Author Rebuttal to Paper Reviewers

prioritize addressing major concerns:

1. Technical Correctness→ Has major problems

a. I invite the author to strengthen the validation part by considering at least the transformer-based model cited in their paper, i.e. Informer and PatchTST, possibly also on the Electricity dataset. It is important to check, though, that PatchTST is still currently the state of the art.

Reply: Thanks for the reviewer's suggestion. In order to make a fair comparison with the Transformer-based timing prediction model, we show the experimental results using RK-block based Transformer in supplementary experiments (See Table I), and our proposed method has the advantages of accuracy, parameter efficiency, and speed of inference over the Transformer-based method. (See RKTrans(RK2-gamma) v.s. Transformer(RK2-gamma) and RKTrans(RK4) v.s. Transformer(RK4)). Unfortunately, due to limited time, we are unable to add more results for the Transformer-based model in the same experimental setup at this time, and the full results will be given in the final submission of the paper.

Table I Results of supplementary experiments. The numbers in the table are the means of 5 experiments, and the standard deviations are in parentheses.

Forecasting step : 48		
RKTrans(RK4)	RMSE	2.687 (0.05)
	MAE	2.07 (0.07)
	#Params	67539
	Inference time	481.69s
RKTrans(RK2-\gamma)	RMSE	**2.615 (0.18)**
	MAE	**1.932 (0.21)**
	#Params	67541
	Inference time	267.26s
RKTrans(Res)	RMSE	2.924 (0.20)
	MAE	2.321 (0.22)
	#Params	67539
	Inference time	140.63s
Transformer (RK4)	RMSE	2.97 (0.23)
	MAE	2.431 (0.26)
	#Params	135137
	Inference time	1057.14s
Transformer (RK2-\gamma)	RMSE	2.717 (0.24)
	MAE	2.079 (0.30)
	#Params	135141
	Inference time	570s

b. To strengthen your paper, I recommend providing a more detailed rationale for the choice of RK blocks. This explanation can be achieved through either theoretical justifications or by citing relevant literature that supports the assertion. Elaborating on the mechanism and advantages of the RK blocks in capturing long-range dependencies would significantly enhance the clarity and credibility of your paper.

Reply: Thanks for the reviewer's suggestion. In fact, we state the motivation for introducing RK-block in the Transformer architecture in the introdution of the paper and cite the relevant references, which give complete experimental results to demonstrate the advantages of RK-block in capturing long-distance dependencies. The original text and corresponding references are given below:

"Based on the above analysis, improving the inference speed and the ability to retain temporal order information of the Transformer model while retaining its advantage of capturing global correlation has become a key issue in applying the Transformer model to the field of power forecasting. In this paper, inspired by [15, 16], we associate residual networks with Euler discretization of solutions of ordinary differential equations (ODE) and propose the Transformer-based model *RKTrans* with improved residual connection units for accurate, real-time forecasting of power time series."

Reference:

[15] Bei Li, Quan Du, Tao Zhou, Yi Jing, Shuhan Zhou, Xin Zeng, Tong Xiao, JingBo Zhu, Xuebo Liu, and Min Zhang, "Ode transformer: An ordinary differential equation-inspired model for sequence generation," arXiv preprint arXiv:2203.09176, 2022.

[16] Xingcheng Zhang, Zhizhong Li, Chen Change Loy, and Dahua Lin, "Polynet: A pursuit of structural diversity in very deep networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 718–726.

2. Experimental Validation→ Insufficient validation

a. Moreover, it would be indicative to run multiple executions based on different seeds to report the mean and standard deviation, thus checking the robustness of the model.

Reply: Thanks your suggestion. We give the means and standard deviations of the repeated experiments of the proposed method and the comparison model in the supplementary experiments.

b. The choice to replace the standard transformer decoder with an MLP for inference acceleration requires a comparison with a full transformer decoder. Assessing the trade-off between speed and performance is essential. A comprehensive comparison of computational complexity, including parameters, memory usage, and inference speed, with existing methods is necessary to provide context. Please clarify how the complexity of RK4 and RK2-gamma, relative to the residual connection version, is managed during inference. This is crucial, especially if speed is a primary concern.

Reply: Thanks for the reviewer's suggestion. We give the results of experiments comparing the

standard RK-block-based Transformer with RKTrans in supplementary experiments (See TableI), where RKTrans achieves an inference speedup with comparable performance (See RKTrans(RK2-gamma) v.s. Transformer(RK2-gamma) and RKTrans(RK4) v.s. Transformer(RK4)); and a comprehensive comparison of number of parameters, and inference speed are also provided; the results of the complementary experiments also verify that RK4 and RK2-gamma achieve higher accuracy with similar inference speeds, relative to the residual connection version (See RKTrans(RK2-gamma) and RKTrans(RK4) v.s. RKTrans(Res)).

Table I Results of supplementary experiments. The numbers in the table are the means of 5 experiments, and the standard deviations are in parentheses.

Forecasting step : 48		
RKTrans(RK4)	RMSE	2.687 (0.05)
	MAE	2.07 (0.07)
	#Params	67539
	Inference time	481.69s
RKTrans(RK2-\gamma)	RMSE	**2.615 (0.18)**
	MAE	**1.932 (0.21)**
	#Params	67541
	Inference time	267.26s
RKTrans(Res)	RMSE	2.924 (0.20)
	MAE	2.321 (0.22)
	#Params	67539
	Inference time	140.63s
Transformer (RK4)	RMSE	2.97 (0.23)
	MAE	2.431 (0.26)
	#Params	135137
	Inference time	1057.14s
Transformer (RK2-\gamma)	RMSE	2.717 (0.24)
	MAE	2.079 (0.30)
	#Params	135141
	Inference time	570s

3. Novelty/Originality→ Minor originality

a. Consequently, it remains unclear what unique contributions your work brings to the field beyond its application to a new dataset.

Reply: Thanks for the comment. We respectfully disagree with the reviewer's comments on our work. Our contribution focuses on the application of Transformer-based time-series models to energy prediction tasks for suitable improvements, rather than being limited to applications on a new dataset.

b. To address these concerns, it would be helpful for your paper to clearly delineate what aspects

you are introducing and to what extent you are building upon previous work and existing methodologies. It is important to emphasize your unique contributions explicitly, making it evident how they extend the state of the art; The only distinct contribution I could discern in your paper concerning [15] was the incorporation of an MLP in the decoder. However, this aspect lacks robust theoretical justification and empirical validation.

Reply: Thanks for the comment. As mentioned in the above reply, our work is based on attempts in the field of NLP to apply RK-block to the Transformer architecture to handle longer input texts; to the best of our knowledge, we are the first to introduce RK-block to the task of temporal prediction of energy data in order to capture long-distance dependencies; furthermore, we investigate the importance of performing seasonal modeling. In summary, we believe that the reviewers failed to understand the focus of our work.

focus on addressing major concerns

Reference to Prior Work→ Does not cite relevant work; Some minor formatting issues and grammar issues.

Reply: In the latest camera-ready version, we could update the references and proofread the formatting of the paper.