Hempel et al. recognise an important problem of quantile mapping and other bias correction methods aiming to correct for misrepresented variance. However, the solution presented by the authors still suffers from the same fundamental mis-application of these methods that actually cause the described problem: quantile mapping cannot downscale - in the sense that quantile mapping does not add small scale variability. Quantile mapping is a deterministic transformation and therefore designed to correct systematic errors, it does not add small scale random variability not explained by the uncorrected numerical model. This problem is exactly the same as that caused by the ination of (perfect prog) statistical downscaling, which has been described more than 10 years ago (von Storch, 1999). The problems arising when quantile mapping is used to downscale have recently been described (Maraun, 2013): because the temporal structure is not that of the local scale but still that of the grid box scale, apart from the ination of trends, area aggregated extreme events are overrepresented and are aggregated dry days are over-corrected. These problems should remain even after the application of the modied approach presented by Hempel et al. (this should be checked). Imagine downscaling to several stations within a GCM gridbox: in reality, there will always be random variability between these stations, yet since quantile mapping is deterministic, they will always be perfectly correlated. A physical rather than a statistical argument is the following: the aim of downscaling is to add energy on small scales not resolved by the RCM. Yet quantile mapping instead increases the energy on resolved scales and thus causes the problems described (see gure). So my guess is that the method proposed here does not work for precipitation, as it does not tackle the underlying problem. However, it might work for temperature, where small scale variations are often rather systematic than random (e.g., caused by orography). The easiest way to investigate whether my criticism is correct is to consider area aggregated QQ plots as in Maraun, 2013, and also the length of dry spells of area aggregated precipitation.

We thank Douglas Maraun for these very helpful comments. We have modified the manuscript to underline that the described method represents a deterministic transformation that does not add the small scale random variability present on small scale in comparison to the higher aggregated GCM data:

"In many cases differences in the variance or even higher moments of the simulated data are adjusted to the observations by parametric or non-parametric (empirical) quantile mapping. Bias correction, however, must not be confused with a spatial downscaling. The correction of the misrepresented local variability is limited, since a disaggregation of the simulation data cannot be performed by a purely deterministic approach. If the resolution of the simulations and observations are considerably different high extremes are usually exaggerated while low events are overcorrected."

"While a simple interpolation to the finer grid would not account for the increase in variability expected for the higher resolution data,

an appropriate increase can be achieved by a bias correction method that adjusts the variance. In the general case, however, this adjustment will be limited since the temporal structure is still determined by the dynamics represented in the larger grid box and do not describe local phenomenons, e.g., small scale turbulence."

We have also tested our data for the introduced error that is now also reported in the paper.

"We focus in our sensitivity study on the range between between the 10% and the 90% quantile. For this central range bias correction methods are expected to perform well, while the correction in the outer ranges of the distribution is typically worse, since there are less events (Maraun 2013). In general bias correction methods tend to exaggerate extreme events, since the limited number of data points prohibits a robust analysis of the relationship between observations and simulations, potentially resulting in an overestimation of these events. In addition, the extreme events always cover the whole grid-box area, i.e. their spatial extent is typically too large. However, since we introduced an upper bound for the bias-corrected values in ISIMIP, the impact of this effect is not arbitrarily large. On the global scale the bias-corrected variables show good agreement with the observational data even in the tails of the distribution (cf. Supplement Fig. 4)."

In the Supplement Fig. 4 we provide aggregated QQ plots for temperature and precipitation which illustrate that the correction is worse for the outer ranges of the distribution, although this is not a crucial issue on the global scale.

The influence of the overestimated spatial correlation on the impact simulations clearly depends on the degree to which impacts at one grid cell are influenced by impacts or meteorological events in the neighboring grid cells that might be particularly relevant with regard to the hydrological models.

It is the primary goal of this paper to describe a bias correction method that preserves the long-term trends as an essential prerequisite for efforts like ISI-MIP and to report on limitations and shortcomings of the ISI-MIP data set that might be relevant for the interpretation of the generated impact projections. Therefore we are very grateful for the comments. The new evaluation of the artificially generated spatial correlation of the high resolution data (map of the variances in Supplement Fig. 5) will allow for a more realistic assessment of uncertainties that may arise even with or directly from the bias correction.

Furthermore, a row of relevant papers should be included in the discussion:
* Eden et al, 2012, discuss were bias correction directly from a GCM might in principle make sense. Not surprisingly, there are regions where the GCM simulated precipitation is so wrong that it cannot be taken as input for a bias

correction. * The authors mention the problem of bias nonstationarities. This has already been investigated by Raisanen and Raty, 2012 and Maraun, 2012, in a pseudo reality. * Much of the discussion in cited Ehret et al., 2012, is based on the review by Maraun et al, 2010. Therefore, the latter paper should be mentioned here, also as it provides a state of the art discussion of bias correction methods.

Thank you very much for these hints. We included the mentioned references at the following places in the introduction:

"If the resolution of the simulations and observations are considerably different high extremes are usually exaggerated while low events are overcorrected (Maraun,2013)."

"A review of state-of-the-art bias correction methods is given by Maraun et al. (2010)."

"Stationarity in the bias in the historical data with respect to the future data is assumed when appling the bias correction to future periods, which introduces additional uncertainty (Raisanen et al. 2012, Maraun et al. 2012)."

"However, a complete bias correction directly from the simulations may not be advisable everywhere (Eden et al. 2012), since there are regions where, e.g., the simulated precipitation is so wrong that a statistical bias correction with a transfer function may result in an even worse data set as some extremes are very much amplified in order to adjust the parameters of the distribution. Different thresholds are incorporated in our bias correction algorithm to restrict the modifications in such cases."

Eden, Widmann, Grawe and Rast, Skill, correction and downscaling of GCM-simulated precipitation, J Climate, 2012.

Maraun et al., Precipitation downscaling under climate change. Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 2010.

Maraun, Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums, Geophys. Res. Lett., 2012.

Maraun, Bias correction, quantile mapping and downscaling: revisiting the ination issue. J. Climate, Online First, 2013.

Raisanen and Raty, Projections of daily mean temperature variability in the future: cross-validation tests with ENSEMBLES regional climate simulations, Clim. Dynam., 2012.

von Storch, On the use of "ination" in statistical downscaling, J Climate, 1999.