Review ESSD-2018-79, reconstructed met data for France

Authors and others have published prior applications, authors themselves have published a separate prior description of SCOPE. Now they provide a careful well-written well-justified description of the SCOPE data product. Sets a good example: country-specific re-analysis followed by careful downscaling of a global reanalysis to reconstruct a long high-resolution meteorological / hydrological record. Useful and necessary expansion of Appendix B in 2017 HESS paper (as the authors explicitly state in line 29 on page 3 of this manuscript). Propose as a 'remedy' to sparse - in duration and location - data, but of course the reanalyses themselves, e.g. 20CR or ERA-20C, depended originally on the same sparse (in space and time) observational networks. Description of the downscaling probably belongs in a different journal but in this case the authors took the initiative and opportunity to validate against, e.g., long local time series.

Two potential uses: for research on hydrology in/over France and, somewhat neglected, as an example for other similar applications in other countries. How would or will SCOPE work in data sparse areas, e.g. Canada or Russia with interesting and vital hydrology in the frozen north but with data and research focus on small agricultural areas of the south? Or Brazil or China for similar reasons - rare reliable data time series from mostly-urban locations needing temporal and spatial extrapolation to larger areas covering a range of land surface types? (Additional similar comment below.)

Overall a good description of an impressive effort. Recommend publication subject to some modest changes / improvements.

Specific comments.

Page 2 line 18: Here the authors use the word 'general' - "The most general choices have been made …" 'General' in this context can unfortunately imply casual, or avoiding specific information. I think the authors mean 'broadly relevant', e.g. that they carefully and intentionally constructed this product to serve a wide range of research users. Replace the word 'general' to better describe their intent?

Page 2 line 19: "the ensemble aspect". The authors clearly intend this as an advantage of this product but for an observational audience in ESSD, they may need to elaborate about those supposed advantages. Later they compare random-selected ensemble members and in one case focus on one specific ensemble member. They even show statistical uncertainties clearly derived across 25 ensemble members, so they clearly regard the ensemble aspect as an asset. Authors need to share their confidence a bit more explicitly here?

Page 2 line 21: "… provides homogeneous time series that will ensure the spatial consistency required for all studies." These authors may themselves 'require' spatial consistency for their own work and may hope that other researchers follow this example. However, one can imagine a range of applications and publications based on this product that will not need or acknowledge consistency with other uses imposed by shared use of one product. Rather than 'required', I think the authors mean 'encouraged' or 'enabled'? Also, spatial consistency in this case derives from 'spatial' homogeneity but as written the sentence allows confusion between spatial and temporal homogeneity and consistency, e.g. a homogeneous time series ensures spatial consistency?

608 climate zones but across the area of France that represents an average climate zone extent of roughly every 30x30km? Thus, at 8 km resolution, perhaps 12-16 grid points in an average climate zone but more likely only one or two grid points in small alpine zones with perhaps 50 grid points per broad climate zones of the Atlantic coast, Mediterranean coast or central agricultural areas (e.g corresponding to the HER regions Armoricain, Mediterraneen or Tables calcaires in Appendix C of the HESS paper)? But only 22 HER (Hydro-ecoregions) used in the HESS paper? Where did the 8 km resolution come from: computational limitations, geospatial considerations? Why focus on 608 regions here but only 22 in the hydrology application? What advantages?

Page 2 line 28 and following: Confusion about Safran temporal extent. Here we read " August 1958 onwards" which implies up to present day. At the end of the paragraph, however, we read " August 1958 to July 2008". But, if Safran relies on "first guess from the ERA-40 reanalysis",

ERA-40 covers only to 2002? Later, on page 4 line 29 we read "using the August 1982-July 2002 period". But, two lines later we again read August 1958 to July 2008. Later still, including in Figure legends, a reader sees 1958 to 2008, 1958 to 2007, 1958 to 2002, etc. At one point later the authors present a set of 2011 data, described as outside of Safran but valid for comparison with Safran? Please can the authors specify the precise data sources and assimilation processes of Safran and consequently its exact temporal extent?

Page 3 line 6 and following: here a reader learns that the time extent of SCOPE comes from the time extent of V2 of 20CR, e.g. 1871 to 2012, because (as stated on Page 4 line 8) "the large-scale reanalysis is then the only dataset that must be available over the period to reconstruct." This information should have come earlier, to justify the SCOPE time period? Authors correct about 20CR depending on SLP but 20 CR V2 also assimilated SST and perhaps sea ice?

Page 3 line 21 and following: the authors used only a single Atlantic SST? Given the maps and the inclusion of Corsica, why did they not also include a Mediterranean SST data point? In the phrase "optimised grid cell", what does 'optimised' mean? I understand why they include an ocean predictor but I do not understand why only this one? Perhaps the only one with a long-enough time series record, and even then they had to interpolate from monthly to daily? So far as I can tell, Appendix B of the HESS paper (cited in line 23) makes no reference to ocean data, no reference to monthly to daily interpolation, and no mention of SST as a valid predictor? Later (line 1 of Page 6) the authors justify this single SST point based on "consistency over France and parsimony of parameters" but additional SST points would not change the number of large-scale predictors and subsequent importance and uncertainty assigned to Mediterranean regions and convective precipitation suggests perhaps the benefit of including Mediterranean ocean SST. No longer possible for this data product, but something to recommend for next version or future efforts?

Page 5 lines 4,5: SANDHY set up for a single predictand, precip. Authors want to extend SANDHY to two additional predictands, temperature and evapotranspiration. Therefore, authors developed the following steps, e.g. 2.2.2, 2.2.3 and 2.2.4? Step 2 (SANDHY-SUB) adds SST and T2M as predictors? Step 3 corrects a 10% dry precip bias by iteratively removing driest analog years (?) followed by resampling? Step 4 involves a rank correlation shuffle within ensemble members to improve spatial cross-correlation but because of shorter archive time period compared to longer target period, the shuffling can only improve spatial correlations, not temporal correlations? That fact that I needed to write this summary for myself suggests that the authors, in their accurate and detailed sequential description, have left this reader a bit in the dark about how each step leads to the overall goal. Perhaps in the final paragraph of the Introduction or as the initial outline of Section 2?

Page 6 Figure 2: is identical to Figure B1 from the HESS paper and should be cited as such?

Page 16 Figure 9 (and applies to other figures): Authors mix percentage differences for precip and evapotranspiration with absolute differences for temperature. But, guessing at an average temperature for France of 10C, 0.2C difference would represent 2%, much better than but now at least in the same units as precip and evapotranspiration? The authors have some reason for using this mixture of relative and absolute units?

Page 17 line 4,5: Valid caution here about application of SCOPE evapotranspiration to specific events but this caution should appear or should also appear in the data limitations section, e.g. page 21 starting from line 25? That paragraph seems to hint at this issue, but this statement provides a more specific example.

page 17 line 7,8: Here the authors provide precip bias estimate "median of annual precipitation bias between Safran and SCOPE Climate shows an absolute value under 5% for the entire France , …". But on Page 6 line 16 the authors claimed "retrieve a near-zero bias in mean interannual precipitation over France." Respectable results in either case, but do these two statements coincide or differ?

Page 17 line 10,11: Confusing! Does this statement "… spring is the only season when there are a few dry analogue dates after SANDHY-SUB …" indicate that the subsetting removed dry days

from the spring, leaving fewer dry days and a larger proportion of normal or wet days? Or, does the statement indicate - as written but seems unlikely - that only spring has a few dry days? Please revise and clarify.

Page 22 line 6: convective precipitation. For this reader, the authors have offered several hints throughout the manuscript of this combined temporal, resolution and geographic weakness for convective precipitation, vis. weaknesses in Mediterranean regions, weaknesses in abrupt topographic regions, weaknesses in Atlantic SST as applied to Mediterranean coastline, weaknesses in autumnal precip and temperature patterns. These various factors, if highlighted and combined here, would add some specificity to the convective precip issue? Rather than weakening the outcome, identification of this process as difficult would in fact strengthen the reader's sense that the authors know their product and their French geography. A sentence about this difficult-to-resolve issue should appear in the abstract?

Somewhere, probably in the conclusions, the authors could / should include a sentence or two about application of the SCOPE approach, and specifically of the three 'improvements' imposed after the SANDHY step, to other settings. The authors have noted the positive aspects of high-resolution time series available for the UK. Could they here make a comment about the opposite situation: larger areas with scarce or no data? Could they or anyone who wishes to reproduce their effort for other regions even imagine this work without the existence of Safran? How would a country of large area and broad range of land surfaces develop information and skills necessary to repeat a SCOPE-like reconstruction? Works nicely for France. Could it work elsewhere?