Review of ’Reply to Comment on ’Bias Correction, Quantile Mapping and Downscaling. Revisiting the Inflation Issue’ ’ by D. Maraun

General remarks

My comment focused on a specific erroneous argumentation in [Maraun, 2013] about quantile mapping and trends. The comment also had a few remarks about the inflation vs. randomization controversy, with a clear viewpoint that both are imperfect. Although my main critique about trends is unrelated to inflation Maraun’s reply revisits the inflation issue full scale, not without introducing extra issues as outlined below. The main critique, however, remains unresolved.

The inflation issue, whose role in [Maraun, 2013] is questionable anyway (see below), should perhaps have been toned down in my original comment, and has been so in the revised version. The now very large discrepancy between comment and reply is of no help to any reader and is just confusing. Accordingly, a reply should be published only if it addresses and focuses on the mentioned trend issue. Extra remarks on inflation are nevertheless attached below.

Specific remarks

The trend argument in [Maraun, 2013] is as follows: To dismiss quantile mapping (QM) an example is given where grid-scale trends differ from QM downscaled local trends, concluding that “as trends are affected, changes in future mean and extreme precipitation [...] are likely to be misrepresented”. I have three specific remarks: First, since RCM grids represent areal averages, to disprove QM this way obviously requires to show that (i) a single station trend is representative of the areal average trend and (ii) the areal average differs from the RCM grid. By giving a number of examples my comment demonstrates that (i) is in fact wrong, invalidating Maraun’s argument that QM misrepresents trends. This main point is not addressed in Maraun’s reply. Second, his reply turns this demonstration into me claiming that refuting QM would entail the equality of local and large-scale trends. This is but one of several rhetoric figures that I found quite annoying (see more below). And finally, neither [Maraun, 2013] nor the reply mention that the trend issue is unrelated to inflated regression.

Towards the end Maraun presents a regression exercise with an unknown driver ($u$). Here for the first and only time trends are discussed explicitly. The example serves two purposes (indicated by corresponding line numbers): to deal with my criticisms on trends (l. 42) and to address the potential climate signals mentioned at the end of my comment (l. 197). These are completely different things, however, which becomes apparent from the simple fact that the first depends on the presence of multiple stations and second does not. The former case is actually quite revealing for orographic precipitation when the grid cell ($x$) is mostly windward and the station ($y$) in question is leeward, as
follows. Suppose $x$ has a positive trend, caused by a slowly shifting main flow, and the unknown driver $u$ representing the leeward part of the cell has a negative trend. Using regression one does not even need $u$ because the negative trend in $y$ can roughly be predicted via some negative $x$ coefficient. Calibrating regression on trended data is of course problematic, but it is all one has in this case. The example does not reveal anything, though, in terms of inflation or randomization.

**General remarks pertaining to inflation**

In his reply, much more than in [Maraun, 2013] itself, Maraun revisits the inflation issue quite broadly and tries to dismiss the idea on several grounds. Maraun’s main point is that for regression-based downscaling, the missing variance should be dealt with by employing probabilistic methods. To prove his point he introduces a simple stochastic model and calculates the probabilistic Brier skill score for the uninflated, inflated and the “correct probabilistic” version of regression. In a second point Maraun tries to address my suggestions how to deal with the missing variance problem in general. For this an unknown trend predictor is added to the original regression example, and the sensitivity of the regression to the strength of the trend are discussed.

The reasoning about the stochastic example is circular (see details below). Of course, if one employs a stochastic model with a well-defined noise characteristic and obtains probability estimates from that same model, then the entire estimation problem is circumvented and the results must be perfect. In a real world case, a model always differs from nature, and simulated probabilities have to be estimated and are imperfect. The ASD model [Hessami et al., 2008] represents a fairly advanced probabilistic regression approach of the type Maraun has in mind, but its performance to reproduce present climate is less than perfect [Bürger et al., 2012].

**Minor comments**

27: *inflation “a concept that has been known to be flawed for a long time”* - Glahn and Allen (1965) do not claim nor show that inflated regression is flawed. By analyzing the utility matrix of a categorical forecast situation they clearly state that it depends on the user needs whether over-prediction (inflation) is asked for or not. Their main conclusion is: ’it cannot be said whether “inflated” forecasts are more (or less) useful than uninflated ones.’ - von Storch (1999), on the other hand, claims but has never ’shown’ that inflation is flawed (or more flawed than randomization).

33: *Bürger (2013) “criticizes the concept of randomisation”* - That is again misleading since I clearly advocate the view of both inflation and randomization being imperfect.

61ff: The allusion to various statistical tools here conveys an impression of a magic bullet. The real-world practical downscaling context is one of imperfection: time series are too short (and error-prone) to establish with certainty their true distributions (Gamma? log-normal? Weibull?), the non-linearity or optimum number of predictors, their coupling, etc. - As a result, the basic regression assumption is always an approximation.

77-99: *prediction vs. simulation:* - It is unclear why one needs to stress the difference between prediction and simulation here. As acknowledged even by [Maraun et al., 2010], prediction is one
cornerstone in the *validation* of downscaling, and *applying* the downscaling model is simulation. So what?

111ff: Here the general problem of precision vs. bias is addressed in a somewhat cumbersome and misleading (inflation is not a 'mis-calibration') way. Accuracy and reliability are equivalent here.

120ff: **probabilistic assessment** - This is now very convoluted because Maraun mixes at least three independent errors into the probabilistic assessment of uninflated, inflated, and the “correct probabilistic” version of the regression:

1. He refers to the Brier score (BS) as a proper score that discourages hedging (or ’cheating’), but actually uses the Brier skill score (BSS) which is not proper [Murphy, 1973].
2. The results would be trivially true for a proper score since, by definition, the true probabilistic forecast scores best (so that hedging becomes impossible).
3. The entire argument is circular: using Eq. (1) as a stochastic process and predicting it as such must give the best results and proves nothing.

Moreover, all this is a mere repetition of Tables 1 and 2 of [Glahn and Allen, 1965] (i.e. overprediction).

124: Now X is suddenly a random variable? - I suppose (X,Y) are from a 2-dim Gaussian distribution.

135: **Regression worse than climatology** - This can only come from a badly estimated regression coefficient, otherwise this is inconsistent.

141: This is not ironical but a triviality, even more so since inflation and regression are not probabilistic forecasts.

154: **“First, correlation is not a measure of model error”** - I really find statements like these quite annoying. In the relevant Table A1 of [Bürger et al., 2012], we describe model error in terms of modeled variance, correlation, and explained variance. In that particular case with variance being adjusted, correlation, model error (rms), and explained variance are all equivalent, and we mention it.

161: **“The concept of explained variance applies to [...] but not to a simulation from the model.”** - From that it follows that regression models in general can or should not be validated using explained variance or any other (equivalent) measure.

167ff: Maraun somehow refuses to acknowledge that a ’correct representation of the predictand’ simply does not exist. Inflation introduces errors and randomization introduces other errors. Unless we know *exactly* that nature is a process following Eq. (1), with a known random process \( \eta \) (e.g. iid white noise) - and almost all of Maraun’s arguments seem to be based on that - there will never be a perfect solution to the missing variance.

185: **“inflation does not help to solve this problem ”** - This was never claimed, not by me nor by anybody who has used inflation.
"The key arguments of Bürger (2013) relate to trends and apply both for QM and inflated regression ." - Maraun identifies ’trend’ here with what was called ’climate signal’ in my comment, which is very confusing since trend was used differently earlier in his reply. But one cannot reduce regression models to trends only, as that defies crucial features such as synoptic weather types etc. that may form a climate signal even if appearing stationary under present climate. In the example given, since any two linear trends are linearly related, any trend in $u$ can be modeled through the trend in $x$. This applies to present as well as future trends (as long as they don’t change). And again, neither Maraun’s nor my key trend argument applies to inflation.

I don’t see where this has been mentioned in my comment.

"the statement that refuting inflation goes along with the assumption that local scale trends should be the same as large scale trends is wrong." - Here is the core of Maraun’s misunderstanding of my comment. To recapitulate, the relevant result of [Maraun, 2013] reads: “already strong trends [...] tend to get amplified by QM”, which later on serves to conclude: “as trends are affected, changes in future mean and extreme precipitation [...] are likely to be misrepresented”. This latter clause tacitly hinges on the assumption that trends should generally not be affected, and it is this assumption and its invalidity that I was addressing in my comment. Trends can be affected, that is, local trends can be different from the grid-scale average, a fact that is totally unrelated to downscaling. There may be more arguments against QM, but this specific one is being built upon an incorrect assumption and is therefore invalid. This is my main comment on [Maraun, 2013]. But in his reply Maraun reads it, in logical reverse, to mean that refutation of QM hinges on this assumption, and goes long ways to refute this mute point.

"Inflated regression does also not improve predictions such as weather forecasts." - I have never claimed that it does. It was clearly stated that regression forecasts deteriorate when using inflation or randomization.