To All Reviewers: We would like to thank all reviewers for your time and insightful feedback. Our full version (18 pages with appendix) already contains the answers to some of your questions, which can be found in the supplementary zip file in our original submission. We address your major concerns briefly here, and will incorporate all comments carefully in the revised version.

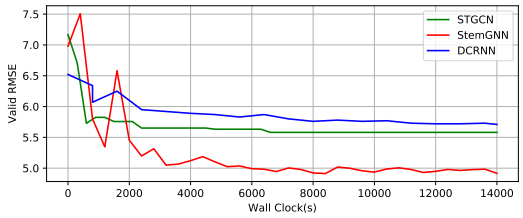

Figure 1: Learning curve comparison on METR-LA.

A common concern is about the time complexity and training/inference costs. A fixed sliding window is used for long sequences, so the time complexity is linear to the length of input data. Although the time complexity is $O(N^3)$ w.r.t. the multivariate dimension N, training process on all the datasets can be finished within a reasonable time. As shown in Table 7 (P15, L570), the training time of StemGNN is acceptable comparing with other SOTAs. For one epoch, StemGNN uses 459s on PEMS07 and 1137s on METR-LA, while the fastest SOTA uses 352s on PEMS07 and 1035s on METR-LA. We can see that StemGNN has acceptable efficiency while achieving better accuracy. When N is extremely large, we can derive more practical solutions by dimension reduction techniques (e.g., randomized SVD), while the experimental study is left to future work. The learning curve on METR-LA dataset is shown in Figure 1, where the x-axis denotes wall-clock time and the y-axis denotes validation RMSE. It shows that StemGNN has an effective training procedure and achieves a superior RMSE at convergence.

To Reviewer 1: We will fix the typos in the text and include a notation table in the Appendix. More prediction figures on COVID-19 datasets are available in Figure 6 (Appendix E.3 (P18)). As shown in the figures, StemGNN is able to make reasonable predictions for the trend. However, fluctuations are much more difficult to be estimated, which may be caused by random noises. We will add more prediction figures on other datasets in the revised version.

After deleting the second StemGNN block, the 15min/30min/45min MAE drops by 5.4%/3.9%/15% on PEMS07, and the 15min/30min/1hour MAE drops by 8.8%/11.9%/11.5% on METR-LA, which indicates the effectiveness of the second StemGNN block.

To Reviewer 2: StemGNN generates a dynamic correlation matrix for each sliding window by the Latent Correlation Layer. We visualize the average correlation matrix from all sliding windows for clearer insights.The dynamic evolution will be analyzed in a future research. Moreover, capturing dynamic correlations jointly in the model is beneficial. You may refer to the ablation results in Table 2 (P7, L233) and Table 6 (Appendix D.2 (P14, L548)), where the "w/o LC" setting utilizes a static covariance matrix from SOTA. Especially for the METR-LA dataset (Table 6), we can see a significant performance drop (8% drop on MAE), showing the advantage of the Latent Correlation Layer.

To Reviewer 3: Removing GFT leads to a significant reduction in performance, as shown in "w/o GFT" of Table 6 (Appendix D.2 (P14, L548)). Especially on the METR-LA dataset, there is a 10% drop on MAE. In Figure 5(c) (Appendix E.3 (P16, L591)), we show the trajectories of countries related to the top eigenvectors. It is clear that the top three eigenvectors correspond to countries in the world, Asia, and South America, respectively. In addition, citation [10] has been moved to Appendix C due to space limitation. We will make the revised version self-contained.

To Reviewer 4: First, we'd like to emphasize our novelty and contributions. (1) The designed model for multivariate time-series forecasting is non-trivial, and many endeavors explore along this direction. Our work is the first that *models both inter-series and intra-series correlations jointly in the spectral domain* and achieves significant performance gains than the best SOTA for various applications (13.3% average RMSE surge on 9 datasets). (2) The model can be generalized on all multivariate time-series without predefined topologies, which solves the pain point of existing SOTAs [16, 33]. (3) As shown in Appendix E.2 & E.3 (P15-18), the learned attention matrices are explainable by humans, which benefits the transparency and credibility of the model.

We have reproduced the LSTNet [15] on all datasets. As a result, StemGNN outperforms LSTNet by 17%/21%/14% for MAE/RMSE/MAPE on the average of nine datasets. Therefore, it does not restrict the potential impact of our model. We have a detailed description of hyper-parameters for StemGNN in Appendix C (P13, L513). For example, the values of H on all datasets can be seen in lines 529-539 of Appendix C. Appendix B (P12, L458) introduces the settings to reproduce SOTAs. We will include a notation table for better presentation in the revised version.

In addition, we give brief answers for clarity here and will polish the corresponding parts in the final version. (1) Y_i in Figure 1 means the i^{th} output channel of Y. (2) In equation 2, B indicates the entire network that generates backcasting outputs (refer to L175). (3) w^q, w^k are learnable parameters in the attention mechanism. (4) The number of eigenvectors used in GFT is equivalent to the multivariate dimension (N) without dimension reduction.