Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-168-RC1, 2016

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "The downscaling and adjustment method ADAMONT v1.0 for climate projections in mountainous regions applicable to energy balance land surface models" *by* D. Verfaillie et al.

Anonymous Referee #1

Received and published: 3 November 2016

Review of

The downscaling and adjustment method ADAMONT v1.0 for climate projections in mountainous regions applicable to energy balance land surface models By D. Verfaillie, M. Déqué, S. Morin, and M. Lafayesse Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-168, 2016

General remark

The present work is potentially relevant, representing a method for adjusting RCM output to the conditions in mountainous environments. I particularly like the approach

Printer-friendly version

Discussion paper

to show the immediate consequences of the method's performance (with respect to the individual meteorological parameters) for energy balance-based land surface models, as done in this study by evaluating snow depth results obtained by driving CROCUS with output by ADAMONT.

Besides, I have several major concerns. The reanalysis data set the authors use as a reference is very specific, as it includes average conditions within different mountain massifs for different altitudes. As such, it is hard to see the relevance of this study in a broader context, for example, when focusing on applications that use more common observation data sets (e.g., local-scale observations, or observations on spatially regular grids). With this regard, the terminology "downscaling method" may also be inappropriate (see my comments below). The authors need to discuss potential implications of using their method for observational data sets other than the SAFRAN. I also have important concerns regarding the evaluation of the results. The authors do not show the performance of the method based on independent data. This affects the entire discussion and the conclusions. Also, the evaluation is not performed at the scale of the application (individual massifs), but at a larger scale (Northern Alps, Southern Alps). However, for the application in energy balance based land surface models we are interested in the skill of the method of reproducing more local-scale conditions. I recommend considering the study for publication in Geoscientific Model Development if the authors perform a major revision.

Major comments

1. The authors do not perform an evaluation based on independent data. They use different learning periods, but as far as I can follow the validation is based on all the available data (thus, including the data used in the training). I acknowledge the importance of considering different learning periods, but in each case, the validation should be based exclusively on data which have not been involved in the learning process. This point affects the entire discussion section and the conclusions (incl. the abstract) of the submitted manuscript. See also my specific comments below.

2. Grid point selection: The role of the SAFRAN massifs' centroids location and elevation is not clear. Apparently, using one location to represent an entire massif represents a simplification. Also, a potential altitudinal bias between the centroids' altitudes and the altitudes of the RCM grid point is not representative for the elevation differences apparent in reality. Not exactly knowing the SAFRAN reanalysis, and the definition and importance of the centroids, it is hard to understand why RCM grid points should be more realistic if they correspond in altitude and location to the SAFRAN centroids of the massifs, given that SAFRAN represent a simplification per se, assuming horizontally homogenous conditions for the entire massifs (thus areas ranging from 500 to 2000 kmˆ2). Overall, it is not recommended in statistical downscaling to use single grid points by atmospheric numerical models as predictors, because single grid point data are affected by numerical noise (see also: the concept of optimum scale, or effective resolution of an atmospheric numerical model). Also, data by at the RCM surface may be outperformed by the respective data extracted from the relevant pressure levels (see, e.g., Räisänen and Ylhäisi, 2011, Hofer et al, 2012, and references therein).

3. The term "downscaling method" for the ALADIN procedure is somehow misleading. The RCM has a horizontal grid point distance of 12.5 kmˆ2, thus the individual grid cells cover areas of approx.. 150 kmˆ2, while the SAFRAN assume spatially homogenous conditions within each 300 m altitude band of each massif, which in turn can cover up to 2000 kmˆ2. To my understanding, it is thus not possible to define a horizontal resolution of the SAFRAN reanalysis, in traditional terms. SAFRAN may represent more realistic conditions, in particular regarding the altitudinal differences within each massif, than the RCM. However, "downscaling RCM data to SAFRAN" is certainly not what is intended by the term "downscaling" used for inferring higher-resolution information by coarser-scale atmospheric numerical models in the scientific community. Please clearly discuss the practical differences between fitting a RCM to the SAFRAN reanalysis vs. downscaling a RCM to, e.g., to higher resolution gridded observations. I don't know any reanalysis data set comparable to the SAFRAN reanalysis. I recommend to discuss the implications of using ADAMONT based on other observational data sets as

are more commonly available. Otherwise this study is important only for a very narrow range of applications (i.e., applications based on the SAFRAN data set).

4. The discussion of the results is too lengthy. I see the importance of showing various evaluation criteria. However, the information should be compressed and presented in a more synthesized manner. Also, there are too many figures and the figure fonts are too small. Try to highlight the important points using figures which summarize the results in a more transparent way.

5. The article needs to be proofread by a native English scientist. The language needs improvement.

Specific remarks

Lines 175-178: This sentence is not clear. At this point, the reader does not know what the authors mean with “adjusted” RCM. Particularly, what do you mean with “there is no risk of introducing any artificial inflation of the simulated series”?

Lines 185-186: This sentence implies an assumption. The sensitivity of quantile mapping to circulation may change in different climates. Please clearly distinguish the terms “weather regimes” and “climate”.

Eq. 1: Indicate the value of N you used. How was N determined?

Lines 199-200: Provide more information about the clustering method you applied, since it is a crucial step in the downscaling procedure. For example, why exactly four weather regimes?

Line 205: What does the definition of the snow year matter for the downscaling procedure at this point?

Lines 208-213: This step is not clear. Is the selection of SAFRAN dates for each RCM date unique? How is this step automatized? Do you consider autocorrelation in the SAFRAN time series to avoid artificial jumps? Thus this step imply a reordering of

the daily RCM time series in order to best correspond to the temporal ordering in the SAFRAN data?

Line 214 and below: The differences between the daily and subdaily ALADIN values should be as small as possible, how is the relation between the SAFRAN subdaily and the ALADING subdaily values? You may provide a formula which describes the transfer. The equations provided concern only air temperature, not the other variables, and it is not clear in which way the SAFRAN sub-daily values are considered.

Line 246: This step is confusing. So you put two quantile mappings on top of each other?

Line 258: Please clearly define what you evaluate. As far as I can follow, you want to evaluate the output of ADAMONT. However, you repeatedly mix "output of ADAMOMT" with the terms "RCM" and" adjusted RCM". For example, line 286: "ratio of the standard deviations between the RCM time series and SAFRAN". Do you really mean standard deviation of the RCM? Then this criterion is not indicative for the performance of ADAMONT, but for the performance of the selected RCM grid point without any adjustment. Further, whenever you use "adjusted RCM" I am not sure if you mean the ADAMONT output or some intermediate step in the downscaling procedure. If you evaluate the hourly output by ADAMONT applied to ALADIN, please say so. All over the manuscript: don't use brackets inside brackets (e.g., in the references, line 265-266)

Line 270: So the method consists of "downscaling" values at a massif scale (with downscaling not necessarily being the appropriate term), but then the evaluation is not performed at the massifs' scale, but at a much larger scale. The evaluation needs to be applied at the scale of interest, in this case, the individual SAFRAN massifs. Then, the resulting (numerous) scores may be synthesized (e.g., box plots of scores resulting for individual massifs for the Northern alps, box plots of scores resulting for the individual massifs for the Southern alps). The same for the altitudinal ranges. There are various ways how to summarize and appropriately illustrate performance metrics

for numerous cases (here, variables, massifs, altitudinal ranges, and variants of the method in terms of learning period, grid point selection, a posteriori corrections, ...). However it is important that the performance metric is applied to the scale of interest (and: that the reader does not need to interpret too many figures, see also my major comment above).

Line 286: seasonal average time series is not an evaluation criteria per se. Mean annual cycle and mean altitudinal gradient: the same. Please be more specific.

Line 287: RCM time series or ADAMONT time series?

Line 290: not an evaluation criterion.

Line 317: Is "analysis of different massifs" is limited to the application of scores to the average conditions in the Northern and Southern alps, correctly?

Line 320: Following the authors description, their evaluation is never independent from the training data. I.e., for the three different learning periods applied, performance metrics are always calculated including the training data. Case 1: Training period: 1980-1994, Case 2: training period 1995-2010, Case 3: training period 1980 to 2010. Evaluation period is always 1980-2010, thus always includes the training data. The performance metrics should be calculated based on split – sample validation (if parametric properties are validated, e.g., biases, distributions, e.g., Cannon), or cross-validation (if the temporal sequencing is validated, e.g., mean squared errors, ect.). Otherwise, the validation has no evidence.

Line 462: term "neighbor selection" technique is misleading. Use "grid point selection" instead.

Figure 3 is a bit confusing. I see only one SAFRAN centroid being linked with the closest grid points in x, y, and z. Again, I am not sure what exactly you mean with "adjusted RCM". Is this the final output by the ADAMONT procedure based on ALADIN? Please be consistent with the terminology for the output.

Figure 16: It is hard to make any conclusions based on a visual inspection of Figure 16. Plotting the deviations from the modelled against the observed cumulative PDFs could help. The same with Figures 17-19.

Figures 20 – 22: It is hard to distinguish amongst the different lines. Again, it could help to plot deviations of the model results to the SAFRAN values. Still, there are too many figures and the information content should be compressed (e.g., in terms of summary statistics for each model option, e.g., boxplots of skill scores).

References:

Cannon, A. J., 2016: Multivariate bias correction of climate model output: Matching marginal distributions and intervariable dependence structure. Journal of Climate, 29 (19), 7045–7064, doi:10.1175/JCLI-D-15-0679.1, URL http://dx.doi.org/10.1175/JCLI-D-15-0679.1, http://dx.doi.org/10.1175/JCLI-D-15-0679.1.

Hofer, M., B. Marzeion, and T. Mölg, 2012: Comparing the skill of different reanalyses and their ensembles as predictors for daily air temperature on a glaciated mountain (Peru). Climate Dynamics, 39, 1969–1980, doi:10.1007/s00382-012-1501-2.

Räisänen, J., and J. S. Ylhäisi, 2011: How much should climate model output be smoothed in space? Journal of Climate, 24 (3), 867–880, doi:10.1175/2010JCLI3872.1.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-168, 2016.

Printer-friendly version

Discussion paper