

The influence of the Atlantic Multidecadal Variability on Storm Babet-like events

Response to reviewer 1:

Review of “The influence of the Atlantic Multidecadal Variability on Storm Babet-like events”

The work of Thompson et al. presents an investigation of the link between Atlantic Multidecadal Variability (AMV) and analogues of storm Babet (October 2023). The authors seek to investigate the sensibility of analogues averaging inside time slices as done by Faranda et al. (2022) to conclude on the influence of anthropogenic climate change on similar storms. The goal of the paper is to quantify the impact of multidecadal variability (AMV in particular) in the North Atlantic on the analogues found using multilinear regression on variability indices.

The paper is well-written and the research question is clearly stated. The latter is an important avenue of research for quantifying the sensibility of the analogues method to the long-term variability of the climate system and understand better what kind of conclusions can be drawn from this method as it is now used for attribution purposes (the authors cite the Climameter tool). Moreover, the work of Thompson et al. is a contribution to the literature on the dynamical evolution of the atmosphere under anthropogenic and natural forcings. This makes this work suited for publication in *Weather and Climate Dynamics*.

However, at this stage I recommend major revisions before accepting the paper for publication. In general, the paper lacks sensitivity tests to convince the reader about the robustness of the results found. But more importantly, the authors seem to claim to have found a causal link between AMV and analogues of storm Babet. I am not convinced that the arguments of the authors are sufficient to defend such a statement. Statements in the conclusion such as L193-194: “For Storm Babet, we show that similar events are more likely during positive AMV phase.” and L196: “The results suggest that if the current trends of amplified warming in North Atlantic sea surface temperatures continue, we should expect to see more events similar to Babet.” are not, or very weakly, supported by the analysis provided here. Moreover, if such statements were to be supported more strongly in a revised version of the manuscript, they should be compared to the scientific literature on this subject. See below for the details of my comments.

We thank the reviewer for the detailed comments, and are pleased we have conveyed the importance of the research. We have revised the manuscript based on the comments, and those of another reviewer, and hope the reviewer finds it improved.

We have made many changes, including further statistical testing to show the robustness of the results, and a more detailed comparison with other scientific literature on the AMV and European storms to help support the case for causality.

Major comments:

1. Sensitivity tests:

a. The authors decided to use an arbitrary number of analogues in each period. While the number of analogues considered will always be more or less arbitrary, I think the authors should investigate the sensibility of their results to the number of analogues chosen (especially for Fig 1 and 2).

Indeed a different number of analogues could be chosen, and for different types of events it could be that this number should be adjusted (depending how rare the event itself is). We have added further explanation of our choice in the methods, and added a figure showing how the quality of the analogues varies with different analogue numbers, based on Fig.1:

Sec2.2: Using too many analogue days would lead to the weaker analogues being too different to the event day to give useful information, too few prevents meaningful assessment. Fig.A2 shows the impact of different amounts of analogues on the analogue composites. The decision to use 1% of days as analogues is based on earlier studies

(Faranda et al., 2022; Thompson et al., 2024; Faranda et al., 2024).

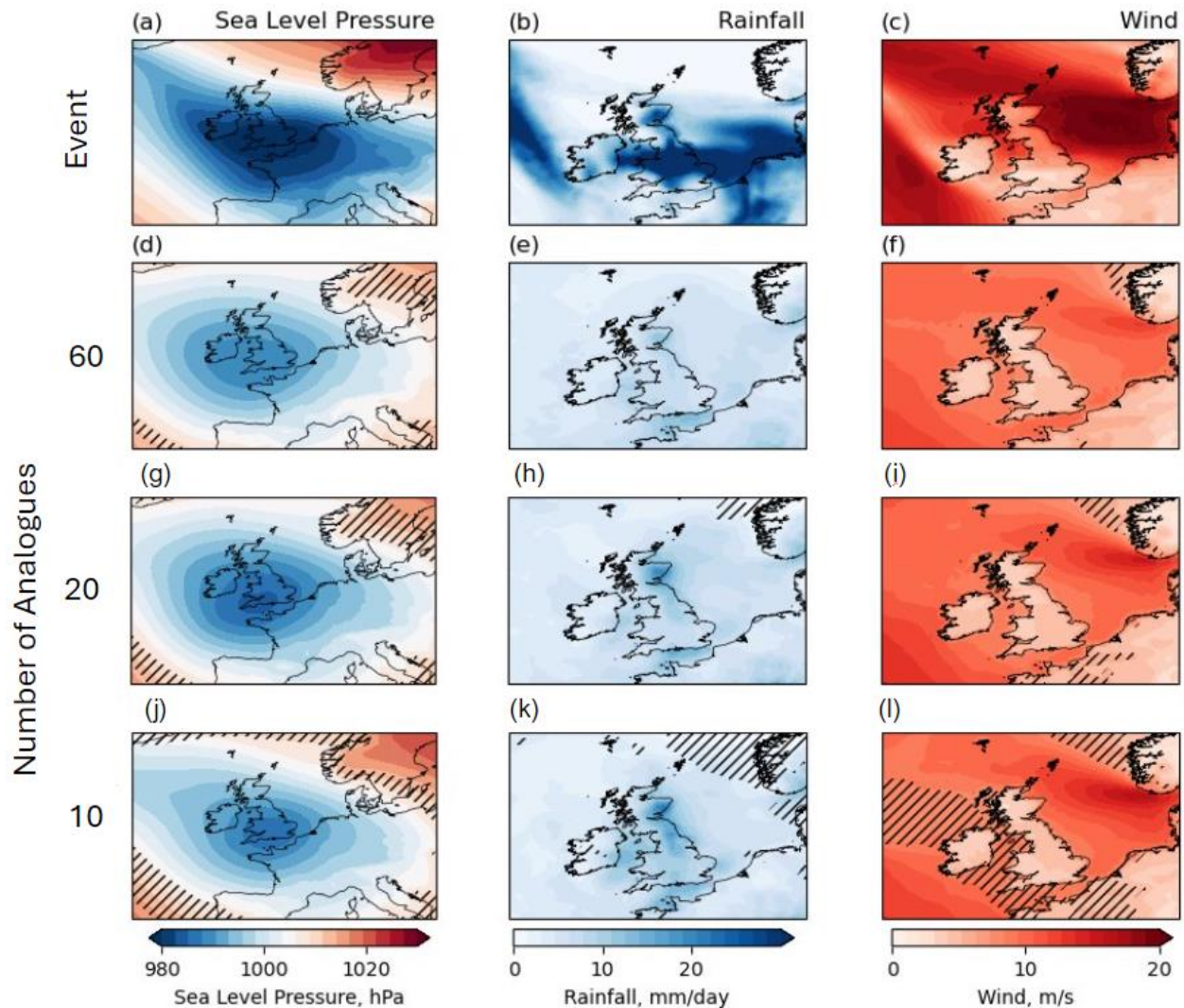


Figure A2: The influence of the number of analogues used. (a-c) The meteorological situation on 20th October 2023, showing (a) sea level pressure (hPa), (b) daily rainfall (mm/day), and (c) daily mean wind field (m/s). (d-f) Composites of the 60 closest analogues from 1950 to 2022, based on sea level pressure, for (d) sea level pressure (hPa), (e) daily rainfall (mm/day), and (f) daily mean wind field (m/s). (g-i) As in (d-f) for the closest 20 analogues. (j-l) As in (d-f) for the closest 10 analogues. All data from ERA5. In (d-l) unhashed regions indicate where the signal in the composite of analogues is statistically different to zero, using a one-sided t-test ($p < 0.05$).

b. Have you assessed the sensibility of your results to the spatial domain chosen for defining the analogues?

As is the case in many climate event assessments – such as the World Weather Attribution rapid attribution studies and the ClimaMeter assessments - we do have to define a spatial domain over which we assess. In the case of this study the domain was chosen to include

the main dynamical features – i.e. the low pressure system. The domain we choose is 40 °N to 65 °N, 20 °W to 20 °E (as shown in Fig1a).

We did explore different domains, for example the region used in Ginesta & Faranda (2003) 30 °N to 60 °N, 20 °W to 20 °E. We found much overlap in event dates (50 % of events show perfect agreement and a further 25 % are likely from the same event, but a day or two out). That study chose a more southerly region to try to include Storm Aline, which impacted Spain at the same time as Storm Babet.

Using too large a region will result in less close analogues being identified. Too small a region may lead to analogues which do not include dynamical features key to the development of the events impacts (such as large scale rainfall). The choice of domain requires expert judgement to ensure the key dynamical features are included, but the domain is not so large that analogues identified are too unusual.

c. In Fig 1 and Fig 2, how the statistical significance is obtained is not clear. The legend in Fig 1 indicates that it is obtained when the mean anomaly is below the standard deviation. I am not sure to understand what that means and I do not think this is a proper statistical test or procedure.

We have changed the statistical significance testing. For figure 1 we assess where the signal in the composite of the analogues is statistically different to zero, using a one-sided t-test ($p < 0.05$). For figure 2 we assess the difference between the analogue sets, the significance is calculated using a two-sided t-test ($p < 0.05$). The methods, figures, and figure captions have been updated to reflect the change.

Sec2.2 changes:

We perform significance testing on the composites. When investigating analogues from one time period we assess where the signal in the composite of the analogues is statistically different to zero, using a one-sided t-test ($p < 0.05$) (Fig.1). When comparing the analogues from different time periods we apply two-sided t-tests to each gridpoint, assuming equal variance, if $p < 0.05$ the distributions are assumed to differ significantly (Fig.2).

d. L155-164: the results of the regression are not reported properly: there is no indication on the statistical significance of the regression, nor on the R^2 , nor on the uncertainties and p values for the coefficients for AMV and GMST.

The statistical significance has been indicated by adding the p-values for each coefficient. We have also now included the R^2 and p value for the multiple regression, which indicate that there is a relationship between the similarity measure and the predictors (GMST and AMV).

Sec 3.2 changes:

The multiple regression model has an R^2 value of 0.45 (uncentred), with a p -value = $1.3e-8$, indicating a statistically significant relationship between similarity and the predictor measures, AMV and GMST. We find coefficients of 0.074 $^{\circ}\text{C}$ (p -value = $1.3e-4$) for AMV and 0.003 $^{\circ}\text{C}$ (p -value = $2.0e-8$) for GMST.

e. Fig 4c: I guess the shadings of the GEV represent the uncertainties. How were they obtained? Why do they begin only for return periods greater than 10 years?

Yes it is uncertainties, at 95% confidence intervals. The caption of the figure has been updated to clarify this. The uncertainty bands have been extended to cover the full range.

f. Fig 4c: The authors should be more explicit in what they fit here. I guess that the location parameter depends linearly on AMV, but does it also depend on GMST? Please explain more clearly what GEV is fitted and with which fitting method.

The method for adjusting the similarity data based on AMV is explained at the end of the methods section (line 114 onwards). We have expanded this to clarify that there is no adjustment made for changes in GMST, only AMV.

Sec 2.3 text:

We find AMV is affecting S_x (see Results), and use extreme value theory to assess the change in return period of S_x with different AMV phase. We use the generalised extreme value distribution, as we are using annual (block) maxima (Coles, 2001; Philip et al., 2020). We assume that the trend shifts with AMV, this is factored into the GEV by allowing the location parameter to be linearly dependent on the AMV. The distribution is then evaluated at two values of AMV: $+0.25$ $^{\circ}\text{C}$ (positive phase) and -0.25 $^{\circ}\text{C}$ (negative phase). We make no adjustment for changes in GMST, only AMV. This is similar to how GEV can be used to assess a factual and counterfactual world, e.g. with and without climate change, but in our case we use only AMV as a covariate (van Oldenburgh et al., 2021; Philip et al., 2020).

g. L173-179: None of the values reported here have a confidence interval associated.

We have added uncertainties at the 95% confidence level - as described in the methods section.

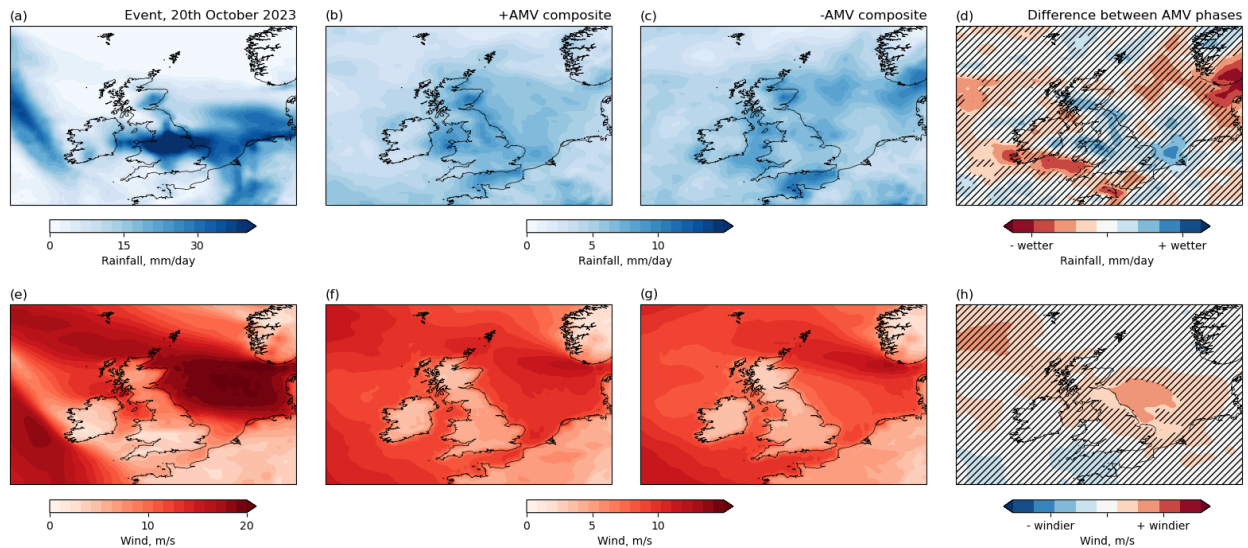
h. Fig 5:

i. It is not clear to me why these two boxes were chosen. How much does this result depend on the region selected? You could rather show a map with the difference in the mean at each grid point to see how this pattern varies geographically.

We have changed Fig.5 to maps following your next comment.

ii. I am not sure I understand the argument behind this figure. Using days for which $S > 0.7$ is not completely equivalent to using days in positive vs negative phases of the AMV. Maybe you could find analogues in the two phases of the AMV and see the difference? This would be similar to what Cadiou et al. (2023) did for ENSO.

Thank-you for pointing out Cadiou et al (2023) - a very relevant paper that I had not yet seen. Indeed, the method used by Cadiou et al makes more sense as we then capture the differences in impacts of the top analogues in each AMV state. Fig.5 has been changed to compare the top 1% of AMV+ and AMV- analogues (as now described in the methods), as shown below:



Text in both the methods and in section 3.3 has been changed to describe the new figure:

Methods:

2.4 Analogues in different AMV phases

We assess the differences between the analogues in different AMV phases by splitting the years into three groups based on AMV. We include the lower third (AMV index below -0.14) as $-AMV$ years and upper third (AMV index above 0.04) as $+AMV$ years. For those years we identify the top 1 % of analogues, using the methods described in section 2.2. Composites of these analogues are calculated (Fig.5). The statistical significance is calculated using a 2-sided t-test between each gridpoint for the AMV+ and AMV- analogues, if $p < 0.05$ the distributions are assumed to differ significantly.

Sec3.2:

We assess the difference between the top 1% of analogues for each AMV phase (Fig.5). For rainfall the AMV+ analogues show greater similarity to the observed event over England and Wales – a region where the analogues over the whole period failed to capture the event (Fig.1b/e). For wind, we also find the AMV+ analogues are more similar to the event than AMV- with statistically significant stronger winds in the southern North Sea during AMV+ years. In agreement with Fig.4, these results suggest events more similar to Storm Babet are more likely during AMV+ years. The spatial pattern of surface impacts – both for extreme wind and rain, are more similar to Storm Babet in AMV+ years (Fig.5).

2. Impacts of AMV and causality statements:

a. The authors use multilinear regression to assess the combined effect of AMV and GMST on Sx, a measure of the quality of the yearly best analogue of storm Babet. I will assume here that using Sx is correct (see below for my comment on this point) for estimating the sensibility of analogues. However, I am not sure how we should interpret the results of this regression. The authors seem to say that a significant AMV coefficient is sufficient to conclude to a causal effect. It may be true but the authors do not give enough arguments to support this claim. First, it is clear to me physically that if one finds a link between AMV and Sx, it is probably because AMV influences Sx rather than the contrary. The authors should explain that more clearly because it is currently implicit in their formulation.

In the introduction we have added further details on the probable mechanisms of a physical link between AMV and Sx, and why we assume AMV is influencing Sx and not vice versa.

Following the comments of another reviewer, we have added a further supplementary figure. This figure assesses correlation of Sx with SST globally, to support the choice of AMV for the multiple regression.

Additions to the introduction:

External radiative forcings may also influence the phasing and magnitude of the AMV. For example, Undorf et al., (2018) and Watanabe and Tatebe (2019) find greenhouse gases and aerosols play a key role in the timing, yet prior to 1950 volcanic eruptions coincide with phase changes (Birkel et al., 2018). With changes to the external radiative forcings on the climate system the drivers of AMV may also be changing. Klavens et al., (2022) find a growing contribution from external drivers. The AMV influences the spatial pattern of the NAO, influencing storm tracks across the North Atlantic (Börgel et al., 2020). Climate model experiments nudged to specified AMV states are often used to investigate the mechanisms. For example, Ruggieri et al (2021) showed that the low-level jet differs significantly between AMV states, with a positive phase leading to a more southerly jet in the eastern Atlantic. This agrees Peings and Magnusdottir (2016), who showed that temperature changes in the extratropics are required to force a shift in the NAO and storm tracks. Msadek et al. (2011) showed that, alongside the southward shift in the stormtrack there is an intensification of the subtropical jet during positive AMV. These mechanisms, found in climate models, are also supported by reanalysis data which shows a stronger relationship between AMV and North Atlantic extratropical storms frequency, than with the accelerated polar warming seen in recent decades (Valino et al., 2019).

Second, to correctly estimate the impact of AMV on Sx, one needs to control for all confounders, i.e. for variables that have a causal impact on both AMV and Sx, and only for these confounders. The authors say in L 215: “We acknowledge that anthropogenic forcings may be influencing AMV, but this will not impact our key findings.” Actually yes it would: any variable that would influence both AMV and Sx (such as anthropogenic

forcings) and that is not taken into account in your regression will bias the estimation of the AMV coefficient. On the other hand, here the authors seem to suggest that GMST is such a confounder, which does not seem correct to me. Moreover the authors do not detail and discard other potential confounders (anthropogenic aerosols for example?). I recommend the work of Kretschmer et al. (2021) for clarifying these points.

We have changed L215:

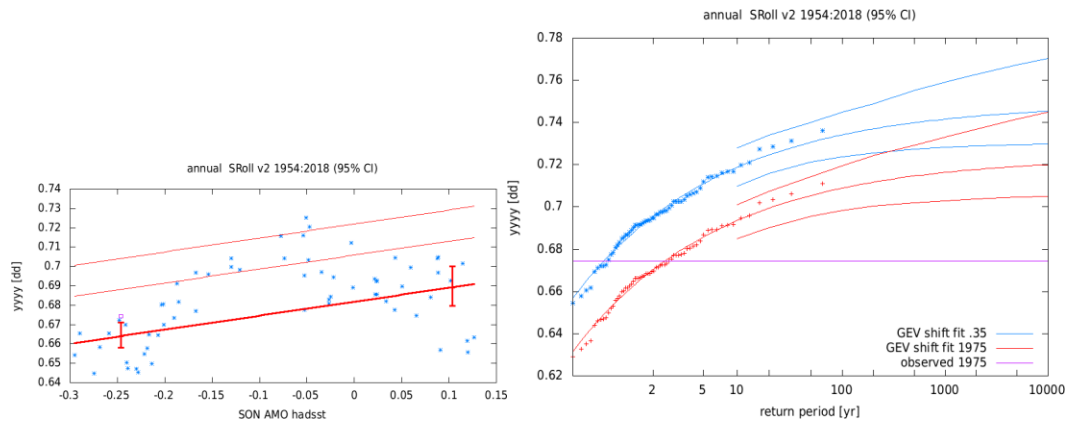
We acknowledge that anthropogenic forcings may be influencing AMV and we do use an AMV timeseries with GMST regressed out, but with only a short observational record available it is difficult to separate the drivers. If applying the method to different events and drivers it is important to consider the relationships between the drivers. Large ensembles of climate models would provide a testbed to investigate the conditional likelihood of events similar to Storm Babet under combinations of drivers, providing a range of states of internal variability.

We have also made changes to the methods text, and added more on causality and the physical links to the introduction (see comment above):

This is done to remove the effect of climate change in the AMV, and will ensure we do not identify spurious correlation between AMV and analogues driven by GMST (Kretschmer et al., 2021).

b. The correlation between AMV and Sx is rather weak (around 0.5 which means that only 25% of the variance is explained). Moreover, Fig 4b is quite worrying: it does not convince me that there is a linear relationship between these two quantities. If anything, it would suggest a second order relationship (ie proportional to AMV^{**2}). Finally, the authors project their results for a value of +0.25 for AMV which is far above what is ever observed in the data set. It seems to me quite problematic: how can you be sure that other data points will not have lower values of the similarity for high values of AMV as the negative trend for points on the right of $AMV=0$ suggests?

We had chosen an AMV of +.25 as it is the highest in the observational record going back further than the Sx record. It does indeed make sense to reduce this maximum to the upper end of what is seen in the Sx record – thus we have altered this down to +0.1. The conclusions are unaltered, and the difference between +AMV and –AMV is increased due to the identification of an error in the script. The new GEV figure has been verified using the Climate Explorer tool shown here (1975 ~ -25K AMV, thus 1975+.35 ~ 0.1 K AMV):



Indeed, with few data points above +0.1K AMV it is difficult to ascertain the relationship between S_x and AMV. We now highlight this issue further in the text:

We note, with only a short observational time series there is still uncertainty in this relationship, ideally more years of +AMV would be included but this is not available using ERA5. Further testing of the relationship with other observational datasets and climate model data would be useful, as highlighted in the conclusions.

c. The key results of this work are based on the analysis of the yearly max S_x . I think this is problematic for several reasons.

i. I do not see the point of using yearly maxima here. It is not clear how these yearly maxima reflect correctly the analogues found: how many of these maxima correspond to analogues used in Fig 1 and Fig 2? How many analogues do not correspond to yearly maxima? (there could be several analogues in a single year and none in other years).

Using yearly maxima allows us to produce an annual timeseries, which we need to allow us to carry out the multiple regression against GMST and AMV. If we instead took the analogues used in Fig1 and Fig2, or some other definition of the top analogues, we would not get values for every year but instead clusters of analogues in certain periods (as shown by the clusters of high similarity events in Fig.A2a). To enable this method to be applied to other events an annual timeseries is required, to allow assessment of the contribution of various modes of variability.

ii. The core of the results lies on the non-stationary GEV fitted on the S_x . Although this method is interesting, I am not sure we can interpret the results as straightforwardly as assumed by the authors. What the GEV gives is the probability that the yearly maximum of S is above a given threshold. It does not say how many analogues there would be per year and whether they will be good: it only gives a probability to the best one being very close. The authors say for example in L174: “events similar to Storm Babet are more likely – the chance of $S_x > 0.7$ is 7.5 times more likely during AMV positive than negative”, I am not sure this assertion is supported by the fit on Fig 4c because I do not think the fit of the GEV on

the S_x can be straightforwardly linked to neither the quantity nor the quality of the analogues in the positive and negative phases of AMV.

We have reworded parts of the interpretation to highlight that we are assessing the probability of the most similar event each year, rather than the similarity of each individual day:

We show a significant shift between phases, when AMV is in a positive phase, events similar to Storm Babet are more likely – the chance of the most similar event in a year showing $S_x > 0.7$ is ≈ 7.5 times more likely during AMV positive than negative

We argue that the fit of the GEV can be linked to the quality of the top analogue of a given year, though we agree it is not able to indicate anything about the quantity of analogues beyond the chance of none above a certain threshold. We use the annual maxima, rather than events above a threshold, to provide a timeseries which allows assessment against annual modes of climate variability.

iii. It seems to me that for what the authors want to achieve, the GPD approach for the extremes of S is more suited because it takes into account the quantity of analogues in addition to the shape of the tail. I want to highlight the fact that using a GPD (or GEV) approach on the distance of the analogues of a point (ie linking extreme value theory and recurrence in dynamical systems) is a mathematical domain that has been well developed in the recent years (see Lucarini et al. 2016 for an exhaustive review) and that it is the mathematical foundation for the local dimension measure provided in Faranda et al. (2022).

The GPD approach would perhaps be more appropriate for this case specifically – but the interannual similarity timeseries of the maximum event each year is required to assess the possible contribution of specific modes of variability or drivers of the change (in this case AMV and GMST). Without using an annual timeseries we would have been unable to use the multiple regression method used to investigate the possible relative roles of AMV and GMST in predicting the S_x variable. For consistency, throughout the study we use the S_x timeseries and therefore block maxima for the GEV.

Minor comments:

1. L27: typo “stormtims”

Corrected to ‘storms’

2. L95: maybe the number of the section is missing

Indeed, I have added ‘2.3’

3. L104: this equation should have a random term epsilon

We have replaced the constant with a random noise term, which would also incorporate any constant:

Linear multiple regression is used to identify the potential relative influence of AMV and GMST on the chance of analogues occurring. S_x can be predicted using AMV and GMST through the equation:

$$S_x = \alpha \text{ AMV} + \beta \text{ GMST} \pm \epsilon$$

Where α is the coefficient for AMV, β is the coefficient for GMST, and ϵ a random noise term. The coefficients can be used to interpret which has greater influence.

4. L179: using the Mann-Whitney test to find differences in the distributions is okay but here you report differences in the averages of these distributions. A Welch's t-test seems more appropriate to me.

As we have changed Fig.5 we no longer assess the differences between these distributions. In the new Fig.5 (and earlier figures comparing two analogue sets) we do use a Welch's t-test, as described in the methods and figure captions.

5. Figure A2: the color of the 95th percentile is indicated to be orange but is green in the figure. Also, why are there no units on the x and y axis of panels b and c ?

The caption has been corrected to state green for 95th percentile, and axis value labels have been added.

References:

Kretschmer, M., Adams, S. V., Arribas, A., Prudden, R., Robinson, N., Saggioro, E., & Shepherd, T. G. (2021). Quantifying causal pathways of teleconnections. *Bulletin of the American Meteorological Society*, 1-34.

Lucarini, V., Faranda, D., de Freitas, J. M. M., Holland, M., Kuna, T., Nicol, M., ... & Vaienti, S. (2016). *Extremes and recurrence in dynamical systems*. John Wiley & Sons.

Cadiou, C., Noyelle, R., Malhomme, N., & Faranda, D. (2023). Challenges in attributing the 2022 Australian rain bomb to climate change. *Asia-Pacific Journal of Atmospheric Sciences*, 59(1), 83-94.