Review comments for “The analogue method for precipitation prediction: finding better analogues situations at a sub-daily time step” by Horton et al.

Recommendation: Major revision

The authors introduced a moving time window (MTW) for the analogue method so that better analogues at a different hour can be found for precipitation prediction in contrast to the use of analogues at fixed hours of the day in standard analogue method. They found that the MTW with the shorter archive on a sub-daily time step improved the analogy criterion values across the entire distribution of analogue dates and the skill of precipitation prediction in comparison with the standard analogue method with longer archive on a daily time step. In particular, the improvement in prediction skill is greater for days with heavy precipitation. The topic is important and has great implications for operational precipitation forecasting and impact studies associated with the hydrological community. The only constraint is that the implementation of such method requires the availability of sub-daily time series, which may not always exist.

I have several major comments. First, some necessary information regarding the the presented analyses should be provided. For example, what season are the results shown in Figures 4, 5,6,7,8, 10 based on?  
We added a description of the calibration and validation periods. The results are for all seasons over the calibration or validation periods, covering several years.  
The authors mentioned in Table 1 that the selection of analogue candidate is limited to the 4 months centered around the target date for every year. However, it is not clear what season the presented analyses focused on.   
As explained above, the analyses are for all seasons over several years. It is true that the search for analogues is restricted to the 4 months centered around the target date, but the target dates cover all days over several years.  
Also, it seems to me that the entire assessment is performed in the prognosis context. The authors mentioned “prediction” several times throughout the paper. No matter for a 47-year archive (1961-2008) or reduced 25-year archive (1982-2007), it is not clear if the authors used part of the archive for calibration and part of the independent period for validation. If it is real “prediction”, what period of data is the prediction performed on? All these details should be clearly described in the method section.   
We now clarified the use of independent data in the section “2.3 Calibration of the analogue method”. It is indeed a prediction over a validation period, but still in the perfect prognosis framework, rather than a forecast.  
Second, the paper, especially the results and discussion sections, is not well structured. These sections are divided into many small sub-sections. The content should be better organized and integrated to convey clear message. One example is, the discussion of Figure 4 and 5 appears in both section 3.1 and 3.3.   
All the analyses were performed again with a more recent dataset. It allowed us to change the workflow of the study and then to restructure the paper to gain in clarity. Some results of secondary importance were moved to the discussion.  
Third, the text needs to be improved in terms of the logic, transition, grammar and wording. Some sentences are really long, confusing, and quite hard to understand (see some specific comments below).   
We tried to improve the language and better explain some points. The paper was corrected by Elsevier’s English editing services.

Specific comments:

1. P1, line 6-7: confusing sentence, how about “the main reason for the use off daily precipitation time series is the length of their available archives, …  
   This was changed.
2. P1, Line 7-9: “However, it is … at a different time of day”. Long and confusing sentence. should rephrase it.  
   This was changed.
3. P2, Line 22-23: “since they are based on observed situations with consistent spatial distribution” – consistent with what? Do you mean between target day and analogue dates? “as long as the analogue dates chosen for a region are the same” – same compared to what? When the target day changes, I think the analogue dates will change accordingly.  
   This sentence was removed as it is a bit out of context.
4. P3, Line 2: “even for much higher orders of magnitude” – do you mean even longer archive?  
   This was changed.
5. P3, Line 2-4: “Hopefully” – better to use “fortunately” based on context. Also, need reference for the statement “it appears that … 10° to 20°”.  
   This was changed and references added.
6. P3, Line 16-19: “Therefore, if the reduction of the archive … to an increase in performance”. – very confusing sentence, please consider rephrasing it.  
   This was removed.
7. P3, line 25: “similar conclusions” – what is the conclusions? - in creasing the prediction skill?  
   This was removed.
8. P5, line 11-12: why MTW can not be applied to the 2nd level of analogy?  
   Yes it can be applied to the 2nd level of analogy. Thus, this comment was removed.
9. P.6, line 9 for Figure 2: why not just keep candidate 24-h precipitation fixed from 6h to 30h, but allowed to choose the analogues on 6h, 12h, 18h, 24h, 30 h for both Z500 and Z1000? That allows you to choose the analogues on multiple time steps but within the 24-h window consistent with conventional method. What is the purpose to have the varying 24h precipitation totals if the main objective is to find the better analogues to predict the same target day precipitation?  
   This section was rephrased for clarity.
10. P6, Line13: confusing sentence “no constraint … in order to restrict.”  
    This was removed.
11. P6, Line 27- 33, it is not clear how the method is implemented. The authors should provide a diagram to show the method. More details are preferred, such as do you just pick one best grid among four, what time lapse is allowed, how the temporal profile of best proxy is used to disaggregate? If you use the proxy variables from NCEP/NCAR reanalyses, why not directly use the precipitation from NCEP/NCAR reanalyses?  
    The importance of this analysis was decreased and it moved to the discussion (Sect. 4.4), without providing all the details. It is indeed not the main message of the paper and brought some confusion. Some additional details on the time lapse and the non-consideration of precipitation were however added.
12. P7, line 13: Is the four points for geopotential height used to calculate the height gradient in both directions?  
    Yes. It has been clarified.
13. P8, Line 9: what does “globally significant” mean? Significant at what level?  
    This is not present anymore in the paper.
14. P8, section 3.2.1: It is not clear to me how the distribution of the analogy criterion for different analogue ranks is constructed. So for any target day, if 50 analogue dates are selected (50 ranks in total), each analogue date should have only one S1 value based on their similarity in geopotential fields.  
    Yes, each analogue date has a unique S1 value. These distributions for the different analogue ranks are obtained when applying the method on a long period. Then, for a considered analogue rank, we have multiple values of S1 corresponding to different target dates. We tried to clarify in the manuscript.
15. P9, line 6: “the number of candidate situations did not increase”, but from table 3, N1 for 2Z-2MI is larger than N1 for 2Z.  
    Yes, but here it was with the original parameters, so without an increase in N1. It has been clarified.
16. P9, Line 10-11: could this because RMSE is not a good metric to assess the similarity for moisture fields?  
    No, it is because the use of an MTW does not increase the sample size in this case, the second level of analogy only subsamples in the dates provided by the first level. A comment was added.
17. P9, Line 13: why it is “prediction”? I think the entire assessment so far is in a prognosis context. Do you reselect the analogue dates for blue bar (MTW algorithm) in Fig. 4 and 5?  
    We use the term “prediction” instead of “forecast” because it is not operational forecast, but we stay in the perfect prog context. Predictions were established for a validation period with independent data. Fig. 4 and 5 do not exist anymore and were replaced by Fig. 9 and 10 that are different.
18. P9, Line 15: It will be good to test if the improvements of MTW and MTW-r over the static approach is significant?  
    The approach changed and the calculation are now done on a calibration period and an independent validation period in order to validate the potential gains.
19. P9, section 3.3.1: Fig.11 also indicates that the spread of difference of the CRPSS performance score is quite larger. It is not correct to say that the performance score was improved for days with high precipitation. The statement should be based on the average performance. Again, for Fig.11, it is not clear to me what each point represents. Do the points represent the analogue dates with precipitation amount in the specific categories? Then the total number of pints in figures are equal to the total number of analogues selected?  
    We improved the description of the analysis. It is indeed a difference on the scores between the conventional approach and the MTW. All points correspond to target dates over the whole period. It means e.g. that over the whole period, the prediction performance of most days with a target value above 30mm/d was improved. The mean is represented by the star and then shows an average improvement.
20. P9, section 3.3 and 3.3.1: when author say “prediction skill”, does the author mean the use of calibrated parameters for independent data set?  
    Now we do use an independent dataset (the validation period, VP).
21. Same as #10, it would be good to show a map about the method 2 to help the reader understand what is concluded in section 3.4 and table 5.  
    As we wrote above, these results were removed to focus on the main results of the study.