Thank you for the detailed review and suggestions on ways we can improve the manuscript. Your feedback is very much appreciated! The following text includes a point by point response to each comment.

Review 1: Line 2: I disagree with the title, in that the downscaling is not to arbitrary time resolutions. It is to 1-hourly resolutions, and the manuscript does not demonstrate the ability to interpolate to finer resolutions.

Response 1: Thank you for your feedback. We appreciate your input regarding the title. We recognize that the current experiments primarily demonstrate downscaling to 1-hourly intervals. However, our model is designed to support interpolation to finer resolutions by leveraging the temporal coherence constraint and the encoder-decoder architecture. We chose to focus on 1-hourly resolutions in this paper as a practical example but are open to clarifying in the title and abstract that the demonstrated application is specifically to 1-hour intervals.

Review 2: Line 28: I am not sure what you mean by precise interpolation. The process of interpolation should inherently be precise by virtue of being a mathematical exercise, but with a limited accuracy inherent to the method. I suggest rewording to “accurate.”

Response 2: We agree that the term "precise" may imply an exactness that is limited by the interpolation method's inherent accuracy. We have revised the wording to "accurate interpolation" to more accurately convey our intended meaning.

Review 3: Line 57: I think it would be more fair to say that reanalysis product temporal resolutions are mostly on the order of hours, not hours to days.

Response 3: We agree that describing reanalysis product temporal resolutions as "mostly on the order of hours" would be more accurate. We have updated this section to reflect this clarification and ensure our wording aligns with current reanalysis product standards.

Review 4: Line 126: I do not understand what you mean by “wind volume.” Please elaborate or rephrase.

Response 4: We realize that "wind volume" may have been unclear. Our intention was to refer to "wind components" or "wind speed" derived from horizontal and vertical wind vector components. We have revised this term for clarity in the manuscript.

Review 5: Line 128: Is there any specific reason to not match the resolution of

topography to that of the ERA5 data?

Response 5: We used terrain data with a slightly coarser resolution (15 km) as a balance between computational efficiency and capturing essential topographic features. This resolution is sufficient to represent major terrain influences on atmospheric processes without significantly increasing computational demands. However, we can discuss this choice more explicitly in the manuscript and explore the potential impact of matching resolutions in future work.

Review 6: Line 131ff: That a region with transitions between climate zones is highly representative of climate change studies is a very strong statement. I believe that this statement does not belong into a figure caption, but should be postulated in the bulk text, discussed further, and above all substantiated with adequate references. In the absence of a thorough discussion, I am not sure what the statement means in detail, and it therefore seems rather dubious. A more thorough discussion would also justify section 2 being a separate section, as currently the section consists of only one paragraph, which could be moved to section 3, which in turn could be renamed to "data and methods."

Response 6: We acknowledge that the statement about the region's representativeness in climate change studies requires further elaboration and should not be confined to a figure caption. We have moved this statement to the main text, within Section 2, and substantiated it with relevant references to clarify its relevance and significance. Additionally, we have expanded the discussion on the region's representativeness in climate studies to avoid ambiguity. Following your suggestion, we have combined Sections 2 and 3 into a unified "Data and Methods" section for coherence.

Review 7: Line 136: Is there a specific reason to use x and y for horizontal and vertical dimensions? It might be more intuitive to use z for the vertical component, as that is more commonly used.

Response 7: We used x and y to denote horizontal and vertical dimensions to align with our grid structure, but we agree that using z for the vertical component would be more intuitive and consistent with common conventions. We have revised the notation accordingly to improve clarity.

Review 8: Line 148ff: I suggest removing the technical description from the figure caption, as this inflates the caption with substance that belongs in the bulk text.

Response 8: We agree that the technical details in the figure caption may be better suited for the main text to streamline the caption. We have moved this information to the bulk of the text and keep the caption concise, focusing on the essential elements of

the figure.

Review 9: Line 157ff: This is not easily understood by me. It appears that you are claiming that processes that can occur over long periods of time can also occur in short periods of time, which I do not agree with. Further, the entire paragraph is worded in a way that implies that it is entirely speculative, and not substantiated. As a result, I cannot confidently agree with anything in this paragraph, and am led to disregard further argumentation that builds on it. I suggest rephrasing the paragraph for more clarity, and to substantiate the claims made therein.

Response 9: Thank you for your thoughtful feedback on Line 157. We appreciate the opportunity to clarify our intentions and provide additional substantiation for the statements made in the paragraph. Our goal was not to imply that atmospheric processes occurring over long periods can directly occur within short periods. Rather, we aimed to highlight that certain atmospheric phenomena exhibit self-similar behavior across different time scales. Specifically, rapid transitions in atmospheric variables can encapsulate dynamic processes that, while condensed in time, share characteristics with longer-term evolutions. This concept is supported by the theory of scale invariance and fractal behavior in atmospheric dynamics, where patterns at one time scale can resemble those at another (Lovejoy & Schertzer, 2013). For instance, turbulence and convection processes exhibit self-similar structures across scales (Schertzer & Lovejoy, 1987). By identifying and utilizing these rapid-transition scenarios, we can extract valuable information that helps the model learn underlying atmospheric dynamics without requiring high-temporal-resolution data. In our methodology, we leverage scenarios with rapid transitions as "proxy" high-resolution data for pretraining. These scenarios are not meant to replicate long-term processes in a shorter time but to provide rich information content that enhances the model's ability to capture complex temporal patterns. We acknowledge that the original wording may have been unclear and appeared speculative. To address this, we propose to include additional explanations and references to substantiate our claims:

- Scale Invariance in Atmospheric Dynamics: Atmospheric fields often exhibit scale-invariant properties, meaning that certain statistical features are preserved across different temporal and spatial scales (Davis et al., 1994).
- Self-Similarity in Meteorological Processes: The concept of self-similarity suggests that small-scale processes can reflect the properties of larger-scale ones, which is a foundational idea in turbulence theory (Frisch, 1995).

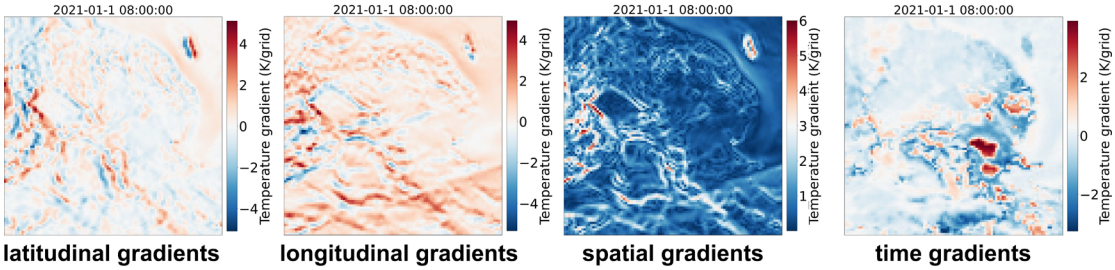
Review 10: Line 162ff: This paragraph states that changes in certain variables can signify changes in weather patterns. While I agree with that, I think it would be good

to also incorporate at least a substantiated estimate of how often this is the case, e.g., how often a change in humidity actually triggers severe convection. Further, it is not clear to me why these events should be filtered out. This is especially the case due to the previous argumentation of an increase in temporal resolution being useful in capturing the timing of changes in weather patterns.

Response 10: Thank you for your insightful comments. The purpose of filtering events with rapid transitions was to enhance the model's capability to capture distinct, representative patterns of atmospheric change, which serve as "pseudo-labels" during pretraining. By training on these cases, the model learns to identify dynamic weather patterns and respond accurately when downscaling at finer resolutions. While incorporating estimates on the frequency of specific events, such as humidity changes triggering convection, could add context, we believe that this filtering approach aligns with our goal of improving temporal resolution by focusing on high-impact scenarios. These representative patterns guide the model's performance in capturing similar transitions across different weather conditions, thus supporting our objectives in a self-supervised framework.

Review 11: Line 211: I do not understand how a temperature gradient can be given in Kelvin, without any reference to physical space (K per meter) or time (K per second). It is unclear whether this gradient refers to a difference in temperature between adjacent grid point and adjacent time steps, or something different. Please rephrase for clarity. Further, the phrasing implies that t2m has vertical gradients, which seems odd to me.

Response 11: Thank you for highlighting this issue. We agree that specifying the temperature gradient in Kelvin alone is ambiguous. We have revised the text to clarify that the gradient refers to temperature differences across either spatial (e.g., K per meter) or temporal (e.g., K per second) dimensions. Additionally, we have rephrased the description to avoid implying that t2m has vertical gradients, as t2m refers specifically to 2-meter air temperature, which typically considers horizontal gradients.



Review 12: Line 237: It might be useful to very briefly state what an encoder-decoder architecture is, and what it aims to achieve, at the beginning of this section, as is done

for the section at line 263.

Response 12: We agree that a brief introduction to the encoder-decoder architecture would be beneficial for readers. We have added a short explanation at the beginning of this section, similar to the description provided later, to clarify its purpose and functionality within our model.

Review 13: Line 301f: Table 1 does not show downscaling from 6-hourly to 1-hourly intervals, and it should be stated that rh and z are only downscaled from 2-hourly to 1-hourly intervals.

Response 13: We have revised the text to clarify that Table 1 does not include downscaling from 6-hourly to 1-hourly intervals, and that relative humidity (rh) and geopotential height (z) are only downscaled from 2-hourly to 1-hourly intervals.

Review 14: Line 303ff: This subsection describes the methods used to evaluate the performance of TemDeep, and therefore belongs in section 3.

Response 14: We agree that the performance evaluation methods should be part of the methodology section for consistency. We have moved this subsection to Section 3, so that all methodological details are consolidated.

Review 15: Line 322ff: The table caption should contain a description of the structure of the table, not an interpretation of the results.

Response 15: We agree that table captions should focus on describing the table's structure rather than interpreting the results. We have revised the caption to provide a clear description of the table's organization and contents, moving any interpretative text to the main body of the manuscript.

Review 16: Line 324: Saying that TemDeep performance approaches performance of TemDeep* seems plausible, but using the word “closely” seems to be an optimistic stretch in some cases (e.g., t2m(2h→1) and t2m(3h→1)). It might be better to rephrase to not overstate the performance, or to argue why this qualifies as “closely approaching.”

Response 16: We agree that “closely” may overstate the performance comparison in some cases, such as for t2m (2h→1) and t2m (3h→1). Corrected.

Review 17: Line 371ff: As diurnal variations are relevant to the evolution of the assessed fields, it would be good to state whether these times are in UTC or the local time within the region.

Response 17: We agree that specifying the time zone is important, particularly given

the relevance of diurnal variations. We have clarified the times mentioned are in UTC.

Review 18: Line 376: The use of the word “exactly” makes this a very strong claim, and I therefore suggest replacing it with “closely,” unless the exactness of the alignment can be clearly demonstrated.

Response 18: We agree that “closely” is a more accurate term, as it avoids implying an absolute alignment that may not be fully substantiated. We have replaced “exactly” with “closely” to provide a more precise description of the alignment.

Review 19: Line 381f: I think the conclusion is worded too strongly. It is only demonstrated that TemDeep can capture non-linear transitions in this specific case, and therefore the conclusion should be that TemDeep is not incapable of capturing these transitions. It cannot be concluded that it is guaranteed to capture all non-linear transitions, as the presented conclusion in the manuscript somewhat implies. I suggest rewording the conclusion to be more careful.

Response 19: We agree that the current wording may overstate the model's capability in capturing non-linear transitions. Corrected.

Review 20: Line 401f: It might be good to state the temporal interval to which the downscaling was performed. This would clearly indicate to which temporal resolution the model is demonstrated to perform well, and above which this has not been demonstrated.

Response 20: We agree that specifying the exact temporal interval for which the downscaling was performed would add clarity. We have stated that the model's performance was demonstrated at the 1-hourly resolution, distinguishing it from finer resolutions that have not been tested.

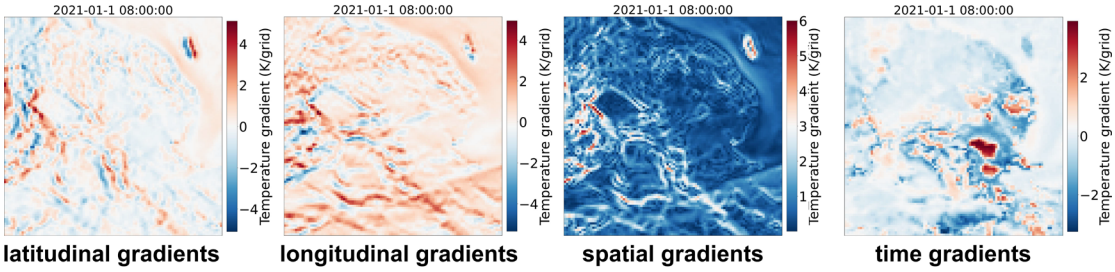
Review 21: Line 403ff: As it has not been demonstrated that the model performs well when downscaling to resolutions finer than 1-hourly, this manuscript strongly motivates further research into assessing how far re-analysis data can be meaningfully downscaled. This is especially the case, as the motivation presented in the beginning is that high-resolution data are very useful for certain applications. While I understand that the validation of finer resolutions is made difficult by the absence of re-analysis data to compare the downscaling results to, I still believe that mentioning this is of value.

Response 21: We agree that exploring the model's potential for downscaling to

resolutions finer than 1-hourly would be valuable, given the applications that benefit from high-resolution data. However, as you noted, our validation is constrained by the temporal resolution of the available reanalysis data—specifically, the ERA5 dataset, which only provides data at 1-hour intervals. We have added a note in the manuscript to clarify this constraint and emphasize the need for further research to investigate the feasibility of meaningful downscaling beyond the 1-hour resolution, should higher-frequency datasets become available in the future.

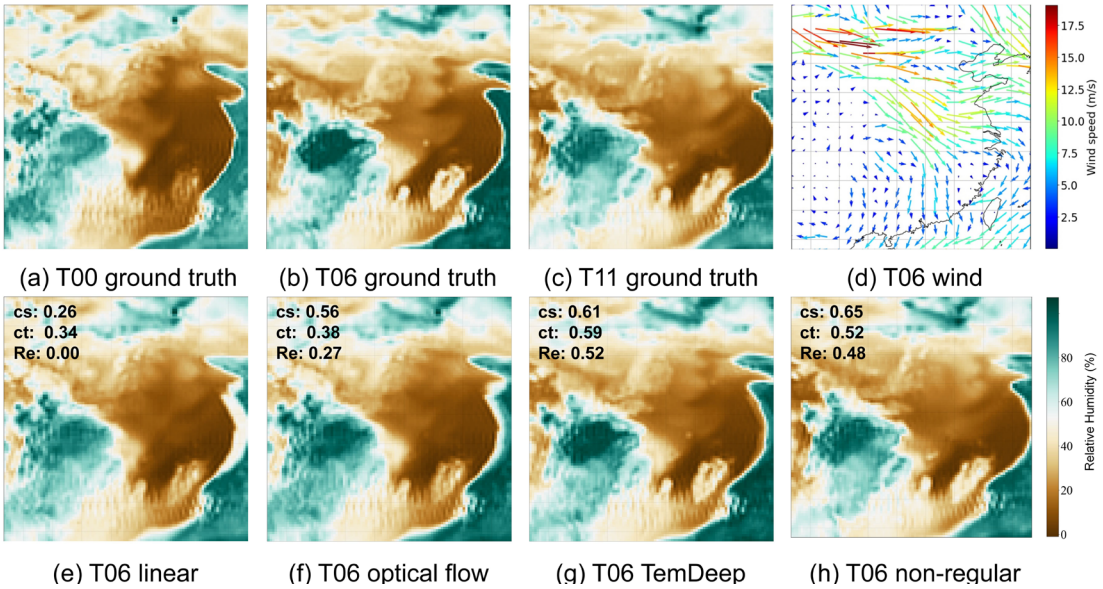
Review 22: Figures 4 and 5: The gradient should not be given in K, as the spatial and temporal dimensions are unclear.

Response 22: We agree that expressing the gradient in Kelvin (K) without specifying spatial and temporal dimensions can be misleading. We have revised Figures 4 and 5 to clarify the units by either specifying the gradient as a rate (e.g., K per unit distance or time) or adjusting the notation to clearly represent the intended spatial and temporal scales.



Review 23: Figure 9: It might be better to label the bottom colorbar as "Relative Humidity [%]," as this would immediately clarify which quantity is being shown. It is already clear from the unit itself that it is a percentage. Also, it might be useful to set the upper boundary at 100%, as relative humidity generally does not exceed 100% by much.

Response 23: Thank you for the helpful suggestion. We agree that labeling the colorbar as "Relative Humidity [%]" would improve clarity, and setting the upper boundary at 100% is indeed more appropriate, given that relative humidity rarely exceeds this threshold. Corrected.



References:

1. Davis, A., Marshak, A., Wiscombe, W., & Cahalan, R. (1994). Multifractal characterizations of nonstationarity and intermitttency in geophysical fields: Observed, retrieved, or simulated. Journal of Geophysical Research: Atmospheres, 99(D4), 8055-8072.
2. Frisch, U. (1995). Turbulence: The Legacy of A. N. Kolmogorov. Cambridge University Press.
3. Lovejoy, S., & Schertzer, D. (2013). The Weather and Climate: Emergent Laws and Multifractal Cascades. Cambridge University Press.
4. Schertzer, D., & Lovejoy, S. (1987). Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. Journal of Geophysical Research: Atmospheres, 92(D8), 9693-9714.