Review of "Everyday weather in a warmer world", by Thomas et al.

The authors use a historical reanalysis system to generate a reanalysis for the year 1903 and then perform "pseudo global warming"-like simulations by increasing SSTs uniformly by 2 K and by additionally changing the CO2 concentration to 530 ppm, while assimilating the same surface and sea-level pressure observations. This allows addressing the hypothetical, pure thermodynamic effects and separate ocean warming and total warming. The paper then discusses the results at different spatial and temporal scales, focusing on the changes in the distributions and always addressing the underlying processes. The paper is well written and interesting and deserves publication. The results are convincingly discussed and relevant for our understanding of climate change processes. The obvious drawback inherent to this approach arises from constraining atmospheric circulation from observations while leaving the thermodynamics to the model. The fear is that it might lead to inconsistencies and it is not clear what the clean separation between thermodynamics and dynamics then actually means (e.g., when the vertical velocity still changes). Or phrased differently: How realistic is the simulated +2K+CO2 weather? Is it an unbiased sample of future weather? I will address that point below. However, I think this is an extremely useful undertaking that should be performed and should be published. We can learn a lot from this paper.

We thank the reviewer for their kind assessment and thoughtful suggestions to improve the manuscript. We answer their comments below. Reviewer words are in blue, our replies and original manuscript text are in black, and new sentences we propose to include in the manuscript are in red. For new Supplementary figures, we will amend the figure numbering upon preparation of the final manuscript.

Comments

- The authors use 1903 as a study year. This happens to be a globally very cold year (one of the coldest in the HadCRUT5.1 data set), which should be stated. The warmest year in HadCRUT5.1 is ca. 1.7 °C warmer. Although 20CRv3 stops a bit earlier, I am missing the argument that the pseudo global warming experiment is not taking the 20CRv3 system far away from what it has experienced and for what it has been evaluated (i.e., future in the paper is the near-future). I think this would be a supporting argument.

- That said: 1903 is the year following the Santa Maria eruption. This also should be stated, as this was one of the largest tropical eruptions of the 20th century.

We thank the reviewer for this helpful comment and note that RC2 raised a similar point regarding setting 1903 in context. We propose to include a Supplementary figure showing annual-mean temperature time series in HadCRUT5.1 for the global (Fig. S6a) and regional cases (Fig. S6b–e), along with 1903’s rank out of the 175 years in the dataset. We also propose to add the following paragraph at the end of Section 2 (Reanalysis Experiments):

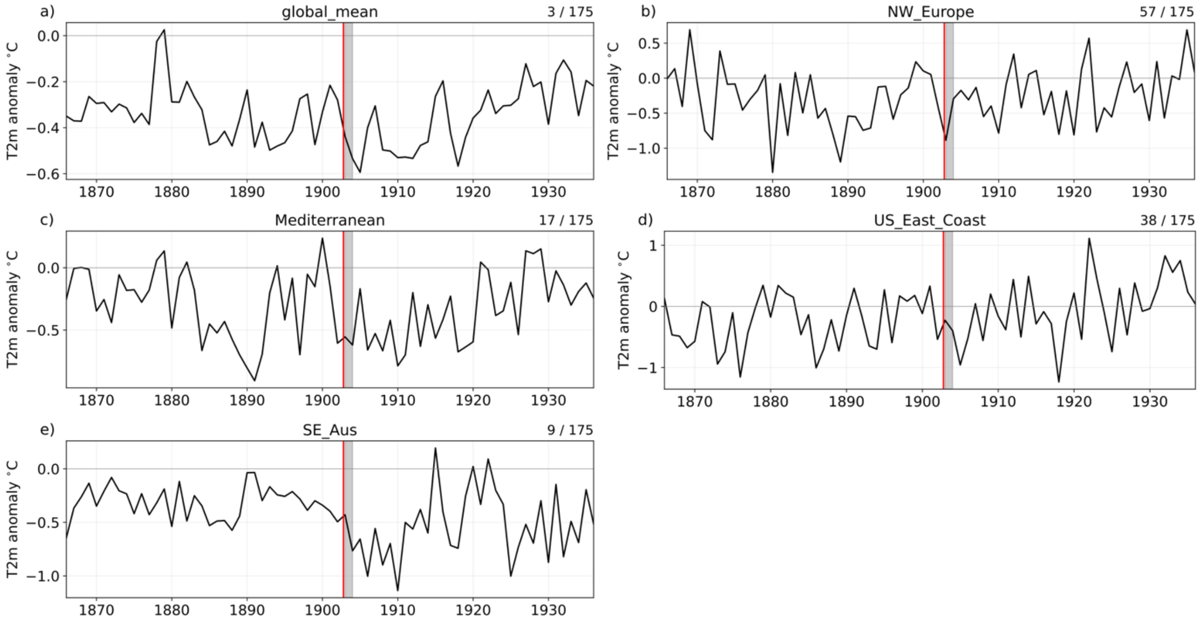


Fig. S6: Annual-mean surface temperature anomalies in HadCRUT5.1 (Morice et al. (2021)) for the global mean (a) and the four well-observed regions in Fig. 1 (b-e). Anomalies are with respect to the 1961-1990 reference period and include temperatures over both land and ocean. Grey shading shows the year 1903 and the red line shows the Santa Maria eruption in October 1902. The rank of 1903 within the 175-year HadCRUT5.1 dataset for each region is shown in the top right of the plot (rank 1 being the coldest and 175 the warmest).

"Focusing on the year 1903 allows us to exploit the existing reanalysis experiments of Hawkins et al., 2023. Globally, 1903 was a cold year: it is the 3rd coldest in the 175-year HadCRUT5.1 dataset (Fig S6a), and it is in at least the coldest third of years in each of the four well-observed regions we analyse in later sections (the lowest rank is 57th / 175 in NW_Europe, i.e. 32nd percentile; Figs S6b–e). The eruption of the Santa Maria volcano in October 1902 – one of the largest tropical volcanic eruptions of the century (Robock, 2000) – presumably contributed to the generally cool conditions (Self et al., 1981) (note that the volcanic aerosol forcing is identical in all of our experiments). The cold baseline conditions in 1903 mean that the perturbed climates in the warmer world experiments are not much warmer than some of the later years in the original 20CRv3 dataset, for which the system's performance has been validated. This builds confidence in the ability of the reanalysis system to plausibly represent the warmer world climates in our experiments."

- The question of the separation of thermodynamics and dynamics is of course key. Does the assimilation of real weather that unfolded in a 2 degree cooler world still give consistent weather in a warmer world? The approach allows the interpretations of individual factors, which is great, but what about the goal of the paper? The Abstract starts with "How would the weather of a year from history be experienced in a warmer world?" and ends with "...how our day-to-day experience of the weather may change in a warmer world", which is not exactly the same. I think the paper does the former, while the latter implies that the simulated weather is (a) plausible and (b) an unbiased sample of the warmer world weather.

We thank the reviewer for this comment.

Regarding the plausibility / consistency of the simulated weather: We acknowledge this potential issue. It is a similar problem faced by all storyline attribution methods (e.g. nudging, PGW). We believe the reanalysis assimilation statistics may offer some insight here, following the reviewer's next comment – see our response below.

Regarding the goals stated in the abstract: We agree with this distinction – we will therefore amend the final abstract sentence as follows:

"By anchoring our analysis in weather patterns that have actually occurred, the reanalysis experiments point to how our day-to-day experience of the same weather patterns may change in a warmer world, even if the weather patterns themselves do not."

- To better judge how consistent this is or how far the system is pushed, some additional analyses might help. It would perhaps be interesting to look into the innovation statistics to see whether the assimilation scheme constantly tries to back-correct thermodynamically induced dynamic effects by the model and vice versa. Or it would be good to know whether the QC now rejects certain observations I did not before (e.g., hurricanes). Do other important factors (localisation, inflation) change? Given the argument above, I doubt that any of this is the case, but would be good to know. The innovation statistics could give an indication into the dynamical changes the model would have liked to do.

We appreciate the reviewer's very helpful suggestions and agree that the assimilation statistics offer useful insight into the system's performance in the perturbed climate. We have therefore added maps and regional averages for each of the following statistics to the Supplementary:

- innovation (i.e. observation minus first guess)
- Localisation
- number of rejected observations

for the 20CRv3+ and +2K+CO2 experiments. We propose to add the following sentences interpreting these figures to the Discussion:

"The approach in this paper makes two implicit assumptions: (1) that the weather patterns that occurred in 1903 could in principle occur in a world that is 2K warmer; and (2) that assimilating the 1903 pressure observations with the 20CRv3 system does not lead to inconsistencies when the system's boundary conditions (SST, sea ice, and CO2 levels) are perturbed.

We can gain insight into both by comparing the reanalysis assimilation statistics from the 20CRv3+ and +2K+CO2 experiments. We find that the innovation (i.e. observed value minus first guess) and localisation lengths do not change substantially in the +2K+CO2 experiment. However, the assimilation system does reject more pressure observations in +2K+CO2 than in 20CRv3+. The relative increase is largest in NW_Europe, the region with proportionally the fewest rejections in 20CRv3+, and smallest in SE_Aus. An increase in the number of rejected observations may indicate that the model's response to the warming causes some inconsistencies with the pressure observations, particularly where the density of observations is high and hence the observational constraint is strongest, as in NW_Europe. Note that tropical cyclone central pressures are assimilated without opportunity for rejection in both experiments.

However, the increased number of rejected observations does not in itself mean that the simulated weather events are necessarily implausible in a warmer world, even if they become slightly less compatible with the model's own forced circulation response. The atmospheric circulation response to warming is both uncertain theoretically and variable among climate models (Shepherd, 2014), which suggests caution regarding how much weight we assign to the dynamical response in 20CRv3. Even if the particular weather patterns of 1903 were to become less likely in a warmer world, it is difficult to argue that they would be impossible a priori, notwithstanding certain features of the 1903 climate (e.g. the eruption of the Santa Maria volcano in 1902)."

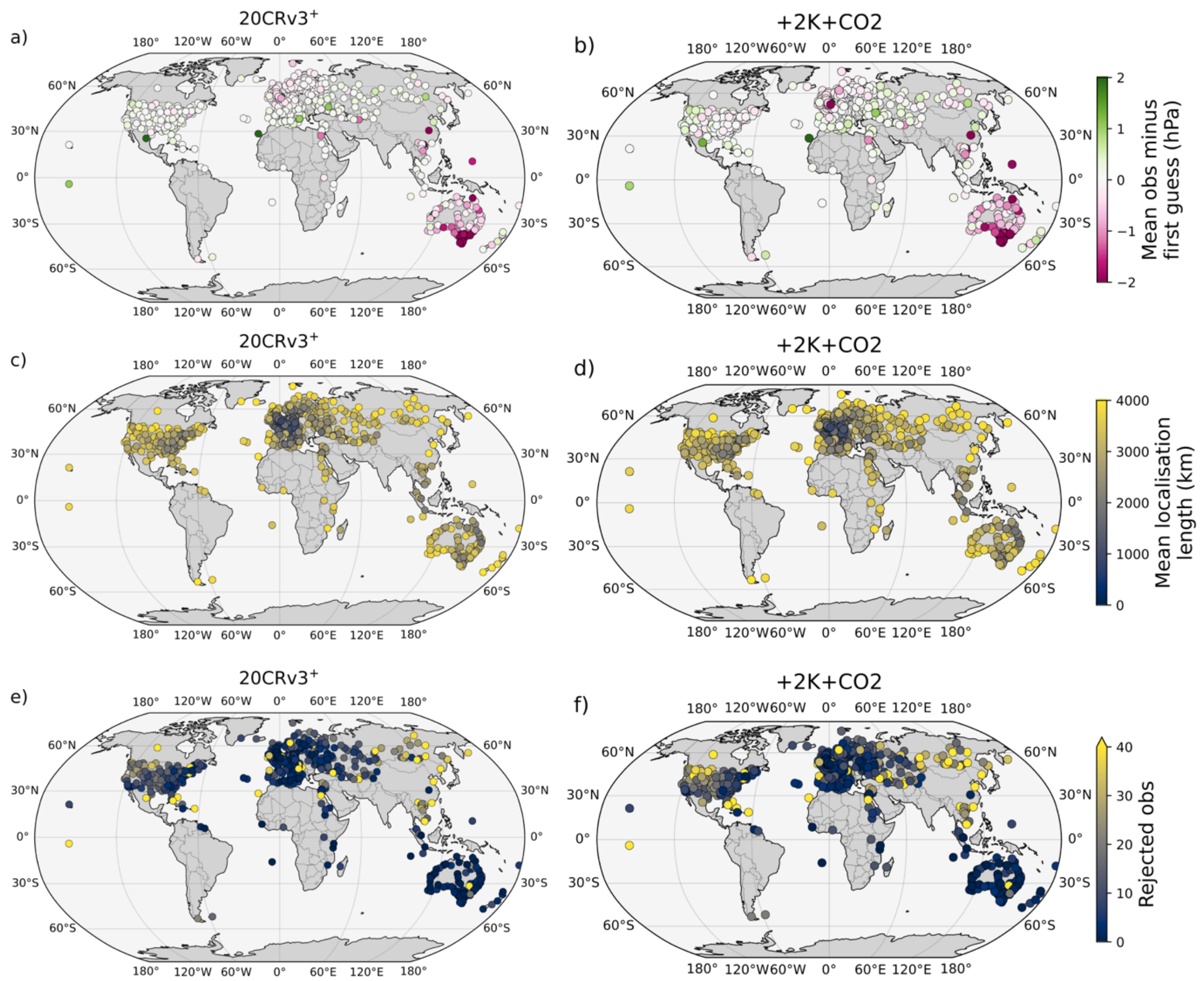


Fig S7: Maps of innovation (observation minus first guess, accounting for observation bias) (a, b), localisation length (c, d), and number of rejected observations (e, f) for land stations in the 20CRv3+ and +2K+CO2 experiments. Innovation and localisation length are average values over 1903 while (c, d) show the total number of rejections across the year.

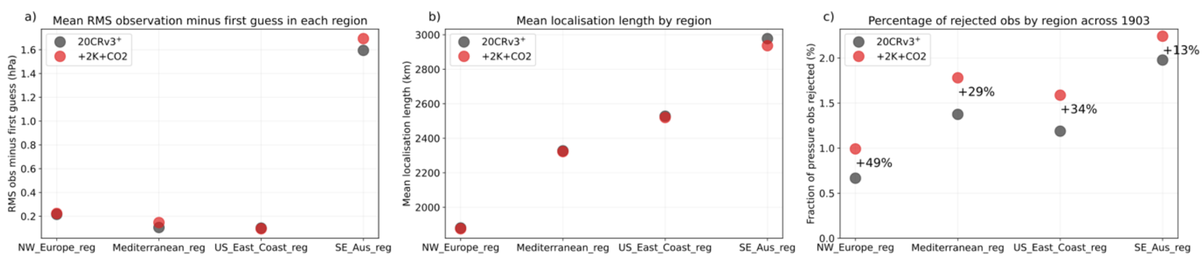


Fig. S8: Regional statistics for the innovation (a), localisation length (b), and fraction of rejected observations (c) shown in Fig. 7. In each case, the values denote the average of all land stations within each of the well-observed regions in Fig. 1. Percentage changes relative to 20CRv3+ are shown next to the markers in (c).

I appreciate the surface energy balance decomposition in the supplement. Still, it might also be interesting to say just a few words about clouds (which appear at many instances in the discussion) and stability or generally the third dimension. I do not think the paper should be substantially longer, however, the clean separation of dynamics and thermodynamics is not so clean as only the surface is constrained and vertical velocity and cloud formation react partly while still being strongly constrained.

We agree and will add the following sentences to the Discussion:

"Assimilating observed surface pressure does not constrain the atmospheric circulation at all time and length scales – for example, the super-CC scaling of short-duration rainfall extremes in October 1903 (Fig. S5) indicate that sub-daily updraft velocities can show large changes, even in well-observed regions. Changes in vertical motion will also affect cloud behaviour, and may contribute to the cloud-mediated amplification of hot days in the SE_Aus region (Fig. S3). Future work will compare where and how strongly the circulation is constrained in the reanalysis experiments versus other established methods such as spectral nudging (Feser and Shepherd, 2025)."

- Please add information about the land cover in the model: How it is prescribed and how does the model depict land-atmosphere interaction (as this is concluded to be a factor in 4.1). Also, is sea ice also changed or just SSTs?

We thank the reviewer for pointing out these omissions. We will add references to the land and atmospheric models used as follows:

L59: We reconstruct the weather of 1903 using the NOAA-CIRES-DOE 20th Century Reanalysis version 3 data assimilation system (20CRv3, Slivinski et al. (2019)). Using an Ensemble Kalman Filter, 20CRv3 assimilates observed pressure values with a NOAA atmospheric model to provide a best estimate of the atmospheric state at each point in time. The atmospheric model used is the NOAA GFS v14.0.1. The atmospheric model is coupled to the Noah land surface model, which incorporates a 4-layer soil model and dynamic fractional snow cover (Ek et al., 2003).

We will also add the following sentences outlining the sea ice methodology to the section describing the reanalysis experiments:

L79: …CO2 radiative forcing. Sea ice boundary conditions are unchanged in the warmer world experiments compared to 20CRv3+. Sensitivity tests suggest that this can cause spurious signals near the ice edge, but these effects do not extend far from the ice edge due to the strong constraint of prescribed SSTs and the assimilation of pressure observations (not shown). The SST, CO2, and sea ice perturbations are idealised and do not reflect observed or projected patterns of change – they are chosen to illustrate the role of ocean warming and atmospheric composition on the weather of 1903 in a simple way.

- For the same reasons as above, it would be very interesting to also analyse some not-so-well constrained regions.

We have recreated the time series in Fig. 2 for a midlatitude region with very few nearby station observations (Patagonia). The day-to-day fluctuations in temperature and precipitation in the 20CRv3+ and +2K+CO2 experiments are still very similar, despite there being many fewer land-based pressure observations compared to the other regions we analysed. This suggests that only a small number of nearby land stations, in addition to ship-based observations (which are sporadic in time), is sufficient to mostly constrain the atmospheric circulation in the reanalysis. We found similar results for different months and different poorly-observed regions (e.g. for boxes over the Amazon and equatorial Africa). In practice, we may therefore have been able to extend our analysis beyond the four original regions, however we still prefer to err on the side of caution.

We propose to (a) include the Patagonia time series in the Supplementary, (b) modify Fig 1 to include a box over Patagonia, and c) add the following sentence under our discussion of Fig. 2:

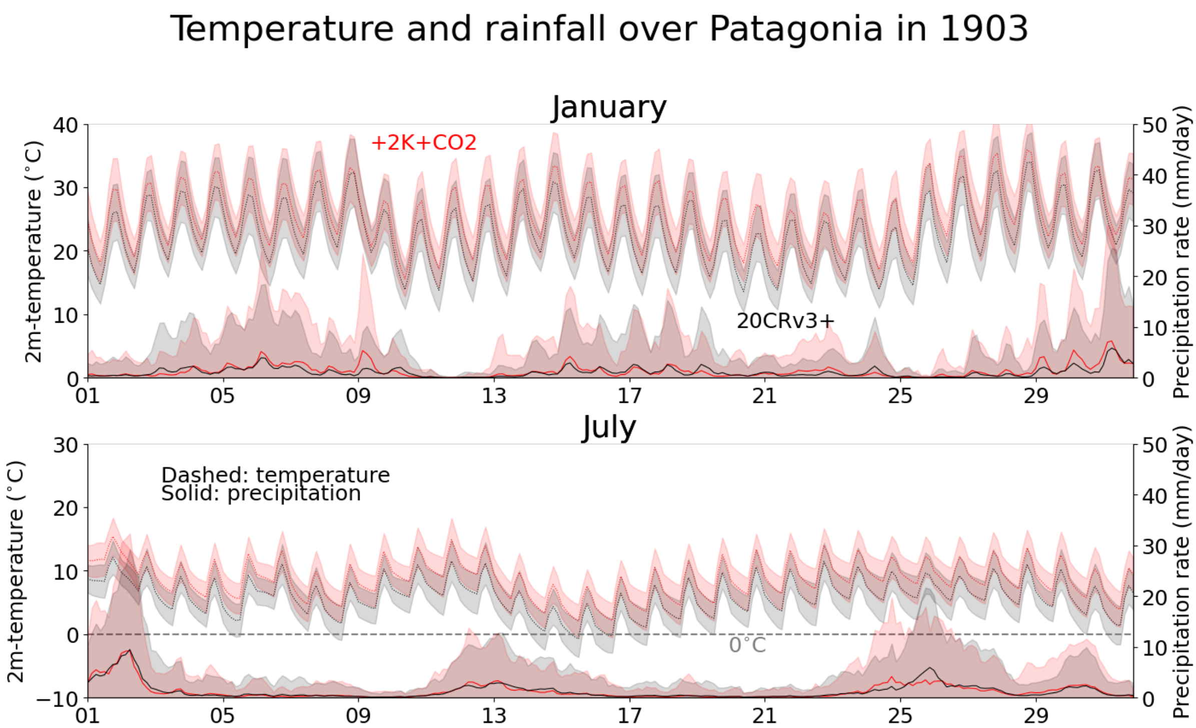
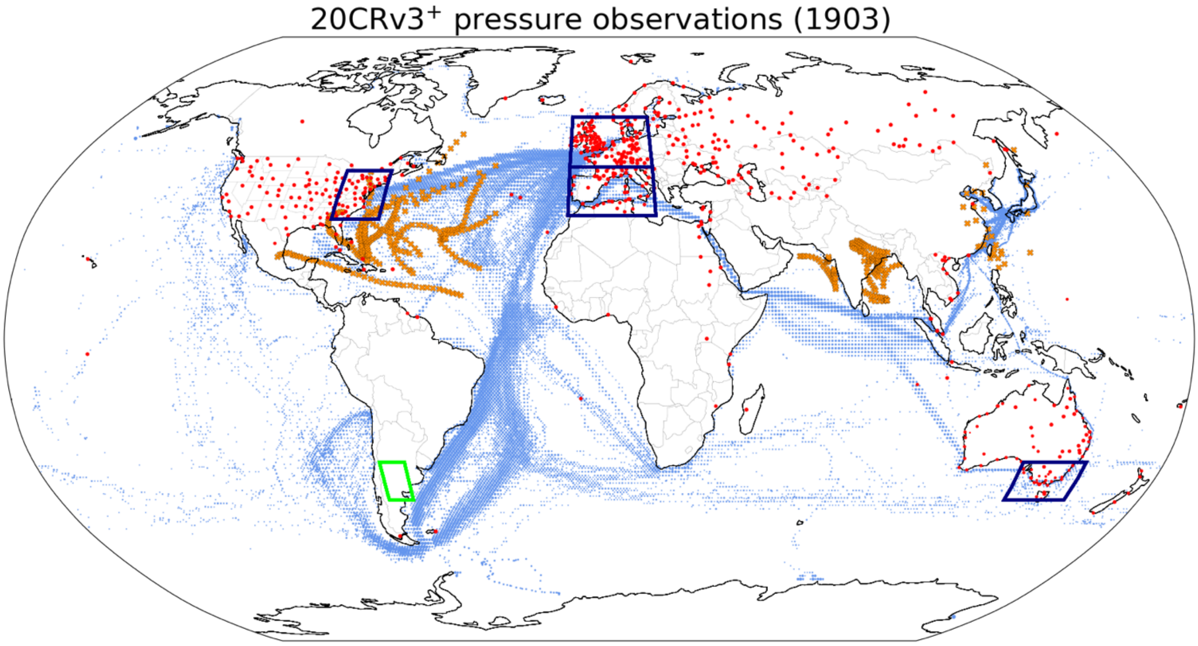


Fig. S9: Temperature and rainfall over a region in northern Patagonia for two months in 1903 (as in Fig. 2, but for a region with fewer pressure observations). The Patagonia region is shown by the green box in Fig. 1.



(New Fig. 1)

"In addition to Central England, we also show a version of Fig. 2 for a region over northern Patagonia with far fewer nearby land-based pressure stations (Fig. S9, the green boxed region in Fig. 1). Day-to-day variations in temperature and precipitation are surprisingly similar in both experiments, despite the much weaker observational constraint than over Central England (Fig. 1). Although we restrict most of the later analysis to well-observed regions, Fig. S9 suggests that even a small number of nearby observations (from ships in this case) can constrain the circulation relatively well in the reanalysis system. This is particularly noteworthy given that the markers in Fig. 1 represent observations across all of 1903; within a single assimilation window, the density of ship-based observations will be much lower than that shown in Fig. 1."

Minor

L. 317: Why can your experiments "be considered a strongly conditioned form of attribution"? What do you attribute?

Broadly speaking, in this study we are attributing changes in 2m temperature and precipitation to increases in ocean warming/atmospheric CO_2 under fixed atmospheric circulation. Parallel experiments will use the 20CRv3 system in a formal storyline "event attribution" sense (i.e. translating a recent extreme event into a 1950s climate, rather than translating historical events into a warmer climate as we have shown here). Compared to other attribution methods that use free-running climate models (e.g. probabilistic methods, Stott et al. 2016), the pressure assimilation strongly constrains how much the event can change in the cooler simulation – the attribution is strongly conditioned on the fixed atmospheric circulation. We will amend the sentence as follows:

"The reanalysis experiments can thus be considered a strongly conditioned form of attribution, as we consider only the effects of warming that are not related to changes in the large-scale circulation (Thompson et al., 2026)."

Supplement, line 10: Fig. S2 -> Fig. S3

Thanks, we will fix this.

Review references not appearing in original manuscript

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G. and Tarpley, J. D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. Journal of Geophysical Research: Atmospheres [online], 108 (D22). Available from: http://dx.doi.org/10.1029/2002JD003296.

Feser, F., Shepherd, T.G. The concept of spectrally nudged storylines for extreme event attribution. *Commun Earth Environ* **6**, 677 (2025). https://doi.org/10.1038/s43247-025-02659-6

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., et al. (2021). An updated assessment of near-surface temperature change from 1850: the HadCRUT5 data set. Journal of Geophysical Research: Atmospheres, 126, e2019JD032361. https://doi.org/10.1029/2019JD032361

Robock, A., 2000. Volcanic eruptions and climate. Reviews of Geophysics [online], 38 (2), 191–219. Available from: http://dx.doi.org/10.1029/1998RG000054.

Self, S., Rampino, M. R. and Barbera, J. J., 1981. The possible effects of large 19th and 20th century volcanic eruptions on zonal and hemispheric surface temperatures. Journal of Volcanology and Geothermal Research [online], 11 (1), 41–60. Available from: http://dx.doi.org/10.1016/0377-0273(81)90074-3.

Stott, P. A., Christidis, N., Otto, F. E. L., Sun, Y., Vanderlinden, J., van Oldenborgh, G. J., Vautard, R., von Storch, H., Walton, P., Yiou, P. and Zwiers, F. W., 2015. Attribution of extreme weather and climate-related events. WIREs Climate Change [online], 7 (1), 23–41. Available from: http://dx.doi.org/10.1002/wcc.380.

Thompson, V., Ermis, S. and Athanase, M., 2025. The need for multi-method extreme event attribution. Weather [online], 81 (2), 40–45. Available from: http://dx.doi.org/10.1002/wea.7779.