Geosci. Model Dev. Discuss., 6, C2684–C2695, 2014
www.geosci-model-dev-discuss.net/6/C2684/2014/

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.

M. Hofer et al.

marlis.hofer@uibk.ac.at

Received and published: 28 February 2014

Reply to the review of Anonymous Reviewer # 1 on our manuscript:

"A priori selection and data-based skill assessment of reanalysis data as predictors for daily air temperature on a glaciated, tropical mountain range"

Revised title:

"An empirical-statistical downscaling method for data-sparse, glaciated

mountain ranges: application to reanalysis data predictors for daily air temperature

(Cordillera Blanca, Peru)"

We would like to thank the editor for obtaining the reviews, and for accepting the delay of our response due to the lead author's maternity leave. We would also like to thank the two anonymous referees for providing very helpful and constructive comments on our manuscript. Without doubt, the comments of both referees helped to improve our manuscript significantly, and we were able to address each of their remarks. We have restructured and reformulated large parts of the manuscript, for more clarity, and to facilitate reading. Note that we have also modified the title of the manuscript (see above). Please consider our detailed responses below, where we refer to our original manuscript (reviewed by the referees) as the "original manuscript", and to the new, modified manuscript as the "revised manuscript".

Marlis Hofer, Ben Marzeion, and Thomas Mölg

General comments by Anonymous Reviewer # 1

"*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."

Reply: We agree with the referee that the structure of our manuscript needed improvement, and have accordingly done a revision. We have put particular efforts into shortening too long sentences, and into eliminating some relative sentences, to facilitate understanding. We address your further concerns in our responses to your specific and technical comments below.

Specific comments by Anonymous Reviewer # 1

"*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?...*"

Reply: In many studies, periodicity is (if at all) handled by subtracting the climatological seasonal cycle from the predictands and the predictors, and subsequently establishing the downscaling model based on the remaining time series (e.g. Madden, 1976). We mention this common alternative to our approach in both the original and the revised manuscripts. As the referee mentions, our approach to use different transfer function (or model parameters) for different months is often found in stochastic weather generators (e.g., Wilby and Dawson, 2007). We have accordingly added a note in the revised manuscript.

"*... 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?*"

Reply: It is true that we assume temporal stationarity of the statistical downscaling transfer functions, which have been established on a monthly basis using only few years of measurements. This implies, as the referee mentions, that the statistical transfer functions change for different months of the year, but remain stationary for different phases of the El Niño Southern Oscillation (ENSO). ENSO, however, is a large-scale phenomenon, which is generally well represented by the predictors (in our case, the reanalysis data, e.g., Garreaud et al., 1999). On the other hand, the seasonally varying effects of the topography (e.g., topographic shading) can not be captured by the predictors, because the large-scale model topography is not realistic. We mention this similarly in Sect. 5 of the revised manuscript.

"*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?*"

Reply: The use of Gaussian target variables in the ESD modelling framework is problematic, because for non Gaussian target variables the normality assumption of the least-squares model error is usually violated. We explain this more clearly in the revised manuscript. Moreover, we discuss measures for including non Gaussian target variables into the ESD modelling framework in Sect. 5 of the revised manuscript.

"*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*

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.)"

Reply: As the referee correctly mentions, the autocorrelation parameter τ applied to the month-separated time series cannot exceed one month. In our study, τ is a parameter in the cross-validation procedure, with the role to guarantee that the validation observations are independent from the model training observations. However, since the time series are month-separated, autocorrelation of larger than one month is automatically eliminated, and is thus not problematic for the cross-validation procedure of Eq. (3). In the revised manuscript, we have added some lines (Sect. 4.1), to avoid confusion between the autocorrelation parameter τ of the month-separated time series, and persistence of the complete time series. The latter could potentially could exceed one month, as the referee correctly mentioned. Further, we have changed the notation in the revised manuscript, to distinguish more clearly between the autocorrelation parameter τ (number of days for which the autocorrelation is close to zero) and the lag-1 autocorrelation. We have replaced τ (original manuscript) with $\tau_{\rho\cong 0}$ (revised manuscript), and τ_1 (original manuscript) with ρ_1 (revised manuscript). This is also consistent with the notation commonly used in atmospheric statistics textbooks (e.g., Wilks, 2006).

"*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.*"

Reply: As we stated already in the original manuscript, the central of the withheld observations in each cross-validation step cv, being separated by sufficient time steps from the training data, " ... can be considered as independent from the calibration process" (p. 2896, l.9 ff). By setting the number of cross-validation steps (cv) equal to the number of observations (n), this procedure allows each observation to be used for the model training, and as validation value. We added more explanation along these lines.

"*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.*"

Reply: We agree. In the revised manuscript, we explicitly show the decomposition of SS.

"*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).*"

Reply: Even if $\tau_{\rho=0}$, and n_{min} in Figure 5 are related to n_{eff}, we would still like to show the two quantities explicitly, because they play an individual role in the skill estimation, and are correspondingly discussed separately in the results section (see sections 4.2 and 4.3 of the revised manuscript).

"*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 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)."

Reply: It is true that our study focuses on daily air temperature time series (as stated in the title). We still consider the application of the presented ESD procedures to time scales beyond the daily time scale a relevant extension of our study, and would like to keep these results in. However, we have shortened the corresponding section 5.3 (original manuscript) and put it together with section 5.5 (original manuscript) into section 4.5 ("Application of the ESD model at varying time resolutions"), such that the reader can find the primary results first (with focus on the daily time scale, sections 4.1 to 4.3, revised manuscript). Further, we mention that the increase of skill with increasing time scale is an expected result (see also our response to your next comment).

"*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, . . .).*"

Reply: The stepwise increase-pattern of SS with increasing time scales appears by averaging the very distinct patterns of SS for the different calendar months. For example, as visible in Fig. 7, values of SS for the January time series first show no significant skill, then increase until the seven-days averages, and remain rather constant thereafter. By contrast, values of SS for the February time series first increase, then decrease, then increase, and remain constant for the ten- to twelve-days averages. Thus, as expected, predictability based on large-scale predictors is - on average

- found to increase with increasing time scales. However, the patterns of SS for the different calendar months are highly variable, and are not examined in detail here, as this would go beyond the study's scope. We have added accordingly more explanation in Sect. 4.4 of the revised manuscript.

"*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.*"

Reply: In Sect. 5 of the revised manuscript, we explicitly discuss the application and transference of the presented ESD procedure in a more general context (i.e., for different predictands, predictors, sites).

Furthermore, we state on p.2897 l.25 ff. of the original manuscript: "Note that, even though not shown in this study, SS as defined above is a powerful goodness-of-fit estimate especially in the case of multiple predictors, because it detects over-fitting (then SS is zero or not significant).". In the revised manuscript, we rewrite the sentence (Sect. 5): " In fact, SS based on cross-validation is a powerful goodness-of-fit estimate especially in the case of multiple predictors, because it detects over-fitting. More precisely, in the case of over-fitting, SS as defined in Eq. (8) is zero, or not significant. For more details about cross-validating multiple predictor-regressions to protect against over-fitting, see Wilks (2006)". This way, it should be clear that it is the cross-validation process which effectively prevents against over-fitting (see Wilks, 2006). We repeat this in the conclusions (p.2907, l.21 ff.): "The validation process is especially useful in multiple predictor fitting because it detects over-fitting.". We insert "(based on cross-validation)" after "...validation process..." to emphasize again that the cross-validation

procedure is the critical step to identify over-fitting.

Technical comments by Anonymous Reviewer # 1

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

Reply: Throughout the entire manuscript, we have split up/shortened too long sentences, in order to make the reading easier.

"*There are a few typos, e.g.: P 2895, L 18 should read "In each cross-validation repetition, . . .";*"

Reply: Revised accordingly.

"*P. 2896, eq. (7), on the right-hand side y_r should read $y[hat]_r$;*"

Reply: Thank you, revised accordingly.

"*P 2987, eq. (9) please specify the meaning of σ_{cv};* "

Reply: $\sigma_{\mathsf{cv}}(t)$ was defined in the original manuscript as "...the standard deviation of a variable (v) over all cv repetitions..." (p.2897 l. 20-21). In the revised manuscript, we try to be more comprehensible: "let ... $\sigma_{\mathsf{cv}}(\hat{y})$ be the standard deviation of $\hat{y}(t)$ over all cv repetitions, for each time step t. Note that both $\overline{\hat{y}}^{\mathsf{cv}}$ and $\sigma_{\mathsf{cv}}(\hat{y})$ are then vectors with n elements. ...".

"*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 are considerably higher (7, 9 and 11 days, respectively);*
"

Reply: We have revised this sentence as follows: "For February, March and April, values of $\tau_{\rho\cong0}$ are considerably higher (7, 9 and 11 days, respectively) than for the remaining months.".

"*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 . . ."); *"

Reply: We also prefer italic for the abbreviated variables. However the journal typesetting team changed this, because after the journal conventions, abbreviations are not set in italic. If the manuscript will be edited again, we hope that the the abbreviated variables will be shown in italic.

"*references, P 2911, L 25. Please check for other typos.*"

Reply: I can not find an error on P 2911, L 25. I have checked the manuscript for more errors.

"*Fig. 3. Although in principle correct, the labels on the x-axes (years) can be misunderstood given the axis legend ". . . time series [days]".* "

Reply: We agree. We have revised the labels of Fig. 4, by adding the exact dates to the x-labels (e.g., "2006/3/1" instead of "2006").

"*Also it is not clear, how the vertical dashed line can "indicate[s] the minimum number of observations . . .".*"

The dashed line in Fig. 4 indicates the minimum number of observations to obtain statistically significant skill for the ESD model, n_{min}, as follows. If both, the March and July time series, would be cut off at the date indicated by the dashed line, the time series would include n_{min} observations. However, we have removed the dashed lines in Fig. 4, since this information was not clear to both anonymous referees. Please note that n_{min} is shown again in Fig. 5 (original and revised manuscript). As also evident in Fig. 5, n_{min} differs considerably for the different calendar months. This is the reason why the dashed lines do not appear at the same dates for the two different calendar months March and July in Fig. 4 of the original manuscript.

References

Garreaud, R., D. David, and S. Battisti, 1999: Interannual (ENSO) and Interdecadal (ENSO-like) Variability in the Southern Hemisphere Tropospheric Circulation. *J. Climate*, **12**, 2113–2123, doi:http://dx.doi.org/10.1175/1520-0442(1999)012<2113:IEAIEL>2.0.CO;2.

Hastie, T., R. Tibshirani, and J. Friedman, 2001: *The elements of statistical learning: data mining, inference, and prediction*. 1st ed., Springer series in statistics, Springer, New York, 533 pp.

Madden, R. A., 1976: Estimates of the natural variability of time-averaged sea-level pressure. *Monthly Weather Review*, **104 (7)**, 942–952, doi:10.1175/1520-0493(1976)104.

Wilby, R. and C. Dawson, 2007: *SDSM 4.2 âĂŤ A decision support tool for the assessment of regional climate change impacts*. Leicestershire, UK, Loughborough University.

Wilks, D. S., 2006: *Statistical methods in the atmospheric sciences*, International Geophysics Series, Vol. 91. 2d ed., Academic Press, 627 pp.

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