Review of “The Analogue Method for Precipitation Prediction: Finding Better Analogue Situations at a Sub-Daily Time Step.” (HESS-2016-246) by Horton et al. 2016.  
  
This is my first time reviewing this manuscript, though it is my understanding that it was previously reviewed by others, with major revisions requested. I have not studied the comments of other reviewers, with the desire to provide guidance as independent as possible.  
  
My recommendation, with regret, is rejection. The article is needs significant reorganization and rethought, above what is required for a major revision. Some of my most significant issues include: (a) it provided detail of a minor improvement to a rather antiquated post-processing methodology; (b) it uses data sets for which there are better alternatives; (c) it doesn’t describe all of the procedures clearly; (d) it doesn’t consider other alternatives as controls against which to evaluate the methodology. In sum, I think the authors need to re-evaluate their research from top to bottom.  
All the analyses were performed again with a more recent dataset. It allowed us to change the structure of the study and then to restructure the paper to gain in clarity. The procedures were then better explained. The other mentioned issues are discussed below.  
  
Here are some more details on my most significant concerns.  
  
Poor organization. Journal articles are generally minor variations on a standard organization, with an introduction, data, methods, results, conclusion. Here are some of the issues with such sections:  
We improved the structure and simplified the overall logic of the study.  
  
Introduction. There is now a rather rich body of literature on the statistical post-processing to produce probabilistic precipitation forecasts. These include Gamma-distribution fitting methods, Bayesian Model Averaging, Extended Logistic Regression, and more. With the use of an older analog approach, one wonders why such an approach is considered, in the first instance, and how one should place in context the results to follow. Without a thorough review of other possible alternatives and some explanation of why your approach is being considered despite these others, the reader is left wondering why they should bother with continuing to read the rest of the manuscript.  
The present use of the analogue method is not a statistical post-processing of the precipitation. It is a statistical adaptation technique, which is classified as a downscaling approach (which can be a language abuse), as it provides a statistical precipitation prediction (or ev. forecast) based on large scale predictors describing the synoptic scale conditions. The precipitation output from the NWP model (or GCM, RCM, … according to the context) is not even considered in the method, as it would be in statistical post-processing methods. The goal of using an analogue method is to predict local precipitation based on the outputs of a global NWP model, without the use of a limited-area model. It brings an alternative approach that can complement the forecast provided by limited-area models in operational forecasting e.g., or the RCMs in the context of climate change impact studies.  
There was a European project (COST VALUE, http://www.value-cost.eu/) that aimed at comparing different downscaling methods. The results of the project are not published yet, but show that there is not an overall best downscaling technique, as the different techniques have some advantages and weaknesses in different characteristics. The analogue method was found to be strong in some characteristics and weaker in others. However, such comparison is out of the scope of the present study.  
As we show in the introduction, there are several articles published in the very last years using the analogue method, and we are aware of its use in different ongoing studies. Thus, the goal of the paper is more to provide an improvement to the users of the method. We refer to the paper of Maraun et al (2010) for alternative downscaling methods.  
  
Data. What information was being used for the forecast was not made clear, and this is crucial information to find out right away. See (2a) below for more.  
We added this information. By the way, the study takes place in a perfect prog context, not a forecasting context.  
  
Methods. Section 2.2 and 2.6 seem to be describing two different methods. There should be one, single, clear, unambiguous description of the methodology to be used.   
Section 2.2 describes the analogue method itself, and section 2.6 (now 2.3) how the parameters of the method are calibrated. We tried to make it more clear and restructured section 2.  
Also in this section, while verification using CRPSS is relatively standard, so are other verification diagnostics like reliability diagrams, which are not presented.  
Reliability diagrams were added (Figure 12).

Choice of data sets.   
  
Forecast data. What is used as the forecast, and why, is not clear. Is another analysis from the NCEP-NCAR reanalysis time series used as a surrogate for a numerical forecast? Is ECMWF, or other model forecast data used? I could not tell. If NCEP-NCAR reanalysis data is used, this then begs the question: why? This would not be a practical forecast methodology, where one needs information in advance of the event. If a numerical weather prediction forecast is used, then there are potential issues of forecast bias; the forecast model and the analysis may be different in character, leading to the issue of whether perfect-prog (your analog) type approaches are suitable or whether more model-output statistics approaches are needed.  
This study takes place in the perfect prog context. There is no forecast here, only prediction on an independent period (the validation period) using the reanalysis data. Then, when applied in real-time forecasting, NWP forecasts are considered. The issue of the biases are known and for this reason: (1) the model used for the NWP forecast and the one used for the reanalysis product should be as similar as possible, and (2) the main predictors are geopotential heights, which are robust and not too much model-dependent (this is more an issue for moisture variables). Moreover, the geopotential heights are compared in terms of their shape and not absolute value. Some MOS use of AMs also exist, provided that a long enough archive is available. In that context, the MTW could also be used and provide improvements. When applied in a climate context, AMs rely on GCM or ev. RCMs outputs. In this study, we stick to the perfect prog context and show an improvement that can then be applied to the forecasting context or to climate impact studies.  
  
Reanalysis data. There are more modern, more accurate reanalysis data sets available now such as ERA-20C, available at higher temporal resolution (3 hourly), which seems to be crucial for an article examining the usefulness of temporal shifts of the data. You dismiss this in your section 4.6 with an older reference, but I think given the focus in this article on temporal shifts, you need to reconsider higher temporal resolution reanalysis data.  
You are right that it is important considering a 3 hourly dataset. We have done all analyzes again using ERA-20C and MERRA-2 reanalysis datasets. Doing so allowed us to consider another workflow and to simplify the analyses and then the paper.  
  
Observation data. While the geographic details of the observation locations are definitely different from many other locations in Europe, there are still other locations in the mountains. Why not consider approaches that supplement the training data with other locations’ observations, potentially allowing you to get more without needing a very lengthy reanalysis? You may have objections to this, but at the least it would be worth explaining your choices.  
We are not sure to get your point. The AM exploit the intrinsic link between the synoptic-scale situation and the local precipitation. This relationship is very location-specific, as the meteorological influences will not be the same when considering a mountainous environment in another country. The location and characteristics of the driving elements, such as the low pressure centers and the fronts, will not be the same, neither the characteristics of the precipitation climatology. We can therefore not exploit remote data.  
  
Methodology. Finding some other reference methodology, e.g., Bayesian Model Averaging, would certainly be desirable. Another logical control would be unadjusted ensemble forecasts from the ECMWF ensemble prediction system. This would allow you to have a point of comparison against which to judge the analog methodology.   
As stated previously, the study does not take place in a forecast context, and thus cannot be compared to ensemble forecasts provided by NWP models.  
  
But even considering the analog methodology in isolation, one wonders why you chose the particular approach to selecting analogs. In particular, I found myself wondering why, for example, you didn’t use canonical correlation analysis approaches to determine what information in the reanalysis data set was most directly related to precipitation variability. Analog approaches, in my experience, are not very “efficient” with their training data; either a previous day’s data is selected as an analog date, or not. This, then, drops on the floor all the other data, which may yet have some useful information. To use an analog approach, then, it seems especially incumbent to have demonstrated that you have chosen the most important predictors that will be used in selecting past dates.  
Your version of the analogue method seems different to what we use here, so it is not easy to answer specifically to your points. However, here are some response elements:

* The choice of predictors is based on previous work (given as references) that made an intensive comparison of many variables. Moreover, the relevance of the geopotential heights and the moisture variables was shown by Horton (2012) to be the most relevant variables for the region of interest.
* A canonical correlation analysis cannot be performed here, as the geopotential heights are analyzed in terms of shape (Teweless-Wobus criteria) over a spatial window rather than absolute values.
* In the AM considered, the selection of analogue dates are not allowed within the same year as the target date. Thus, a selection of the previous date as analogue is not allowed.

In summary, I regret recommending rejection. I hope the authors will constructively use this feedback in the spirit intended.