**Climate
of the Past
Discussions**

# *Interactive comment on* "Reconstruction of high resolution atmospheric fields for Northern Europe using analog-upscaling" *by* F. Schenk and E. Zorita

**F. Schenk and E. Zorita**

frederik.schenk@hzg.de

Received and published: 21 June 2012

Response to Reviewer #2

We thank referee #2 for his/her constructive and thorough review of our manuscript. We reply to the comments as follows:

Referee:

"I think the authors are not being careful enough in the management of the references."

Response:

We accidentally used an older reference list when compiling the paper. We updated four missing references.

Referee:

"Section 2.3. Perhaps I have missed it, but I do not see clear how do the authors arrange for the reconstruction of monthly data. Do they compute monthly averages of predictor (SLP) and use them to find analogs of monthly averaged predictand? Or do they use daily predictors to compute daily predictands to get after this step the monthly average of the predictand? The authors explain the method quite clearly for temperature, but it is not clear for me how do they proceed with other variables, such as the ones shown in figures 6-7."

Response:

All variables with exception of T2M are reconstructed daily using daily SLP as predictor as illustrated in Fig. 1 with the suite of fields from the model (analogs) attached to a single date. Hence, all variables reconstructed for a given day stem from the same day. In the case in which we analyze monthly means/sums, these are calculated from daily values. The sentence is indeed somewhat misleading on page 829, line 16. We change the sentence to "The daily reconstruction is separately run for the twelve months of the annual cycle..."

Referee:

"p 832, lines 9-10. I do not see clear why the slope in the decay of the correlations with rank is the best measure of the density of suitable analogs. I think this must be justified either by a strict mathematical proof or by worked examples. I feel this point should be treated with a deeper discussion by the authors. It is probably true, but I don't think it is evident at all. I think that the authors must better explain the content of the figure. Do they show the correlation of the analog-based forecast with rank using the n-th rank analog? Or, are they using the correlation of the analog-based forecast with rank using

the AVERAGE of the leading n-th best analogs? I think this must be clearly stated."

Response:

To use the n-th neighbour instead of the best analog for the reconstruction is just one test to estimate the similarity (density) of the n-th best analogs. The different test cases with the n-th best analog instead of the best analog are calculated similarly to the standard reconstruction. Hence, it is not the average over the n-th best analogs but just a replacement by a neighbouring analog. The mean field correlations of the different test cases are plotted in Fig. 4. We will add this information to the text and the caption of Fig. 4:

"In this test, we replace the best analog by the next neighbouring analog, then by third... n-th best neighbour. For every next neighbour we calculate the correlation between the reconstruction with increasingly poorer analogs and the reference data (Sect. 3.2.2). The mean field correlations of the different tests are depicted in Fig. 4."

We additionally add this sentence how to interpret the density/slope:

"If the slope is relatively small, the confidence in finding appropriate analogs is higher because the neighbouring analogs are quite similar to the best analog. A steep slope in contrast indicates a lower density of similar anlogs and the AM cannot as easily find analogs that are similarly good as the best one."

Referee:

"Pages 839 to 841. The authors show several maps where they count the amount of times a local significance test is/is not rejected. I am afraid that in some cases, global significance (Wilks, 2006 and references therein) is not warranted. Could the authors evaluate it?"

Response:

The estimation of field significance is computationally very expensive with daily res-

olution and with more than 15.000 grid points over 1550 time steps per month and variable. For daily correlation, all values are globally significant given that the smallest amount of positive local tests is 81.3% for rel. humidity in July. Applying a global significance based on Livezey and Chen (1983) for rel. humidity and cloud cover, the amount of local tests being significant just by chance on daily scale would be 13.7% (January) and 11.7% (July) for rel. humidity and 12.0% (January) and 10.6% (July) for cloud cover, respectively.

For the correlation on daily and monthly scale, we run field significance tests for all variables. To do so, we randomized the time steps of the reconstruction and estimated 1000 times the number of locally significant test cases for correlation, our null distribution. Then we estimate the 95th quantile of these test cases and check whether our original test statistic is over this number. If h(local) < 95th percentile of the 1000 test cases we reject the null hypothesis and thus state global significance at the 5% confidence level. For bias and variance ratio, we test whether h(local) > 95th percentile of the 1000 test cases so that bias/variance is globally not significantly different from 0 (bias) or 1 (variance ratio), respectively. We indicate the results in Tables 1-3 through italics. We added two paragraphs to Sect. 2.5 to explain the field significance test in more detail e.g. taking into account serial correlation etc.

Referee:

"Page 848, lines 20-25. The authors mention that the search of a sequence of analogs improves a better simulation of persistence. However, they also mention that it is harder to find a sequence of analogs than a good analog. However, they do not show any additional verification under this setup of the method. If it is harder to find good series of analogs, I guess that correlation or fraction of variance should probably be worse than in the previous setup. I think the authors should show some results about this, in case it is real."

Response:

The reviewer is right that the correlation should decrease with increasing block length. This follows already from a logical point of view: Searching for a sequence of analogs becomes increasingly difficult with the length of the sequence. Given the same calibration data set (or pool of possible analogs), it will always be more difficult to find an analogue sequence terminating at the target day than an analog just for the target day. Finding a similar sequence requires to identify a good analog for the target day and, additionally, similar proceeding days. Thus the selection criteria are much stricter. This must result in a sub-optimal selection of the analog for the target day, since a compromise of similarity of all days along the sequence, and not only the optimal target-day-analog, has to be struck. The similarity between the target day and its analog will necessarily be lower than in the case in which the search criteria are based only on the target day. This represents in general a trade-off between a good similarity between target day and its analogue on the one side and a better representation of the persistence in time on the other side.

A second effect of increasing block length is that the same day is used several times for neighbouring days as stated in the manuscript. This is directly caused by the difficulty, explained above, to find optimal days for a sequence of days instead of a single day. We run additional tests now again using different lags from 0 to 9 days. We added an example for the trade-off in the correlation as a function of the block-length (lag) to Fig. 10B including a short explanation in Sect. 3.3.5. The new figure is shown at the end of this response showing the trade-off for the correlation with increasing block length. Only in case of T2M there is an improvement for n-lags of around 4 like we have used it for our T2M reconstruction in the manuscript in Sect. 2.3.4.

Referee:

"When mentioning reanalyses (page 822, lines 6-12), I feel that ERA Interim should also be cited by now."

Response:

We added ERA Interim here as another reference (Dee et al., 2011).

Referee:

"Regarding the discussion by the authors in lines 8-10 in page 827, I find this reference quite relevant: Vautard, R., and P. Yiou (2009)"

Response:

We added the reference from Vautard and Yiou 2009 here.

Referee:

"Page 828. Authors should justify/explain why do they regrid fields. Common RCAO grid??"

Response:

Yes, this interpolation was the standard grid for RCA at the SMHI.

Referee:

"I am not completely sure what the authors mean in lines 9-10, page 830. I find easy to accept that the precipitation at days d, d-1 is dependent on the evolution of the SLP field at d, d-1, but I do not find easy to accept which other "several mechanisms" (using the authors' words) can explain precipitation's persistence. I think this should be reworded."

Response:

"several mechanisms" were meant here for any variable with precipitation being just one example. We added "any variable" to make this clear. As illustration of what is meant here by 'other mechanisms' one can think that precipitation on a particular day may depend also on precipitation in the previous days (e.g. accumulated soil moisture in summer).

Referee:

"Page 832, lines 1-3. I guess that the authors use C1 to C5 decades as the analog pool and then check the ability of the analog method to reconstruct the whole period. I guess that the skill score might be a little bit (not too much) optimistic, due to the autocorrelation or decadal variability. I think that leaving out just one day is probably not enough in terms of autocorrelation, although I haven't computed the autocorrelation in the series. Perhaps checking the ability to reconstruct the 40 years not in the C1 to C5 subsets would be a tougher test. However, I don't expect the results would be very different. They already consider the autocorrelation in their significance tests (p833) but, anyway, verification and formulation of the models should ideally be carried out with independent information."

Response:

By the expression "leave-one-out" it is meant that the particular year of the target day is left out in the search of the analogue to fully ensure that the search of the analogue is truly independent of the target day, and avoiding any contamination in the assessment of the skill of the method..We add this information here again for clarity. We also applied a crosswise calibration-validation for different periods (25/25 years, 10 years vs. 10 years from different sub periods) showing also very little differences (see response to referee #1). We modify the sentence to "Using a leave-one-out approach skipping always the actual year..."

Referee:

"Page 834, lines 3 to 7. I think that after this paragraph the authors should discuss somehow the fact that the model, although considered as "true" is probably underestimating the real spatial variability of the field, so that, in my opinion, their estimation of the ability of the method is probably a little bit optimistic. They show some results consistent with this a little bit later in the text, I think they should explain something about it now."

Response:

We test the AM in the surrogate world of the model which is an idealized and simplified estimation of the real world. The resolution with 25 km is quite high but nevertheless neglects many local effects so that the model might be more spatially homogeneous than in reality. We add the following sentence:

"As the model presents only an idealized and simplified estimation of the real world, the idealized testing of the method might yield somewhat optimistic skills e.g. because of a lower spatial variability in the model compared to observations. Nevertheless, this idealized testing of the method provides a good benchmark for the AMs potential skill based on the chosen settings."

Referee:

"Page 836, lines 20-23. Is this phrase complete? Do the authors refer to "n=1 to n=10" at the beginning, when they refer to a 6% decrease?"

Response:

The sentence is complete but we reworded it to

"As an example for reconstructed daily SLP, the explained variance decreases linearly with a rate of around 6% per 100 neighbours in January and 3% for July for neighbours >100 to the best analog. For the first ten neighbours, the slope is larger with already a decrease of 6% in January and 7% for July for 10 neighbours, respectively."

Referee:

"Page 845, line 26. Authors mention "The reconstruction CLEARLY reproduces". I would say it is far from clear in some cases, such as precipitation, for instance. I would recommend a more realistic phrase here."

Response:

The types/forms of the different distributions in general are clearly reproduced which does not mean that there are deviations between bins as in the case of precipitation.

The deviations appear larger as we used a logarithmic scale for the frequency to better depict deviations at the upper (left) tail of the distribution. We mentioned this point on the next page 846, lines 22-29. We add "...the general distribution types are clearly reconstructed..."

Referee:

"Figure 4. Left label states "correlation [-1,1]" but range of plot is [0,1]. Do the authors refer to absolute value of the correlation? Correlation squared? Or is it really correlation but just plotted the [0,1] range? According to the case, the authors might have to change the label."

Response:

Plotted is in all cases the correlation reaching from -1 to 1. As no negative values have been present, the plot only shows the range 0 to 1. We remove [-1,1] from the label for clarity.

Referee:

"Figure 6. I think the authors must rewrite the caption to clearly explain which quantities are represented in each panel. They mention RCAO/ERA40, and I guess that means RCAO (forced by ERA40), therefore, I don't see the point of mentioning here ERA40, since in the previous cases RCAO has been used. Expressions "(left columns)" and "(right columns)" should be added after the mention to daily and monthly in the caption. The authors mention the existence of white shaded lines. I have only been able to see them after using a 300% zoom when reading the PDF file. I suggest them to make the lines thicker."

Response:

We remove RCAO/ERA40 and use consistently RCAO (RCAX for T2M) throughout the manuscript. We added right and left columns to the figure capture. We also increased the thickness of lines.

Referee:

"Figure 9. The caption should mention which panel corresponds to each site. It is very hard to understand this figure as currently"

Response:

The information about the location is given at the bottom of the figure for SLP (de Bilt, Haparanda) and other variables (Bergen, St. Petersburg). This is also noted in the figure caption. We added "solid lines" for HiResAFF and "shaded lines" for RCAO and RCAX.
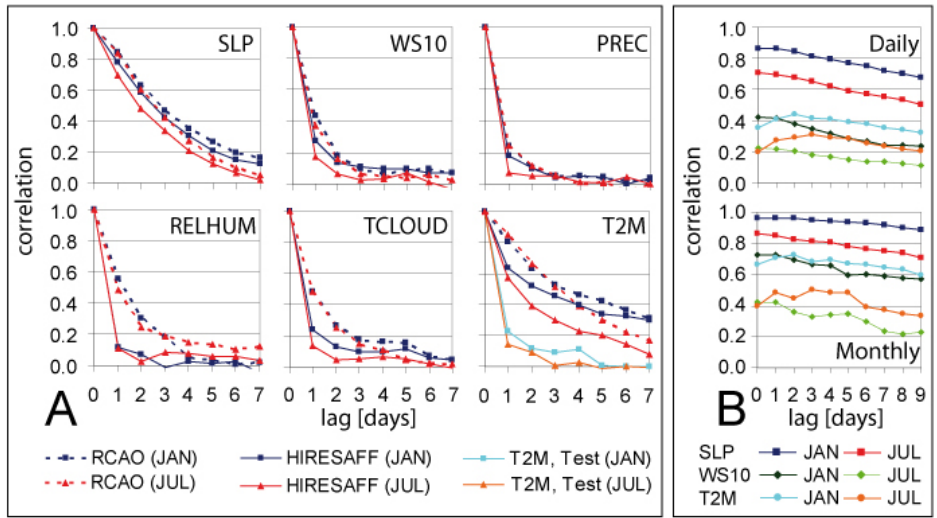
We corrected language issues where applicable. Thank you.

Corrigendum: Please note that we corrected Fig. 7j which was showing again WS10 instead of PREC. Numbers and the description in the manuscript have been correct.

Interactive comment on Clim. Past Discuss., 8, 819, 2012.

**Fig. 1.** A: serial correlations like before in Fig. 10, B: Changes of daily (top) and monthly (bottom) mean field correlation for different variables dependent on n-lag for January und July.