

## ***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 #2**

Received and published: 11 September 2013

### GENERAL COMMENTS

For a number of mountainous areas, the climate data are of high importance to study, for instance, the impact of the climate on available water resources. However, it is difficult to obtain long and continuous observational records - especially in glaciated areas - where data are usually affected by instruments dis-functionalities due to complex topographical features. Thus, empirical or statistical downscaling methods can be seen as a simple and promising alternatives to fill these gaps. Nevertheless, ESD model applications to short time series have been neglected by the scientific community during the last decade making this work relevant to overcome this limitation.

C1423

The paper presents an ESD model tailored for short observational time series (less than ten years of daily data) and discusses various aspects related to ESD calibration and evaluation purposes. In the ESD model, predictand are daily air temperature at automated weather stations located at the Cordillera Blanca, Peru (local scale). The predictors are outputs from global reanalysis data (large scale). Seasonality, predictor selection and model skill estimation procedure are also discussed in the paper. The main finding of this paper is the estimation of the minimum length of observational records which is required to provide a statistically significant model skill. This can definitely help in measuring new climate data at specific sites for shorter periods with minimum cost.

The several analyses performed here are innovative and have the merits to be published. There is, however, a need to reorganize the paper differently which will definitely help the the reader. In addition, the authors must clearly distinguish between results from previous studies and the new key main point findings of this work. The several hypothesis must be discussed and argued in more detail to avoid confusions. Also, in some sections, the titles are not appropriate.

Section 4.3 seems to be the main core of the paper but appears late in the manuscript. The model calibration and cross-validation seems to be the innovative part of the work and need more highlights. However, my big concern is whether the authors are aware about the Jackknife method commonly used for parameter estimation errors and cross-validation purposes. It is stated from the wikipedia page that “The basic idea behind the jackknife variance estimator lies in systematically recomputing the statistic estimate leaving out one or more observations at a time from the sample set. From this new set of replicates of the statistic, an estimate for the bias and an estimate for the variance of the statistic can be calculated.” (please refer to the following WIKIPEDIA link for more details about the method [http://en.wikipedia.org/wiki/Resampling\\_%28statistics%29#Jackknife\\_2](http://en.wikipedia.org/wiki/Resampling_%28statistics%29#Jackknife_2) ; this link also includes a list of useful references). It seems like the “JackKnife” method has been

C1424

used here to give a better estimate of the Skill Score (SS) compared to the  $r^2$  statistic. The authors must definitely discuss on how their methodology is different from a Jackknife method.

The overall linguistic quality of the paper is acceptable. There is a need to 1) shorten long sentences (see detailed comments), 2) avoid unnecessary extensive use of parentheses, be consistent with definitions and acronyms, and review missing and wrong references in the text (e.g. page 2906 – L1 : Sect. 9!) and figure's caption. This will facilitate reading the paper.

To summarize, the paper could be considered for publication provided major modifications and quality improvement of the text by addressing the several concerns previously mentioned and taking into account the following specific and detailed comments.

#### SPECIFIC COMMENTS

Sections 2 and 3 must be combined together and shortened. A distinction between the Study site, the Predictand, and Predictors would be appreciated. ESD model Architecture (Section 4). In this section, there is a need of an introductory paragraph before the subsection 4.1 starts. The general framework of the ESD should be first reviewed, from a general aspect, then details can be given later on in the paper giving specific adaptations to meet with the short time series purposes. Downscaling process (Section 4.3). The statement on Page 2894 L26 needs some clarification. What would the author suggest as an alternative for using their model to downscale precipitation. ESD model Architecture (Section 4). This section should start with subsection 4.3. The authors need to start with general aspects and move to more specific ones, which seems to be reversed here. Also, part of subsections 4.1 and 4.2 refer to earlier results and do not introduce new material. I suggest to split these subsections partly in sections introduction and discussion. Also, subsection 5.1 "Downscaling parameters" should be displaced inside Section 4 as it is part of the ESD framework and not the application. Finally, I can hardly see how the title in section 5 can be appropriate here as it corre-

C1425

sponds more likely to the results. Moreover, Sub-sections 5.3 to 5.3 can be combined together and presented in the discussion. Accounting for seasonal periodicity (Section 4.1). By estimating the model parameters separately for each month, the authors mentioned that the bias introduced by temporal dependencies is removed. By doing so, any dependency between the months, if it exists, will be also eliminated. What would the authors suggest to reconstruct reliable estimates of daily temperature time series based on their ESD model? For AWS1 predictand (Page 2920, Fig. 4), it is clear that ESD model parameter estimates for the wet season show a larger bias than for the other months, which is stated by the author on Page 2899, L22-23. This bias might be in the origin of the high minimum number of required observations as shown in Fig. 5 and for the SS difference between the two Periods 1 and 2 (Page 2900 L28 to Page 2902 L3), for AWS1 as shown in Fig. 6. It seems from the different figures that the wet season require larger sample data for the ESD model to be more stable than the other seasons. This issue might be included and discussed in the present paper. Results for different time scales (section 5.3). It is well known that the model skill increases with averaging the time scales because the variability is diminished in the aggregated values.

#### DETAILED COMMENTS

Page 2884 Line 2: delete the comma after "altitude"...place a comma after "ESD model..." Line 10: place a comma after "Similarly" and remove the one after "temporal resolutions" Line 13-16: run on sentence Line: place a comma after "Ongoing developments in atmospheric modeling" and remove the comma after available choices of long term" Line 19-20: I would say "However, these data..." Line 21-23: Avoid starting the sentence with "especially". Sentence needs to be revised. Line 23: Please add "The" before "so-called..." Page 2885 Line 4: "have emerged" instead of "has emerged" Line 9: Place a period after "the last decades" and start a new sentence with "Concerning the choice..." Line 11 : in "..., and (iii) of ..." remove "of" Line 21: Place a comma after "atmospheric sciences" Page 2886 Line 2-5: Run-on sentence Line 4: put first letter of

C1426

each word to upper case in “automated weather station (AWSs) ...” Lines 6-7: avoid repetition of the word “study” Line 19: Put a comma between “agriculture” and “and households...” Line 23: remove extra “the” Line 29 and Line 1-2: Please rephrase Page 2887 Line 5: add “located at” to “... AWSs at and nearby glaciers...” Line 6: Period after “installed” and start a new sentence with “Primarily...” Line 16- 19: The sentence is too long; try splitting it in two parts. Line 19: I would say “methodology that is based on single...” instead of “methodology, based on single” Line 29: Be consistent : replace (hereafter AWS3) by (hereafter referred to as AWS3) Page 2888 - Line 10: replace “described by ...” by “..., as described by ...”

Page 2889 Line 7 : capitalize first letter in each word in “... numerical weather prediction ...” Line 14 : Please be consistent with names and acronyms as “the National Centre for Environmental Prediction (NCEP) ... “ Lines 20-25 : please rephrase and shorten the sentence ... Line 23 : move CFSR inside the parentheses Lines 25-28 : rephrase sentence and define “interim” as first occurrence. Line 27 : replace “who” by “which” Page 2891 Line 3 : to which model the authors refer to ? Page 2892 : Line 17 : replace “model” by “reanalysis” to be consistent Line 19 : start the sentence with “However, ...” Line 20 : remove “s” in levels Page 2893 Lines 18-21 : The authors must argue their choice here. Line 24 : what does “at the grid points located closest to the study site” means. Do the authors refer to the average of the closest grid points taken from the different reanalysis, respectively? Line 26 : remove “s” in “550 hPa levels” Page 2894 Line 3 : the verb “is” is missing in “More precisely, the coarser the large scale” Line 8 : replace “best” by “better” and capitalize “limited area ... models”, and avoid the use of “certainly” ... Line 18 : replace ‘10-fold’ by “N-fold (N>1)” to be more general. Lines 19-21 : I would say, “. . .because the estimated coefficient of regressions or model parameters are not affected by temporal dependencies ...” Line 24 : I would use “. . .for each month in the time series” instead “. . .month’s time series...” Line 26 : please move this sentence to the conclusions and perspective section and give alternatives of regression relationship to use with non-Gaussian variables such as precipitation. Page 2896 Line 2 : What does index “s” stands for in Equation (1). I guess

C1427

“standardized”. Line 13 : define the right hand side of equation (2). Does it stand for estimated values ? Please give details. Page 2897 Line 2 : replace “step” by “repetition” Line 20 : I think that it would be relevant for the authors to give the expressions of the two penalty terms, here, instead of referring to the work of Murphy (1998). Line 25 : please define n and p Line 25 : The authors must explain and argue this sentence. Otherwise, this sentence has to be removed.

Page 2898 Line 10 : replace “round” by “repetition to be consistent with the previous sections. Line 24 : I would say, “Part of the results in Fig. 3 are also discussed later in Sect. 5.2. Page 2898 Line 1 : Please add an introductory sentence to this section. Page 2899 Line 9 : Start a new sentence with “consequently”. Line 11 : remove parentheses and start new sentence with “For instance, the AWS1 March as well as July time series consists of 186 values ...” Line 15 : The R(sigma) parameter should be defined in the model framework (section 4.3) Line 23 : What does “The largest coefficient uncertainties” mean ? Line 25 : Please add “the” prior to “largest” Page 2900 Line 8 : rephrase the sentence Line 16 : start a new sentence with “However,” Line 23 : “ENSO” must be defined as first occurrence. Page 2901 Line 3 : remove the second “for” in “...for April and for September ...” Line 12 : This sentence is too long. Please rephrase and try to split the sentence. Also, in “we suggest that variability must...” should be used with precautions. The sentence seems to be more likely a hypothesis than a suggestion Line 20 : It is not evident from Fig. 3 that the day-to-day variability is smaller for dry season than wet season. Line 27: please avoid repetitions further in the text “of where data is available for all three AWSs” and use Period 1, instead. Also capitalize “p” in period 1. Page 2902 : Line 8: what do the authors mean by “positive” in “... the cross validation results are more positive.” Do they mean less biased? Line 27 : replace “time length” by “duration” Page 2903 : The authors must define  $r^2$  at first use. Line 8 : The authors must discuss this issue in the concluding section rather than here.

Line 13 : replace “available” in “observations n available” by “required” observations ”

C1428

Line 25 : replace "time scale" by "temporal resolution" to be more consistent with the text. Page 2904 Line 1 : To what values the authors refer to ? I guess SS values ! And be consistent, use "SS values " instead of "values of SS". - Lines 13-15 : This is a well-known behavior from previous studies ! Line 22 : please use italic font type for abbreviated variables. Line 24 : I am concerned about this choice because in the re-ens-air the authors have chosen the closest grid point to the study area instead of the 4 surrounding grid points. This may introduce difficulties in comparing the results as the main hypotheses differ. Line 28 : How can the authors "observe" inhomogeneity. Use "found" instead of "observe" Page 2905 Line7 : I could not find the results. What is the statistical test used here? Line 24 : Please correct the reference here. "Sect. 9" does not exist! Page 2906 Line 1: Again, "Sect. 9" wrong reference Lines 16-20 : long sentence. Please split it in two Page 2907 Line 3 : remove "s" in "emphasizes" Line 5 : be consistent, use "the model skill" in "...differences in skill..." Lines 15-17 : What does "model surface" means ?

Page 2914 Use ERA-interim instead of ECMWF interim.

Page 2915 In the table caption, the authors must be consistent with the text in defining the different periods. Period 1 and Period 2 should be clearly defined in the text and referred to in the table caption.

Page 2916 There is no section 9. Use ERA-INTERIM instead of interim. And please provide information about the optimum spatial domain in the table and refer to the work of Hofer et. Al (2012)

Page 2917 The authors must provide a title in the caption figure first. Then give details after. Please add "as" to "mentioned in the text"

Page 2918 Please replace "for each moth of the year" by "Monthly statistics of daily temperature..."

Page 2919 Replace "in March" by "for March". The authors should be aware that the

C1429

shaded area is a moving area and precise to what cv values the figure correspond. I do not see from the figure how the vertical dashed lines indicate the minimum number. It is confusing to have one line between 2007 and 2008 in March (top) and between (2010 and 2011) in July (bottom).

Page 2920 The caption of the figure must read "Box-plots of the downscaling model parameters. . .). Then, explain the different components in the caption detail.

Page 2921 The x-label must be  $n_{min}, n_{eff}$  [days] to be consistent with notations in the main text. The authors may also include the results for the other AWSs using different colors and line types.

Page 2922 Period 2 has not been defined in the text! And remove "respective"

Page 2923 The color scale bar in this figure is confusing. I wonder whether the authors must use 12 different color instead of using a gradually color bar. It is really difficult to distinguish between the colors for Apri-May-June within the different time scales.

Page 2924 Section number is missing!

Page 2925 Same as previous figure, the section number is missing from the caption. Please replace figure x-label with  $r^2$  to be consistent with the text.

---

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

C1430