

Reply to Reviewer #02

We thank the reviewer for the suggestions and comments to improve our manuscript. We have accordingly revised our manuscript to address these concerns. Please find detailed point-by-point responses (in blue) to the reviewer's issues below.

The authors develop AtmoNAT by taking a NA/NATTEN type model and then supplementing it with GRPE and SAFN. The work is promising and substantive, and extensively evaluated/benchmarked (though this can be further improved). I read RC1 and I agree with all of it. I recommend publication after the authors address RC1 and the comments below.

- ✧ Mirroring RC1: Clarify contributions and novelty more clearly. In my mind, you have two new pieces (GRPE and SAFN) and you add them to established architecture to come up with a new model (AtmoNAT).
- ✧ I think you could do more to highlight GRPE and its power/usefulness relative to other PE methods.
- ✧ I applaud the ablation study performed, but due to challenges mentioned in the manuscript, it is not really possible to isolate the contributions cleanly. Right? Can you say more about this? Any tricks to improve isolation?
- ✧ I think it would help to have more discussions/results on how the architecture additions/innovations (GRPE, SAFN) connect to physics mechanisms, but this could also be left to future work... some more discussion may help here though.

Reply: Thank you for your comments.

1. We have revised the manuscript to clarify our contributions. Details can be accessed at the first reply in our reply to the first reviewer.
2. We compared GRPE with other PEs, including Relative PE and RoPE-Mixed, which are used in FuXi and FengWu. We also have done an ablation study to isolate GRPE's contribution compared to other modules. In discussion, the visualization of GRPE further elucidate its working principle. We think more experiments on GRPE's performance is not that urgent.
3. You are right. It is not possible to isolate the contributions cleanly. For example, the order of ablation study may affect the results. GRPE's design is not compatible with Swin Attention, so the first ablation choice should be the backbone comparison between Swin Attention and Neighborhood Attention. Finetuning of AtmoNAT should be placed at the final stage since it is about training strategy rather than model structure.

In principle, the order of (2) adding the Multi-Scale Head, (3) incorporating GRPE into the Neighborhood Attention, (4) substituting the standard Transformer feed-forward network with SAFN, and (5) increasing the latent dimension to 128, can be changed freely. But that would result in 16 different combinations of ablation study. The training of AtmoNAT is time consuming. "The model was trained on two NVIDIA RTX

4090 GPUs for 28 days”. Even the 96-dimensional model without fine-tuning still take 7 days. Doing all 16 experiments is not possible under limited resources.

On the other hand, we think even all the 16 experiments are conducted. The averaged results will not diverge so much from our presented results. To improve the isolation, we think there is hardly any trick to cut corners but to do more experiments, since the interaction of different modules may still affect the results.

4. We do state a lot about our physical motivations of GRPE and SAFN in the introduction, including:

“Both approaches impose geometry-based positional priors that are independent of the instantaneous atmospheric state: identical spatial configurations in attention windows across locations and forecast times receive the same relative bias or rotation rule. Such static representations may be restrictive for Earth-system forecasting, where interactions among atmospheric variables are spatially heterogeneous and dynamically modulated by geographical setting, circulation regime, season, and evolving weather conditions.” (L59-L63 of the submitted manuscript).

“These findings motivate the development of positional encodings that can represent atmospheric dynamics more efficiently while adapting to both location and flow state.” (L73-L75 of the submitted manuscript).

“Rather than prescribing a fixed positional relationship, GRPE dynamically adjusts relative positional biases according to the evolving atmospheric state. Specifically, GRPE represents the positional bias as a gated linear combination of learnable prototypes, providing a compact yet flexible representation of flow-dependent spatial relationships. It further retains coordinate-dependent biases to capture persistent geographical heterogeneity, analogous to the role of ESPB, but with only a negligible increase in model size.” (L78-L82 of the submitted manuscript).

“Interactions among atmospheric variables are central to weather evolution, motivating the use of gating to represent these dependencies in feature space.” (L92 of the submitted manuscript).

“SAFN combines channel expansion, parallel branches, and depth wise convolution with multiplicative gating to model feature interactions. We further introduce a global positional bias to capture spatial context associated with terrain.” (L99-L102 of the submitted manuscript).

As you can see, GRPE and SAFN’s designs are deeply rooted in its physical motivations (or physics mechanism in a broad way), which is fully elucidated in the introduction. There is no need to repeat them in the discussion.