Reply to the reviewers comments

The manuscript presents a gridded reconstruction of daily precipitation and temperature in Switzerland over the last 140 years approximately, based on available but sparse station observations. The methodology is based on the Analogue Resampling method , with post-processing applied Quantile Mapping and Ensemble Kalman Filter. The study is indeed very interesting – the idea of combining the Kalman Filter and the analog method is I think novel. The applied methodology is valuable. The manuscript is generally clearly written and well structured. Therefore, I m happy to recommend the manuscript for publication after some revisions, which I hope that the authors may want to consider.

We'd like to thank the reviewer for this positive feedback and for all the helpful comments and suggestions to improve the manuscript.

General comments:

1) The manuscript discusses at length the success and deficiencies of the reconstructions, both with the ARM and the post-processed reconstructions. This discussions is focused on the replication of the mean, variability and extreme events. I have one general comment in this regard. The ARM using just one analogue is in principle unbiased and should also replicate the correct variance, since it is simply a re-sampling of observations. Therefore, deficiencies in the replication of aggregated statistical measures, such as mean and variance, found in the same 'pool' period 1961-2017 can only be originated in the predictand field, the girdled temperature and precipitation products. (Of course, the skill in replicating the temporal succession and extremes is another question). Thus, the evaluation of the ARM by the leave-one-out method is actually not only a validation itself but also in combination with the gridded temperature and precipitation fields. Since the construction of these fields always involves some sort of regression or averaging of station data, the extremes and in general the variability is reduced compared to station data.

This is an important point. We agree with the reviewer that in principle, the ARM using only the best analogue has the advantage over e.g. simple interpolation methods to reproduce natural variability and mean values. However, as the reviewer points out, certain methodological choices as the coupling of temperature and precipitation reconstructions or also the application of a distance measure over all station data can result in reduced variance and biases. The best analogue thus represents a best compromise to optimally satisfy all criteria described in section 3.1 of the manuscript. Some causes of reduced variance and bias (limited size of the analogue pool, availability of station data) are discussed in the manuscript. In section 3 of the revised manuscript, we will state more clearly that the ARM generally should be unbiased and reproducing natural variability and we will add a few sentences on possible impacts that methodological choices may have, limiting the capability of the ARM to do so.

2) I understood why the station predictor data need to be de-seasonalized and standardized, as temperature and precipitation have different variation ranges. However, I did not understand why the gridded predictand fields also need to undergo this preprocessing. In theory, once the ensemble of n analogues is identified, the same days can be selected from the pool of un-preprocessed predictand fields. Perhaps, the Kalman Filter algorithm requires that preprocessing, but it is not obvious to me. A short explanation, if that is the case, would help the reader.

Thank you for this remark. Station data are standardized and temperature measurements also de-seasonalised. To the gridded data, however, no standardisation is applied, but only a de-seasonalising of temperature fields. For precipitation, the best analogue dates are selected directly from the pool of un-processed predictand fields (absolute values). As analogues are calculated using temperature deviations from a mean seasonal cycle, reconstructed temperature fields are accordingly taken from pre-processed gridded data (temperature anomalies). The mean climatology

is then added again to get absolute temperature values. This procedure might not be entirely clear from the formulation in section 3.1 and we will try to clarify it in the revised manuscript.

3) Through the manuscript, especially in the beginning I had problems to figure out which data are the ‘predictors’ and which the ‘predictand’. It becomes clearer later in the manuscript, but perhaps the authors would like to use this terminology or a similar one from the start. It will help those readers that are not that acquainted with the analog method
Thank you for this suggestion. For the analogue method, the spatial fields of a given day of interest is the predictand and all data used to look for the best analogue (station data, weather types) are used as predictors. In the revised version, we try to make this distinction clearer for better understanding.

Particular comments:

Some refer to the English usage, but I am not a native speaker, so the authors may want to double-check

Thank you for these comments. We changed the order of the reviewer's comments in the following to first answer remarks regarding language and then go into detail with comments regarding content and understanding of the manuscript.

line 3 ‘ whereas prior to that local station observations ‘
the sentence is ambiguous : whereas prior to that year, local stations observations..

line 115 The day of interest and possible analogue days are required to be of the same WT to assure similar synoptic-scale
to ensure

Line 338: From this, we can conclude, that reconstructions provide accurate precipitation fields for low to moderate precipitation
delete comma after conclude

Thank you for these suggestions. We will correct the errors and adjust the wording in the manuscript for better understanding.

line 37 ‘The analogue approach makes use of this statistical relationship between large-scale and local weather or meteorological patterns, while the former is used to predict the latter. ‘
what is ‘the former’ and which ‘the latter’ ?
This is a good point. In fact, the analogue method can be applied in both ways: for downscaling large-scale weather data to a local scale, as well as to predict large scale weather data from local scale information. As in the introduction we want to keep the description of the method general, the wording will be changed in the revised manuscript as follows: "The analogue approach makes use of this statistical relationship between large-scale and local weather or meteorological patterns, while one can be used to predict the other". Further details are given in section 3.

Line 81: ‘Errors are estimated to be in the order of factor 1.7 for precipitation on) and 1.3 for precipitation above the 90% quantile.
I guess units are mm/day
Thank you for this comment. In the description of the RhiresD dataset (MeteoSwiss, 2016a), standard errors of the dataset compared to local point observations are indicated to be in the order of a factor between 1.3 to 1.7 (dimensionless).

line 104 data to predict the spatial fields and a record of the spatial data from which the reconstructions are drawn. the spatial fields we used daily station observations, while the RhiresD and TabsD datasets for 1961–2017 from M
I would set here which are data are the predictors and which the predictands. Many readers would refresh their understanding of the method by going directly to this section
Thank you. In accordance with comment 3), we will clarify the terminology in the revised manuscript.

line 158: where x denotes the updated state vector (analysis), x and y as described above and K is the Kalman gain or innovation matrix calculated from the ensemble. In this and the following equations, H describes the Jacobian matrix of H(x) and extracts
I am not sure this is the Jacobian matrix. In my undesrstanding the Jacobian of a vector function of several variables is constructed by taking the partial derivatives along the vector dimensions. Here, I think the authors mean the projection operator or the selection operator
Generally H(x) is the symbol for the operator and H is the Jacobi matrix of the H(x), so Jacobi matrix would be correct. However, in this particular case, H describes indeed a simple selection operator, as no transformation of the data is done.

line 191 I think that QM becomes necessary because of the use of the Ensemble Kalman Filter. The ARM (best analog) would deliver the correct pdf (unbiased, correct stdev, etc). Also an ARM based on an ensemble of analogues would need QM. Is that true ? Perhaps the authors may want to discuss this point.
Thank you. The Ensemble Kalman Filter is only applied as post-processing of temperature fields, but was discarded in favor of the simpler method of quantile mapping for precipitation. The necessity of post-processing of precipitation is mainly related on the reasons discussed in the answer to the reviewer's comment 1).

line 340 ‘Extreme events, however, are underestimated by ARM reconstructions and show large errors also for post- processed data. As extreme events by definition occur more rarely, the number of suitable analogues is limited. As argued in for upper and lower extreme values. In general, errors could be significantly reduced with Kalman fitting. The average bias reveals, that while analogue reconstructions tend to overestimate negative extreme values and underestimate extremely high’
I am not sure that I completely agree (see previous comments). The ARM (best analogue) would automatically produce the correct pdf, It would miss extremes, and produce them at the wrong point in time, but the pdf should be the best possible (it is simply a re-sampling of the observations). I agree that the Kalman filter, and in general an ‘ensemble ARM’ would produce smaller RMSE at the expense of reduced variability, but trade-off belongs to the general statistical trade-off between bias and variance of an estimator.
Thanks for this point. Referring to the authors' response to comment 1), there are certainly various reasons that limit the capability of the ARM to correctly reproduce the correct pdf, which will be discussed in more detail in section 3. As shown in figure 6 (over space) and figure 7 (for stations over time) such deviations between ARM reconstructions and observed distributions exist and can be associated with upper (and for temperature also lower) extreme values. Lines 340ff (figure 6) refer to the validation over space; in this particular case, also uncertainties originating from a sparse station coverage play an important role. While the ARM assuming an unlimited pool of possible analogues would produce a correct pdf over time, uncertainties regarding spatial patterns in regions without measurements would still persist.