RC1

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):

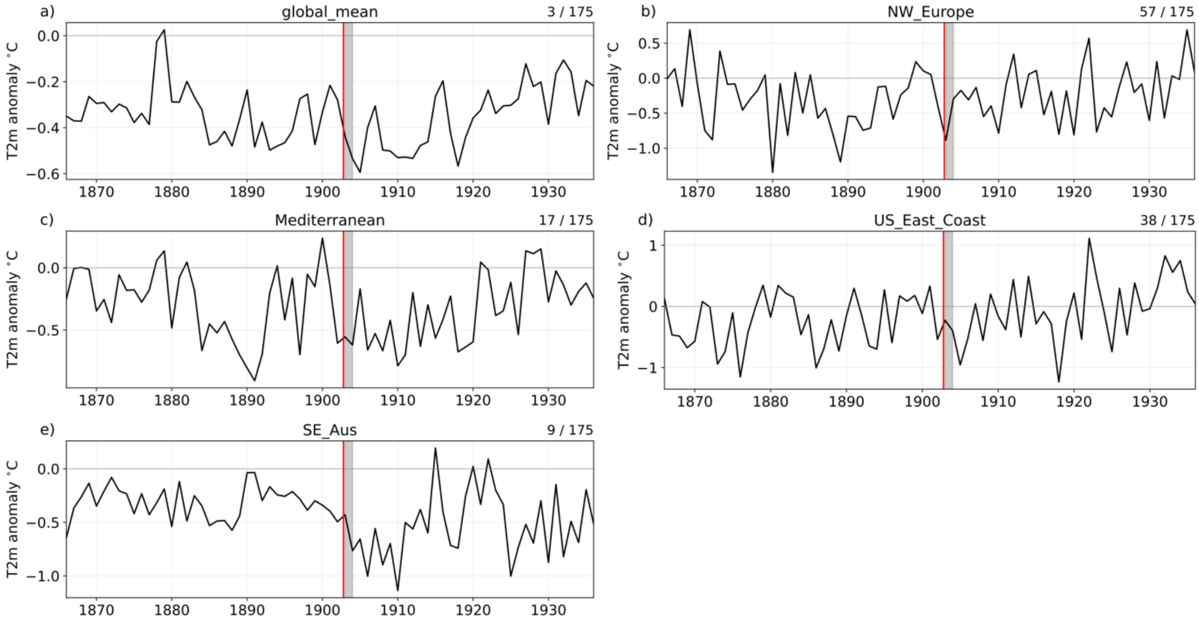


Fig. S3: Annual-mean surface temperature anomalies in HadCRUT5.1 (Morice et al. (2021)) for the global mean (a) and the four well-observed regions in Fig. 1 (b-e). Anomalies are with respect to the 1961-1990 reference period and include temperatures over both land and ocean. Grey shading shows the year 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).

L118 "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:

L12 "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:

L346 "Our approach shares some of the limitations of other storyline attribution methods. 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 (Figs. S11a–d and S12a–b. However, the assimilation system does reject more pressure observations in +2K+CO2 than in 20CRv3+ (Figs. S11e–f and S12c). 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. S3)."

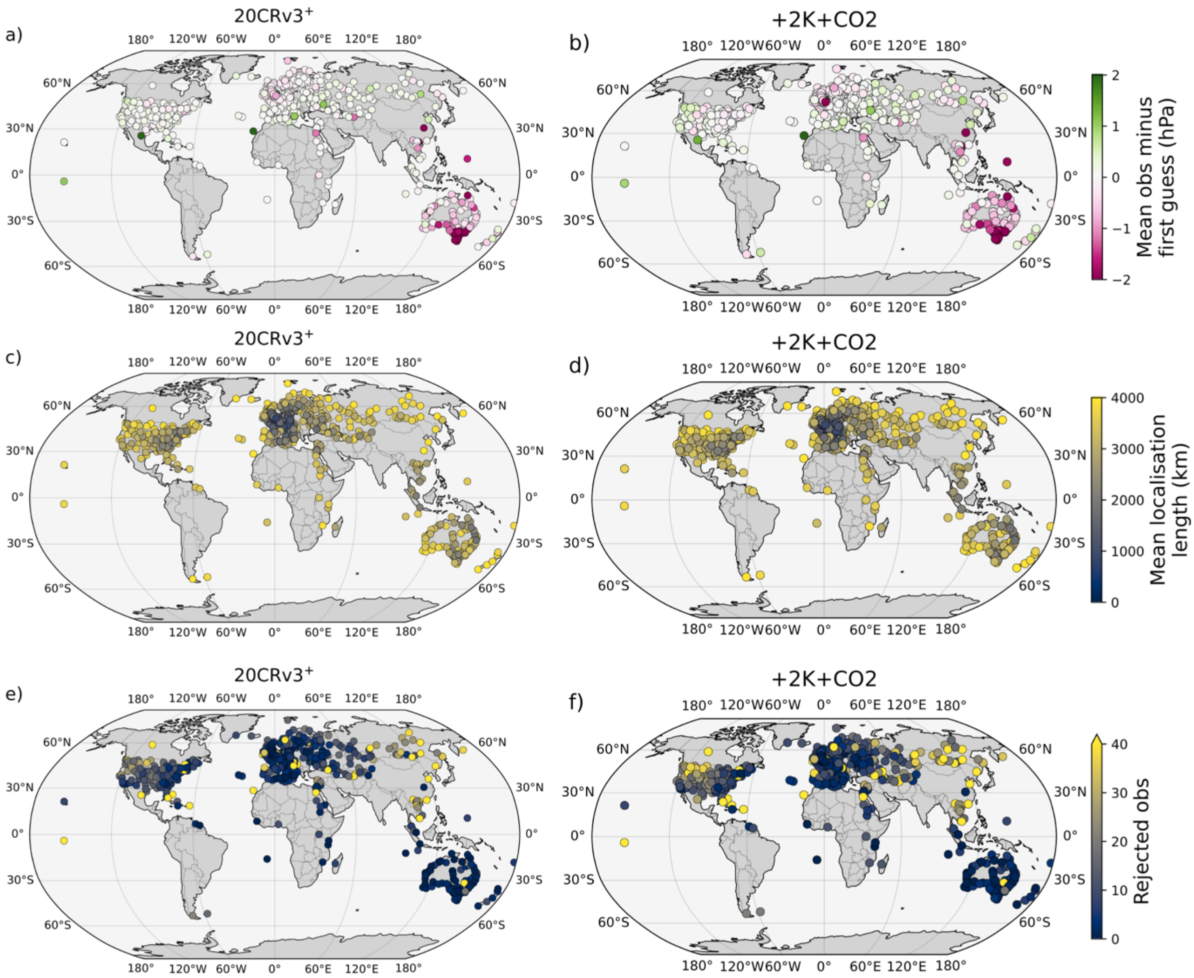


Fig S11: Maps of innovation (observation minus first guess, accounting for observation bias) (a, b), localisation length (c, d), and number of rejected observations (e, f) for land stations in the 20CRv3+ and +2K+CO2 experiments. Innovation and localisation length are average values over 1903 while (c, d) show the total number of rejections across the year.

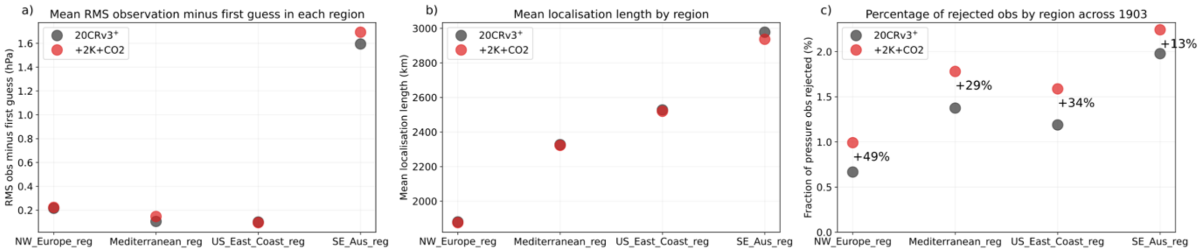


Fig. S12: Regional statistics for the innovation (a), localisation length (b), and fraction of rejected observations (c) shown in Fig. 7. In each case, the values denote the average of all land stations within each of the well-observed regions in Fig. 1. Percentage changes relative to 20CRv3+ are shown 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:

L372 "On the other hand, 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. S10) 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. S5). 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:

L67: 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:

L86: …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:

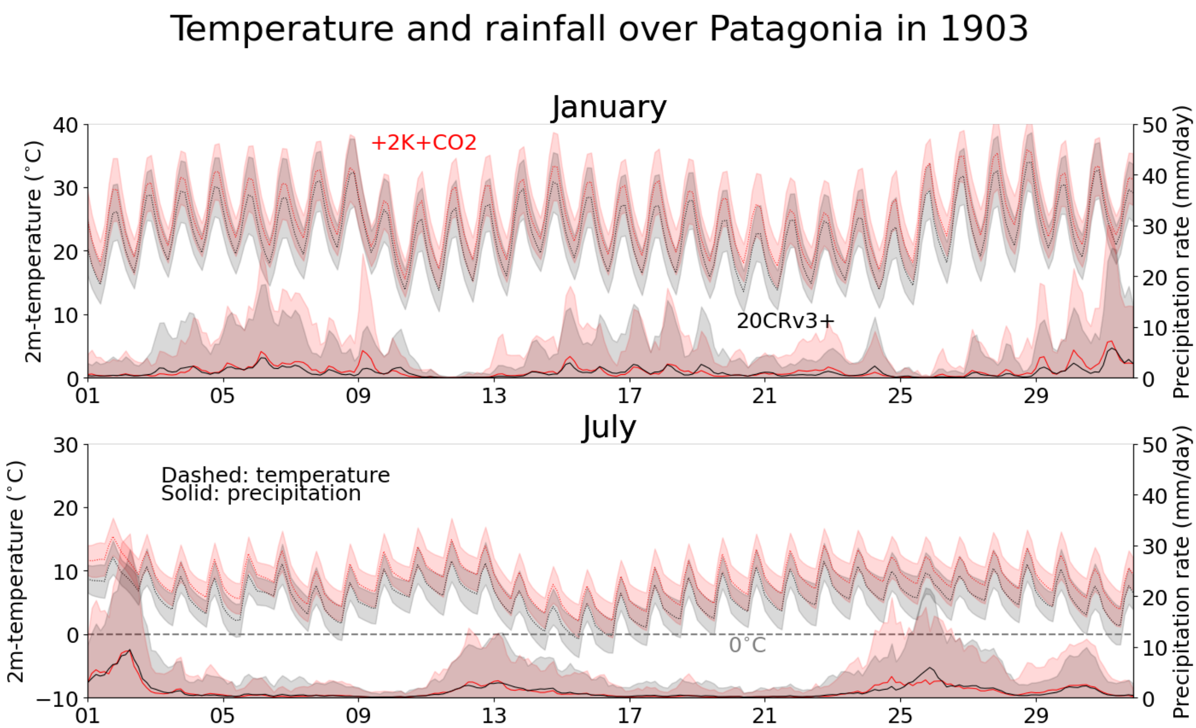
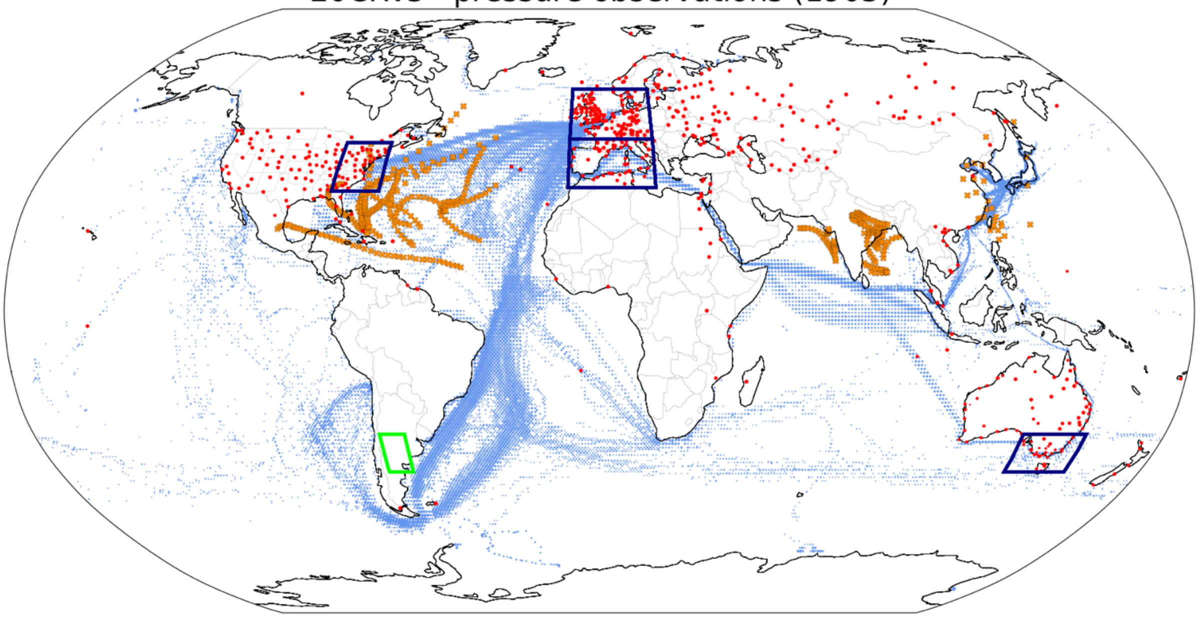


Fig. S2: Temperature and rainfall over a region in northern Patagonia for two months in 1903 (as in Fig. 2, but for a region with fewer pressure observations). The Patagonia region is shown by the green box in Fig. 1.



(New Fig. 1)

L111 "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:

L367 "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, this has been fixed.

RC2

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:

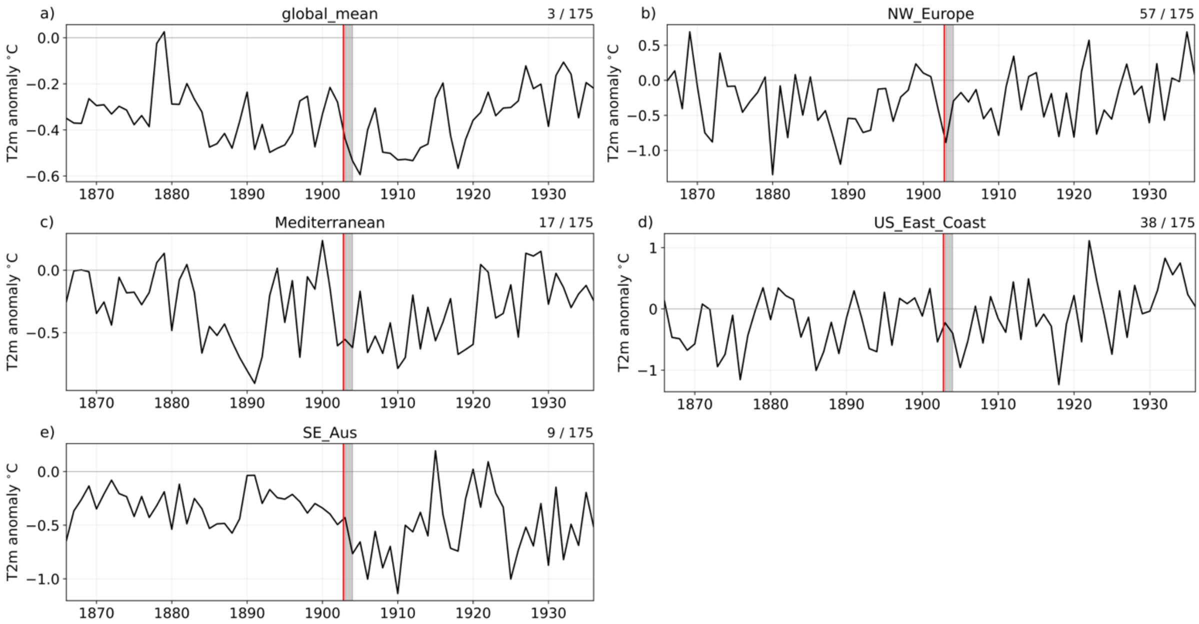


Fig. S3: Annual-mean surface temperature anomalies in HadCRUT5.1 (Morice et al. (2021)) for the global mean (a) and the four well-observed regions in Fig. 1 (b-e). Anomalies are with respect to the 1961-1990 reference period and include temperatures over both land and ocean. Grey shading shows the year 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).

L118 "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:

L365 "If changes in the atmospheric circulation do occur in future, they could lead to much larger changes in extreme temperature and precipitation, for example if persistent weather patterns were to occur more frequently (Petoukhov et al., 2013; Mann et al., 2017). 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). This also 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."

And later:

L379 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 will add the following sentences in the text:

L288…Output from the reanalysis could be downscaled using regional atmospheric models to better represent rainfall over mountainous regions if desired. Historical daily gridded rainfall observations are rarer outside the UK, so we also validate 20CRv3+ rainfall against station observations from each of the well-observed regions in the Supplementary (see Section S2 of the Supplementary).

And the following supplementary section S2 (including Figs. S7–9):

In this section we validate the precipitation statistics of 20CRv3^{+} against observations. The fairest comparison is to a daily gridded observational dataset, as in Fig. 8, but this is more difficult outside the UK. We therefore validate 20CRv3^{+} against daily station precipitation observations for a small number of sub-regions within each of the four well-observed regions in Fig. 1. 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. S7, the daily time series in Fig. S8, and mean and daily precipitation variability are shown in Fig. S9.

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 the sub-regional average from the gridded observational dataset over Central England, the reanalysis performs much better (compare Figs. 8e and S8c). 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 compared to the gridded observations in Fig. 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 well-observed regions.

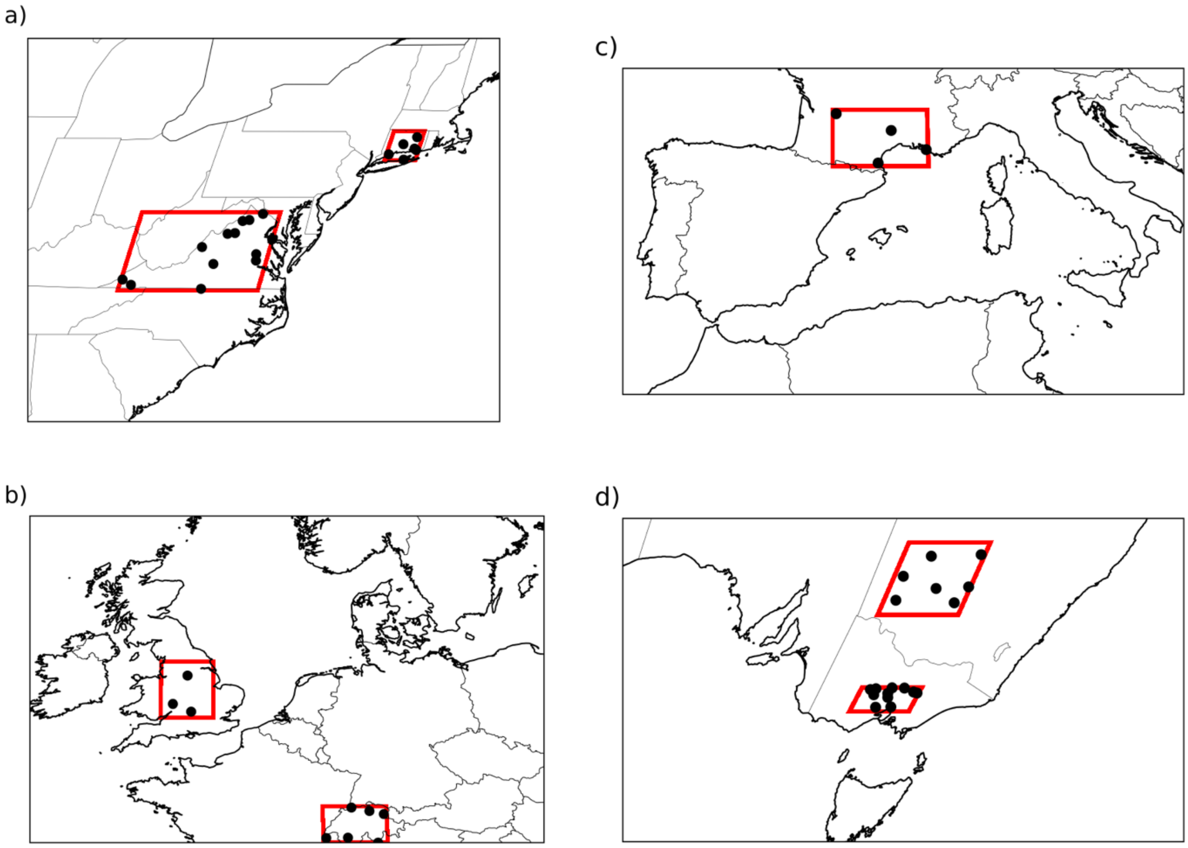


Fig. S7: Subregions used for validating 20CRv3+ precipitation in Figs. S11 and S12. Subregions are chosen within each of the main well-observed regions in Fig. 1: (a) US_East_coast (subregions: CT and VA), (b) NW_Europe (subregions: Central England and Switzerland), c) Western_Med (subregion: France), 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.

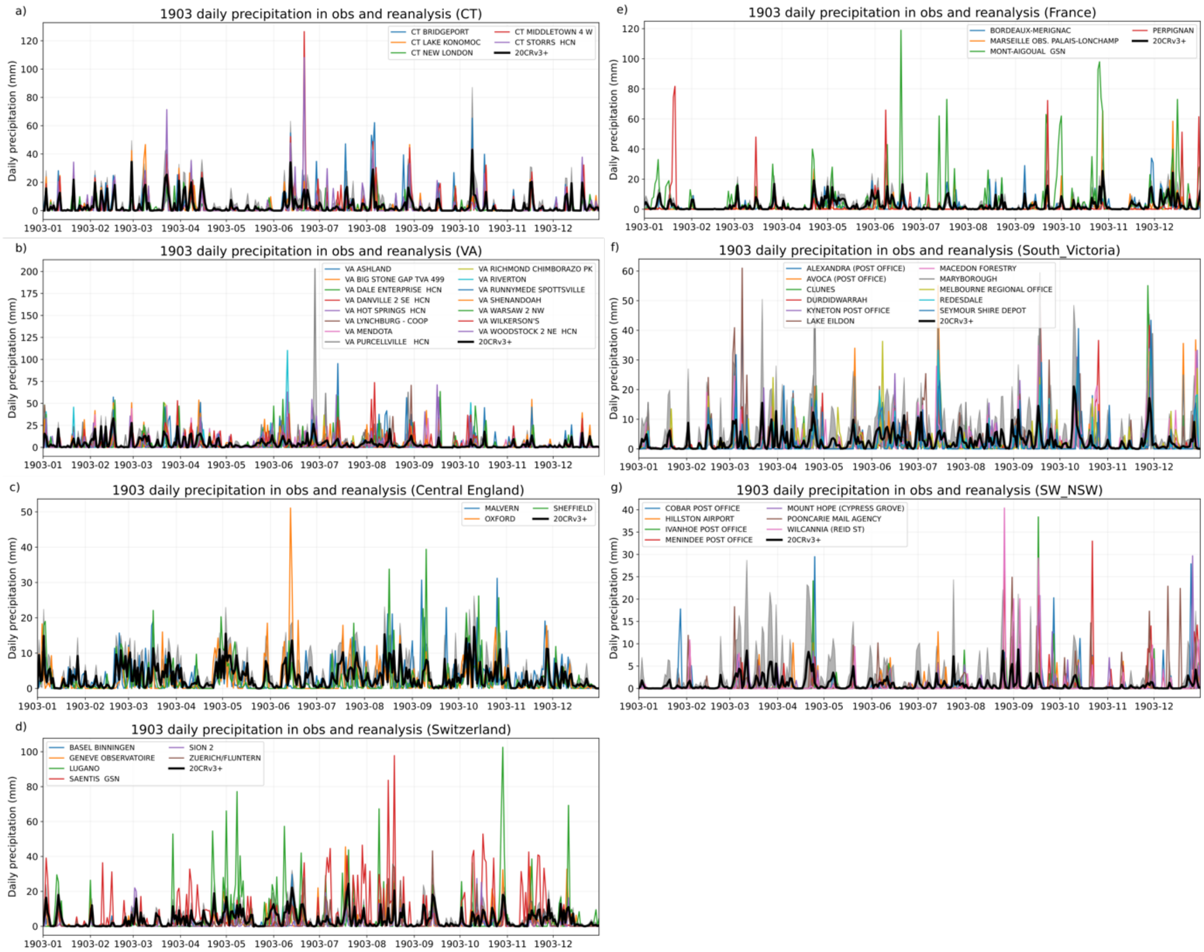


Fig. S8: Daily precipitation time series for the validation stations in Fig. S10 and for 20CRv3+. The dark black line shows the ensemble-mean 20CRv3+ precipitation averaged over the whole subregion. The grey shading indicates the 5-95% range across the 20CRv3+ ensemble.

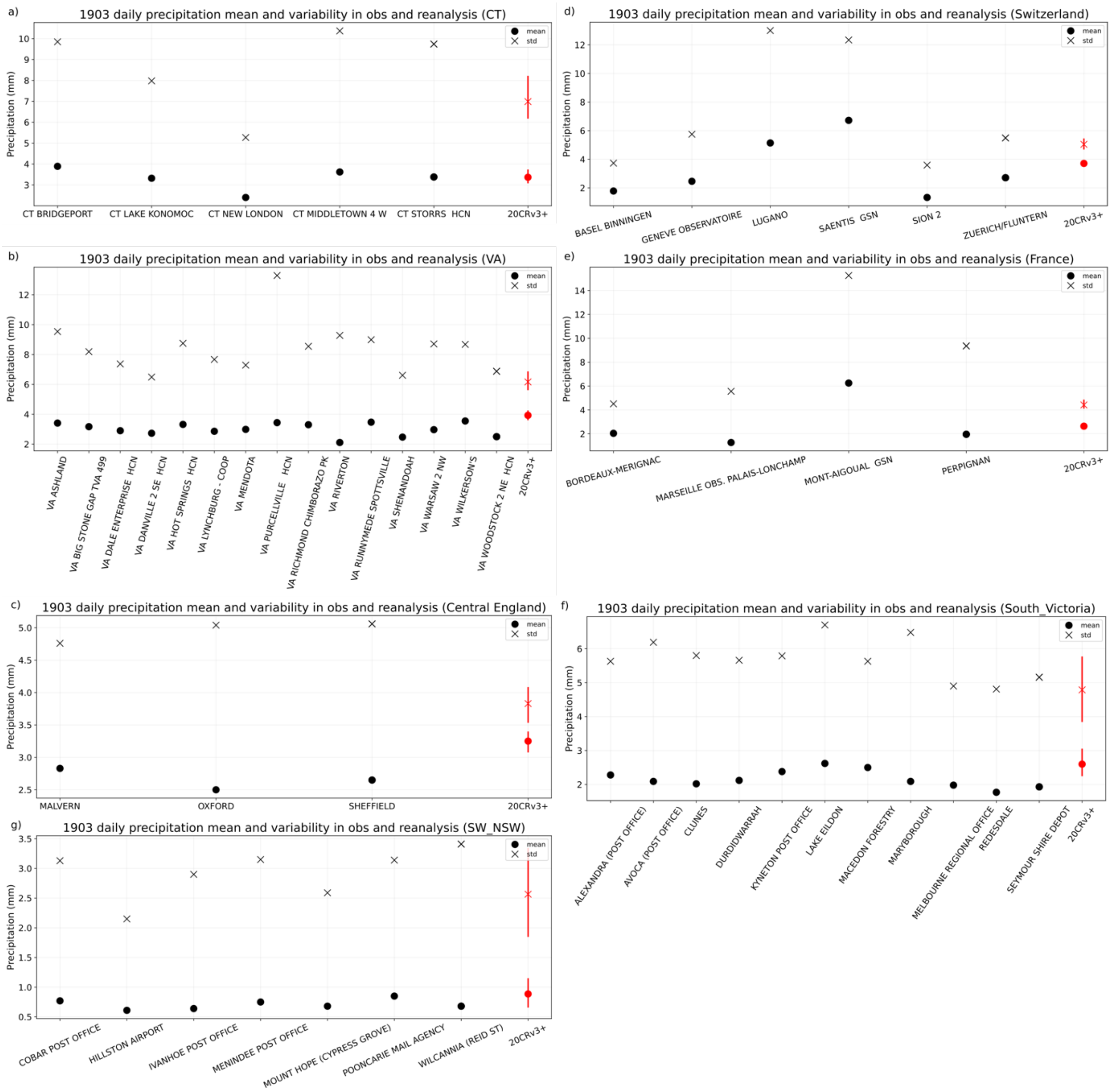


Fig. S9: Mean (circles) and daily variability (crosses) in precipitation for the validation stations in Fig. S10 and for 20CRv3+ (red). Statistics are calculated over the full 12 months of 1903. As in Fig. S11, the 20CRv3+ values are averages over the whole subregion. Vertical bars on 20CRv3+ indicate 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:

L44 "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:

L142: …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 (Pepin et al. 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

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

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

Pepin et al. 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.

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.

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).