Dear Dr. Bárdossy,

The authors would like to thank you for your valuable comments and suggestions. Below we respond to each point.

Reviewer Comment: This paper treats an important problem - how to bias correct RCM output to be used for possible impact assessment. The authors argue that the frequently preferred quantile mapping is under circumstances leading to unreasonable results - therefore another more robust method is needed. The scaled distribution mapping (SDM) suggested by the authors is a sophisticated version of the classical alteration of sequences by multiplication of precipitation and linear scaling of temperature. The method is reasonable but it is not proved that it is really better than others. Some artefacts are removed - therefore some others (not detected or not presented) are introduced.

Author Response: Following the logic that we have outlined in the paper, our method is proven to be better. To sum up our logic from the paper again: We first illustrate that it is not safe to assume that the error correction functions used in quantile mapping (QM) are stationary. And since the inflation/deflation of the climate change signal can be attributed to this stationarity assumption, we argue that QM lacks justification to alter the raw model projected changes. Furthermore, we propose that until a bias correction method can unequivocally demonstrate that it is beneficial to alter raw model projected changes, better performing methods will more effectively preserve these raw model projected changes after bias correction. This is precisely what we have demonstrated. We show how well different methods preserve the raw model projected changes to the leading three moments of the distributions of temperature and precipitation. However, we acknowledge that it would have been useful for us to provide statistical significance measures for the performance metrics. Using a t-test, we find the improved performance of SDM to be statistically significant for the leading three moments. We even find this to be the case for skewness which visually appears very close, though our sample size is quite large (~294,000 mean absolute error values = 3500 grid cells * 12 months * 7 outlook periods).

Yes, it is clear that QM has statistical artifacts as a result of the stationarity assumption. We have developed a method that is free of the stationarity assumption and thus removes the known statistical artifacts inherent to QM. We do concede that some statistical artifacts of a much smaller magnitude could be introduced in our method when scaling between distributions that are poorly fit to the underlying data (this is discussed in more detail in a later comment). Our revised paper will more clearly state the assumptions related to the fitting of the distributions.

Reviewer Comment: The section describing the impossibility of comparing bias correction methods is interesting, but at the same time the example is discouraging. If models for present climate have such biases in predicting the changes than how can we believe their changes for the future are reasonable? I missed more discussion on this problem.

Author Response: Unfortunately, the reviewer misinterpreted the results of this section. First, consider the following two types of bias: 1) the bias between modeled and observed values in some calibration period (the case in the paper is 1951-1980), and 2) the bias between the climate model projected changes and observed changes between the validation period and the calibration period for the leading moments (the case in the paper uses the calibration period 1951-1980 and validation period 1981-2010, for example observations show 1 degree of mean warming over this time period in a specific grid cell while the model shows 1.5 degree of mean warming over the same time period and the same grid cell). The first type of bias is something we can know, though we cannot be sure how it will change under non-stationary. The second type of bias is something we know in the example split-sample validation test. However, in real applications of bias correcting an actual future period (e.g., 2050-2100), we do not know which regions the model will more closely simulate the observed changes that have yet to happen. We do not know if there will be persistence in the ability of the model to correctly simulate observed changes in climate. Likewise, the literature provides no clear evidence that certain model projections should be weighted more than others. This is related to model selection and is beyond the scope of this paper. It is this second type of unknown bias that influences the performance of a bias correction method when using a split-sample test. As a result, we argue that a different type of test should be used to measure the effectiveness of a bias correction method.

Reviewer Comment: The problem physical consistency of bias correction was discussed by the other referee Uwe Ehret. I am surprised that the authors do not consider the problem of consistency on different spatial and temporal scales. From the title I expected discussions in this direction. Hydrological applications require spatial data, therefore spatial correlations should also be considered. Individual corrections do not mean that the correction is correct over different spatial scales. In a previous paper (Bardossy and Pegram 2012) we investigated the spatial coherence of RCMs and found significant differences between model and observations. This problem becomes more severe if the bias corrected output is applied in hydrology. The same also applies for temporal aggregations. Bias correction on a single scale at a single location is an interesting exercise, but may be biased on other scales.

Author Response: The authors gratefully acknowledge the reviewer for providing the valuable reference to the Bardossy and Pegram (2012) paper. It is true that we have proposed a bias correction method that performs grid cell by grid cell. Some users may choose to trust a model with respect to the spatial patterns of meteorological variables, and as a result, they would not match the modeled correlation matrix to that of the observed correlation matrix. However, since we perceive the goal of bias correction is to properly reflect the statistical properties of observations in the calibration period at a variety of scales, we do believe that prior to bias correction it is a wise choice to recorrelate the data as outlined in the reference. This recommendation and the citation will be added to our paper.

Reviewer Comment: The choice of 0.1 mm daily precipitation threshold is in my opinion not appropriate. Precipitation amounts between 0 and 1 mm are very inaccurate in measurements.

They may even contaminate the estimation of the precipitation distributions. I would suggest to use a mixed approach fitting only above the 1 mm limit.

Author Response: We have used a threshold which is the threshold of non-zero precipitation amounts in the E-OBS data set. Different users have different needs. We have used 0.1 mm as a threshold, but this is not a fixed assumption of the method. We will more clearly state in the paper that an appropriate threshold (e.g., 0.1 mm, 0.5 mm, or 1 mm) can and should be chosen by the user to fit their needs.

Reviewer Comment: The gamma distribution for daily precipitation is a reasonable choice, but not for the extremes. It is anyhow very unusual to use the name return periods for relatively frequent events. Please use another notation.

Author Response: Our primary focus was to present a method that preserved raw model changes across the entire distribution. While the method is succeeding in this respect, the choice of a gamma distribution may not always be appropriate. In particular, for extreme events the scaling between return periods can and will often be different depending on the distribution used (e.g. using a weibull instead of a gamma distribution). We need to more clearly state that we are illustrating the method by using a gamma distribution for precipitation, though a user should use the distribution that is most appropriate for their case.

With respect to the name ‘return periods,’ our method deals with scaling all parts of the distribution and not only the extremes. Technically, we are talking about and using return periods. However, if this is too confusing for the reader, perhaps we could use ‘recurrence interval’ instead.

Reviewer Comment: The paper is very difficult to read. In my opinion the methods presented in the paper are described in a difficult to follow manner. I spent a lot of time to understand, and finally found that the methods are not very complicated themselves. I personally do not like equations written in a programming language style. There is a correct mathematical description with equations and not using words like sort etc.

Author Response: We specifically wrote the paper and the equations in a style that attempted to be transparent and makes it as easy as possible for a reader to implement themselves. In the end, our method is simply scaling the observed distribution by the amount by which the future modeled distribution is scaled from the historical modeled distribution. Though this is simple in theory, the application becomes a little less elegant. We have outlined the method in a way that we perceived to be appropriate and reader-friendly. If there are more specific suggestions that the reviewer has to improve the readability of the paper, we would enjoy receiving this feedback.

Reviewer Comment: In conclusion I found the paper interesting but not well presented and not addressing important issues. Therefore I suggest a major revision.

Author Response: We respectfully disagree with the reviewer that our paper is not addressing important issues. The authors took the time to understand and implement the spatial recorrelation method described in Bardossy and Pegram (2012). We did this for each month in a 30 year historical time period using bias corrected model precipitation data in a 10 by 10 grid (seen outlined in Figure 1). Similarly to the paper, we find that the model we used underestimates the average correlation between all grid cells with respect to observations. The correlation matrices of the observations, bias corrected model, and the recorrelated bias corrected modeled are shown in Figure 2. After adjusting the bias corrected modeled values using the recorrelation method outlined in the paper, the correlation matrix aligns with the observed correlation matrix (left and right subplots of Figure 2). The spatially averaged daily amounts of modeled precipitation are underestimated for high values and overestimated for low values (though the mean of all these daily values are the same because here we are using bias corrected data). We found the most extreme case of underestimating high values to be for the month of May. Figure 3 shows the sorted observed values versus the sorted modeled values for the month of May in this 10x10 grid before and after the recorrelation. The left subplot shows the shows a systematic underestimation of the uppermost quintile of the spatially averaged modeled precipitation versus observations prior to recorrelation (underestimation because it is below the black 1-to-1 line, the red line is a least squares fit to the data). The amount of the model underestimation averaged across the quintile is approximately 8%. The right subplot shows that with the adjustment, there is no longer this systematic underestimation for this upper quintile (the red line is now obscured by the black line). To reiterate, we find that by not having the correct spatial scale of precipitation events, which can be characterized by the correlation matrix, we would underestimate the spatially averaged precipitation in the upper quintile by 8%. In our paper, it can be seen in Figure 7d, that QM will overestimate the spatially averaged May precipitation by 9% purely as a statistical artifact due to non-stationarity. The magnitude of the problem addressed by recorrelation and the problem of non-stationarity is on the same order. Our proposed method does not rely on this stationarity assumption, and as a result, we remove the systematic inflation/deflation that is directly the result of QM statistical artifacts.

The authors would like to thank the reviewer again for your thoughtful comments.

Yours sincerely,

Matthew Switanek and coauthors

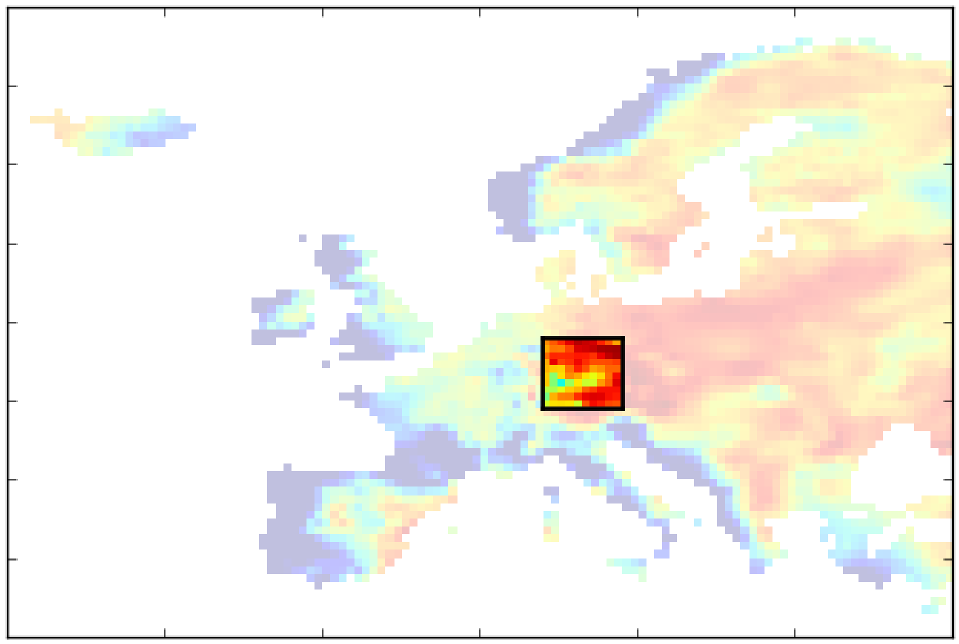

Figure 1: The subregion of 10 by 10 grid cells that we use to implement the recorrelation methodology.

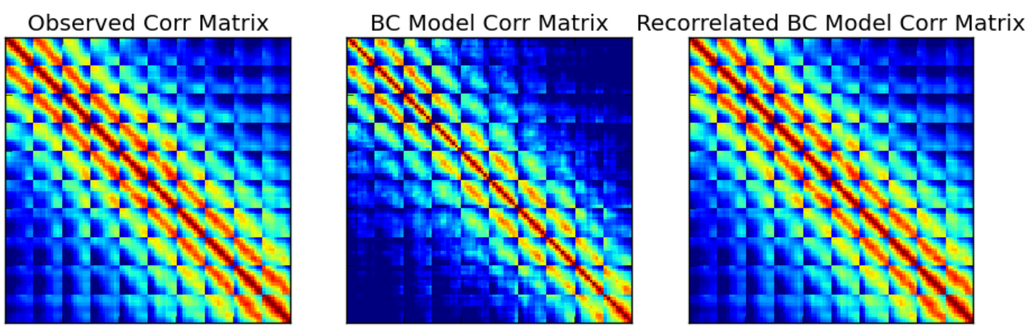

Figure 2: Correlation matrices shown with the colorbar showing correlation coefficients between 0.2 (blue) and 1.0 (red).

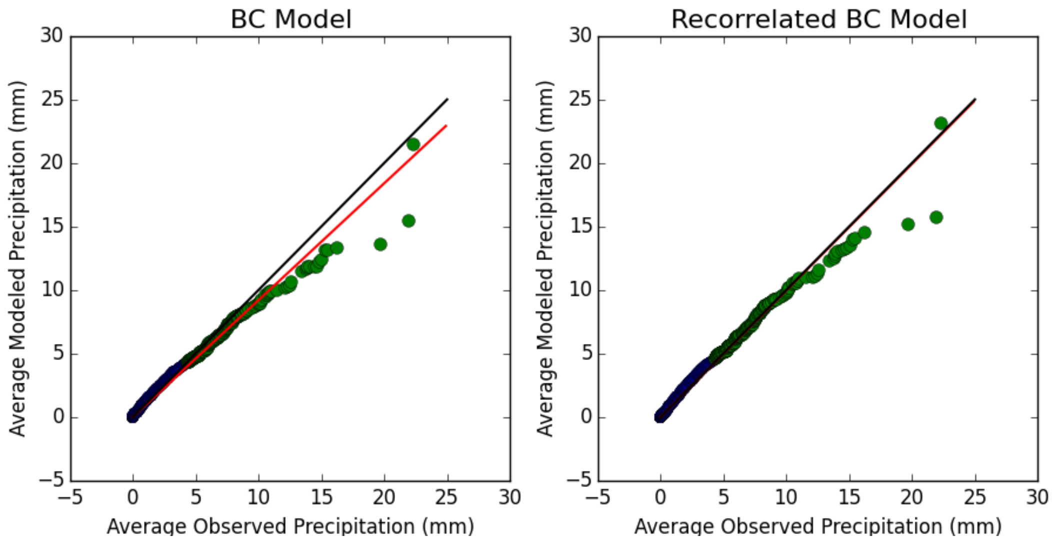

Figure 3: Upper quintile of values prior to and after recorrelation. Black lines are 1-to-1, and red lines are fitted least squares lines.

Bárdossy, A. and G. Pegram, Multiscale spatial recorrelation of RCMprecipitation to produce unbiased Climate Change scenarios over large areas and small, Water Resources Research, 48, W09502, doi:10.1029/2011WR011524, 2012