictions and vice versa. To capture this contextual semantic, self-supervised learning serves as a pretext task aiding the model in understanding this bi-directional property, as* *evidenced in Table III. Comparison between models **w/o self-supervised learning** and 3S-TBLN highlights the clear advantage of the latter in traffic speed prediction. For the horizon 24, **w/o self-supervised learning** resulted in an increase of 1.707% in MAE, 1.715% in RMSE, and 73.651% in MAPE in Ningxia-YC; similarly, for the horizon 48, the increases were 1.600% in MAE, 1.201% in RMSE, and 72.973% in MAPE. These comparative outcomes verify the significance of self-supervised learning in long-term traffic speed prediction.*

These enhancements have been incorporated in **Section [V]**, where we provide additional explanations to clarify the influence of each essential component in the results.

[7] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial- temporal graph convolutional networks for traffic flow forecasting," in Proceedings of the AAAI conference on artificial intelligence, vol. 33, no. 01, 2019, pp. 922–929.

[40] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," in Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019, pp. 1907– 1913.

We deeply appreciate your constructive feedback. We believe this response significantly enhances the clarity and interpretability of our work. In future research, we can refer to the latest work on interpretability in the field of artificial intelligence to help us solve the interpretability problems in some current traffic prediction models. Please let us know if there are any further refinements that could improve our manuscript. Notably**, we have uploaded this response to personal repositories** to solve the doubt of readers with similar problems.

Reviewer: 2

Recommendation: Reject

Comments:
This paper proposed a self-supervised spatio-temporal bilateral learning network for highway traffic speed prediction. Here are some comments:

(1) Compared to the authors' previous work, the novelty enhancement in this paper is insufficient. (1) MT-STNet: A Novel Multi-Task Spatiotemporal Network for Highway Traffic Flow Prediction. (2) When Will We Arrive? A Novel Multi-Task Spatio-Temporal Attention Network Based on Individual Preference for Estimating Travel Time.

Response: Dear Reviewer, we sincerely appreciate your thoughtful and constructive feedback regarding the novelty enhancement of our manuscript in comparison to our previous works, particularly the two publications you highlighted:

1. *MT-STNet: A Novel Multi-Task Spatiotemporal Network for Highway Traffic Flow Prediction*, and
2. *When Will We Arrive? A Novel Multi-Task Spatio-Temporal Attention Network Based on Individual Preference for Estimating Travel Time.*

We value the opportunity to clarify the unique contributions of our current study and to distinguish it from our prior work.

1. Novelty Different From MT-STNet and MT-STAN

While our earlier works focused on multi-task spatiotemporal networks, the current study significantly extends these foundations in the following key aspects:

- **Introduction of a New Problem Domain:** Unlike our previous focus on highway traffic flow prediction and individual travel time estimation, this paper tackles highway traffic speed prediction, which is a critical aspect of intelligent transportation systems. This shift in research focus addresses a novel set of challenges, particularly involving multi-step traffic diffusion extraction, traffic pattern modeling, and reconstructing historical variables based on self-supervised learning.
- **Algorithmic Innovation and Enhanced Methodology:** The current study proposes 3S-TBLN, which is bilateral architecture, where both the encoder and the decoder consist of the semantic transformer, multiple spatio-temporal blocks (ST-Blocks), and bridge transformer (BridgeTrans). Especially, to capture the multi-step traffic diffusion process, a novel multi-step graph attention mechanism called MSGraph-Att, based on the Top-k multi-heads graph attention (TK-MGAT), is introduced. This mechanism enables the adaptive modeling of the most significant spatial relationships among multiple time steps of the road network, disregarding redundant information from neighboring segments. To address the traffic patterns in the highway network, particularly the observed similarity during the same time intervals between adjacent weeks, we introduce a novel approach called BridgeTrans. Instead of directly using the past weeks' traffic speeds as input for modeling, BridgeTrans independently learns the traffic pattern and utilizes the learned information as a reference to generate hidden prediction representations in a single step. To fill the gap in existing traffic prediction approaches by forecasting the future based on the past and then reconstructing the historical observations using the predicted values. By

designing a self-supervised learning method, the model can effectively utilize unlabeled data and learn more comprehensive bi-directed time series representations. This approach ensures that the predicted values closely approximate the observed data, enhancing the model's accuracy and capturing important temporal dependencies. These innovations are different from the tricks used in prior works, MT-STNet and MT-STAN.

2. Empirical Contributions

- **Data Scope and Complexity:** The current study employs a richer and more diverse dataset, encompassing Ningxia-YC, METR-LA, and PEMS-BAY datasets. This allows us to validate our model under scenarios with greater complexity and variability, significantly extending the applicability and robustness of the proposed approach beyond what was demonstrated in our earlier works.
- **Performance Improvements:** The results section (**Section [V]**) demonstrates clear performance enhancements over baselines. These include the performance of long-term traffic speed prediction and detailed case studies illustrating the practical implications of our model. Such empirical evidence reinforces the originality and effectiveness of our contributions.

We hope that these clarifications adequately address your concerns regarding the novelty of our study. Thank you once again for your valuable comments. We remain available to address any further questions or suggestions you may have.

(2) This paper emphasizes "limited input variables" in the title, referring to three data types: timestamp, road segment index, and traffic speed. However, these data types, collectively known as spatial-temporal traffic data, cannot be used as a highlight. The reason is that almost all traffic prediction studies use spatial-temporal traffic data as input. Even some studies can achieve excellent predictive performance without the need for the historical periodic data mentioned by the authors.

Response: Thank you for your careful review of our manuscript. We appreciate your valuable comments and suggestions, especially regarding the emphasis on "limited input variables" in our title. We have given this issue thorough consideration and would like to address your concerns in detail.

1. Clarification on “Limited Input Variables” in the Title

We acknowledge your concern regarding the phrase **“limited input variables”** in our title. You are right that spatial-temporal traffic data, including timestamp, road segment index, and traffic speed, are commonly used in traffic prediction studies. However, the novelty of our work lies not in the types of data used but in how we utilize these data to achieve accurate predictions with a minimal set of input variables. Our research focuses on developing a new model that can effectively extract and utilize the inherent spatial-temporal patterns from these basic data types.

In our study, we have demonstrated that by using only these three basic data types, our model can achieve competitive prediction performance compared to other models. This is due to the unique architecture and algorithms we have designed, which enable the model to better capture the complex relationships and dependencies in the spatial-temporal traffic data. Moreover, we concentrate on the problems in traffic data modeling. For example, our model uses a novel graph neural network structure to model the multi-step spatial relationships between road segments and a temporal attention mechanism to capture the temporal patterns of traffic speed changes. These innovations allow us to make full use of the limited input variables and obtain accurate prediction results.

2. Justification for Using Only Spatial-Temporal Traffic Data

While many traffic prediction studies utilize extensive feature sets, including weather, roadwork data, and traffic incidents-based information, our approach deliberately restricts the input variables to only fundamental spatiotemporal traffic data. The key motivation for this design choice is:

- **Generality and Scalability**: Many real-world transportation systems lack comprehensive historical or external data. By relying only on fundamental traffic data, our model can be widely applied to different highway networks without requiring extensive data collection efforts.

3. Addressing Performance Concerns With Historical Periodic Data

We acknowledge that some existing studies have achieved excellent forecasting performance without explicitly relying on historical periodic data. Additionally, several studies have incorporated historical periodic data as supplementary information to enhance model performance [7]. However, these approaches generally do not focus on learning the specific patterns of traffic speed variations. In contrast, our model takes a different approach by employing an **independent structure (BrigeTrans)** that specifically learns the **changing patterns** of historical traffic data. This learned representation is then utilized as a reference during the current prediction inference phase. Therefore, our model does not simply treat historical periodic data as input. Instead, it fully models these patterns independently, allowing it to capture intricate variations in traffic speed that are not merely periodic but also contextually influenced by real-time factors. This results in a model that is not only more comprehensive in its representation of traffic dynamics but also more competitive in real-world applications, where such complex patterns are often present. We believe this distinction significantly enhances the robustness and accuracy of our model in real-world traffic prediction scenarios. In the **ablation experiments** (**Section [V] E**), the variant without considering the traffic pattern learning component also presents comparative performance with baselines.

Moreover, we have also compared our model with some studies that do not use historical periodic data. Although these studies may achieve excellent predictive performance in certain scenarios, their models often **rely on other data preprocesses (e.g., STWave [46]) or assumptions** that may not be applicable in all situations. Our model, on the other hand, is more general and can be easily applied to different traffic prediction tasks with only the **basic spatial-temporal traffic data**.

[7] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in Proceedings of the AAAI conference on artificial intelligence, vol. 33, no. 01, 2019, pp. 922–929.

[46] Y. Fang, Y. Qin, H. Luo, F. Zhao, B. Xu, L. Zeng, and C. Wang, "When spatio-temporal meet wavelets: Disentangled traffic forecasting via efficient spectral graph attention networks," in 2023 IEEE 39th international conference on data engineering (ICDE). IEEE, 2023, pp. 517–529.

We truly appreciate your thoughtful comments, which have prompted us to refine the title to: "How to Accurately Forecast Highway Traffic Speed? A Self-Supervised Spatio-Temporal Bilateral Learning Network." We believe that the revised manuscript more effectively articulates the motivation behind our approach and its relevance to real-world traffic prediction challenges. Thank you once again for your valuable feedback, and we welcome any further suggestions for improvement.

(3) "Semantic Transformer" is a standard input data embedding operation, and it does not need to be highlighted in the abstract.

Response: We sincerely appreciate your valuable time and insightful feedback, which have significantly contributed to improving our manuscript. Below, we address your concern regarding the emphasis on **Semantic Transformer** in the abstract.

1. Clarification on the Role of "Semantic Transformer" in Our Study

We acknowledge that **Semantic Transformer** is a commonly used embedding technique in deep learning-based traffic prediction models. Our intention in highlighting it in the abstract was not to claim novelty in its usage but rather to emphasize its role in **enhancing spatial-temporal feature representations** within our model. The real detailed contributions of our research work have been listed at the end of **Section [I]**.

However, we understand your point that the **Semantic Transformer** is a standard embedding operation and does not constitute a unique contribution that warrants emphasis in the abstract. Given this, we have revised the abstract to ensure that it better highlights the novel aspects of our work, particularly:

- The **essential elements** should be considered in traffic speed prediction of highway networks,
- The **key components** and **key tricks** in the proposed approach,
- The model's ability to achieve high-accuracy traffic predictions with **minimal input variables**,
- The proposed method and baseline models are **reproducible** and have been uploaded to a personal **GitHub repository**. Reviewers and readers can download these models to their local devices and **deploy them in intelligent transportation systems**.

2. Revision of the Abstract

Following your suggestion, we have carefully revised the abstract by **removing the emphasis on the Semantic Transformer** and instead focusing on the methodological innovations and practical implications of our approach.

Accurately predicting traffic speed is critical for traffic system scheduling, management, and optimization. In highway traffic speed prediction, three essential elements should be considered: (1) the complex spatial diffusion process of traffic over time, (2) the influence of traffic patterns on prediction, and (3) the crucial role of bi-directed learning mechanisms in time series forecasting tasks. A novel self-supervised spatio-temporal bilateral learning network (3S-TBLN) for long-term traffic speed forecasting is proposed to address the above challenges. 3S-TBLN adopts an encoder-decoder, which is a bilateral architecture, where both the encoder and the decoder consist of the semantic transformer, multiple spatio-temporal blocks (ST-Blocks), and bridge transformer (BridgeTrans). Especially, the ST-Blocks are designed to model multi-step dynamic spatio-temporal correlations in both encoder and decoder; in the encoder, BridgeTrans is applied to learn forward traffic patterns from the last week's observations, and vice versa in the decoder; a bilateral architecture-based self-supervised learning method is proposed for reconstructing historical variables as a pretext task, combined with a speed prediction task to learn the bidirectional context. The experiments were evaluated on three real-world highway *datasets. Experimental results demonstrate that the proposed 3S-TBLN model significantly outperforms state-of-the-art baselines and can efficiently solve the problem of long-term highway speed prediction. The source code is available at https://github.com/zouguojian/Traffic-speed-prediction/tree/main/3S-TBLN.*

We appreciate your thoughtful suggestion, which has helped us refine our abstract to more accurately reflect the key contributions of our work. Thank you once again for your constructive feedback. We look forward to any further recommendations you may have.

(4) The description of the "Temporal-Att" module in the paper is not sufficiently clear. It is necessary to express it using equations.

Response: We sincerely appreciate your insightful feedback regarding the clarity of the **Temporal-Att module** description. Your suggestion to express it using equations is highly valuable, as it can significantly enhance the precision and comprehensibility of our methodology. Below, we address this issue in detail.

1. Enhancing the Mathematical Formalization of the Temporal-Att Module

We acknowledge that our initial description of the **Temporal-Att module** lacked sufficient mathematical rigor, which may have made it challenging for readers to fully grasp the mechanism. To address this, we have now introduced explicit equations to formally describe the computation process of the Temporal-Att mechanism.

Specifically, we define the temporal attention mechanism as follows:

For the node $v_{t_j}^i$ at time step t_j, the correlation coefficient between time steps t_j and t at the l-th layer of the temporal attention mechanism can be computed using equation (12).

$$\beta_{v_{t_j}^i, v_t^i}^{l,m} = \frac{exp\left(\delta_{v_{t_j}^i, v_t^i}^{l,m}\right)}{\sum_{r=1}^{\mathcal{T}} exp\left(\delta_{v_{t_j}^i, v_{t_r}^i}^{l,m}\right)} \tag{12}$$

where $\delta_{v_{t_j}^i, v_t^i}^{l,m}$ represents the correlation between time steps t_j and t, $\mathcal{T}$ denotes the length of the input time steps, and the subscript r ranges from 1 to $\mathcal{T}$.

The correlation $\delta_{v_{t_j}^i, v_t^i}^{l,m}$ can be computed by taking the dot product of the query vector of node $v_{t_j}^i i$ at time step t_j and the key vector of node v_t^i at time step t.

$$\delta_{v_{t_j}^i, v_t^i}^{l,m} = \frac{\left\langle f_Q^m\left(hdt_{v_{t_j}^i}^{l-1}\right), f_K^m\left(hdt_{v_t^i}^{l-1}\right)\right\rangle}{\sqrt{d}} \tag{13}$$

After obtaining the correlation coefficient $\beta_{v_{t_j}^i, v_t^i}^{l,m}$ in the m-th attention head, the temporal correlation representation $hdt_{v_{t_j}^i}^{l}$ of node $v_{t_j}^i$ at time step t_j in the l-th layer of the temporal attention can be computed using equation (14).

$$hdt_{v_{t_j}^i}^{l,m} = \sum_{r=1}^{\mathcal{T}} \beta_{v_{t_j}^i, v_{t_r}^i}^{l,m} f_V^m\left(hdt_{v_{t_r}^i}^{l-1}\right) \tag{14}$$

$$hdt_{v_{t_j}^i}^{l} = \mathrm{BN}\left(\Vert_{m=1}^{M} hdt_{v_{t_j}^i}^{l,m} \mathrm{W}_T^{l,m}\right) \tag{15}$$

where $\mathrm{W}_T^{l,m} \in \mathrm{R}^{d_m \times d_m}$ *represents the learnable weight matrix. Additionally, the final temporal correlation* $hdt_{v_{t_j}^i}^l \in R^d$ *of node* $v_{t_j}^i$ *at time step* t_j *can be computed using formulas (12) to (15). The initial input of Temporal-Att is* $\mathrm{XS}^{C'}$*, and the output is* HDT^l.

These equations have been added to **Section [IV] C** of the revised manuscript to provide a clear and formalized explanation of the Temporal-Att module.

We appreciate your constructive feedback, which has greatly helped improve the clarity and rigor of our manuscript. With the newly incorporated equations and enhanced explanations, we believe the **Temporal-Att module** is now more comprehensible for readers. Thank you again for your valuable suggestions, and we welcome any further recommendations.

(5) The variables and equations in the paper need to be standardized and italicized. Additionally, some equations have issues, such as the format of equation (1). Please check for any writing errors in equations (4)-(6).

Response: We sincerely appreciate your detailed review and your valuable suggestions regarding the standardization and formatting of variables and equations in our manuscript. Your comments have helped us refine the presentation of mathematical expressions, ensuring greater clarity and consistency. Below, we outline the improvements we have made in response to your feedback.

1. Standardizing Variable Notation and Equation Formatting

We acknowledge that some variables were not properly standardized or italicized, which may have affected readability and consistency. To address this, we have carefully reviewed all variables and equations throughout the manuscript and have ensured that:

- **All scalar variables and vectors are italicized** (e.g., t, h, f).
- **All sets and matrices are uprighted** (e.g., X, W, and H) to distinguish them from scalars.
- **Mathematical operators and functions** such as softmax(·), log(·), and $\sum$ are correctly formatted according to IEEE LaTeX standards.

These changes have been systematically applied to all equations in the revised manuscript.

2. Revising Equation (1) Formatting

We appreciate your observation regarding the formatting of **Equation (1)**. Upon review, we identified that the alignment and notation could be improved. The revised **Equation (1)** is now properly formatted to enhance readability and mathematical clarity.

$$e_{v_i,v_j} = \begin{cases} 1 & \text{if nodes } v_i \text{ and } v_j \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

3. Checking for Errors in Equations (4)–(6)

Following your suggestion, we carefully examined **Equations (4)–(6)** for any typographical or conceptual errors. The specific revisions include:

- **Correction of missing or misplaced subscripts and superscripts** to ensure consistency with the notation used elsewhere in the paper.
- **Verification of mathematical operations** to confirm that all terms are correctly defined and aligned with our methodological framework.
- **Ensures functions and variables are consistent with the algorithm** so that the matrix and vector operations remain valid in the algorithm.

$$TK-MGAT-MultiHead(\mathrm{Q},\mathrm{K},\mathrm{V}) = \mathrm{Concat}(head_1,\cdots,head_M)\mathrm{W}_O \tag{4}$$

$$Top-k-Attention(\mathrm{Q}',\mathrm{K}',\mathrm{V}')=\mathrm{SoftMax}\left(Top-k\left(\frac{Q'K'^{\top}}{\sqrt{d}}\right)\right)\mathrm{V}' \tag{5}$$

$$head_m=Top-k-Attention\left(\mathrm{QW}_Q^m,\mathrm{KW}_Q^m,\mathrm{VW}_Q^m\right) \tag{6}$$

All identified issues have been corrected in the revised manuscript. Additionally, we have checked similar problems in other equations.

We sincerely appreciate your meticulous review, which has greatly improved the precision and clarity of our mathematical expressions. With the implemented revisions, we believe that the variables and equations are now **fully standardized, properly formatted, and free of typographical errors**. Thank you once again for your invaluable feedback, and we look forward to any further suggestions that may enhance our manuscript.

(6) In the experiments, the authors validated their approach on the Ningxia-YC and METR-LA datasets. However, there are many other benchmark traffic speed prediction datasets, such as PEMS-BAY. The authors could consider validating their approach on additional datasets.

Response: We sincerely appreciate your insightful suggestion regarding the evaluation of our approach on additional benchmark datasets. Below, we provide a detailed response addressing your concern.

1. Justification for Selecting Ningxia-YC and METR-LA Datasets

In this study, we selected the **Ningxia-YC** and **METR-LA** datasets to validate our proposed approach due to the following reasons:

- **Diversity in Traffic Conditions**: The **Ningxia-YC dataset** represents traffic conditions with a relatively high fluctuation, while the **METR-LA dataset** captures traffic dynamics in a highly changing regularity. This selection allows us to evaluate our model's adaptability to different traffic environments.
- **Spatial and Temporal Heterogeneity**: The two datasets exhibit different spatial network structures and temporal patterns, making them suitable for assessing the robustness of our method.
- **Real-World Application Focus**: The **Ningxia-YC** and **METR-LA** datasets are collected from real-world intelligent transportation systems in China and the US, respectively, to evaluate traffic prediction models in highway environments, aligning with the objectives of this study.

2. Consideration of Additional Benchmark Datasets (e.g., PEMS-BAY)

We acknowledge that there are several widely used benchmark datasets for traffic speed prediction, such as **PEMS-BAY**, which features an extensive road network with a high density of sensor stations. While evaluating our model on **PEMS-BAY** could further validate its generalizability, we also considered the computational cost and data preprocessing efforts required to incorporate additional datasets within the scope of this study.

Nevertheless, to address your suggestion, we have included a discussion in **Section [V]** of the revised manuscript, highlighting the potential benefits of evaluating our approach on **PEMS-BAY** and other large-scale datasets in future work. We anticipate that testing on a broader range of datasets would provide further insights into the model's scalability and effectiveness across diverse traffic networks.

3. Changes to the Revised Version

We have added the PEMS-BAY dataset description to Section **[V]** (A. Datasets). The results, comparisons, and analysis have been added to question (7)'s response, the reviewer can reference the next question.

*(3) **PEMS-BAY** This traffic dataset is collected from loop detectors of the highway in the Bay Area, California, United States, from January 1, 2017, to May 31, 2017 [42]. For the experiment, we selected 325 sensors and collected 3 months of data. The traffic speed of each monitoring station is recorded every 5 minutes.*

Table RI

DATASET INFORMATION

Dataset	*Data type*	*Station*	*Granularity*	*Time range*
Ningxia-YC	*Speed*	*108*	*15 min*	06/01/2021 - 08/31/2021
METR-LA	*Speed*	*207*	*5 min*	03/01/2012 - 05/31/2012
PEMS-BAY	*Speed*	*325*	*5 min*	01/01/2017 - 03/31/2017

[42] F. Li, J. Feng, H. Yan, G. Jin, F. Yang, F. Sun, D. Jin, and Y. Li, "Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution," ACM Transactions on Knowledge Discovery from Data, vol. 17, no. 1, pp. 1–21, 2023.

We greatly appreciate your constructive feedback, which has helped us enhance the discussion of dataset selection in our study. While we focused on **Ningxia-YC** and **METR-LA** due to their contrasting characteristics, we recognize the importance of further validation on additional benchmark datasets such as **PEMS-BAY**. Therefore, we incorporated the **PEMS-BAY dataset** to evaluate and compare the performance of the proposed method against baseline models. Thank you once again for your valuable insights, and we look forward to any further recommendations you may have.

(7) In the comparative experiments, the authors used a relatively recent model, MTGNN (2020), as a baseline. It is recommended to include baselines from the past two years to enhance the competitiveness of 3S-TBLN.

Response: We sincerely appreciate your valuable suggestion regarding the inclusion of more recent baseline models in our comparative experiments. Your feedback is highly valuable in ensuring the robustness and competitiveness of our proposed **3S-TBLN** model. Below, we address this concern in detail.

1. Justification for Selecting popular methods, such MTGNN (2020), as Baselines

We acknowledge that **MTGNN (2020)**, while relatively recent, may not fully represent the latest advancements in traffic prediction models. Our primary motivation for selecting DCRNN (2018), AGCRN (2020), ASTGCN (2019), Graph-WaveNet (2019), GMAN (2020), ST-GRAT (2020), and MTGNN (2020) as baselines was due to their strong performance in **spatiotemporal graph-based traffic prediction**, as well as their widespread adoption in recent research. In addition, these methods' codes have been opened in GitHub and can be downloaded and reproduced. Moreover, these methods were designed to efficiently capture spatial dependencies while reducing computational complexity, making it a relevant benchmark for comparison. Therefore, we choose these methods as baselines in our research.

2. Consideration of More Recent Baseline Methods

We agree that incorporating more recent models from the past two years would further enhance the comprehensiveness of our experimental comparisons. To address this, we have taken the following steps:

- **Expanded Baseline Selection**: We have reviewed the latest literature and identified **[DGCRN (2023)]**, **[STWave (2023)]**, and **[BigST (2024)]** as additional competitive baselines that align with our study's focus on spatiotemporal traffic prediction. These models leverage advanced deep learning architectures, such as spatiotemporal graph neural networks, and provide state-of-the-art performance in recent studies.
- **Additional Experimental Comparisons**: We have conducted new comparative experiments incorporating these recently published models and updated **Tables [II] and [IV]** in the revised manuscript. Our results demonstrate that **3S-TBLN** remains competitive, consistently outperforming or achieving comparable performance to these newer baselines while maintaining high efficiency.

3. Discussion on Competitiveness of 3S-TBLN

To further highlight the advantages of **3S-TBLN**, we have revised **Section [V] E**, where we provide a more detailed discussion on how our model compares against the newly included baselines. In particular, we emphasize:

- The ability of **3S-TBLN** to maintain high accuracy for long-term traffic speed prediction does not rely on extensive external features.
- The trade-off between model complexity and real-world applicability, where **3S-TBLN** achieves a balance between prediction accuracy and computational feasibility.

4. Changes to the Revised Version

We have now incorporated additional baselines and added detailed comparisons and analysis in **Section [V]**.

Baseline Methods:

STWave*, is a disentangled fusion framework that uses a dual-channel spatio-temporal network to decouple traffic data into stable trends and fluctuating events, addressing distribution shifts, employing a query sampling strategy and graph wavelet-based positional encoding to capture dynamic spatial correlations.*
DGCRN*, is a dynamic graph convolutional recurrent network that generates dynamic graph topologies through hyper-networks to extract dynamic node features at each time step, employing a training strategy that limits decoder iterations.*
BigST*, is a spatio-temporal graph neural network that leverages a scalable feature extractor to encode long-range inputs into low-dimensional representations, while a linearized global spatial convolution network distills time-varying graph structures to facilitate efficient message passing.*

Predicting Performance Comparison:

Table II compares the proposed method with various baselines across different prediction horizons in the Ningxia-YC, METR-LA, and PEMS-BAY datasets. In METR-LA and PEMS-BAY, prediction spans of 25-30, 55-60, 115-120, and 235-240 minutes (equivalent to horizons of 6, 12, 24, and 48) are considered, while in Ningxia-YC, spans of 75-90, 165-180, 345-360, and 705-720 minutes are analyzed.

Table II. Performance Comparison of Different Approaches for Traffic Prediction on Ningxia-YC, METR-LA, and PEMS-BAY Datasets ('⁎' Means Top-k Is Canceled in PEMS-BAY Because of the High Computation Cost on Sort, and and '-' Denotes High MAPE Value)

Data	Method	Horizon 6			Horizon 12			Horizon 24			Horizon 48		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Ningxia-YC	ARIMA	5.892	9.252	12.845%	5.933	9.305	12.909%	6.039	9.418	13.052%	6.177	9.586	13.267%
	SVR	5.799	9.608	12.890%	5.886	9.704	13.047%	5.995	9.811	13.226%	5.841	9.668	12.904%
	LSTM_BILSTM	7.683	10.897	29.201%	7.855	11.083	29.384%	8.028	11.281	29.565%	8.024	11.286	29.489%
	ST-GRAT	5.685	9.295	12.696%	5.773	9.332	12.873%	6.256	9.711	13.107%	6.594	10.001	13.744%
	GMAN	5.424	9.007	14.295%	5.409	9.006	14.163%	5.416	9.014	14.252%	5.454	9.057	14.442%
	MDL	5.529	9.154	14.296%	5.570	9.219	14.558%	5.647	9.361	14.203%	5.778	9.461	14.288%
	DCRNN	5.342	8.957	12.151%	5.411	9.033	12.210%	5.455	9.104	12.250%	5.554	9.203	12.460%
	ASTGCN	5.721	9.277	15.518%	5.789	9.380	14.460%	5.853	9.446	14.657%	5.682	9.260	12.807%
	AGCRN	5.291	8.896	12.351%	5.398	9.009	13.717%	5.415	9.052	13.545%	5.515	9.166	14.600%
	MTGNN	5.335	9.026	12.191%	5.386	9.082	12.419%	5.424	9.147	12.314%	5.524	9.265	12.563%
	DGCRN	5.267	8.865	**12.005%**	**5.313**	**8.922**	12.058%	**5.352**	**8.982**	**12.138%**	5.572	9.145	12.595%
	Graph-WaveNet	**5.252**	8.880	12.196%	5.316	8.978	**11.984%**	5.457	9.221	13.441%	5.498	9.215	12.630%
	BigST	5.253	**8.868**	13.035%	5.325	8.966	12.421%	5.416	9.097	13.144%	5.450	9.132	12.964%
	STWave	5.292	8.935	12.083%	5.366	9.059	12.234%	5.373	9.065	12.203%	**5.323**	**8.971**	**12.044%**
	3S-TBLN	**5.190**	**8.952**	**11.473%**	**5.195**	**8.984**	**11.439%**	**5.213**	**9.037**	**11.431%**	**5.250**	**9.078**	**11.444%**
	Gains	1.181%	-0.947%	4.431%	2.221%	-0.695%	4.548%	2.597%	-0.612%	5.825%	1.371%	-1.193%	4.982%
METR-LA	ARIMA	7.799	13.710	20.554%	7.658	12.695	19.913%	7.971	12.646	20.298%	8.668	13.584	21.832%
	SVR	3.761	7.984	9.880%	4.492	9.246	12.945%	5.452	10.974	17.417%	6.074	11.968	20.407%
	LSTM_BILSTM	5.213	9.723	18.629%	5.636	10.339	19.696%	6.316	11.273	21.437%	7.094	12.195	23.192%
	ST-GRAT	3.229	6.708	9.188%	4.029	8.490	12.023%	5.405	10.960	16.608%	7.147	13.433	22.048%
	GMAN	3.539	7.452	10.065%	3.760	7.997	10.828%	3.885	8.281	**11.264%**	**3.952**	**8.412**	**11.489%**
	MDL	3.402	6.893	10.003%	3.802	7.993	11.704%	4.129	8.720	12.984%	4.505	9.060	13.599%
	DCRNN	3.993	9.579	-	5.069	11.664	-	6.652	13.892	-	9.082	15.974	-
	ASTGCN	3.576	7.116	10.853%	4.328	8.734	13.990%	5.032	9.639	15.666%	5.870	10.767	18.580%
	AGCRN	3.232	6.474	9.444%	3.612	7.471	11.032%	3.898	8.151	12.331%	4.131	8.699	13.047%
	MTGNN	3.239	6.819	9.268%	3.648	7.878	10.996%	3.987	8.657	12.397%	4.317	9.142	13.489%
	DGCRN	3.190	6.420	8.958%	3.536	7.386	10.339%	3.864	8.172	11.686%	4.152	8.718	12.514%
	Graph-WaveNet	**3.098**	**6.242**	**8.362%**	**3.494**	**7.314**	**10.083%**	**3.823**	**8.054**	11.396%	4.101	8.466	12.344%
	BigST	3.352	6.719	9.590%	3.824	7.769	11.521%	4.067	8.419	12.495%	4.217	8.637	13.164%
	STWave	3.184	6.565	9.056%	3.547	7.463	10.577%	3.844	8.136	11.758%	4.071	8.658	12.650%
	3S-TBLN⁎	**3.127**	**6.499**	**8.844%**	**3.497**	**7.425**	**10.091%**	**3.751**	**7.987**	**10.755%**	**3.915**	**8.261**	**11.082%**
	Gains	-0.936%	-4.117%	-5.764%	-0.086%	-1.518%	-0.079%	**1.883%**	**0.832%**	**4.519%**	**0.936%**	**1.795%**	**3.543%**
PEMS-BAY	ARIMA	5.379	10.893	13.507%	4.827	9.599	12.310%	4.444	8.198	11.274%	4.757	8.093	11.751%
	SVR	2.050	4.603	4.692%	2.605	5.908	6.341%	3.283	7.439	8.741%	3.957	8.877	11.405%
	LSTM_BILSTM	2.328	4.954	5.570%	3.106	6.630	8.055%	4.034	8.676	11.459%	4.606	9.674	12.925%
	ST-GRAT	2.162	4.821	4.829%	2.976	6.722	7.038%	4.089	9.058	10.123%	5.498	11.248	13.979%
	GMAN	1.944	4.414	4.606%	2.160	5.002	5.322%	**2.286**	5.280	5.699%	**2.374**	**5.444**	**5.918%**
	MDL	1.886	4.148	4.257%	2.200	4.911	5.063%	2.474	5.515	5.795%	2.771	6.051	6.763%
	DCRNN	2.078	4.540	5.072%	2.686	5.981	7.032%	3.144	6.623	8.129%	4.074	8.241	10.613%
	ASTGCN	2.102	4.656	4.857%	2.680	5.944	6.683%	3.070	6.593	7.812%	3.680	7.510	9.167%
	AGCRN	1.875	3.985	4.167%	2.191	4.686	4.994%	2.433	5.146	5.628%	2.910	6.387	7.222%
	MTGNN	1.791	4.019	**4.009%**	**2.086**	4.741	**4.754%**	2.393	5.411	5.506%	2.562	5.667	6.303%
	DGCRN	1.819	3.960	4.068%	2.103	4.655	4.774%	2.365	5.170	**5.351%**	3.255	7.148	7.638%
	Graph-WaveNet	**1.787**	3.971	4.062%	2.134	4.849	5.115%	2.375	5.438	5.921%	2.676	6.074	6.594%
	BigST	1.816	**3.912**	4.129%	2.122	**4.634**	5.055%	2.353	**5.054**	5.632%	2.613	5.548	6.286%
	STWave	2.267	5.239	5.064%	2.928	6.763	6.640%	3.481	7.844	7.931%	3.735	8.028	8.681%
	3S-TBLN⁎	1.881	4.226	4.414%	**2.062**	4.682	4.959%	**2.170**	**4.898**	**5.229%**	**2.267**	**5.029**	**5.393%**
	Gains	-5.260%	-8.0275%	-10.102%	1.151%	-1.0365%	-4.312%	5.074%	3.087%	2.280%	4.507%	7.623%	8.871%

The performance of ARIMA is comparatively lower than that of other baseline models, highlighting the challenges associated with long-term traffic speed prediction. Incorporating non-linear temporal correlations into recurrent neural networks, exemplified by LSTM_BILSTM, is crucial for accurate speed prediction. For example, in horizon 12 prediction, LSTM_BILSTM reduced MAE, RMSE, and MAPE by 26.404%, 18.558%, and 1.090%, respectively, compared to ARIMA in METR-LA, and by 35.654%, 30.930%, and 34.565% in PEMS-BAY. Notably, in the Ningxia-YC, METR-LA, and PEMS-BAY datasets, SVR outperformed LSTM_BILSTM across all metrics. This outcome resulted from LSTM_BILSTM's dynamic prediction generation, while SVR used a one-by-one training approach for each horizon task, thereby avoiding prediction error propagation.

Spatial correlation plays a crucial role in traffic forecasting. For the horizon 48 prediction, MDL outperforms LSTM_BILSTM, reducing MAE, RMSE, and MAPE by 27.991%, 16.170%, and 51.548%, respectively, in Ningxia-YC; by 36.496%, 25.707%, and 41.363% in METR-LA; and by

39.839%, 37.451%, and 47.675% in PEMS-BAY. However, the non-Euclidean nature of the speed prediction task highlights the limitations of CNNs in extracting spatial correlations, as demonstrated by existing studies. Graph neural networks (GNNs) are utilized for modeling traffic diffusion in spatial dimensions within non-Euclidean structured data. For instance, for the horizon 48 prediction, MDL underperforms compared to AGCRN, Graph-WaveNet, GMAN, MTGNN, DGCRN, STWave, and BigST, with increases in MAE by 4.769%, 5.093%, 5.941%, 4.598%, 3.697%, 8.548%, and 6.018%, respectively, in Ningxia-YC, and by 9.053%, 9.851%, 13.993%, 4.355%, 8.502%, 10.661%, and 6.830% in METR-LA. These comparisons confirm the capability of GNNs in modeling spatial properties within non-Euclidean space, offering insights for future advancements in forecasting performance. Notably, in the METR-LA and PEMS-BAY datasets, ASTGCN, DCRNN, and ST-GRAT perform worse than MDL, possibly due to (1) dynamic decoding and (2) the coupling of spatial correlation and temporal dependency.

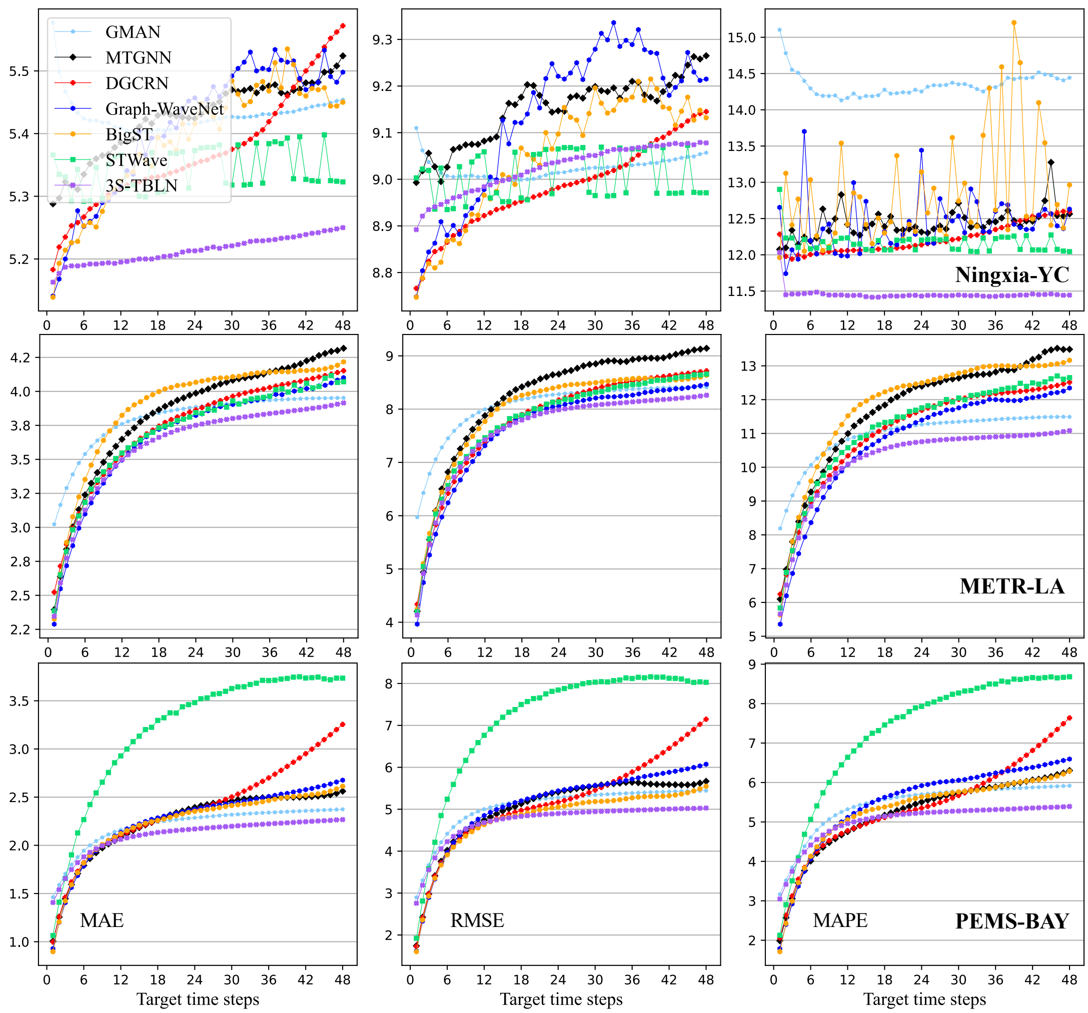

Fig. 7 Long-term highway traffic speed prediction ability: the first row presents the prediction error of each step in Ningxia-YC; the second row reflects the prediction accuracy of each step in METR-LA; the third row reflects the prediction accuracy of each step in PEMS-BAY.

This paper introduces a self-supervised spatio-temporal bilateral learning approach for traffic speed prediction, drawing on insights from GNNs to model spatial correlation. As shown in Table II, our method outperforms all baseline models in terms of MAE, RMSE, and MAPE across three datasets. Notably, compared to spatio-temporal methods like DCRNN, AGCRN, ASTGCN, Graph-

WaveNet, GMAN, ST-GRAT, MTGNN, DGCRN, STWave, and BigST for the horizon 24 prediction, our 3S-TBLN approach achieved reductions in MAE by 4.436%, 3.730%, 10.935%, 4.471%, 3.748%, 16.672%, 3.890%, 2.597%, 2.978%, and 3.748%, respectively, in Ningxia-YC; by 43.611%, 3.711%, 25.457%, 1.883%, 3.449%, 30.601%, 5.919%, 2.924%, 2.419%, and 7.770% in METR-LA; and by 30.980%, 10.810%, 29.316%, 8.632%, 12.288%, 5.074%, 9.319%, 8.245%, 37.660%, and 7.777% in PEMS-BAY. For the horizon 48 prediction, 3S-TBLN reduced MAE by 5.474%, 4.805%, 7.603%, 4.511%, 3.740%, 20.382%, 4.960%, 5.779%, 1.371%, and 3.670%, respectively, in Ningxia-YC; 56.893%, 5.229%, 33.305%, 4.535%, 0.936%, 45.222%, 9.312%, 5.708%, 3.832%, and 7.161% in METR-LA; and 44.354%, 22.096%, 38.397%, 15.284%, 18.188%, 4.507%, 11.514%, 30.353%, 39.300%, and 13.241% in PEMS-BAY. Furthermore, as illustrated in Fig. 7, the performance of 3S-TBLN gradually presents apparent superiorities across different horizons as the target time step increases.

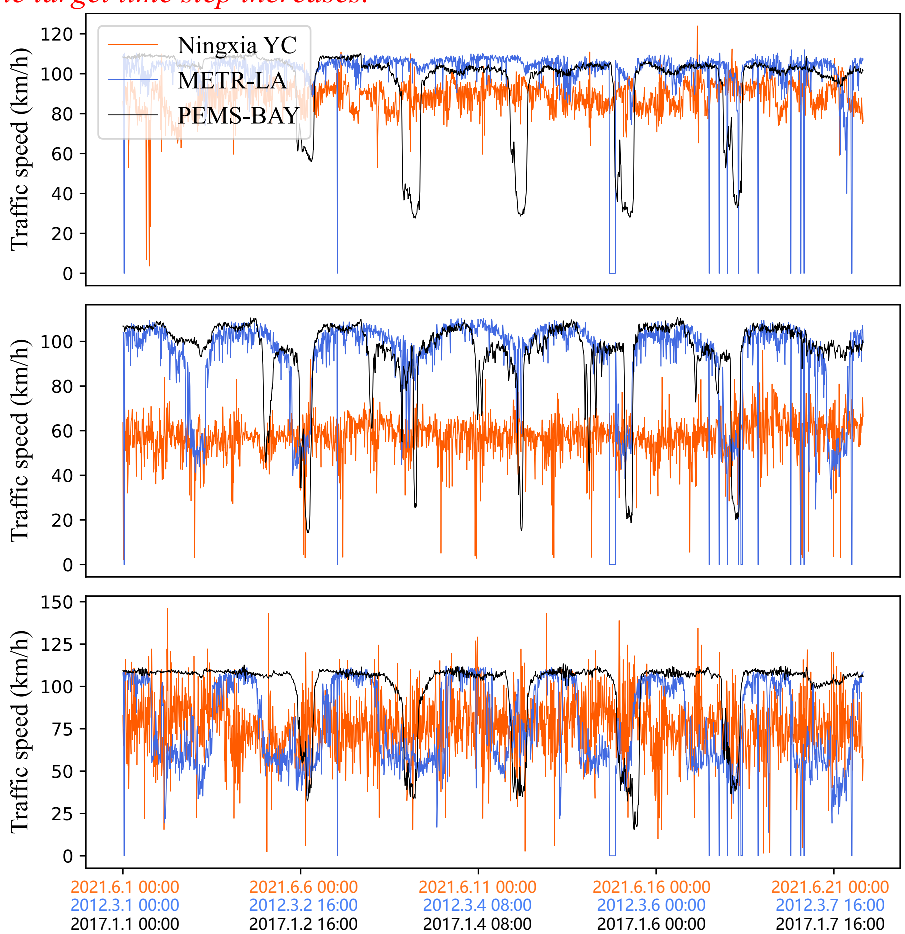

Fig. 8. Examples of traffic speed distribution in Ningxia-YC, METR-LA, and PEMS-BAY datasets, nine road segments or sensors are selected randomly.

The comparisons above yield three key findings: (1) Graph neural networks demonstrate superior capabilities in extracting spatial correlation within non-Euclidean spaces, as evidenced in Tables II; (2) The traffic speed distribution exhibits greater disorder in Ningxia-YC compared to METR-LA and PEMS-BAY, posing challenges for forecasting, as illustrated in Fig. 7 and 8. Despite this, the accuracy of 3S-TBLN remains unaffected, showcasing clear advantages specifically in Ningxia-YC; and (3) 3S-TBLN attains state-of-the-art predictive performance, notably more pronounced in longer-term horizons, as indicated in Fig. 7. We assert that long-term traffic speed forecasting holds more significant practical benefits, enabling transportation agencies to optimize traffic proactively based on predictive insights.

Influence of Each Essential Component:

Table III Performance of the Next Forty-Eight-Time Steps Prediction for Different Variants on Ningxia-YC Dataset

Data	Method	Horizon 6			Horizon 12			Horizon 24			Horizon 48		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Ningxia-YC	W/o MSGraph-Att	5.202	**8.879**	13.540%	5.206	**8.913**	13.576%	5.219	**8.950**	13.630%	5.258	**8.983**	13.672%
	W/o fusion gate mechanism	**5.177**	9.044	16.736%	**5.185**	9.098	16.517%	**5.208**	9.147	16.325%	5.261	9.183	16.316%
	W/o traffic patterns	5.341	8.966	16.780%	5.349	9.001	16.649%	5.431	9.129	16.744%	5.609	9.256	16.931%
	W/o self-supervised learning	5.268	9.101	19.991%	5.283	9.156	19.939%	5.302	9.192	19.850%	5.334	9.187	19.795%
	3S-TBLN	5.190	8.952	**11.473%**	5.195	8.984	**11.439%**	5.213	9.037	**11.431%**	**5.250**	9.078	**11.444%**
	Gains	-0.251%	-0.822%	15.266%	-0.193%	-0.797%	15.741%	-0.096%	-0.972%	16.134%	0.152%	-1.058%	16.296%

W/O Multi-Step Graph Attention Compared to 3S-TBLN, a decline in performance was observed. For the horizon 24 prediction**, w/o MSGraph-Att** resulted in increases in MAE, RMSE, and MAPE by 0.115%, -0.963%, and 19.237%, respectively, in Ningxia-YC, and by 0.152%, -1.046%, and 19.469% for forecasting the horizon 48. These results confirm the beneficial role of the multi-step graph attention (MSGraph-Att) mechanism in traffic speed prediction. Specifically, effective traffic diffusion modeling is crucial, and the design of MSGraph-Att is tailored to address this challenge.

W/O Fusion Gate Mechanism Spatial and temporal properties are essential elements that need integration in existing spatio-temporal methods. This paper introduces an adaptive fusion gate mechanism to achieve this integration. The performance of **w/o fusion gate mechanism** is lower than 3S-TBLN regarding the three metrics in Ningxia-YC. For instance, concerning horizon 24 in Ningxia-YC, **w/o fusion gate mechanism** exhibited increases of -0.096%, 1.217%, and 42.813% in terms of MAE, RMSE, and MAPE, respectively, compared to 3S-TBLN; similarly, it showed increases of 0.210%, 1.157%, and 42.573% for the horizon 48. The experiments underscore the significant impact of adaptively fusing spatial and temporal correlations on traffic speed prediction, highlighting its crucial role in the methodology.

W/O Traffic Patterns Illustrated in Fig.2, neighboring weeks exhibit similar traffic patterns, reflected in speeds and timestamps. Timestamps encompass weekly and daily periodic information, serving as indicators for traffic speeds. Subsequently, the speeds from the previous week and the current week's timestamps are embedded into the proposal. Regarding the three metrics on all datasets, the results of **w/o traffic patterns** are inferior to 3S-TBLN. For instance, in Ningxia-YC, **w/o traffic patterns** leads to increases in MAE, RMSE, and MAPE by 4.182%, 1.018%, and 46.479%, respectively, for the horizon 24, compared to 3S-TBLN; by 6.838%, 1.961%, and 47.947% for the horizon 48. In addition, compared with all other variants, **w/o traffic patterns** produces the worst forecast. The experiments validate that traffic pattern has a crucial impact on long-term speed prediction, alleviating forecasting errors and shifts.

W/O Self-Supervised Learning Traffic speeds exhibit continuous variations across different time steps, creating a bi-directional property in speed forecasting, where historical data influences future predictions and vice versa. To capture this contextual semantic, self-supervised learning serves as a pretext task aiding the model in understanding this bi-directional property, as evidenced in Table III. Comparison between models **w/o self-supervised learning** and 3S-TBLN highlights the clear advantage of the latter in traffic speed prediction. For the horizon 24, **w/o**

self-supervised learning *resulted in an increase of 1.707% in MAE, 1.715% in RMSE, and 73.651% in MAPE in Ningxia-YC; similarly, for the horizon 48, the increases were 1.600% in MAE, 1.201% in RMSE, and 72.973% in MAPE. These comparative outcomes verify the significance of self-supervised learning in long-term traffic speed prediction.*

Computation Cost:

Computation cost is a critical metric that reflects the application value of the proposed model, especially the inference time. Table IV shows the proposed method's and baselines' training and inference costs on Ningxia-YC, METR-LA, and PEMS-BAY datasets, and some exciting findings can be summarized.

Table IV Computation Cost of the Proposed Method and Baselines on Ningxia-YC, METR-LA, and Datasets ('' Means the Model Trains Once on the Whole Training Set)*

Data	Model	Parameters	Training / (iteration=100 and batch=8)		Inference / (batch=1)	
			Time cost	GPU usage	Time cost	GPU usage
Ningxia-YC	**ARIMA**	-	14196.343s	-	493.992s	-
	SVR	-	14,991.264s	-	4,218.156s	-
	LSTM_BILSTM	294,017	45.554s	1,516MiB	179.967s	524MiB
	ST-GRAT	2,238,849	26.204s	10,162MiB	1,037.873s	1,666MiB
	GMAN	907,201	37.750s	16,892MiB	72.284s	620MiB
	MDL	285,377	16.767s	5,026MiB	70.516s	5,018MiB
	DCRNN	372,353	138.466s	2,710MiB	1,034.471s	1,574MiB
	ASTGCN	255,686	26.125s	2,336MiB	45.718s	1,970MiB
	AGCRN	753,000	32.251s	4,376MiB	170.429s	2,578MiB
	MTGNN	1,111,744	6.913s	1,994MiB	16.600s	1,902MiB
	DGCRN	191,521	81.103s	2,862MiB	765.551s	2,028MiB
	Graph-WaveNet	493,312	10.998s	2,678MiB	48.779s	1,904MiB
	BigST	83,344	10.072s	1,638MiB	37.805s	4,752MiB
	STWave	226,930	11.705s	3,210MiB	35.649s	1,626MiB
	3S-TBLN	183,170	30.251s	8,692MiB	43.021s	748MiB
METR-LA	**ARIMA**	-	263,085.404s	-	8,024.562s	-
	SVR	-	314,609.572s	-	104,069.664s	-
	LSTM_BILSTM	312,641	61.896s	2,540MiB	627.576s	556MiB
	ST-GRAT	2,238,849	66.126s	20,068MiB	3,176.253s	1,774MiB
	GMAN	925,825	92.392s	30,748MiB	371.886s	1,004MiB
	MDL	342,017	39.769s	5,986MiB	283.443s	5,090MiB
	DCRNN	372,353	151.000s	3,656MiB	3,434.786s	1,576MiB
	ASTGCN	387,059	38.025s	4,456MiB	223.063s	3,554MiB
	AGCRN	753,990	59.787s	9,198MiB	668.584s	6,952MiB
	MTGNN	1,803,952	20.357s	2,274MiB	80.517s	3,404MiB
	DGCRN	199,441	84.789s	4,376MiB	4,884.951s	4,016MiB
	Graph-WaveNet	495,292	21.943s	3,652MiB	181.476s	3,444MiB
	BigST	92,656	10.340s	1,648MiB	139.888s	18,626MiB
	STWave	889,142	33.951s	6,384MiB	135.195s	1,716MiB
	3S-TBLN	201,794	74.977s	16,884MiB	172.138s	1,004MiB
PEMS-BAY	**ARIMA**	-	372,426.659s	-	8,781.825s	-
	SVR	-	479,571.950s	-	154,214.775s	-
	LSTM_BILSTM	320,193	79.129s	4,588MiB	684.415s	556MiB
	ST-GRAT (bat=4)	2,238,849	77.163s	19,002MiB	4,014.681s	1,968MiB
	GMAN (bat=4)	933,377	114.333s	30,748MiB	558.545s	1,644MiB
	MDL	394,881	105.961s	2,898MiB	405.336s	1,506MiB
	DCRNN	372,353	155.586s	4,640MiB	3,608.678s	1,578MiB
	ASTGCN	646,069	62.671s	6,126MiB	267.431s	4,726MiB
	AGCRN	755,170	69.153s	13,814MiB	817.691s	10,110MiB
	MTGNN	2,629,008	31.398s	2,750MiB	102.782s	4,442MiB

	DGCRN	*408,209*	*122.638s*	*6,556MiB*	*4,855.911s*	*5,402MiB*
	Graph-WaveNet	*497,652*	*33.481s*	*4,972MiB*	*215.526s*	*4,508MiB*
	BigST	*96,432*	*15.297s*	*1,664MiB*	*172.921s*	*29,250MiB*
	STWave	*889,146*	*37.816s*	*9,128MiB*	*210.510s*	*1,812MiB*
	***3S-TBLN** (bat=4)*	*209,346*	*64.558s*	*30,746MiB*	*183.104s*	*1,516MiB*

*For the **model size**, only deep learning considers the parameter scale. Compared with baselines, the proposed 3S-TBLN outperforms all baselines while utilizing fewer parameters, as evident in Tables II and III. However, a highly complex model may enhance performance at the expense of increased computational costs. The **training phase** for the proposed 3S-TBLN takes more time than baselines, except for ARIMA and SVR, owing to its dual optimization for prediction and reconstruction. Similarly, **inference cost** is relatively low, close to optimal benchmarks or lower than partial baselines, such as GMAN, BigST, and STWave. Considering the significant performance enhancement for long-term prediction (as shown in Table II and Fig. 7), the computation cost of 3S-TBLN is moderate. In addition, **GPU memory usage** is a critical metric for assessing application capabilities. Due to the substantial data load, the proposed model incurs higher memory utilization during the training phase but operates at a lower capacity during inference. Significantly, inference cost holds greater importance in real-world scenarios. Considering these observations, the prominence of 3S-TBLN in long-term speed prediction tasks becomes manifest.*

Case Study:

To better demonstrate the performance of 3S-TBLN, we compare it with the other six optimal spatio-temporal baseline models and visualize the fitting results.

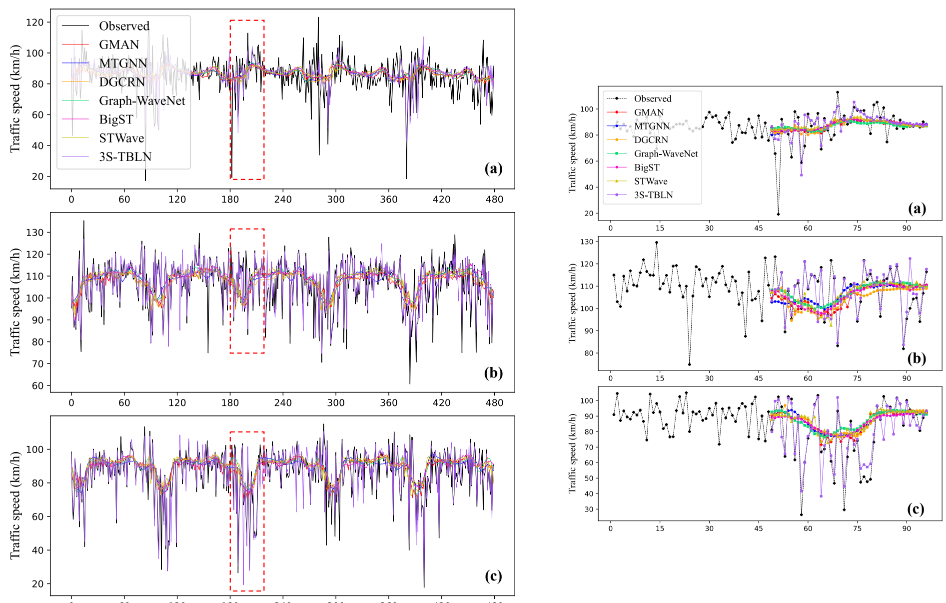

Fig. 9. Left: The visualization of fitting results in the Ningxia-YC dataset. Right: example of predicting the next 48-time steps (1-48 is historical time steps, and 49-96 is predicted horizons).

Figures 9, 10, and 11 show nine road segments exampled from the three datasets, respectively, and visualize the prediction results for the forty-eight-time steps. In the experiment, ten continuous samples are randomly sliced from the test set in each dataset. Figures 9, 10, and 11 show that 3S-TBLN can accurately fit the changing trend of traffic speed and adapt to complex speed fluctuations, compared with optimal baseline models, GMAN, MTGNN, DGCRN, Graph-WaveNet, *BigST, and STWave. For example, the traffic speeds present huge differences between different types of road segments, such as in Fig. 9 and 11; however, the performance of 3S-TBLN remains steady compared with baseline models. This suggests that the 3S-TBLN method more effectively captures the dynamic characteristics of traffic flow when modeling traffic speed variations over continuous time steps.*

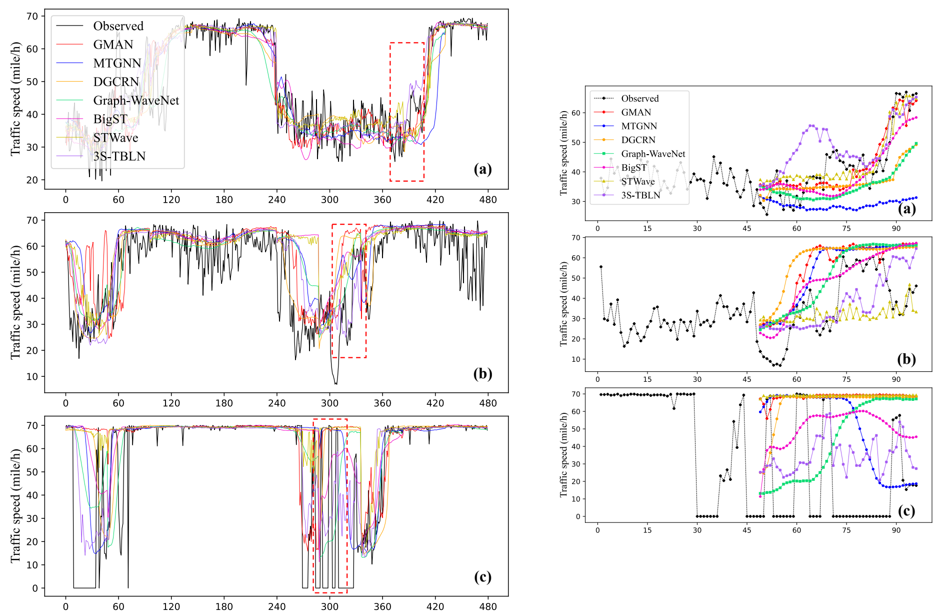

Fig. 10. Left: The visualization of fitting results in the METR-LA dataset. Right: example of predicting the next 48 time steps.

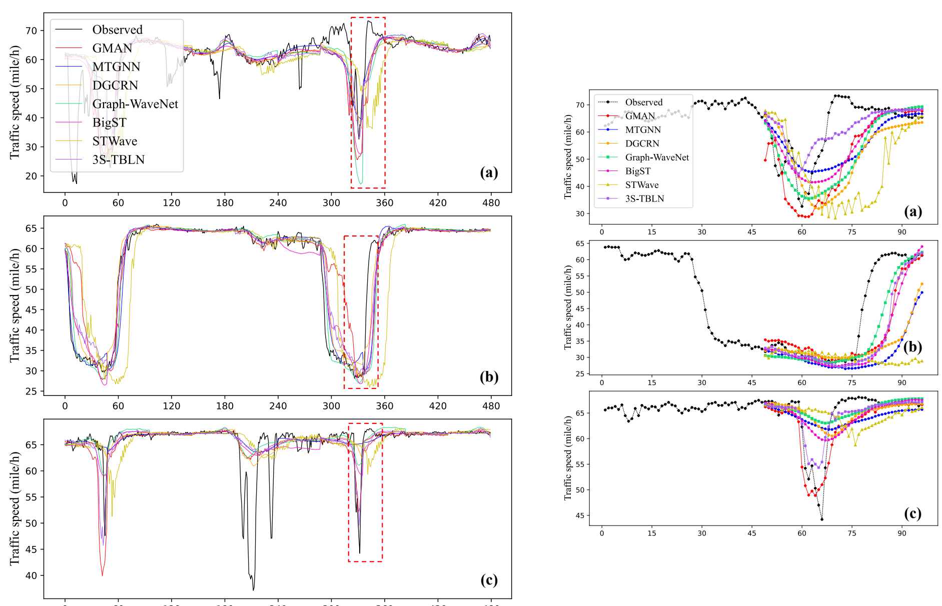

Fig. 11. Left: The visualization of fitting results in the PEMS-BAY dataset. Right: example of predicting the next 48-time steps.

Moreover, the proposed 3S-TBLN method predicts the traffic speed trend and speeds normally under lacking historically observed values, such as Fig. 10 (c). 3S-TBLN consistently outperforms other baseline models by producing predictions that closely align with actual observations and trends. It effectively learns traffic trends even without relying solely on real historical observations

and effectively handles speed fluctuations. These properties play a vital role in future travel services and traffic control.
We deeply appreciate your insightful feedback, which has led to a stronger experimental validation of our model. With the newly added baselines and comparative analysis, we believe our manuscript now provides a more comprehensive and up-to-date evaluation of **3S-TBLN**. Thank you once again for your constructive comments. We look forward to any further suggestions you may have.

(8) To my understanding, 3S-TBLN is a long-term prediction model, and long-term prediction is related to the temporal granularity of the dataset. In the experiments, the longest prediction horizon is 12, which is a conventional horizon in the field of traffic prediction. The authors could attempt longer horizons, such as 24, 36, or 48, to better highlight the model's long-term prediction capabilities.

Response: We sincerely appreciate your insightful comments regarding the long-term prediction capabilities of **3S-TBLN** and your suggestion to extend the prediction horizon beyond the conventional 12-step setting. Below, we provide a detailed response addressing your concern.

1. Justification for the Current Prediction Horizon (12 Steps)

In our original experiments, we adopted a **12-step prediction horizon**, which is widely used in traffic prediction studies. This choice was made based on the following considerations:

- **Comparability with Existing Works**: Most benchmark studies in traffic forecasting report performance at 3, 6, and 12 steps, making it easier to compare our results with state-of-the-art methods.
- **Maintaining a Balance Between Accuracy and Stability**: Longer prediction horizons (e.g., 48 steps) often introduce greater uncertainty due to compounding errors, making it challenging to maintain stable and reliable predictions.
- **Dataset Characteristics**: The temporal granularity of the datasets (5-minute intervals for METR-LA and Ningxia-YC) means that a 12-step horizon already corresponds to a **one-hour-ahead** prediction. Beyond this, the traffic management department and traveller can make accurate traffic control or trajectory planning in advance according to the 12-step horizon.

2. Extending the Prediction Horizon 48 Steps

We agree that testing longer horizons would better demonstrate **3S-TBLN's long-term forecasting capabilities**. In response to your suggestion, we have conducted additional experiments with **6-step, 12-step, 24-step, and 48-step horizons** and updated our results in **Tables [II] and [IV]** of the revised manuscript.
Our key findings from these extended experiments are as follows:

- **3S-TBLN remains competitive for longer horizons**: Even with **24-step (2-hour-ahead)** and **48-step (4-hour-ahead) predictions**, our model maintains reasonable accuracy compared to conventional 12-step settings.
- **Performance degradation is expected but controlled**: As with all deep learning-based traffic prediction models, errors accumulate as the prediction horizon increases. However, **3S-TBLN's multi-scale spatiotemporal learning structure effectively mitigates this degradation**, making it more suitable for long-term forecasting than many existing baselines.

- **Discussion on long-term limitations**: In **Section [VI]**, we now include an expanded discussion on the challenges of long-term forecasting in traffic prediction and propose future work that could integrate Knowledge Transfer Learning (that means using similar traffic patterns from other road networks to complement the insufficient prediction performance in the target road due to lacking traffic data) to further enhance predictive stability.

3. Changes to the Revised Version

The results, comparisons, and analysis have been added to question (7)'s response. In addition, the prediction length in the context of research has been modified from 12 steps to 48 steps. While **3S-TBLN** demonstrates strong performance for long-term forecasting, we acknowledge that further enhancements could be explored. We have outlined these directions in **future work** (**Section [VI]**) to provide insights into possible extensions of our research.

Several limitations should be acknowledged. This study does not account for the influence of external factors, such as traffic accidents and meteorological conditions, on traffic speed. Moreover, the long-term forecasting performance obtained is modest, indicating a need for further refinement to enhance the generalization capacity of our 3S-TBLN. Considering these identified limitations, as part of future endeavors, we aim to explore knowledge transfer learning to integrate knowledge from different highway networks with similar traffic patterns, thereby improving prediction accuracy. Accessing multi-source data in real-world highway applications presents challenges; for instance, some rural areas lack monitoring (e.g., meteorological data), and certain records, such as traffic accidents, are often delayed. Despite these constraints, future efforts will incorporate external factors into the traffic prediction process, particularly for urban traffic networks.

We sincerely appreciate your constructive feedback, which has prompted us to extend our experimental analysis and provide a more comprehensive evaluation of **3S-TBLN's** long-term forecasting capabilities. With the inclusion of **6-step, 12-step, 24-step, and 48-step** results, as well as an expanded discussion on long-term prediction challenges, we believe our manuscript now provides a stronger validation of our model. Thank you again for your valuable suggestions, and we welcome any further recommendations.

(9) Some figures need further optimization. For example, the curves in Figures 9 and 10 are too thick, making it visually difficult to distinguish the fitting effects of different models.

Response: Thank you for your valuable feedback on our manuscript. We appreciate your suggestion regarding the optimization of Figures 9 and 10. You are right that the thick curves in these figures make it difficult to distinguish the fitting effects of different models. In addition, there are so many baselines in Figures 9 and 10, resulting in the curves not being clear. We have addressed this issue by optimizing the figures as follows:

1. Optimization of Figures 9 and 10

To enhance the clarity and visual distinction between different models, we have made the following optimizations:

1. **Reduced Line Thickness:** We have decreased the line thickness of all curves in Figures 9 and 10. This makes the curves thinner and easier to visually compare.

2. **Adjusted Line Styles and Colors:** To further enhance the distinguishability, we have varied the line styles (solid, dashed, dotted) and used a color palette with higher contrast for different models. This helps in clearly identifying each model's curve even when they are close to each other.
3. **Added Markers at Key Points:** We have introduced markers at specific intervals along the curves. These markers serve as visual anchors and make it easier to track the performance of each model across different prediction horizons.
4. **Legend and Label Improvements:** We have refined the legends and axis labels to ensure clarity. The font size has been adjusted appropriately to ensure readability without overcrowding the figure.
5. **Reduced Baselines in Figures:** We chose six optimal baselines to visualize the fitting performance in the Figures, which could provide good vision so that the reviewer and audience can read clearly.

Here are the revised Figures 9 and 10:

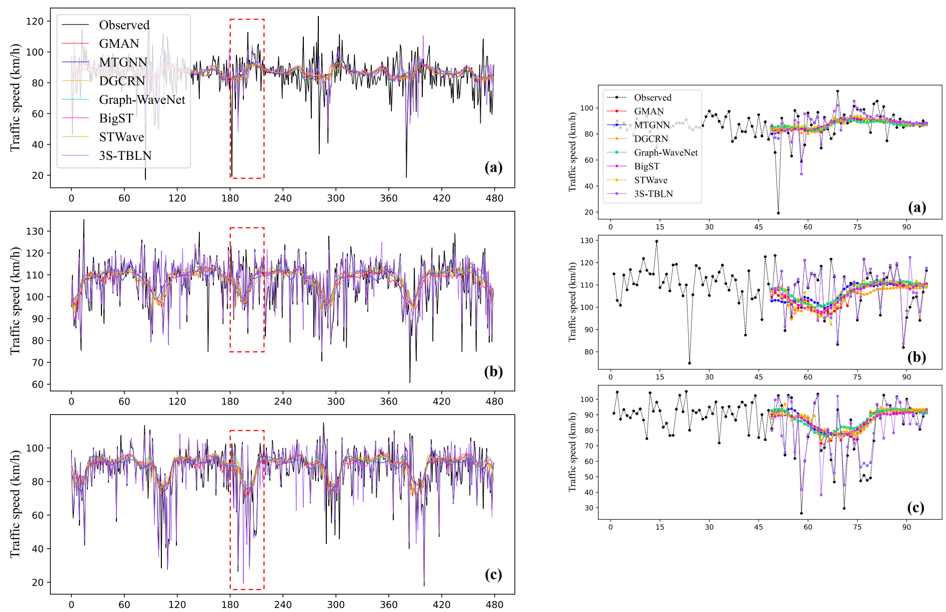

Fig. 9. Left: The visualization of fitting results in the Ningxia-YC dataset. Right: example of predicting the next 48-time steps (1-48 is historical time steps, and 49-96 is predicted horizons).

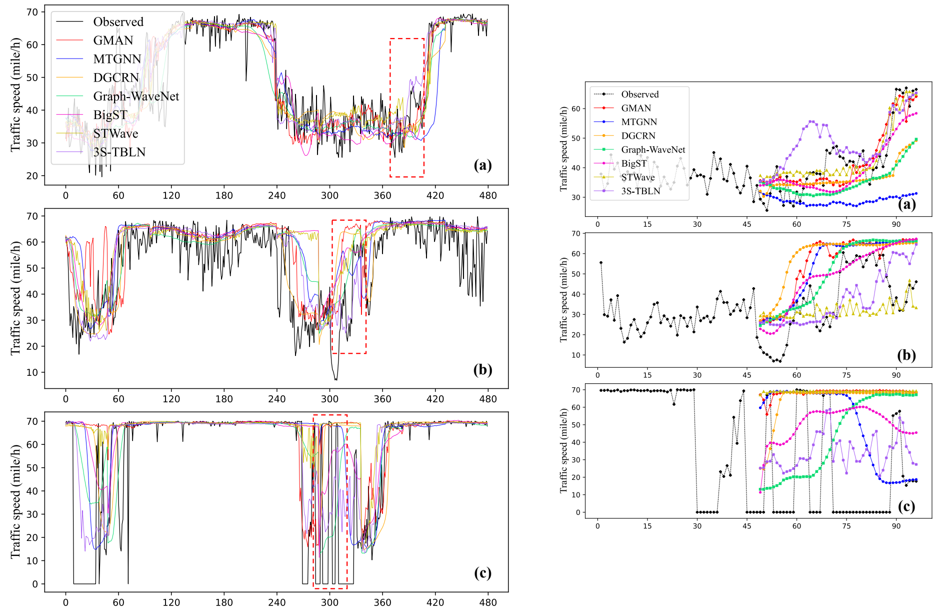

Fig. 10. Left: The visualization of fitting results in the METR-LA dataset. Right: example of predicting the next 48 time steps.

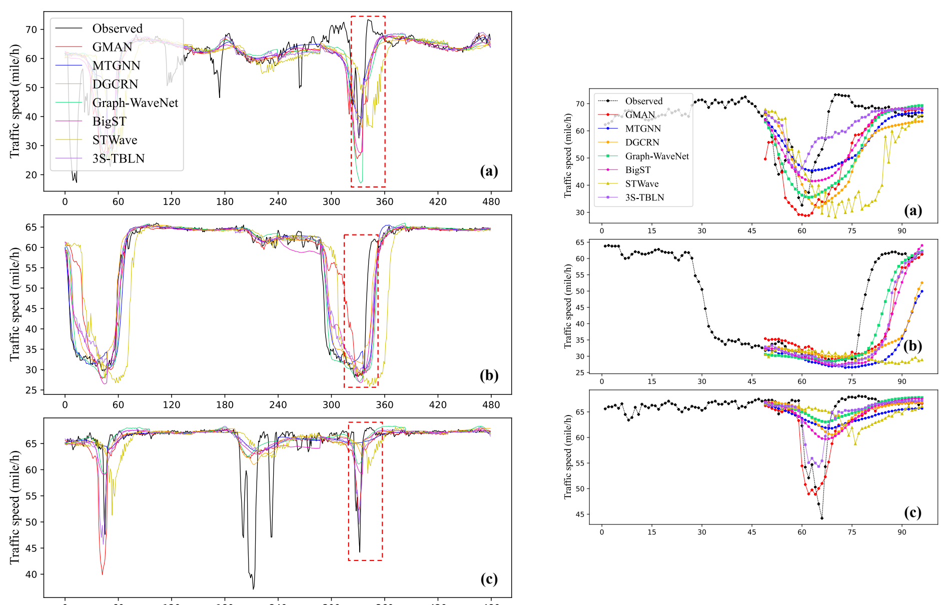

Fig. 11. Left: The visualization of fitting results in the PEMS-BAY dataset. Right: example of predicting the next 48-time steps.

We believe these optimizations have significantly improved the clarity and readability of the figures. The thinner curves, varied line styles and colors, and added markers now allow for a clear comparison of the different models' fitting effects across various prediction horizons.