The article presents a dataset of reconstructed daily meteorological data. Gridded precipitation, temperature, and reference evapotranspiration were reconstructed over 142 years by means of an analogue approach named SCOPE. The article presents the construction of the dataset as well as different welcomed quality evaluations. The dataset can find various useful applications. The figures are of very good quality and the results are interesting. This is an original work worth being published after consideration of some corrections.

The authors would like to thank Referee 1 for his/her strong positive comments on the manuscript. We also thank him/her for the specific and technical comments (in italic below) that will lead to improve the manuscript. The detailed answers to the specific comments are presented below.

• *It was highlighted in different studies that 20CR has larger errors that other reanalyses, which is expected due to the small amount of data assimilated. This will have an impact on the results of the analogue method. Of course, to cover this period, you have no choice but to use 20CR. It is stated that using another reanalysis would results in different predictions, but you should discuss the consequences of using 20CR on the quality of the prediction. This dataset aims at reconstructing accurate past meteorological conditions, not only statistically correct, but also with the correct chronology (am I correct?). Has it been compared with a reconstruction using ERA20C on a shorter period?*

Correct chronology is indeed aimed at in SCOPE Climate. Literature findings about inhomogeneities of 20CR predictors and associated uncertainties that may affect SCOPE Climate are provided in Sect. 6.1 of Caillouet et al. (2016). Inhomogeneities have been found in the ensemble mean of 20CR before the 1930s, because of the too few data assimilated in this period. Consequently, SCOPE Climate may be less accurate before the 1930s. Some sentences will be added in the discussion section to specify this. ERA-20C has been released at the end of our work, so no comparison between these two global reanalyses has been done. Nevertheless, Dayon (2015) provided a quick comparison of ERA-20C and 20CR for downscaling purposes, only mentioning lower scores for spring temperatures reconstructed with 20CR in comparison to ERA-20C. A more recent study focused on the impact of the driving reanalysis on downscaled precipitation over Switzerland with various analogue methods, including SANDHY (Horton et al., 2018)

• *Safran (& 20CR): comment on the quality of the products. What are the known errors and uncertainties? What can be their impact on the final product?*

See the above response for 20CR. Vidal et al. (2010) performed a detailed validation of the gridded Safran dataset with both dependent and independent data. They showed that the errors on precipitation are low and constant over the 1958–2008 period. Errors on temperature are decreasing with the increasing number of available surface observations. This will be added to the main text. We therefore believe that possible temporal inhomogeneities of SCOPE Climate mainly come from 20CR, as possible inhomogeneities in Safran are smoothed out through the multi-member analogue resampling.

• *Different periods: we get a bit confused with the time periods (archive, target, calibration). Can it be summarized clearly? Is there a period for independent validation?*

This will be rephrased. As there is no calibration period, there is no validation period. SANDHY needs two periods: a target period, which is the period to reconstruct, and an archive period, which the period when the meteorological situations are picked up to reconstruct the target period. Large scale reanalysis (20CR) should be available on both periods whereas local scale reanalysis (Safran) should be available on the archive period. For a date in the target period, large-scale situation will be analysed. The most similar situations will be found in the archive period and correspondent local meteorology will be used to reconstruct the meteorology of the picked up date in the target period. To produce SCOPE Climate, target period is 01/01/1871-31/12/2012 and archive period is 01/08/1958-31/07/2008. Moreover, in SANDHY, geopotential analogy domains are optimised with an algorithm of growing rectangular domains over the 1982-2002 period. SCOPE Climate has been validated using homogenized time series on the 1900-2000 period.

• Sect. 2.2.2: Please restructure the section. It starts with "The stepwise subselection . . ." as we are supposed to know about it, but it is explained in the next paragraph. The definition should come earlier.

Indeed, this will be restructured.

• Sect. 2.2.3: When you remove a precipitation analogue and duplicate another one, do the same happen with the temperature and the ET for the same dates? If not, how do you keep the physical consistency between variables? Please specify.

Indeed, as analogue days are the same for precipitation and temperature, if the day is removed after correction of the precipitation bias, it will also be removed for temperature. Physical consistency is therefore kept. This will be added to the text.

• Sect. 2.2.4: Please explain if the reordering is the same for all variables (P, T, ET) when the ensemble members are reordered, so that their ranks stay consistent. If so, is the order based on the precipitation and applied to the rest? If not, as previously, how do you keep the physical consistency between variables? Please specify.

For the Schaake Shuffle part, reordering is done independently for each variable. Nevertheless, observed rank correlations are derived from the Safran multivariate meteorological fields and applied to the reconstructed ensemble. It thus ensures a spatial and intervariable coherence of any single ensemble member. This physical consistency is therefore at least as good as in Safran.

• Sect. 3.1.1: Do the analogues to 1910 correspond to other dates with flood events? It would be interesting to know.

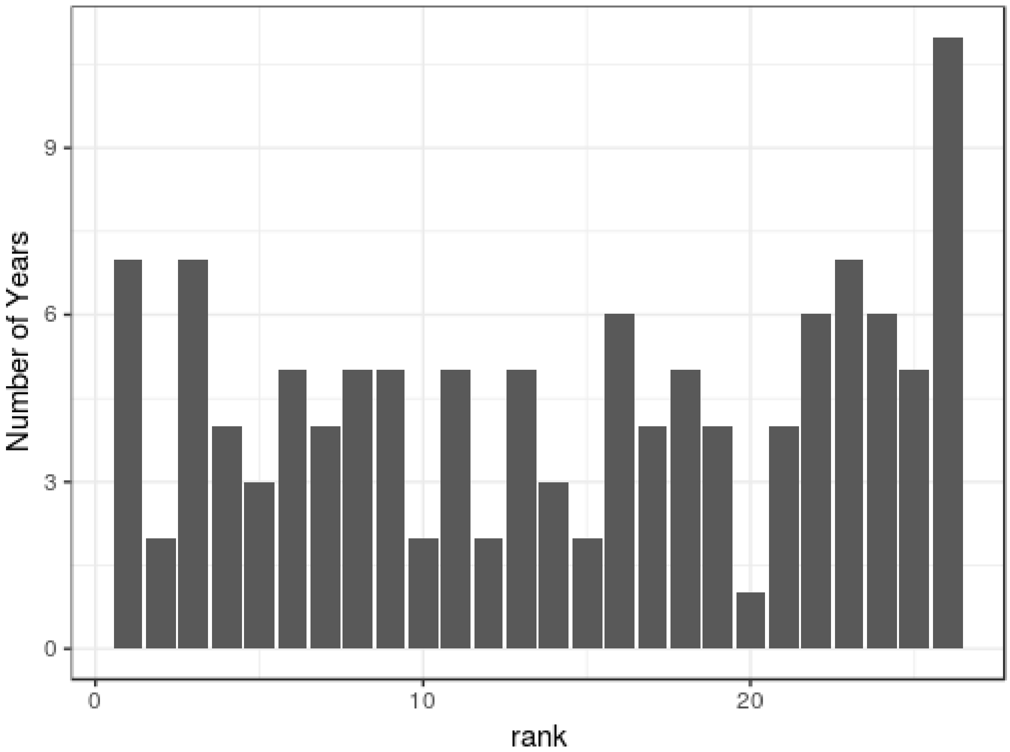
This would indeed be very interesting to look out. This is presumably the case, as it is the basis for analogue downscaling, provided that analogue dates are sufficiently grouped in time to generate a flood on the very slow-reactive Seine basin. Such a verification is of course outside of the scope of this study.

• Sect. 3.1.2: Comparison to the station precip: It should be mentioned at the beginning of the work if Safran's gridded precipitation is point precipitation or areal mean precipitation. In case of areal mean precip, comment on the fact of comparing areal mean and station precip.

Indeed, it is mentioned p.2 l.29 that Safran is a gridded product at an 8-km resolution, representing mean precipitation over the grid cell. Reality is actually a bit more complicated than that: gridded precipitation (and other variables) in Safran are values at the altitude of the grid cell, and an average over the climatically homogenous zone containing the grid cells (see Vidal et al., 2010 for details) is done. Comparison between gridded products like Safran or SCOPE Climate and station data are of course not fair. This fact will be added to the text.

• Sect. 3.1.2: On the plots of Fig 5, the precip seems to be under-dispersive in summer. It would be desirable, when concluding that the observation falls well into the range of SCOPE climate, to support it with rank histograms.

Summer precipitation from SCOPE might be under-dispersive, but rank histograms are not necessarily well suited in this particular case, given the low sample size of years. Below is plotted the rank histogram for summer precipitation with respect to the homogenized series. It is very noisy and interpreting it is not straightforward. We would therefore prefer not to include such rank histograms in the manuscript.



• Sect. 3.2.1: How did you select the four different cell? Is this setting representative of the rest of the dataset?

The four different cells have been selected in four zones with different climatic influences. Oceanic for Finistère, mountainous for Haute Savoie, meridional oceanic for Corrèze and Mediterranean for Cévennes. Regimes from different cells in these zones have been analysed and the ones being the most different from each other have been kept. Not all the cells in France have been studied but we believe that these 4 cells represent a good subset of the rest of the dataset.

• Sect. 3.2.1: Your text sounds like the precipitation in Fig. 8 is good, when you are actually missing most of the main events and are producing peaks when no precip was observed. You state that the sequence of dry and wet periods is well represented, and the bias was fixed. However, if the actual chronology is not accurate, can users really use the dataset to analyse past events, or should it rather be used as a climate simulation (not real chronology)? Is Safran a reliable reference here? Please better discuss the results.

Well, the figure in its present state may be misleading in that respect, and we'll be working on adapting it to better convey the actual performance of SCOPE Climate. The actual figures on wet/dry days with a threshold of 0.1 mm are the following, for the year 2011 and this cell: 131 potentially wet days in SCOPE (with at least one member being wet) out of 141 (i.e. 92 %), and 221 potentially dry days in SCOPE (with at least one member being dry) out of 223 (i.e. 99%). So one may say that the chronology is rather accurate. On the question of Safran being a reliable reference: Safran is the gridded reference product over France, with its advantages and limitations of being a gridded product. In this specific comparison, the comparison is rather fair, as we compare two gridded products, SCOPE being a resample of Safran data from the 1958-2008 period.

• You try by different ways to reduce the selection of analogues from other seasons. This exchange of seasons causes problems with ET (P18 L7-11). What about coming back to a fixed calendar preselection (moving temporal window)? Does the preselection on temperature really justify adding such complexity to the method (SST and T2m) and having issues with ET? I do not expect a full analysis on this, but it should be discussed.

Indeed, coming back to a seasonal calendar would fix issues about evapotranspiration. Nevertheless, Ben Daoud (2011) showed that allowing a selection of analogues in other seasons for SANDHY is leading to improvements in the downscaling of precipitation. Moreover, setting fixed seasons would not be necessarily ideally suitable for our case – a long-term historical reconstruction – as there could be season shifts. Additionally, tests have been done in a previous paper for having a calendar subsetting in lieu of the stepwise subsetting, and we showed that the stepwise selection leads to higher rank correlations and lower mean errors between reconstructed and observed temperatures (Caillouet et al., 2016).

Technical corrections

• P1 L15-17: Long sentence. Please rephrase.

This will be rephrased.

• P2 L1-4: Please rephrase.

This will be rephrased.

• P2 L8: Be more specific

This will be rephrased.

• P2 L13: Mentioning "one" of the resulting dataset let us wondering what the others are. Are they equivalent climate reconstructions? If so, why is this one better?

The other resulting dataset is SCOPE Hydro. It has been created using SCOPE Climate in a hydrological model to provide a reconstruction of streamflow for France over the 1871-2012 period. Another paper will make SCOPE Hydro available.

• P2 L15-16: "appropriate space and time resolutions for hydrological applications": it depends on the catchment size and the goal of the application! As any other dataset, it cannot fit all purposes (e.g. flash floods). Please be more specific on which applications are possible.

This will be specified.

• P2 L18: "the most general choices": what kind of choices? Be more specific.

The methodological choices made to create SCOPE Climate (sub selection, bias correction, Schaake Shuffle) never favorited the reconstruction of specific types of events, such as floods or droughts. This will be clarified.

• P2 L18-21: This paragraph sounds more like a conclusion than an introduction.

It appeared important for the authors to highlight the innovation introduced by SCOPE Climate in the introduction

• P3 L16: ". . .spatially interpolated on the 2.5 deg. grid required. . ." How did you do the interpolation? Why is it required?

It is required as the SANDHY method had been set-up for inputs at 2.5 deg grid. The interpolation is the commonly used bilinear interpolation.

• P4 L7-9: Not clear. Please rephrase.

This will be rephrased.

• P4 L12: Please rephrase.

This will be rephrased.

• P4 L13: "four analogy levels": Better, explain that these are consecutive subsampling steps.

This will be added.

• P4 L14: 4-day window: is 2 days sufficient for the independence of the geopotential height?

XXX

• P5 L2: "independently for the 608 climatically homogeneous zones": What do you mean? Please be more specific.

This means that the SANDHY method is applied 608 times, each time corresponding to one zone. There is no spatial consistency between zones, the latter being added with the Schaake Shuffle process.

• P5 L6: Improve in what aspect?

This is specified l.7, to improve the fact that precipitation is over estimated in spring and under estimated in autumn for Mediterranean areas. As suggested before, this section will be re-ordered.

• P6 L11: It is not clear when stating "the lowest precipitation" if zeros are included or not.

They are included. This will be added.

• P6 L12: "resampled": are they duplicated?

Indeed, they are duplicated.

• P6 L13: How is the value of N chosen? Why is three the maximum?

Median rank of Safran precipitation in the range of SCOPE Climate was 0.55 before removing the bias. To get a rank back at 0.5, this meant removing N=1.25 days. Tests have been done by removing N=1, N=2 or N=3 days for the entire France. Results show that removing these numbers of days -- depending on the zone in France -- allowed retrieving an annual precipitation bias around 0%. Removing more than 3 days for the zones with N=3 would have transformed the slight positive bias in slight negative bias.

• P6 L27: Why "Julian day"?

This is simply the day-of-the-year.

• P7 L12: Specify which region

This will be specified.

• *P10 L7: “heavy amounts”: Please rephrase.*

This will be rephrased.

• *P16 L3-4: Not clear. Please rephrase.*

This will be rephrased.

• *P17 L11 – P18 L1: Please explain.*

Annual precipitation is under estimated, this is the reason why the bias correction consists in removing dry days in the 25 ensemble member of SCOPE Climate. Summer, autumn and winter are indeed under-estimated. Nevertheless, the spring season shows an overestimation of precipitation. This means than removing dry days in spring will potentially exacerbate the precipitation overestimation during this season. This could be managed by adapting the number of dry days to remove depending on the season, as it is done for the zones. Zero dry days would be removed in spring, whereas 1, 2 or 3 days will be removed in the other seasons. This has not been done for the sake of parsimony as the number of days is already adapted to each of the 608 zones.

• *P18 L21: The CRPSS is normalized by the climatology, not the CRPS.*

This will be corrected.

• *P18 L22: Which climatological reference did you use?*

The climatology is calculated using data from Safran over the archive period (1958-2008). Meteorological ensemble data is selected randomly from ±60 days around the target date to take seasonality into account.

• *P18 L31: What is responsible for the irregular patterns and the negative CRPSS at the annual time step?*

These patterns interestingly come from temporal non-homogeneities in the Safran data. More precisely, specific zones with negative CRPSS show temperature trends that are not consistent with surrounding areas, and even sometimes negative over the last decades. This results from changes in the observation network, with more stations being installed at higher altitudes, hence with a lower mean temperature. As shown by Vidal et al. (2010), Safran is not suited for trend analysis. These irregular patterns thus therefore suggest that interannual variability and trends are more spatially homogenous in SCOPE Climate than in Safran. This is however still to be shown.

• *P19 L6-7: Please rephrase.*

This will be rephrased.

• *P21 L30-33: Not clear. Please rephrase.*

This will be rephrased.

References

Ben Daoud, A., Sauquet, E., Lang, M., Bontron, G., and Obled, C.: Precipitation forecasting through an analog sorting technique: a comparative study, Adv. Geosci., 29, 103–107, doi:10.5194/adgeo-29-103-2011, 2011.

Caillouet, L., Vidal, J.-P., Sauquet, E. and Gradd, B.: Probabilistic precipitation and temperature downscaling of the twentieth century reanalysis over France. Clim. Past, 12, 635–662, doi: 10.5194/cp-12-635-2016, 2016.

Dayon, G.: Evolution du cycle hydrologique sur la France au cours des prochaines décennies. PhD thesis, Université Paul Sabatier, Toulouse, 2015.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, Int. J. Climatol., 30, 1627– 1644, doi:10.1002/joc.2003, 2010