The manuscript by Thomas et al. uses the 20CRv3 reanalysis system to assess the effect of a uniform 2°C SST warming and a combined 2°C SST warming plus a CO_2 concentration increase on the global and regional weather of the year 1903. The very same pressure and cyclone track patterns are assimilated in both the control and the sensitivity experiments, allowing an approximation of the pure thermodynamic effect (or at least some part of it) of the imposed warming/CO2 increase for identical weather situations. The authors find a regionally dependent atmospheric warming of typically a little more than the imposed +2°C. In four investigated sub-domains, the largest warmings (much more than 2°C) are found at the tails of the temperature distribution, i.e. on cold days and on warm days. Daily rainfall becomes more variable, and the thermodynamic signal alone can cause seasonally dependent precipitation changes.

I have to admit that after reading the abstract and introduction I have not been too convinced on the usefulness of the work presented, especially of its motivation. While reading the full manuscript, however, more and more interesting points became evident, and I believe that the manuscript could be a valid contribution to the scientific community after a few adjustments were made. The quality of presentation, in general, is high. The results are nicely illustrated for most parts. The conclusions openly and appropriately mention the limitations of the work.

My suggestion is to return to the manuscript to the authors for major revision, during which the points listed below should be addressed.

With kind regards.

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

MAJOR ISSUES

Motivation of the work: I do certainly understand that it is tempting to use existing simulations (Hawkins et al.) and to make more complete use of them. But the choice of the year 1903 for this study needs to be much better motivated in my opinion. Where is this particular year located in the climatology of the early 1900s? I doubt that it was regular/mean year in all parts of the world. Presenting such an analysis is important when interpreting the results as regular/mean weather. At the end of the day, simulating and analysing a 20- or 30-year period would be much more informative.

Regarding the short time period: The strong constraint of the pressure assimilation (and hence the large-scale circulation) means that the signal-to-noise ratio is much higher when comparing the reanalysis experiments than if we were comparing a year's worth of data from two unconstrained climate models, for example – this is the same rationale used to justify small ensembles in nudged storylines (van Garderen et al., 2021). Ideally we would use a longer period for this analysis as suggested, but we are limited by data availability. In future we aim to produce paired reanalysis experiments over a longer multi-year period and we will use the present work to focus our future research questions using that data.

Regarding 1903: We agree with this comment and note that RC1 also called for more context regarding the year 1903. As in our response to RC1, we have added a Supplementary figure showing annual-mean temperature time series in HadCRUT5.1 for the global (Fig. S6a) and regional cases (Figs. S6b–e), along with 1903's rank out of the 175 years in the dataset. We also propose to add a paragraph at the end of Section 2 (Reanalysis Experiments). These changes, also shown in our reply to RC1, are reproduced below:

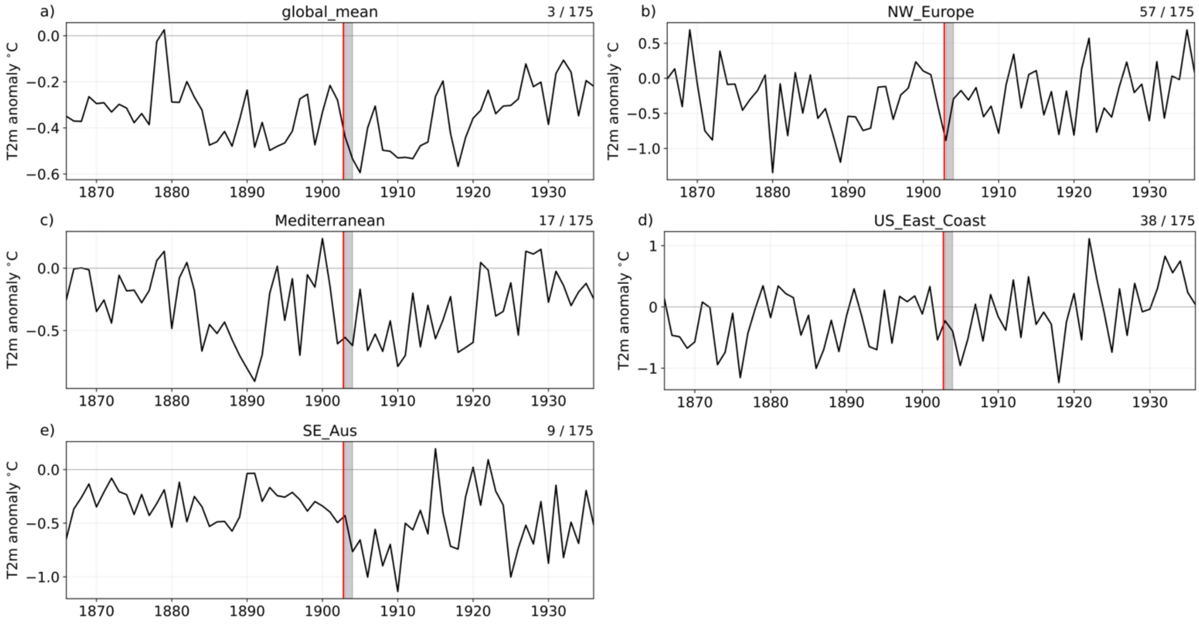

Fig. S6: Annual-mean surface temperature anomalies in HadCRUT5.1 (Morice et al. (2021)) for the global mean (a) and the four well-observed regions in Fig. 1 (b-e). Anomalies are with respect to the 1961-1990 reference period and include temperatures over both land and ocean. Grey shading shows the year 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."

Motivation of the technique: Relating to the previous point, interpreting the results of the one-year long simulations as thermodynamic effects on that year's weather is partly critical in my opinion. The applied technique (assimilation of pressure observations) is regionally extremely strong I'd say and would not allow regional feedbacks (within a general large-scale weather situation) and their effects on day-to-day weather variability to develop in a perturbed setup. Could the authors provide some more evidence that the system is not too constrained in this respect? Also, imposing the SST warming without adapting the pressure observations that are assimilated can be critical in this sense. Buth would go hand in hand for pure thermodynamic reasons. Could the authors comment on this?

We agree with the reviewer that the assimilation of pressure observations constrains the atmospheric response quite strongly – in reality, both dynamic and thermodynamic aspects would respond interactively to an imposed perturbation. However, separating these aspects to varying degrees is an established practice in storyline attribution (e.g. nudging and PGW approaches discussed in the Introduction) in order to make progress on the parts that are better represented in models.

- **Regarding strength of the constraint:** we refer to our reply to RC1 where we showed and discussed the reanalysis assimilation statistics. Little change is seen in the innovation or localization length scale in the warmer world experiments. However, we do see more observations being rejected by the assimilation scheme, suggesting that there may be some tension between the model's response to the warming and the imposed observations. We will add sentences discussing this to the Discussion (see reply to RC1). The ability to analyse the assimilation statistics in this way is a unique feature of using a reanalysis system.

- **Regarding the SST perturbation:** we apply a uniform SST warming to avoid altering the SST patterns from 1903. However, we appreciate that in reality warmer SSTs could alter local heat sources that could feed back onto the circulation. As above, neglecting this by not adjusting the pressure observations is a result of the deliberate design choice of the experiment.

We propose to add the following sentences to the Discussion outlining an example of a process neglected by our method and clarifying our motivation for choosing this approach:

"As in other storyline attribution approaches, the warmer world reanalysis experiments take the observed atmospheric surface circulation patterns as fixed and seek to quantify the thermodynamic impacts of global warming. This means that feedbacks from the warming onto the large-scale circulation are neglected. For example, some studies have suggested that in heat waves arising due to a persistent atmospheric block, soil moisture anomalies can, in certain circumstances, feed back onto the atmospheric circulation pattern itself (Fischer et al. 2007, Merrifield et al. 2019). Temperature increases on the hottest days may thus be higher than estimated in Fig. 5 if the circulation were allowed to respond. In reality, the dynamics and thermodynamics of the system cannot be cleanly separated, but in studies like ours it allows us to focus our investigation on processes that we think are better represented by the current generation of climate models (thermodynamic effects) than others (such as persistent weather regimes)."

Validation: In general, only few validations are presented, only at the Central England scale, and focusing on precipitation. However, if changes in day-to-day precipitation variability are analysed, for instance, some decent performance of the 20CRv3 control run for these aspects would have to be shown (although, as the authors correctly state, precipitation conditions are probably more poorly constrained).

We agree that it is helpful to validate 20CRv3+ precipitation more thoroughly than in the original manuscript. The fairest comparison is to a daily gridded observational dataset, as in Fig. 8, but this is more difficult outside the UK. We have therefore validated 20CRv3+ against daily station precipitation observations for a handful of sub-regions within each of the four main regions in the original manuscript. The sub-regions are chosen to sample potentially different climatological conditions within each region (e.g. mountains or plains, coast or inland). Observations are taken from the Global Land and Marine Observations Database (GLAMOD) v8.1 (Noone et al., 2021). The sub-regions are shown in Fig. S10, the daily time series in Fig. S11, and mean and daily variability are shown in Fig S12.

In most sub-regions, the largest localised peaks at individual stations are not captured by the reanalysis. This is not a surprise, as 20CRv3+ values represent averages over the sub-regions – when compared to sub-regional average from the gridded observational dataset over Central England, the reanalysis performs much better (compare Figs. 8e and S11c). The reanalysis also struggles with orographic precipitation, for example at MONT-AIGOUAL GSN station (1567m elevation) compared to the other French stations where it performs much better (below 100m elevation). This is consistent with the poor performance seen over Scotland in Fig. 8b vs 8a. Otherwise, the mean and variability in 20CRv3+ is broadly within the range of values at individual stations for most sub-regions. This builds confidence in the capability of the reanalysis system to represent precipitation variability over the larger regions in the manuscript.

We propose to include the above paragraphs, along with the figures below, in the Supplementary.

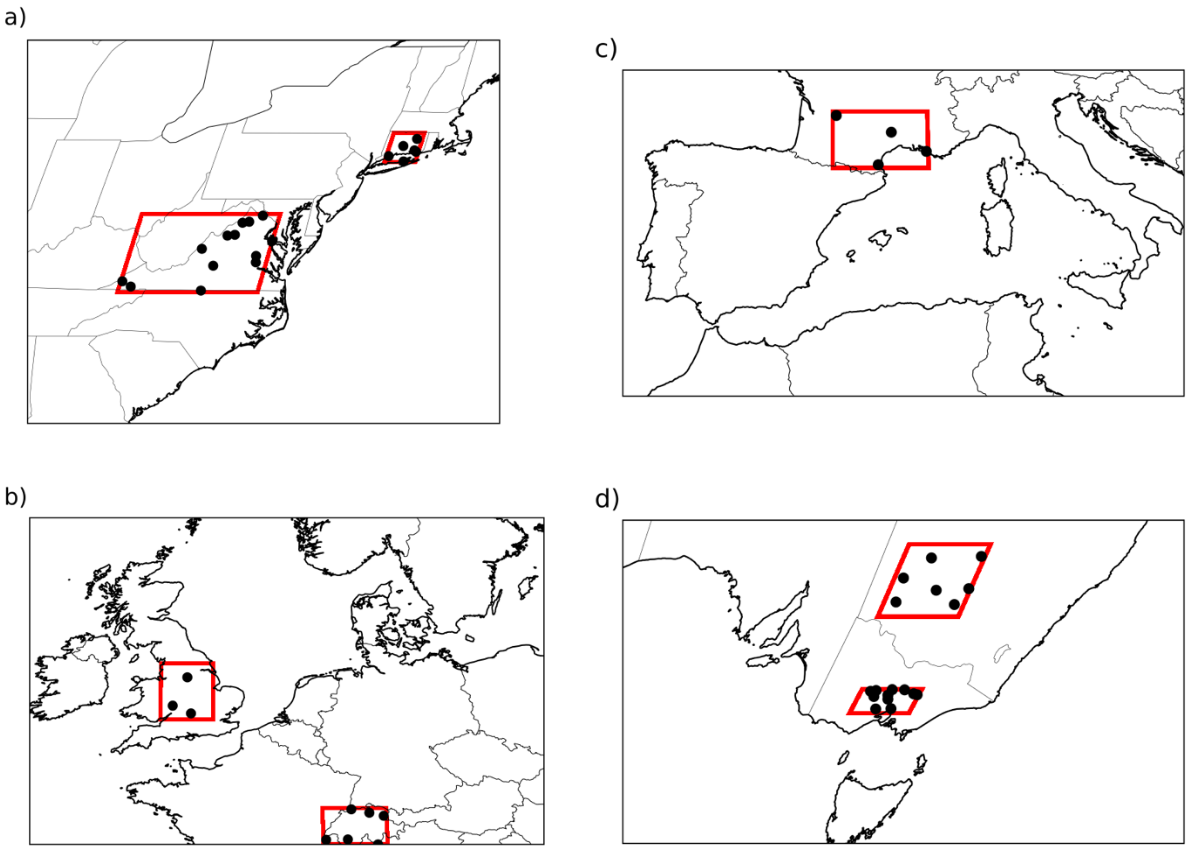

Fig. S10: Subregions used for validating 20CRv3+ precipitation in Figs. S11 and S12. Subregions are chosen within each of the main well-observed regions in Fig. 1: (a) US_East_coast (subregions: CT and VA), (b) NW_Europe (subregions: Central England and Switzerland), c) Western_Med (subregion: France), and (d) SE_Aus (subregions: South_Victoria and SW_NSW). Markers show the GLAMOD (Noone et al., 2021) stations within each subregion for which daily precipitation records from 1903 are available. Only one subregion was chosen in Western_Med due to lower data availability.

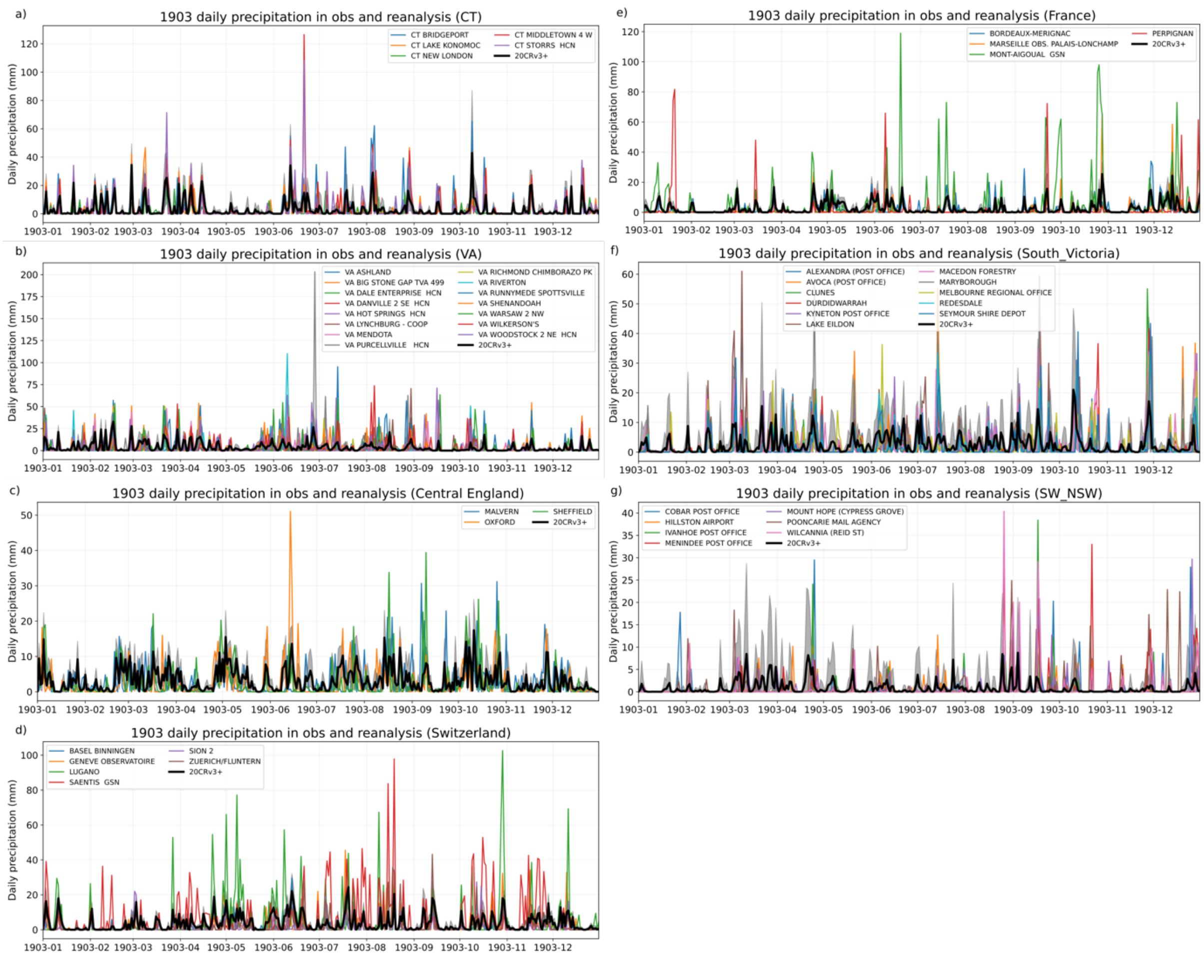

Fig. S11: Daily precipitation time series for the validation stations in Fig. S10 and for 20CRv3+. The dark black line shows the ensemble-mean 20CRv3+ precipitation averaged over the whole subregion. The grey shading indicates the 5-95% range across the 20CRv3+ ensemble.

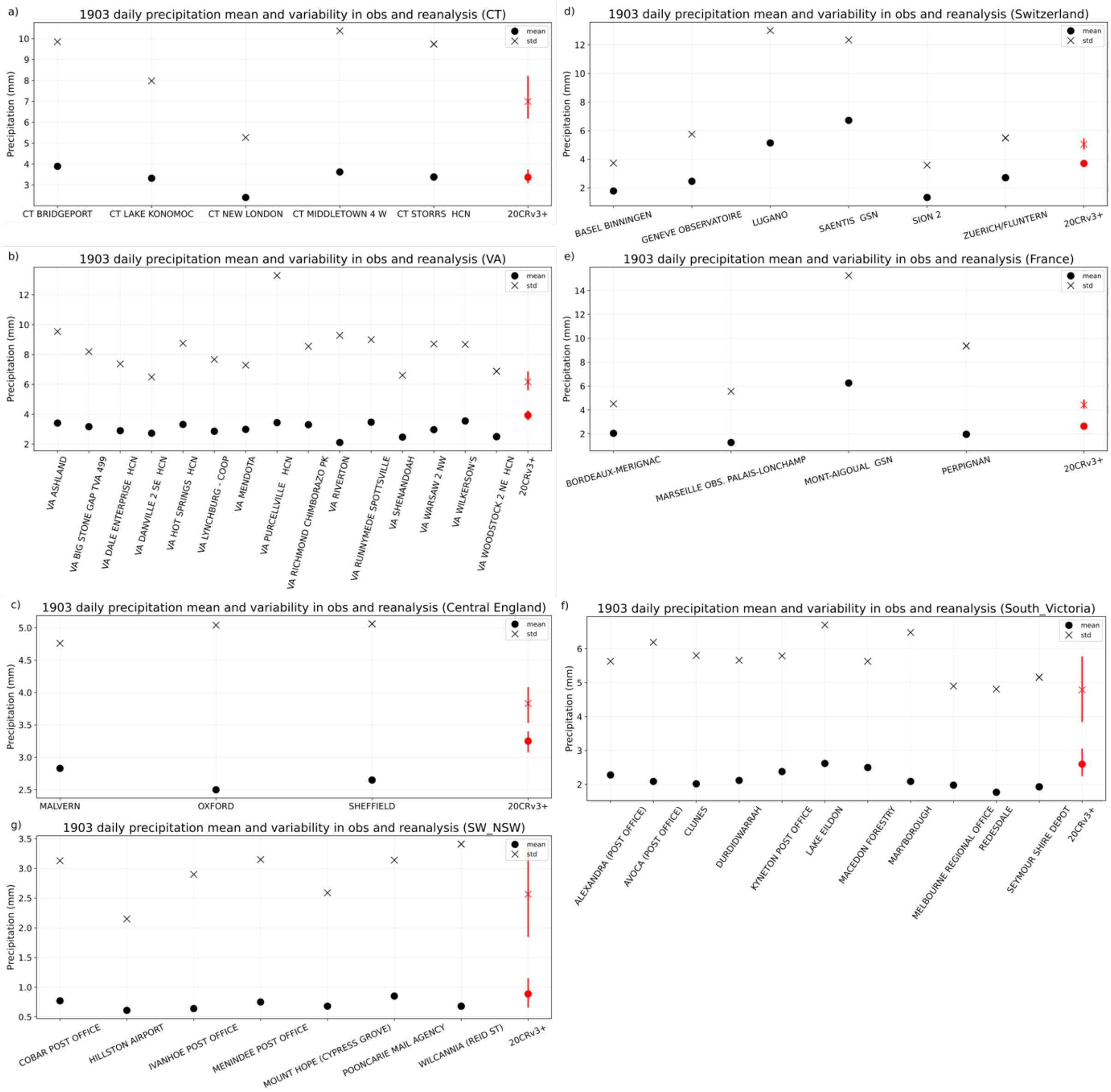

Fig. S12: Mean (circles) and daily variability (crosses) in precipitation for the validation stations in Fig. S10 and for 20CRv3+ (red). Statistics are calculated over the full 12 months of 1903. As in Fig. S11, the 20CRv3+ values are averages over the whole subregion. Vertical bars on 20CRv3+ indicate the 5–95% range across the ensemble.

Literature review and comparison: The literature review covers the PGW technique, but misses out further possibilities (e.g. large scale nudging). Also, the results obtained should somehow be set into the context of existing works (not targeting specifically the year 1903 of course, but more general the thermodynamic effects of global warming on weather variability).

Regarding nudging literature: We agree with the suggestion to include the nudging literature and have amended the relevant paragraph as follows:

"Our approach is similar to other storyline attribution methods, such as nudging experiments where the large-scale winds are relaxed towards specified values (Wehrli et al., 2020; van Garderen et al., 2021; Sanchez-Benitez et al., 2022; Feser and Shepherd, 2025), and the pseudo global warming (PGW) method, where the boundary conditions for regional climate models (RCMs) are adjusted to reflect the effects of climate change (Schär et al., 1996; Brogli et al., 2023; Lenderink et al., 2025). These adjustments, often referred to as PGW "deltas", are chosen such that the imposed dynamical driving at the RCM boundaries is unchanged in the warmer world simulation. Similarly to the reanalysis experiments, there is no change in intraannual variability between the control and "counterfactual" (i.e. nudged or PGW) simulations, which allows robust signals to be extracted from shorter simulations than are typically required with free-running climate models (Sato et al., 2007). The main differences between these methods are in how the circulation is constrained: in PGW, typically the 3D dynamics are imposed only at the edge of the RCM domain, while large-scale nudging can be performed globally or across a smaller regional domain in the free troposphere (Wehrli et al., 2020; van Garderen et al., 2021); in the reanalysis experiments, the synoptic circulation at the surface is constrained through the assimilation of surface pressure observations. As in some nudged simulations, the reanalysis experiments are also global, allowing multiple locations to be studied in a single run. The reanalysis experiments in this paper thus offer an additional and computationally efficient line of evidence regarding the thermodynamic effects of warming on regional weather and climate."

Regarding thermodynamic effects on weather variability: Many studies have separated the thermodynamic and dynamic consequences of global warming, e.g. in the storyline attribution literature, however these often focus on specific events rather than changes in the *statistics* of day-to-day weather (i.e. changes across the distribution) – filling this gap is partly the motivation of our study. We have referenced literature that discuss thermodynamic mechanisms that may explain some of our results, e.g. land surface processes leading to amplified warming on hot days (L180, L184), humidity and land/sea contrast changes leading to Mediterranean drying in the summer months (L197), and changes in the number of dry days related to increased CIN (L240).

Sea ice snow: Parts of the temperature sensitivity and probably also parts of the global patterns presented in Fig. 4 could be explained by snow and/or sea ice changes. A brief supporting analysis of these points would help the interpretation of the results.

We agree with the reviewer that this design choice could explain some of the later results and appreciate their helpful suggestion. We will add the following sentences to our discussion of the respective figures:

L115: …match the real world values precisely. Another idealised design choice is the unchanged sea ice fields in the warmer world experiments, which is likely the reason for the weak Arctic warming in Fig. 4a. Keeping sea ice fixed in the warmer world experiments will result in a smaller increase in global-mean temperatures than if more plausible reductions in sea ice concentration had been specified (Fig. 3a). Enhanced warming also occurs over regions of high elevation including the Andes and Himalayas (Figs. 4a–b), consistent with thermodynamic mechanisms such as a reduction in snow albedo and Planck feedbacks (Mountain Research Initiative EDW Working Group, 2015).

MINOR ISSUES

Language: The authors frequently refer to the issue how the 1903 weather conditions would be “experienced” in a warmer world. This closely follows the storyline concept, but in the end only meteorological standard variables are analysed and no relation is made to the actual perception of the weather conditions. I’d hence suggest to adapt the wording here.

We appreciate the reviewer’s comment, but we disagree. The focus of the paper is on near-surface temperature and precipitation, variables that are familiar to most people (e.g. they feature in every weather forecast) and strongly shape how the weather is perceived. We think it’s fair to use “experienced” in this context.

Line 265: You probably mean Figs. 8c and 8d here.

Thanks, this will be amended.

Review references not appearing in original manuscript

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

Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D. and Schär, C., 2007. Soil Moisture–Atmosphere Interactions during the 2003 European Summer Heat Wave. Journal of Climate [online], 20 (20), 5081–5099. Available from: http://dx.doi.org/10.1175/JCLI4288.1.

Merrifield, A. L., Simpson, I. R., McKinnon, K. A., Sippel, S., Xie, S. -P. and Deser, C., 2019. Local and Nonlocal Land Surface Influence in European Heatwave Initial Condition Ensembles. Geophysical Research Letters [online], 46 (23), 14082–14092. Available from: http://dx.doi.org/10.1029/2019GL083945.

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

Mountain Research Initiative EDW Working Group. Elevation-dependent warming in mountain regions of the world. *Nature Clim Change* **5**, 424–430 (2015). https://doi.org/10.1038/nclimate2563

Noone S, Atkinson C, Berry DI, et al. *Progress towards a holistic land and marine surface meteorological database and a call for additional contributions*. *Geosci Data J*. 2021; 8: 103–120. https://doi.org/10.1002/gdj3.109

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

Sánchez-Benítez, A., Goessling, H., Pithan, F., Semmler, T. & Jung, T. The July 2019 European Heat Wave in a warmer climate: storyline scenarios with a coupled model using spectral nudging. *J. Clim.* **35**, 2373–2390 (2022).

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.

van Garderen, L., Feser, F. and Shepherd, T. G., 2021. A methodology for attributing the role of climate change in extreme events: a global spectrally nudged storyline. Natural Hazards and Earth System Sciences [online], 21 (1), 171–186. Available from: http://dx.doi.org/10.5194/nhess-21-171-2021.

Wehrli, K., Hauser, M. & Seneviratne, S. I. Storylines of the 2018 Northern Hemisphere heatwave at pre-industrial and higher global warming levels. *Earth Syst. Dyn.* **11**, 855–873 (2020).