Clim. Past Discuss., 9, C1766–C1770, 2013
www.clim-past-discuss.net/9/C1766/2013/

Climate of the Past Discussions

Interactive comment on "Statistical downscaling of a climate simulation of the last glacial cycle: temperature and precipitation over Northern Europe" *by* N. Korhonen et al.

Anonymous Referee #2

Received and published: 20 August 2013

The subject of the paper is interesting and rarely addressed - downscaling coarse EMIC climate simulations to obtain regional climate change information over the long periods that higher resolution dynamic models cannot yet simulate. The specific interest comes from the fact that the method makes use of data from two high-resolution time slice simulations, including the LGM, offering a potential for a valuable calibration and evaluation of the statistical model. As written by the other referee, the inclusion of ice sheet related data in the predictors, based on an ice sheet model coupled to the EMIC, is a novel and useful research.

However, some aspects of the statistical model and/or its description are somewhat confusing, and its evaluation does not yet appear fully convincing.

General comments

The basic principle appears right - selecting a model by trying to get the best fit for a given number of degrees of freedom. However, the choice will be among the tested predictors, and while it is interesting, I am surprised with what came up as equations 3, 4, and 5.

All have a sort of "map", the function of latitude and longitude $s_j(x, y)$. For the method to be valid, any contribution of these maps to the final result needs to be fully independent of time: it needs to be limited to the correction of a sort of "local bias" in the interpolated EMIC result. The potential to achieve this probably depends on the "diversity" in the calibration data : in the extreme case which would only use one calibration time-slice, such a map could, depending on the available degrees of freedom, represent most of the data. We are evidently not in this case, as there are 2 glacial time-slices + the recent past: the risk of exaggerating the contribution of this time-independent term (which might not be time-independent in the real world) is limited. However, this suggests that the model needs to be carefully validated, because there is a potential to obtain a very good fit for the calibration data, but much less good results when predicting other time periods - even in the "calibration range" of each predictor. A related difficulty is that the model includes quite complex functions, in particular $s_7(x, y, T_{CLI})$. This particular contribution (s_7) suggests that some locations are more "sensitive" to the EMIC's output than others (could an interpretation be provided?). This is perhaps true, but again, there is a potential for this contribution to appear independent of time for the calibration cases, while it is not clear that it is more generally true (in the real world).

The formulation of the GAM used for precipitation is relatively surprising. It appears that the model for log (P) is the addition of splines including the precipitation from the EMIC, P_{CLI} (would $log(P_{CLI})$ provide the same results?). Thus, Figure 2(a) suggest that for high (monthly) precipitation, the expected regional rainfall includes a term increasing exponentially with the precipitation from the EMIC, while for medium precip-

itation amounts, there is relatively little influence of the precise EMIC simulated value. It seems possible but not evident that this will remain valid under different climate conditions. It does not seem easy to provide a clear interpretation for the results in figure 2, especially panels a) and c), therefore this also suggests that the validation should be very careful.

However, I do not have the impression that the validation is sufficiently careful in the current version of the manuscript: was there an attempt at evaluating the GAM by comparing its results to the observations that were not included in the calibration data? I apologize if I missed something - I did not find such a validation in the manuscript. It could probably be based on 1) calibrating the model on the "present" and LGM data, then comparing it to the 44 kyr data, or 2) calibrating the model on some month(s) and validating it on other(s). Achieving good results with (1) would be impressive. Could such a validation be added? (if something is already done, then please clarify what is the data used for calibration and what is the data used for validation).

The calibration method would probably benefit from a clarification - is the calibration performed independently for each month, using monthly values for each grid point and all the 3 time slices?

In summary, I do not have the impression that the results from tables 1 and 2, as well as figures 3 and 5, are currently sufficiently convincing regarding the validation of the statistical model: they are interesting as they show that an acceptable fit is achieved for the calibration period, but it would be useful to find a way to confirm that the choice of predictors is appropriate and that the GAMs do provide good predictions for other periods.

In the current version, the most important result in this regard could be figure 6. However, I do not understand why the main difference seems to be a systematic error (bias) over all the period. Are the GAMs performing better than a simple correction of the present-day bias in the interpolated result from CLIMBER? Again, I apologize if I

am missing something - I would feel it strange that the comparison can be completely changed by a simple bias correction and that it was not done. The changes shown on figure 6 also appear relatively small, hence the potential to compare the observations to the statistical model is limited : could more locations and/or a longer period be provided?

Specific comments

Page 3376 - line 12: I do not understand the sentence referring to a "stepwise screening of the data".

Equation (1): Remark: $+\epsilon$ is probably not required (as this expression provides the expected value, including an error term (residuals) does not seems appropriate, please check)

page 3377 - line 6 - 9: How can predictors be extracted from the RCA and CCSM models ? I though that those models were used to calibrate the GAMs?

Equation (2): Please add a reference, including for the "cost weight" $\gamma = 1.4$

Equation (6), (7) and (8): why are there different notations for expected values - what is the difference between the overline in equation 6 and the brackets in equation 8, and what are the differences between the definitions provided in equations 6 and 8 ?

page 3383 - line 3: I would expect the wording "skill scores" to apply to model "predictions", that is, not for the calibration cases. If this is about predictions, please clarify; otherwise I think that the wording should be changed.

page 3383 - line 14: The temperature change with altitude seems rather small, at ~2°C/km. Could you comment?

page 3384 - line 11: What is meant by "general error" ? Is it systematic or random?

page 3385 - line 27: Remark: be careful regarding simulations for the future, as it would require a different calibration and validation, including simulations with much more GHGs.

References: please check the status of the papers mentioned as “submitted”, in particular “Martin et al., 2013a, 2013b, 2013c”. If these are not accepted, they should not be used as references.