

Review of manuscript nhes-2018-199 entitled "On the use of Weather Regimes to forecast meteorological drought over Europe" by C. Lavaysse, J. Vogt, A. Toreti, M. Carrera, and F. Pappenberger

This study proposes and evaluates a novel approach to predict meteorological drought on monthly time scales based on the forecast of large-scale flow patterns. The authors show that in the ECMWF extended forecasting system drought forecasts based on weather regime (WR) occurrence outperform drought forecasts based on direct precipitation forecast in most regions of Europe. Particularly those regions (British Isles, Scandinavia, NE Europe) benefit from WR-based drought forecasts, which show a strong link of drought and the large-scale weather regimes. The linkage of WR and drought as well as its predictability is thoroughly investigated; stratified according to seasons, and sensitivities to drought intensity and previous drought conditions tested.

It is shown that the WR approach has even more benefit for stronger droughts. Furthermore, the linkage is stronger in winter than in summer. Finally it is shown that the forecast captures well the linkage between WR and drought, but has difficulties in correctly representing WR frequencies.

Overall this study presents a very important contribution to research on monthly and sub-seasonal predictability and novel applications of now existing operational NWP data. It thoroughly documents that forecast products based on atmospheric fields that are easier to predict in NWP (e.g. geopotential, temperature) than more complex variables (e.g. precipitation, wind) can be effectively used to predict weather impacts due to the linkage of flow patterns and surface weather. The paper is well organised, clearly written in most parts and the figures carefully designed and chosen to support the storyline. Only at few places I struggled to follow and some references (to tables) were misplaced. Despite the long list of comments (which are all minor) I recommend to accept after one round of revisions.

First, we would like to thank the reviewer for his/her positive and encouraging response highlighting the importance of this scientific paper. We have replied to all his/her comments in red.

Broader Comments:

1. Several studies document that predictability on monthly time scales primarily arises from predictability in week 1 and week 2, while it vanishes in week 3-4 (e.g. Ferranti et al. 2018, Vigaud et al 2017, 2018). Did you check weekly or two-weekly forecast skill for the drought events? How would a two-weekly stratification look like?

Thanks for the references which point to an important topic. We focus on specific drought duration that could be relevant for users and decision makers (WMO No.1090). The scores shown in the study are obviously related to the ones at weekly or two-weekly time scales and skill is influenced by the performance of these initial weeks. However, we have not quantified this contribution as the objective of this study was the assessment of specific drought events on monthly time scale. As this topic deserves further investigation, we plan a dedicated study on it.

2. To me the usage of the term "teleconnection" is misleading. I understand under this term large-scale linkages from e.g. the Madden-Julian-Oscillation or SST or ENSO on weather regimes. In this paper I would talk of a linkage between the weather regimes and smaller-scale local weather/precipitation.

We agree with the reviewer.

As suggested, 'teleconnection' has been replaced by 'linkage' throughout the document.

3. Please carefully revise and check how you introduce your terminology. Sometimes different terms are used for similar items, some terms are poorly or not introduced. This makes the paper in partly difficult to read. Details are given in the line-by-line comments.

Thank you for this helpful comment. Please see the responses of the line-by-line comments.

4. Please also explain some of the methodology in more detail.

As requested, we have substantially modified the explanation of the methodology. All the steps have been displaced from SI to the main document, and we have clarified the entire section.

5. Some more literature could be cited: E.g. studies by Lavers et al. 2016ab and Ferranti et al. 2018, also support the idea that large-scale fields provide more predictability for a local weather phenomenon than trying to predict the phenomenon itself. Linkage to climate change could be mentioned e.g. with Santos et al. 2016 or Schaller et al. 2018 in the outlook. Linkage of weather regimes to other surface variables e.g. wind could be mentioned (e.g. Grams et al. 2017).

Thank you for these references. Some of them have been added as suggested. Nevertheless, according to the comments of Reviewer 1, the impacts of WRs onto other surface variables have been removed and because the link with climate studies is not straightforward, we prefer to not discuss that.

6. Table references are mixed up. Also order these in their order of occurrence in the paper. I found it difficult to directly understand tables and figures, due to too little information in the caption - in particular for tables. All Supplemental Figures should also be cited in the main text in their order of appearance.

We are sorry about the mistakes with the order of tables and references. All the figures are now cited in the main document in good order. We have also improved the captions of some figures/table for clarification.

Detailed comments:

reference order: Does NHES require stating the most recent literature first? If not please revert.

Done as suggested.

I51: refer also to Ferranti et al. 2018 as a recent study on WR and cold extremes. For wind Grams et al. 2017 might be an appropriate reference

Thank you for providing these interesting references. But following the comment of reviewer #1, the impacts of WRs on other variables have been removed.

I52: avoid talking of being teleconnected -> linked/associated with

See previous comment

I55: up to here you nicely introduced into the WR concept. It becomes confusing (I55-60) to now talk of NAO+-, without having clarified the differences between the NAO and WR concepts and without having clarified that the two NAO phases are two of the 4 winter regimes. So consider to first contrast NAO (as only describing part (i.e. 30%) of the large-scale variability on monthly/seasonal time scales) to WR (as describing most of the variability (i.e. 75%) on monthly time scales).

To clarify this point, a sentence has been added

“First and principally studied in Winter time, when they are more stronger, 4 main states have been defined namely the positive North Atlantic Oscillation phases (NAO+), the negative NAO (NAO-), the Blocking and the Atlantic Ridge.”

Section 2b: Do you do the k-means clustering in physical or phase space? E.g. is an EOF analysis performed? Do you use time-filtering? You should provide few (2-3 sentences) more details on the WR definition and also more details on how individual days are attributed to a WR (e.g. in physical or phase space I130ff). You repeatedly state why you only use WRs based on a k-means clustering in ERAI. Once justifying this approach is sufficient and convincing! Remove the redundancy.

The GA-K means (Genetic Algorithm k-means) has been performed on the anomalies of geopotential height at 500 hPa while the attribution for the other cases has been done by minimizing the distance from the centroid. Redundancies have been removed and the methodology section deeply clarified.

I140/I146: You refer to Table 2 here!? I found it difficult to directly understand Table 2. Re-Reading these lines, I understand it now, but it would help if you repeat the definition of a WR combination in the caption of Table 2. "WR combinations are defined as either additions or subtractions of monthly WR frequencies" Also in Table 2 I would replace WRa, WRb, ... by a, b

We apologise once more for all the mistakes on tables numbering that have now been corrected. We suppose the reviewer refers to Table 1. As suggested, we have modified the caption and the names of the WRs.

Table 1 is really helpful but hardly referred in the text. (as Table 6?). Definitely keep it.
See previous comment.

I113: season definition appears confusing. Please explicitly state if winter is NDJF or DJF, ...
As mentioned already in the text, winter is defined from December to February.

Section 2c: some more details on how SPI is computed would help please indicate a current (WMO) reference.

The WMO reference has been added. As this method is really common we consider brief description as sufficient enough. Nevertheless, we have clarified some descriptions.

I178: should state Table 3? Table 3 is quite standard, needed?

The references have been corrected. Because all the scores are based on the different values of the contingency table, we prefer to keep it.

I198: please specify in the text what is meant by "best correlation criterion". Do you compute at each grid point for each of the 16 WR/combination freq. time series the correlation to the SPI-1 time series? Then the WR/combination with highest correlation is shown in Fig. 1? This procedure needs to be stated.

This section has been completely reorganized and clarified according to previous comments and comments from reviewer #1

Section 3: clearly introduce the names/terminology used for the different setups/methods presented in Table 1 and subsequently please use it consistently. The definition of the fourth vs. third method remains obscure.

We use now the names of the different experiments introduced in table 1.

The differences between experiment 3 and 4 are better explained (see previous comment).

I212: please name this first method as the "Reference" method (as in Table 1)

Please see previous comment.

I213: Table 1!

Please see previous comment.

I217-219: the sentence in brackets is not needed / redundant.

This sentence has been modified (see the following comment).

I216: Suggestion for a slightly clearer formulation: "The third forecasting method, called "operational" in Table 1, computes MOAWR from ensemble data attributed to the WRs based on ERAI (see Section 2b)."

Modified as suggested.

I219ff: Here I am confused: Do you still compute WR-k-means clustering based on ENS data, or do you mean, that WR anomalies are computed wrt. ENS climatology not ERA-I climatology.

As mentioned previously, this section has been deeply clarified. The WR-k-means clustering is, for all the experiments, done by using ERA-I climatology. ENS is used during the attribution phase to define which WR is forecasted.

I228: The term "Reference" is not introduced, yet.

Reference is now clearly introduced before this paragraph.

The bulk of the result sections is well-written, therefore I have less detailed comments in the following.

I281: it would be nice to show evidence for these findings on the other seasons in the Supplement.

The results for the others seasons have been added in SI

Section 5a: The WR evaluation section is a bit weak. Please refer to studies by Ferranti et al. 2015, 2017 or Matsueda and Palmer for approaches of WR evaluation. E.g. how are weekly WR freq. evolving with time?

Thank you for the references, that have been added and discussed in the section.

The purpose of this study is not to deeply evaluate the forecasted WRs. Only the errors that could influence the results, and so could be considered as source of uncertainties of the drought forecasts, are analysed. That is why this evaluation is not complete but adapted to that specific discussion section.

I303: the statement about blocking is vague. Do you refer to the Blocking regime (so one of the 4 WR) or blocking anticyclones in general?

Here we refer to the Blocking regime. This has been clarified in the text.

I304: The sentence starting in line 304 and ending in line 309 is complicated. Perhaps directly start with what is shown in Figure 7 then go into the interpretation. I do not understand the argumentation of causes and effect for the anomalies. Try to rewrite lines 298-312 in clearer language. I wonder if Fig. 6 and 7 show really different things, or if only one of the two is sufficient. I prefer Fig. 7 but would elaborate on its description and interpretation.

According to this comment and the comment from reviewer 1, we have clarified the text. These sentences:

“The WR-distributions as given by the forecasts are characterized by a higher degree of similarity than the ones given by ERAI, with a peak of occurrence at around 5-8 days in winter (blue bars, Fig.6). The same holds for the other seasons (not shown). The lower spread of the forecasted WR occurrences, associated with reduced tails (i.e. reduced occurrences for durations exceeding 20 days), could be explained by the underestimation of the long-term blocking. A further examination of the temporal evolution of these occurrence anomalies suggests that the distribution of forecasted drought occurrences (previously shown) could mainly explain the overestimation of low occurrences using the reanalysis (i.e., larger number of forecasted events compared to those derived from ERAI with durations shorter than 5 days) and the underestimation of longer duration events (i.e., lower events with durations longer than 15 days using ENS than ERAI, red dotted lines in Fig. 7).”

have been replaced by :

“The WR-distributions as given by the forecasts are characterized by a higher degree of similarity between the WRs than the ones given by ERAI, with a peak of occurrence at around 5-8 days in winter for the four distributions (blue bars, Fig. 6a-d). The same holds for the other seasons (not shown). The lower spread of the forecasted WR occurrences, associated with reduced tails (i.e. reduced occurrences for durations exceeding 20 days), could be explained by the underestimation of the longer blocking episodes. A further comparison of the MOAWRs from ERAI and ENS (scatter plots in Fig. 7) suggests that : i) the distribution of forecasted drought occurrences could be explained by the overestimation of low occurrences using ENS than the reanalysis (i.e., larger number of forecasted events compared to those derived from ERAI with durations shorter than 5 days) and, ii) the underestimation of longer duration events (i.e., lower events with durations longer than 15 days using ENS than ERAI, red dotted lines in Fig. 7). ”

I311: you mean Table 4. But except here, the Table is hardly used. Do you really need it?

Please see previous comment.

I404: Ferranti et al. 2015, Weisheimer 2016, Magnusson 2017 and/or Grams et al. 2018 could also be cited to highlight the challenges in predicting WR. Furthermore you could insert a statement on the evolution of WR under climate change e.g. Santos et al. 2016, Schaller et al. 2018.

Thank you for these references that have been added in the manuscript.

As mentioned earlier, the link between this study to those on climate change is not straightforward. Thus, we prefer to not mention this statement.

References:

Ferranti, L., L. Magnusson, F. Vitart, and D. S. Richardson, How far in advance can we predict changes in large-scale flow leading to severe cold conditions over Europe? *Quarterly Journal of the Royal Meteorological Society*, 0, doi:10.1002/qj.3341.

Ferranti, L., S. Corti, and M. Janousek, 2015: Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Q.J.R. Meteorol. Soc.*, 141, 916–924, doi:10.1002/qj.2411.

Grams, C. M., R. Beerli, S. Pfenninger, I. Staffell, and H. Wernli, 2017: Balancing Europe's wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, 7, 557–562, doi:10.1038/nclimate3338.

Grams, C. M., L. Magnusson, and E. Madonna, An atmospheric dynamics' perspective on the amplification and propagation of forecast error in numerical weather prediction models: a case study. *Quarterly Journal of the Royal Meteorological Society*, 0, doi:10.1002/qj.3353.

Lavers, D. A., F. Pappenberger, D. S. Richardson, and E. Zsoter, 2016a: ECMWF Extreme Forecast Index for water vapor transport: A forecast tool for atmospheric rivers and extreme precipitation. *Geophys. Res. Lett.*, 43, 2016GL071320, doi:10.1002/2016GL071320.

Lavers, D. A., D. E. Waliser, F. M. Ralph, and M. D. Dettinger, 2016b: Predictability of horizontal water vapor transport relative to precipitation: Enhancing situational awareness for forecasting western U.S. extreme precipitation and flooding. *Geophys. Res. Lett.*, 43, 2016GL067765, doi:10.1002/2016GL067765.

Magnusson, L., 2017: Diagnostic methods for understanding the origin of forecast errors. *Q.J.R. Meteorol. Soc.*, 143, 2129–2142, doi:10.1002/qj.3072.

Santos, J. A., M. Belo-Pereira, H. Fraga, and J. G. Pinto, 2016: Understanding climate change projections for precipitation over western Europe with a weather typing approach. *J. Geophys. Res. Atmos.*, 121, 2015JD024399, doi:10.1002/2015JD024399.

Schaller, N., J. Sillmann, J. Anstey, E. M. Fischer, C. M. Grams, and S. Russo, 2018: Influence of blocking on Northern European and Western Russian heatwaves in large climate model ensembles. *Environ. Res. Lett.*, 13, 054015, doi:10.1088/1748-9326/aaba55.

Vigaud, N., A. W. Robertson, and M. K. Tippett, 2017: Multimodel Ensembling of Sub-seasonal Precipitation Forecasts over North America. *Mon. Wea. Rev.*, 145, 3913–3928, doi:10.1175/MWR-D-17-0092.1.

Vigaud, N., A. w. Robertson, and M. K. Tippett, 2018: Predictability of Recurrent Weather Regimes over North America during Winter from Submonthly Reforecasts. *Mon. Wea. Rev.*, 146, 2559–2577, doi:10.1175/MWR-D-18-0058.1.

Weisheimer, A., N. Schaller, C. O'Reilly, D. A. MacLeod, and T. Palmer, 2017: Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *Q.J.R. Meteorol. Soc.*, 143, 917–926, doi:10.1002/qj.2976.