Geosci. Model Dev. Discuss., 6, C1376–C1381, 2013
www.geosci-model-dev-discuss.net/6/C1376/2013/

Geoscientific Model Development Discussions
Open Access

Interactive comment on "A priori selection and data-based skill assessment of reanalysis data as predictors for daily air temperature on a glaciated, tropical mountain range" *by* M. Hofer et al.

Anonymous Referee #1

Received and published: 2 September 2013

GENERAL COMMENTS

For many areas of the world, assessments of climate effects and climate change impacts on geo- and biophysical systems suffer from the lack of specific data, data gaps in observational records, etc. Provided that sufficient data is available to develop a reliable model for relating a target variable or predictand with one or more predictors, empirical statistical downscaling (ESD) represents an efficient approach for overcome this problem and has therefore become popular among the impact assessment community. Nevertheless, systematic considerations of the setup of ESD are not very common, and the work presented here certainly fills in a gap.

The paper presents an empirical statistical downscaling approach that is in principle suitable for use in situations when the available database is limited. It addresses various aspects of the model development and verification, and shows how the approach can be used to infer daily air temperature in a glaciated mountain area given a minimum of observational material.

The paper is by and large well written. There is, however, a need to reconsider its setup, which is necessary to achieve a better balance between general considerations and new findings and keep the attention focused on the fundamentals. The paper touches upon different facets of ESD, but eventually the core of the paper is to be found in section 4.3 and related results, and this is what should be emphasized. While sufficient material is available to understand the methodology, some of the key points and main assumptions need to be addressed in more detail (see specific comments).

In contrast, part of section 4.1 and even more so section 4.2 refer to the results of earlier studies by the authors and do not introduce new material. These sections could therefore be substantially shortened. Similarly, section 5.4, while per se pertinent, distorts somewhat the attention from the main issue. It should be better integrated or perhaps, recalling section 4.2, inserted in advance of the main results. The same holds true for section 5.5, which deals with an topic that does not seem to be essential in the present context (sub-daily scale variability is otherwise not an issue in this paper).

Also, much of the discussion is implicitly tailored to fit the specific application, as stressed in the title, but in the final section the paper lacks somewhat of a broader perspective (e.g. a discussion of how to deal with variables that cannot be assumed to be normally distributed, applications to other geographic areas or other topographic conditions within this same study area, etc.). It is therefore not clear, how much additional work is needed if one wishes to use the approach in a different context.

The linguistic quality is good, although there is an unnecessary (in my view) abundance of relative sentences in parenthesis.

In summary, the paper can be considered for publication provided that some efforts are undertaken to address the above concerns and the specific comments that follow.

SPECIFIC COMMENTS

- Seasonality (section 4.1). As stated on P 2890, L 11 ff., one of the key assumptions is that seasonal atmospheric variability leads to changing relationships between predictors and predictand throughout the year. It is therefore in order to look at statistical models for each month separately (this is actually a common approach for instance in downscaling exercises that rely on stochastic weather generator). But, are there alternatives? Would it be reasonable to first develop a model for the seasonal cycle and then consider only the residuals as predictands? Moreover, the model assumes that the statistical relation between predictors and predictand does not change in time. However, it is mentioned on p. 2888, L 26 ff., that consideration of the ENSO is important to understand the climatic variability of the study area. The characteristic time scale of ENSO phases is of the order of 10 years. Wouldn't it be reasonable to assume that ENSO affects the relation between predictand and predictors, even if this is established on a monthly basis?

- Gaussian target variables (section 4.3). As stated by the authors on P 2895, L 25, the linear model (1) is valid only for Gaussian target variables. What would a suitable model for non-Gaussian variables be? What other steps in the design of the ESD model are critically dependent on the normality of the target variable?

- Composite time series (section 4.3). The analysis is carried out on the basis of composite time series obtained by collating the monthly time series of individual years into a single time series for each month of the year. This step is not crucial for the validity of the linear model (1) because (P 2895, L 4 ff.) "least-squares regression does not account for the time ordering in data series, and is therefore not affected by the use of discontinuous (month-separated) time-series". However, it is not obvious that this is not problematic in relation to the determination of autocorrelation, in particular if it

GMDD

6, C1376–C1381, 2013

cannot be assumed a priori that the autocorrelation vanishes at time lags of the order of 10 days to one month. Consequently, the fact that time series are composite is not much of an issue in the context of equation (8), because in this instance only the lag-1 autocorrelation is considered, but could be a topic in relation to equation (3) if this ESD method is applied to variables characterized by a pronounced persistence (see also P 2900, L 18 ff.)

- Downscaling process (section 4.3). Please justify equations (3) and (8). Please justify why it is in order to use only the "central value" of the "withheld observations" (P 2896, L 9) as a target for the validation.

- Skill score (section 4.3). The possibility of decomposing the skill score (P 2896, L 19 ff.) according to the scheme presented by Murphy (1988, op. cit.) is used but not shown explicitly. Although the decomposition is readily derived, it could help the reader not familiar with Murphy (1988) to have it shown explicitly.

- Skill assessment and significance analysis (section 4.3 and 5.2, and Fig. 5). Figure 5 presents three inter-related quantities that are needed in relation to skill assessment and significance analysis introduced in section 4.3 and discussed in section 5.2. Although this is in order, I was wondering whether it would not be more straightforward to only show the effective size n_eff, which is the one needed to perform the left-tailed t-test (section 4.3).

- Results for different time scales (section 5.3). It is a well know result (for instance in relation to seasonal weather predictions), that skill scores tend to increase with increasing averaging time and this supports the conclusion that (P 2907, L 9 ff.) "For the same number of observations at different temporal resolutions (i.e. from one-day to twelve-day averages), values of skill averaged over all months increase with increasing averaging time windows. Consequently we suggest switching to lower temporal resolutions when the ESD model skill is low, given that long enough data series are available". On the other hand, whether this is feasible/acceptable or not depends as

GMDD

6, C1376–C1381, 2013

well on the primary questions that need to be answered. In the limiting case that there is skill only for averaging times > 10 days or so, switching to this lower resolution may prevent achieving the original goal of reconstructing daily time series (as stated in the title).

- In the same context, it is perhaps worthwhile to try to consider the reasons why (P 2904, L 7 ff.) “the increase is not monotonously, but rather stepwise, with a considerable increase from the three- to four-day averages, with rather constant values thereafter and another considerable increase from the nine- to ten-days averages” through a more detailed consideration of Fig. 7 (sharp increase in the skill for averaging times > 3 days for the months of August and September; sharp increase in the skill for averaging times > 6 days for the month of January, . . .).

- Model discussion (section 6). Apart from critical elements already mentioned (seasonality, etc), it is worthwhile reiterating that there is a need to suggest how the present study could be integrated in a broader context, and to explain why certain assertions are valid. For instance it is stated on P 2907, L 21 ff, that “The validation process is especially useful in multiple predictor fitting because it detects over-fitting. The method is not restricted to reanalysis data and can be applied to any atmospheric model predictors”. Unless I missed something important, my impression is that nothing has been said in the paper concerning how over-fitting could/would be detected.

TECHNICAL ISSUES

- Generally speaking, there is an abundant use of parentheses that perhaps could be avoided. This is the case for instance in section 4.3.

- There are a few typos, e.g.: P 2895, L 18 should read “In each cross-validation repetition, . . .”; P 2896, eq. (7), on the right-hand side y_r should read y[hat]_r; P 2987, eq. (9) please specify the meaning of sigma_cv; P 2900, L 8 ff. should read “Values of tau are of the order of 2 or 3 days for all month except the wet season-months of February and March and the transitional-season month of April, for which values of tau

GMDD

6, C1376–C1381, 2013

are considerably higher (7, 9 and 11 days, respectively); sections 5.4 and 5.5; please write acronyms of the variables in question in italic to avoid misunderstanding (e.g. P 2906, L 3.". . . the predictor aid shows the highest covariance . . ."); references, P 2911, L 25. Please check for other typos.

- Fig. 3. Although in principle correct, the labels on the x-axes (years) can be misunderstood given the axis legend ". . . time series [days]". Also it is not clear, how the vertical dashed line can "indicate[s] the minimum number of observations . . .".

Interactive comment on Geosci. Model Dev. Discuss., 6, 2883, 2013.