Clim. Past Discuss., 11, C3111–C3118, 2016
www.clim-past-discuss.net/11/C3111/2016/

Climate of the Past Discussions
Open Access
EGU

Interactive comment on "Probabilistic precipitation and temperature downscaling of the Twentieth Century Reanalysis over France" *by* L. Caillouet et al.

L. Caillouet et al.

laurie.caillouet@irstea.fr

Received and published: 1 February 2016

The referee comments are recalled in italics and followed by the authors' responses.

The paper presents a novel downscaling technique to obtain daily temperature and precipitation data over France back to 1871. The authors use an analog re-sampling approach within an existing daily data set for France and discuss various strategies of how to choose the analogs in the best way (from the Twentieth Century Reanalysis data set) in order to obtain an accurate estimate of precipitation and temperature variability on daily to interannual time scales. The paper is scientifically sound and fits well into the scope of Climate of the Past. Similar approaches may be applied in other regions

or settings in order to extend our daily data record.

The manuscript is at various instances rather technical and it is not easy for the reader to always grasp the idea. Particularly, the reader is not well guided in the methods section. However, the method is promising. Overall the paper merits publications, but some revisions to enhance the readability will make it more valuable for the larger community.

The authors would like to thank Referee 1 for his/her positive evaluation and for his/her constructive comments and suggestions.

Major points:

- Better explanation of the methodological steps: I had to read the manuscript several times to be able to understand what is done. Please add at least one schematic figure that shows the individual steps. This would help the reader. Perhaps it might also help to express some of the steps in the form of equations.

We will add a synthetic diagram in an appendix to hopefully improve the overall understanding of the different steps of the analogue selection and their succession for all three methods developed in the manuscript.

- Daily vs. seasonal scale: The method provides daily output, but most of the evaluations as well as the given example refer to monthly or seasonal averages. This is a bit puzzling and the reader is left wondering whether the daily output is not useful (or, conversely, whether other methods would work equally well on the monthly or even seasonal scale).

The daily output is actually useful and is indeed currently used for daily continuous hydrological applications. As pointed out by the referee, the manuscript indeed focused on the annual and the seasonal time scale as they were the ones where improvements to the initial SANDHY method were first needed. Of course, further validation at the daily time scale has been carried out, but on the contrary to the seasonal regime it

CPD

11, C3111–C3118, 2016

does not allow discriminating between the different refinements and developments to SANDHY (First domain, Calendar and Stepwise), and the choice was made to not include them here. We may for example provide some additional plots showing the (very small) differences in cumulative distribution functions at the daily time scale for the two case study zones if the referee and the editor recommend it.

Minor:

Abstract: The performance on daily scales is not mentioned.

As the most part of the assessment and intercomparison of methods is done at a monthly/seasonal/annual time scale, we did not mention the daily time scale in the abstract.

P. 4428, l. 15-20: Have these studies produced long, continuous data sets or cases? Could this be seen as an advantage of this method?

Indeed, the studies referenced in the introduction have all produced long and continuous datasets. As mentioned in the manuscript P.4428 l.25- P.4429 L.4, the advantages and innovative features of our study are (1) the probabilistic aspect taking into account the uncertainties of the downscaling step; (2) the local optimisation of the predictor domains; (3) the application to a whole country.

P. 4430, l. 18: Were the data used 6-hourly or daily fields?

6-hourly predictors are actually used, and this is indeed not explicitly mentioned in the manuscript. Such information is however already given in Radanovics et al. (2013) and Ben Daoud et al. (2016), both referenced in the manuscript. This will be clarified in the main text, and the different and specific times of the day at which the predictors are retrieved will be added to Table 1.

Concerning T2m: Could also station observations be used here or would this make the procedure worse?

CPD

11, C3111–C3118, 2016

Indeed, a local station observation can be used, as it was done by Kuentz et al. (2015) for a single catchment . However, the only existing long-term station for Kuentz et al. (2015) was located outside of their case study catchment. In our context, it would be much more difficult to find different stations available since 1871 for each of the 608 zones paving France. Moreover, the temporal homogeneity of local observations is not guaranteed and using a predictor from the 20CR reanalysis (as for the other predictor for SANDHY) ensures a completeness and a relative homogeneity over the 140 years.

How was the seasonal cycle treated?

We are not sure what the referee meant by "treating the seasonal cycle". While some statistical methods use anomalies with respect to the mean seasonal cycle, raw daily predictors are used for all downscaling methods described in the manuscript, including SANDHY. As the objective is to reconstruct daily temperature fields, a corresponding variable (here the large-scale T2m) is chosen as a predictor.

P. 4431, l. 5: It is surprising that a monthly data set is used for obtaining daily data. I think this should be explained in a bit more detail.

First, the authors have to apologize for an error in the text of the manuscript: the interpolation that has been applied to generate daily series from monthly series is a spline interpolation, and not a linear interpolation as indicated. This will be corrected. To answer the above comment, to the authors' knowledge, no daily dataset that would cover the entire period (1871-2012) is available, while daily predictors are required for the analogue subselection. The NOAA ERSST v3b is one of the most recently released datasets, but with a monthly resolution. As the sea surface temperature varies slowly, a spline interpolation of monthly means seems rather relevant to obtain continuous daily time series.

P. 4431, l. 7: Safran is capitalized earlier on.

This will be corrected.

CPD

11, C3111–C3118, 2016

P. 4431, l. 21: Is daily 0-24 UTC or something else?

The WMO convention for daily precipitation measurement is used here: 6-6 UTC.

P. 4433, l. 17: Perhaps say a few words about the time of day.

The authors will add the information in Table 1. See also comment on 6-hourly fields above.

P. 4433, l. 22: "Previous applications ... considered all zones individually as target location" This implies that the presented approach does not. But in fact it does, as I take from P. 4434, l. 17 ("combining analogue days independently from one zone to another and from one day to the next"). This is a crucial point and should be clarified.

Indeed, the formulation is not correct. The sentence should read: “The spatial domain for a predictor where the analogy is looked for (spatial extent and position) should be defined and potentially optimized for each of the climatically homogeneous zones shown in Figure 2, considered as individual target locations, following Radanovics et al. (2013).”

P. 4434, l. 3-4: The procedure is repeated 5 times, but could results in almost the same selection of analogue days, right? So the 125 days are not 125 different analogue days?

This is right. The analogue selection is done independently for the five near-optimal domains. As a consequence, some identical analogue dates may be picked up through different predictor domains, resulting in less than 125 different analogue dates. The general idea is that the most relevant analogue dates will be picked up by several of the predictor domains.

P. 4434, l. 17 "combining analogue days independently from one zone to another and from one day to the next": This is crucial. So there is not necessarily a continuity from one zone to the next, but on the other hand you have probabilistic information. It would be very important, particularly for applications, to have a little more information here.

CPD

11, C3111–C3118, 2016

This is indeed right and is definitely not stressed enough in the manuscript. For example, the spatial coherence between zones is not guaranteed during the downscaling step and we are currently working on different and very promising methods to improve this coherence.

The reader might get the impression that this is the reason why the focus later is on the monthly or seasonal scale. Is this a statement you want to make? Should the product only be used on that scale?

This product may actually be used at different time scales – including the daily time scale – depending on the target application. Using a monthly scale balances the random sampling effect observed at the daily scale. We understand that there is a need to clarify this. The properties of the dataset described in the manuscript as well as suitable ways to actually use it will therefore be detailed further in a dedicated subsection of the discussion.

P. 4435, l. 14: How big are these domains? How different are they for neighbouring climate zones?

The domain size varies depending on the target location considered. More information about the size and the shape of these domains is given in Radanovics et al. (2013) (with some examples). The 5 domains are quite similar between neighbouring zones in most French regions. Specific climate borders like the Cevennes area (southern ridges of the Massif Central facing the Mediterranean) generate very different domains between zones from different sides, reflecting the different meteorological influences, in this case Atlantic and Mediterranean.

P. 4435, l. 20: I think "keeps the N2 analogues closest to the target calendar day" would be more clear

Thank you for the suggestion. The sentence will be modified accordingly.

P. 4436, l. 10: Again with respect to (monthly) SSTs: Is the ambition to include a

CPD

11, C3111–C3118, 2016

large-scale predictor that also captures longer time scales?

That was actually not the first objective. The SST indeed captures the seasonal variability that was missing in SANDHY precipitation reconstructions for the Cevennes. However, as mentioned in the manuscript (P. 4448 L25-27), having SST (but also the large-scale T2m) in the set of predictors allows including low-frequency temperature signals coming from either natural variability or anthropogenic warming.

P. 4438, l. 13: +/-60 days is quite a lot if the seasonal cycle is not removed. This will certainly affect the evaluation.

This is indeed right if one considers (1) the temperature variable and (2) the absolute value of the CRPSS. Here, the aim is only to compare the CRPSS between methods. The use of the CRPSS (and therefore of the climatological CRPS) was driven only by the need to normalize results among zones with different magnitudes of the variables (and therefore different magnitudes of the error represented by the CRPS). It allows having scatter plots in Figure 9 that would roughly fit in the [0-1]-[0-1] domain. In brief, choosing a rather crude climatological CRPS for temperature does not change the relative evaluation of the different methods.

But my question is more general: Why is the seasonal cycle not removed from temperature (or also GPH, omega, humidity)?

We are not sure what the referee meant by removing the seasonal cycle from temperature. If this comment refers to performance indicators, the above comment should bring a start to an answer. If it refers to the method for searching the analogue dates, and as mentioned in an earlier comment above, raw predictor values are consistently used in all steps of all downscaling methods. Indeed, the general aim of our downscaling approach is to find dates when predictor values (and not predictor anomalies with respect to an average seasonal cycle) are close to predictor values from the target date. We are quite unsure of the benefits that would be brought to results by removing the seasonal cycle in the predictors prior to the analogue search. It would anyway make

CPD

11, C3111–C3118, 2016

the physical interpretation of the selected analogue dates much more difficult.

P. 4446, l. 3: "Minimum, median and maximum" of what? Of the ensemble or of the daily precipitation sums within the month? The figure suggests it's the former but this is not fully clear.

This is actually "Minimum, median and maximum of the ensemble of the daily precipitation sums within the month", this will be clarified.

P. 4447, l. 21: I think the members are available, but maybe I am wrong.

To the authors' knowledge, the individual member outputs are not available for the predictor vertical velocity at 850 hPa.

References

Ben Daoud, A., Sauquet, E., Bontron G., Obled, C. Lang, M.: Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin. Atmospheric Research, 169, 147-159, doi: 10.1016/j.atmosres.2015.09.015, 2016.

Kuentz, A., Mathevet, T., Gailhard, J. Hingray, B.: Building long-term and high spatio-temporal resolution precipitation and air temperature reanalyses by mixing local observations and global atmospheric reanalyses: the ANATEM model. Hydrology and Earth System Sciences, 19, 2717-2736, 2015.

Radanovics, S., Vidal, J.-P., Sauquet, E., Ben Daoud, A. Bontron, G.: Optimising predictor domains for spatially coherent precipitation downscaling. Hydrology and Earth System Sciences, 17, 4189-4208, doi: 10.5194/hess-17-4189-2013, 2013.

Interactive comment on Clim. Past Discuss., 11, 4425, 2015.