EGU peer-review training 2025

Review of "The contribution of circulation changes to summer temperature trends in the northern hemisphere mid-latitudes: A multi-method quantification" by Pfleiderer et al. (egusphere-2025-2397)

Review by Arundhati Kalyan, Anjali Thomas, and Jan Zibell*

*J.Z. declares being employed at the same institute as one of the co-authors of the study.

In the above study, the authors assess long-term (multidecadal) circulation-induced changes in summer temperature trends in the northern hemisphere midlatitudes (30–60°N). They employ four different statistical/machine learning-based methods – ridge regression, atmospheric circulation analogues, direct effect analysis, and a convolutional neural network – to isolate temperature trends driven by circulation changes over historical and future time periods in free-running climate model simulations (using CESM2) and ERA5 reanalysis. The relative performance of these methods is evaluated against a benchmark comprising nudged climate model experiments (also using CESM2) which include forced and internal components of circulation variability, but no forced thermodynamic component. The four methods are found to be effective on climate timescales, albeit with biases. The paper also highlights regional differences in dynamically forced trends, revealing alternating wavelike patterns of warming and cooling throughout North America and Eurasia. Finally, the authors discuss in depth the challenges and limitations in using statistical methods to decompose circulation vs. thermodynamically driven trend signals.

The study makes robust choices relating to data and methodology, in that four different decomposition methods are used and multiple statistical evaluation metrics are analysed for each. The authors transparently present a summary of the climate of the wind-nudged simulations which form the reference for their assessment of the decomposition methods. The presentation of challenges (multidecadal timescales, differing climate components included in the circulation-related decomposition, creation of a suitable benchmark, need for multiple models) in the discussion section is strong and may serve as useful reference for future studies. The study completes the stated target of quantifying circulation-driven trends across the NH midlatitudes and validating the different methods against a suitable climate model benchmark. In addition to highlighting features with low and high skill for each method, the study adds some degree of confidence to the findings of earlier studies that examined regional patterns of circulation-driven temperature trends. The article is concise and generally well-written. The title is informative and appropriate to the study, although the abstract could be improved as suggested below. The relevance of the study, objectives, and scientific challenges are introduced well. Even so, there are portions of the manuscript that are hard to follow and hinder conceptual understanding and interpretation of the results. In particular, the explanations of the statistical methods, the illustration of statistical significance/uncertainties associated with the results, and the dynamical interpretation of the study could be improved. We recommend publication of this manuscript in Weather and Climate Dynamics after minor but necessary revisions as outlined in the comments below.

Main comments:

- Statistical methods: The description of the four methods used is not straightforward and quite hard to follow for a non-expert reader. We are not experts in the statistical/ML methods used in this study and hence, cannot comment on the strengths and weaknesses of the design and implementation of these methods. However, here are a few suggestions to make the methods understandable to a broader dynamics audience, such that we can better appreciate the significance of the study's findings:
    - Add an introductory sentence or two in plain language in all of the sections 2.3.1–2.3.4 to explain what the method and/or equation aims to achieve
    - Clearly define all variables and constants introduced in each equation
    - Consider also whether certain details in the method descriptions can be moved to supplementary text
    - Here a few line-specific comments:
        - Line 134: Please explain why a 40°×40° rectangle around each grid cell was chosen. Could other sizes change the results?
        - Line 156: Please discuss briefly how you choose the number of analogues (Ns = 50) and repetitions (Nr = 100). Does changing these choices affect the results?
        - Line 200: It might help to briefly explain the physical meaning of the Yperp component in plain language in this context.

- It would be good to see more rigorous discussion of the statistical significance and uncertainties associated with the results presented. The need for this arises due to the following reasons:
    - "… observed trends are falling out of the range of model-simulated expected trends" (Line 35)
    - There is "ambiguity" on the magnitude of the historical circulation-induced trends, and all methods likely underestimate these magnitudes (Sec. 3.2)
    - In light of the above, it would be helpful to better understand the likely range/distribution of trends for each method and how they differ. Are some methods more likely to include the observed/nudged values than others?
    - The use of a very small sample size (only three ensemble members) for the free-running and nudged simulations limits the ability to show robust uncertainties. Could you provide more information on whether the initial conditions for the three nudged ensemble members were chosen at random or based on certain criteria? Could you comment on how large of an added value more members would be and why you decided on three only?
    - Table 1 would also benefit from the inclusion of confidence intervals and/or p-values to reject the null hypothesis that the correct sign or correlation of temperature trends was obtained by random chance.

- There are instances where we think that the discussions in this study may benefit from a more dynamical perspective:
    - It is clear from the introduction and methods that the wind-nudged simulations are viewed as a ground-truth benchmark. The limitations of this approach are

discussed as, for instance, in line 335: "However, there may be factors of residual climate variability (such as ocean variability) or feedbacks between circulation and other factors such as land-atmosphere coupling that could still affect thermodynamical processes on climate over land." Aren't there even more concrete examples for a physical relationship that is not a priori captured, such as the time mean thermal wind balance?

  - Line 36–37: Please consider to slightly reframe "... may indicate ... that a forced change in circulation is missing in the models". This framing could make one think of unresolved/parameterized processes in climate models, such as latent heating in deep convection in the midlatitudes. For those it is the consensus that a forced circulation change *is* missing in the models. Should you actually refer to this, the sentence could be reframed as "that circulation changes due to unresolved processes, which are not captured by the models, turn out to be meaningful / non-negligible".
  - It is certainly not the purpose of this study to discuss all limitations of wind-nudging in detail and of course every alternative also has its limitations, but we suggest that the authors at least address the point that in the real atmosphere, the development of a heatwave is not the linear sum of a thermodynamic component plus a dynamic component. There are many non-linear dynamical feedbacks from thermodynamical processes, e.g. from coupling with radiation via clouds, surface fluxes depending on soil moisture, or latent heat release (which, if speaking of heat waves, within a warm conveyor belt may increase blocking intensity (Pfahl et al. 2015)). Possibly, addressing this is related to emphasizing more prominently that the authors regard GMST as the representative variable of thermodynamic changes.
  - Overall, we suggest that 1) the assumption that a trend can be decomposed into thermodynamical and dynamical components and 2) the use of wind-nudged simulations (as introduced in Sect. 2.1) to achieve this are discussed a bit more critically. This could be done very briefly when introducing the wind-nudging experiments in the Introduction and then in a more elaborate way for instance as a sixth discussion point in Sect. 3.3.

- The main objective of the study is to investigate long-term temperature trends. Therefore, the prominent motivation based on heatwaves and extremes seems somewhat out of place. Motivating this research with individual events is fine per se, but this study does not investigate trends in heatwaves. It would be helpful if the authors could discuss a more concrete example of how their estimated trends allow a better understanding of extreme events (as indicated in lines 31-34)?

- The use of the dynamical vs. thermodynamical separation of trends could be even further strengthened by a discussion of the geographical variability of the thermodynamically induced trends. In Figure 4, over Eurasia your methods disagree whether the thermodynamical change is rather uniform or dependent on longitude or latitude. Are there any physical arguments in the literature for what the thermodynamically-induced pattern of warming should look like? For instance, it is observed that the Mediterranean region is warming faster (Brogli et al., 2019). Alternatively, it seems also fine to note that this is left for future study or refer to discussions in other studies.

Minor comments:

- Line 3–4: “Over the northern hemispheric mid-latitudes, considerable regional differences in summer temperatures have been observed.” → Presumably, you mean differences in summer temperature *changes.*
- Line 5–6: We think the general readability of the abstract would benefit from a brief description of ‘decomposition method’, i.e., what you decompose the trends into. If one is not very familiar with the topic or similar literature, this is not obvious but only (and well) presented in the introduction.
- Line 10–11: “Most decomposition methods show skill in estimating the sign of circulation-induced trends but all methods underestimate the magnitude of these trends.” This statement contains the fact that you use the wind-nudged simulations as your benchmark and that you assume that the nudged simulations contain 100% of the dynamical component of the trend. This should be presented more clearly as was done in the Introduction, for instance, around line 60.
- Line 16: Consider changing: The intensity of heatwaves “increases globally” to “has been increasing globally”.
- Line 18–19: Consider adding a reference/s here to show that intensification of heat waves occurs in a warmer climate due to thermodynamic factors (perhaps an attribution study?).
- Line 19: “However, heat waves are not only...” is a long sentence and could benefit from restructuring.
- Line 22: It might be worth mentioning here whether land-atmosphere interactions are more or less important than circulation changes as a factor in driving summer temperature trends.
- Line 23: It might be worth commenting on whether regional trends are more pronounced in the NH midlatitudes than elsewhere. May also be good to include a sentence citing studies that examined trends in the tropics or Southern Hemisphere.
- Line 24–27: This is a long sentence and could be split into two separate sentences.
- Line 31: Consider including more specific references that show why forced changes in circulation are small compared to internal variability.
- Line 48: Consider whether there are any other limitations of the nudged circulation experiments and include that here.
- Line 49: Add “e.g. the circulation not being in thermal wind balance” to clarify that you don’t mean unresolved processes, which are also mechanisms not represented in the models used.
- Line 49–51: On the other hand, most of statistical decomposition methods ...” can be rewritten/shortened or split into two sentences to enhance readability.
- Line 54–56: “Moreover, benchmarks for circulation-induced long-term trends have not been available so far, and to our knowledge no systematic comparison of dynamical adjustment methods has been performed.” It would be good to clarify if this applies globally or just to mid-latitudes and to briefly mention if any emerging efforts exist. This would make it clearer why the study is filling an important gap.

- Line 57: Add the specific NH latitude range being examined here.
- Section 2: Adding a sentence or two linking each data/methods subsection to the main aim (separating thermodynamic vs circulation-driven temperature trends) would help the reader understand why each method or simulation is being used.
- Line 69: You could specify the ERA5 years here as well.
- Line 72: "First, three standard historical and future forcing experiments ....". At first reading, this sounds like three different forcing scenarios or the like. Please specify that you mean three ensemble members.
- Line 80: How about specifying some of the main features of the nudging, e.g., whether your nudging is done at the model grid or involves some spectral transformations, and to which vertical level it is done? This way the reader gets a good first impression without having to refer to Topal and Ding (2023) to find out what is "similar" to their approach and what is different.
- Line 80: Is CAM6 an abbreviation of something? If yes, please mention.
- Line 81: "These simulations will be henceforth referred...". Simplify this sentence for readability.
- Line 87: The phrasing of thermodynamic forcing being represented by "surface temperature" somewhat suggests that using this metric of forcing is not a choice. In reality, temperature change is non-uniform throughout the atmosphere with implications for the midlatitude circulation. We suggest that the authors instead say 'commonly approximated by GMST' or similar.
- Line 89: Good point, but it would help to add a short explanation of why it is hard to evaluate these methods in a coupled system.
- Line 95: Will the residual internal variability (e.g. from the ocean) influence the evaluation of the decomposition methods, and how do you account for that?
- Line 111: "However, we assume that the effect...": This assumption is reasonable, but you might want to add a reference or a short justification for this assumption.
- Figure 2: It is not clear what is meant with the cooler versus warmer histograms in panels b), c). Please clarify.
- Line 115: Consider splitting the paragraph into two shorter ones: one describing the experimental setup and another explaining its implications and limitations. This would improve readability.
- Line 120: Briefly define AMIP in the text for clarity.
- Line 132: Could split into two shorter sentences for clarity.
- Line 216: Consider providing more details of the transient CESM2 simulations used to train the UNET.
- Lines 219–220: Clarify why the training is done on CESM2 first and then fine-tuning on ERA5—why does this improve performance or robustness?
- Line 220–222: Rephrase to sound more concise and formal.
- Line 227–237: Consider presenting these two sets of bullet points together instead of separately to make the text more concise and easier for the reader to associate each skill metric with what it represents.

- Section 3.1: The discussion mentions how the methods differ (DEA captures magnitude, UNET conservative), but the rationale behind these differences could be explained more clearly. For instance, why does UNET underestimate magnitudes?
- Line 276: This is a long sentence. Consider splitting it into 2–3 smaller sentences to enhance readability.
- Line 289–291: In “The ridge regression … up to 0.6 K/dec”, change “where” to “were”, and add “*suggest* stronger circulation induced trends …”
- Fig. 4: Which wavelength could approximate the wave-pattern change that you find? Can you relate this to other studies?
- Line 305: This paragraph sounds like a re-introduction of dynamical adjustment from zero. A bit of repetition is appreciated for the flow, but at the current stage this introduction of dynamical adjustment is even clearer than in the introduction (using even more references). Please consider streamlining this or, otherwise, stating more clearly if you mean something different than in the introduction or possibly moving some of this material into the introduction.
- Section 3.3 is purely a discussion. Why not make it a new section called Discussion? There is no new result in this section.

Technical comments:

- Line 35–36 and onward: Check for the use of citet vs. citep and citep[][]{} throughout the paper.
- In Figure 3, the kernel density maps could be enlarged with axes labels shown.
- Figure A2 is not explained or referenced in the text. Change.
- Figure B2 is not explained or referenced in the text. Change.

References:

Brogli, R., N. Kröner, S. L. Sørland, D. Lüthi, and C. Schär, 2019: The Role of Hadley Circulation and Lapse-Rate Changes for the Future European Summer Climate. J. Climate, 32, 385–404, https://doi.org/10.1175/JCLI-D-18-0431.1.

Pfahl, S., Schwierz, C., Croci-Maspoli, M. et al. Importance of latent heat release in ascending air streams for atmospheric blocking. Nature Geosci 8, 610–614 (2015). https://doi.org/10.1038/ngeo2487