Paper egusphere-2024-1253

How relevant are frequency changes of weather regimes for understanding climate change signals in surface precipitation in the North Atlantic-European sector? – a conceptual analysis with CESM1 large ensemble simulations

by Luise J. Fischer, David N. Bresch, Dominik Büeler, Christian M. Grams, Robin Noyelle, Matthias Röthlisberger, and Heini Wernli

Replies (2nd round)

We are grateful to the editor and reviewers for their additional comments that helped us to further improve the manuscript. Below we provide a one-to-one response to the specific points raised by the editor and reviewers. The comments from the editor and reviewers are in black and our replies in blue. Line numbers refer to the revised version of our paper.

We also would like to apologize for the long time it took us to do the (seemingly) minor revisions. The reasons were (i) lack of understanding of the first and last authors of the request for statistical testing, (ii) technical difficulties with re-accessing some of the data, and (iii) the busy schedule of the first author (who is now working outside academia). For the statistical testing, we invited advice from Robin Noyelle, a colleague in our group, and we added Robin to the list of co-authors. He helped us understanding the issue with the statistical testing raised by the 2nd reviewer and the editor, and he performed the testing (see below) and therefore contributed essentially to this new version of the paper. All co-authors agree with adding Robin to the list of co-authors.

Editor

Both of the original reviewers have provided an assessment of the revised manuscript, and appreciate all the work the authors have put in. I tend to agree with reviewer 2 that the two outstanding issues they mention are worth addressing. For significance, I think there has been a slight misunderstanding, but the reviewer clarifies this well – it is not about boosting weak signals by increasing the ensemble size but rather evaluating the significance of any/all identified WR-specific precip signals.

We addressed this point by performing a bootstrapping test (as explained in detail in our reply to the 2nd comment of reviewer 2.

If the authors could complete these minor revisions (clarification requested by reviewer 1, plus adding the significance assessment to the relevant figures and a few lines of discussion of how/whether the methodological choices affect the conclusions), I would see the manuscript as ready for publication. I plan to offer reviewer 2 the chance to see the manuscript one last time, but only if they have the time.

A few additional comments that may be helpful for the authors in finalizing the manuscript:

L59: multi-daily → multi-day or daily?

Changed to “multi-day”.

L70: Suggested edit: Distinguishing between thermodynamic and dynamic climate change effects not only improves…, it also helps…

Thank you, changed as suggested.

L73: I'd remove “among others”, or if absolutely necessary to keep it, move it to the end of the sentence.

Removed as suggested.

L79: “more positive phases” doesn't necessarily/exactly mean an increased frequency of NAO+.

Changed to “an increase in the frequency of the positive NAO phase”.

General comments, applicable throughout manuscript:

- historic → historical. Thank you, changed in several places.
- incorrect usage of “respectively” (e.g., Fig. 2 caption). Deleted “respectively” in the caption of Fig. 2.
- a different colour scale for $\gamma_{overall}$ might help show differences better. The main intention is to show that values for $\gamma_{overall}$ are small, and this comes out best if we use the same colour scale as for the other γ_i plots.

Reviewer 1

The authors have addressed my comments, and I have no further comment than the one below.

Caption of Figure 2: “Units days per day” Both the sentences and the units of the shading in panels (e) and (f) are not clear. The caption says number of wet days in which case the unit is days. However, the text in line 300 says that is a wet day frequency in which case the unit should be a % or a ratio between 0 and 1. Could the authors clarify the unit and text (caption or text)?

Thanks for spotting these inconsistencies. What we show is wet day frequency with values from 0 to 1. We clarify this in the captions and text.

Reviewer 2

Despite answering in detail to most of my comments, the authors did not fully understand the main issues I reported. I respect the authors' original methodological choices, but I still think some of the requests made in the first round of revision are legitimate and can be tackled without modifying the general approach. This does not change my overall opinion of the paper, which I consider an excellent piece of work and worth of publication. Nevertheless, I would like these two points to be more carefully addressed in the final version.

1) Significance. I do not agree that performing significance tests with ensemble data is not meaningful since "there is no limit in generating more ensemble members", nor that "even tiny differences would appear as significant": in general, we do not know which changes are significant and which are not before performing a statistical test. It could well be that most of the changes are indeed significant, but this can't be taken for granted without a proper calculation. The test I proposed is quite simple: take the ensemble standard deviation of 10-year averages in the historical period and, if the future-hist difference at a specific point is larger than that, that is an indication that the change is significant. I would apply this to regime composites of the change in Figs. 4–7: is the future-hist change larger (in absolute value) than the historical standard deviation of 10-year averages? Of course, this is only one possible measure of significance, but alternative ways are also possible. I don't think that this repeats Yettella and Kay (2017), since they did not investigate regime-specific changes.

We first struggled to develop an actionable plan to address this concern. Thanks to the expertise of Robin Noyelle, who joined the author team, we found a way to implement a bootstrapping approach to assess the statistical uncertainty in estimating the different terms induced by the finite size of the large ensemble. Our approach works as follows and this explanation is included in the revised paper at the end of Sect. 3: "In order to test the sampling sensitivity of the decomposition of $\Delta\varphi$ into WR-specific values to the 35 members used in each period, we implement a bootstrap procedure. For a given season and parameter φ we resample randomly with replacement over all members the same numbers of days as in the original data set. We then compute the $\Delta\varphi_i$, the decomposition (Eq. 5) and the $\gamma_i(\varphi)$ parameter in this resampled data set for each WR at each grid point. We repeat this procedure 1000 times in order to give an estimation of the sampling distribution of the different terms. We then compute the probability in this resampled distribution that the value 0 (for $\Delta\varphi_i$ and the terms in the decomposition of Eq. 5) or 1 (for $\gamma_i(\varphi)$) is exceeded or subceeded. This gives a bootstrapped p-value for significance at each grid point. We then use a false discovery rate test of 0.1 to take into account spatial correlations (Wilks, 2016) and flagged as significant the grid points that pass this test."

2) Interpretation. Your claim is quite strong, namely that regime frequency changes do not matter for the understanding of the impacts of future circulation changes on seasonal precipitation. However, I think the results you show are tightly linked with some of your methodological choices, and this should be discussed more in depth. In particular, these two choices impact the value of gamma:

- since you separately removed the climatology on the historical and future ensembles (also, the calculation of the normalization factor is done separately), part of the climate change signal on the circulation is removed, thus reducing the amplitude of regime frequency changes.

  We understand your point of view but still think that there is no other straightforward way to address regime frequency changes between different climates. We see the value of weather regimes in describing the variability of the flow relative to the respective climatological mean, and therefore we separately remove the climatology in both climate periods.

- on the other side, you do not remove the mean climatological change in the precipitation fields, so that regime intensity changes contain that signal (most of which is independent from the regimes, as you observe).

  This is because we would like to "understand", i.e., decompose the climate change signal in precipitation. We explained in the last replies that removing the mean climatological change in precipitation would lead to a different study addressing a different research question.

As far as I understand, the first choice reduces the "frequency effect", while the second enhances the "intensity effect". This is not to question your methodological choices – which are legitimate and are always to some extent "subjective" – but to better inform the community that it is actually quite complicated to disentangle the two effects, and the results you get depend – at least to some extent – on some of your assumptions.

We are grateful for your critical perspective on our decomposition, which made us think again about what we regard as the most appropriate way of addressing the main question of our study (as phrased in the title of the paper). Without someone implementing your alternative approach (removing climate change signal in precipitation and calculating the regimes in the future climate not relative to the future climatology), we cannot fully clarify the pros and cons of our conceptually different viewpoints. We therefore add a brief remark in the conclusions as follows: "During the review process the concern was brought up by one of the reviewers that the values of $\gamma_i(\varphi)$ are relatively small because we separately removed the climatology on the historical and future ensembles and therefore removed part of the climate change signal on the circulation, thus reducing the amplitude of regime frequency changes. However, since we regard weather regimes as patterns describing the circulation variability relative to the respective climatological mean, we don't see how regime frequency changes between different climates can be calculated in a way that is different from the one we implemented. We trust that future studies can further elaborate on these important conceptual questions."