Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-399-RC1, 2016

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "A Combined Statistical Bias Correction and Stochastic Downscaling Method for Precipitation" *by* Claudia Volosciuk et al.

S. Kotlarski (Referee)

sven.kotlarski@meteoswiss.ch

Received and published: 9 October 2016

The work by Volosciuk et al. presents, applies and evaluates a two-step bias correction and downscaling method for daily precipitation. The method consists of the separate application of a parametric quantile mapping (QM) implementation to correct for systematic biases at the spatial scale of the underlying climate model. Second, a vector generalized linear model (VGLM) is applied for stochastic downscaling to the point scale. The approach is applied to a set of 86 stations covering the entire European continent to bias-correct precipitation simulated by a reanalysis-driven regional climate model (RCM). Both steps of the procedure are evaluated separately and also in their combined application. For the latter, the five-fold cross validation framework developed

Printer-friendly version

Discussion paper

within the COST Action VALUE is employed. The results indicate an improvement of simulated precipitation characteristics – measured by a CDF score, a score targeting high-intensity events and by the spatial autocorrelation function – by the new method for many cases, but also apparent problems and a deterioration with respect to raw RCM or only bias-corrected RCM output for others.

The manuscript is an excellent piece of work that is highly relevant for the growing climate downscaling community as it introduces a new bias correction and downscaling method that could be employed in a larger frame. The concept to separate bias correction from downscaling is likely to provide advantages compared to simpler approaches, although these advantages are actually not shown (see below). The methods, the underlying data sets and the results are for most parts described appropriately. The quality of presentation is high. The conclusions are – with minor exceptions – properly based on the results obtained. The are no language issues. The work, however, suffers from a number of minor and one major issue that should be improved before final publication of this work. My main concern relates to the missing comparison of the new approach to the widespread “standard” application of QM to both bias-correct and downscale in one step. Please see the listing below for further details. I’d therefore recommend to return the manuscript to the authors for minor revisions. I hope my comments are considered as constructive and will help to improve the (few) weak parts of the paper.

With kind regards, Sven Kotlarski

MAJOR ISSUES

The main motivation of the study relates to the deterministic nature of standard MOS techniques such as QM and their inability to accurately reproduce local-scale variance that is not explained by the actual large-scale predictor. The developed approach is designed to improve on this by separating the bias correction from the downscaling step and by introducing stochasticity into the latter. The performance of both steps and of the combined scheme is evaluated and compared to the performance of (1) raw RCM output and (2) bias-corrected RCM output (bias correction at the resolution of the RCM). What is missing, however, is a comparison to a “standard” QM application that directly bias-corrects and downscales from the grid cell to the point scale (i.e., the first step of the approach directly targeting the stations series instead of the EOBS grid cell). In my opinion, only such a comparison can show the advantages of the new approach compared to standard applications. This is essential in the light of its apparent problems (at many stations the performance of the two-step procedure is worse than raw RCM or bias-corrected RCM output). I'd therefore suggest to include a fourth dataset in the evaluation of the combined approach, i.e., QM of step 1 directly applied to the station series. This might require to rethink the choice of performance metrics, as I'd expect neither the CvM score nor the 95% score to reveal the advantages of introducing stochasticity. Only the spatial autocorrelation might show such improvements. I'm aware that the suggested extension will to some extent be covered by upcoming VALUE papers, but given the motivation of the two-step approach this comparison is essential for the present paper in my opinion.

MINOR ISSUES (p: page, l: line)

Temporal scale of the model calibration: Evaluation results are presented for both winter (DJF) and summer (JJA), but the temporal scale of the calibration of the two steps remains unclear to me. Have all mdoels be fitted separately for winter and summer, or for the full year, or for every doy-of-the-year with a moving window? This information is a detail, but should be provided.

Extension to further variables: Climate impact studies often require more downscaled variables than precipitation only. Could the presented approach be applied to other variables (e.g., temperature) as well?

Performance measures and model selection: An overview of the performance measures (CvM, 95% score, spatial autocorrelation) is not provided, and most information

Printer-friendly version

Discussion paper

has to be taken from the appendix. It would be helpful for the reader to clearly state in Section 2 or 3 which performance measures are applied. Details can still be covered by the appendix. Also, the “model selection” description is a little scattered and hidden. An additional paragraph in the main part of the manuscript clearly outlining the selection strategy (step 1, later on also the best-performing overall approach) would be helpful.

p1 l19: I'd suggest to replace “precipitation data” by “precipitation projections” as this statement primarily concerns future scenario series.

p2 l18: “leading to too smooth variance in space and time” -> I'd suggest to slightly extend this part and to describe the problems of inflation a little more detailed. Readers not familiar with this issue might have problems to understand this limitation. In my opinion, the variance in time (if measured by the SD) should be rather well captured by QM through inflation.

p4 l13: Please check: This 0.44° RACMO simulation might not be the one that is used in the VALUE perfect predictor experiment mentioned before (there it is the corresponding 0.11° simulation to my knowledge). If so, it would be beneficial to briefly mention this fact as it would prevent an inconsistent comparison with the VALUE results.

p4, l27-32: The new method is presented as being versatile and applicable to free-running GCMs/RCMs. While this is true in general, it does not apply to the grid box selection step that relies on temporal correspondence. This should at least briefly be mentioned.

p6 l26: “bias corrected RCM precipitation” (and further occurrences in this section) -> I'd suggest to more generally speak of “coarse-scale precipitation” as not always bias-corrected precipitation is used as predictor (in the evaluation of the second step it is EOBS precipitation, and in the evaluation of the full setup it might also be raw RCM precipitation depending on the selected model).

p8 l5-18: In short, can the remaining inaccuracies of QM be related to non-stationary correction functions (which would show up in the applied cross validation framework)? Section 5.2 (Evaluation of the second step): To evaluate the second step, the modelled precipitation CDFs are used for the QQ plots which requires a standardization to a stationary gamma distribution. Why not using the same framework as for the evaluation of the combined approach, i.e., drawing 100 realizations of precipitation series and then computing the respective percentiles? This would be more straightforward. Also, why is the evaluation of the second step not carried out within the cross validation framework but within a scheme where calibration and validation periods are identical?

p12 l21-34: It remains unclear how the spatial autocorrelation has been computed. Based on seasonal means? Or separately for each day and then averaged?

p14 l14-16: This statement is a little overconfident given the results previously shown. In many regions VGLM is comparable, sometimes even inferior to raw or bias-corrected RCM (as said above, an extra comparison to QM directly onto the station series would be very beneficial here). Concerning the spatial autocorrelation, at least in DJF I wouldn't speak of an improvement by VGLM.

p14 l16-18: This sentence requires a proper reference (this aspect is not covered by the work present).

User guidance: Given the remaining apparent problems of the new two-step approach at many sites (inferiority even against raw RCM data), some summarizing guidance would be helpful on when to use the new scheme and when this is considered critical. This guidance should account for the fact, that in a free-running setup which is the setup for climate scenarios, not all parts of the presented evaluation can be carried out beforehand (no temporal correspondence which prevents the use of the same strategy for grid box selection and to some extent also the cross-validation setup with short time slices).

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-399, 2016.