Response to referee #2

We thank the anonymous referee for his/her comments and thoughts on our paper. Below we address all the comments. Please note that referee comments are shown in *italics* and our responses in plain font.

Major comments

1. The introduction is very long and can be shortened substantially. For example, the section on orography effects on precipitation is too long and detailed, and it also lacks references. For example, there is no mention of the all the regional climate models that downscaled ERA40 data (van der Linden and Mitchell, 2009, http://ensemblesrt3.dmi.dk/). Also, a paper should not have more than one main aim. The main aim of this paper is to develop and test methods to interpolate station data with the aid of downscaled reanalysis. The hydrological modelling is an application and evaluation of the data. However, I would urge the authors to stress more in the introduction what is the main purpose of the paper and why it is important. What is the novelty of the methods?

Reply: This study aims at evaluating precipitation estimates in data sparse mountain regions. In such regions, where the gauges are often not in representative locations in the catchment, both deriving areal precipitation estimates and cross validation of different precipitation estimates is difficult. In particular, one aspect of our study is to investigate whether monthly accumulated precipitation fields from downscaled reanalysis data can be used for the interpolation of gauge observations and to compare this approach to other methods. The second aspect is to evaluate different precipitation estimates by comparing simulated and observed discharge, which has the advantage that the precipitation estimates are evaluated at the same scale at which they are needed for hydrological modelling. Here the novel aspect is to further extend this approach in order to differentiate between errors in the overall bias and the temporal dynamics, and by taking into account different sources of uncertainties. Both aspects are strongly linked as the evaluation framework by hydrological modelling is required for the performance evaluation of the developed precipitation estimates. Therefore we argue that they should be treated in one paper.

Following the reviewer's suggestion, the introduction has been shortened in several places. It has been revised in order to emphasize the link between the different aspects of our study as mentioned above. In this respect and also in view of the comments of Reviewer 1, we now put more emphasis on making clear that, in this paper, we consider the calibration performance of a hydrological model to be an essential part of evaluating the different precipitation estimates.

2. You state in section 2.3 that a direct comparison between station data and model output from models are not comparable, but you still carry out the comparison. What does this bring to the analysis? Is there any value in such a comparison? In this comparison, I would also suggest to look at the distribution of the data series, as well as linear trends and anomaly correlations.

Reply: In the manuscript we do not state that station data and climate model output are not comparable, but that there are limitations to such a comparison. Even if the model was perfect, the model values would deviate from the gauge observations due to the fact that the model output refers to a 12 km x 12 km area, while the station data are point measurements. This does however not mean that such a comparison is meaningless. There are no suitable gridded observational data available for a comparison, as any gridded observational data in this region can also only be based on this very sparse station network. Thus a comparison to station data seems like the most reasonable approach for a comparison to observational data. Such a comparison could for example reveal very large biases or an inacceptable representation of precipitation seasonality. It would thus indicate that the model is not suitable for the study area so that using these data for the interpolation of station data would not be useful. Results for the comparison of the pdfs are not shown in this paper to keep this section shorter. In addition, we think that this would not really contribute to the objectives of this study, as for the interpolation method using downscaled reanalysis data only monthly data are applied. Concerning trends, please refer to our reply to major comment 5.

3. The discussion on using hydrological model to assess the quality of precipitation (2.4.1) is not clear to me. What is meant by the statement: "precipitation data set with larger differences from the true precipitation might lead to lower mean deviations between simulated and observed flow and would therefore be classified as the better one." And also the following statement on systematic errors is very strange.

Reply: We agree that without referring to Heistermann and Kneis (2011) these two sentences do not give sufficient explanation and expanded this part. It now reads:

"The Monte Carlo approach is for example applied by Gourley and Vieux (2005) and Heistermann and Kneis (2011). In this approach Monte Carlo simulations are carried out for each precipitation data set and the selected goodness of fit measure is calculated for each simulation. For each precipitation data set one then evaluates the mean goodness of fit over the whole or subsets of the Monte Carlo ensemble and ranks the precipitation data sets according to this value. An advantage of this approach compared to the calibration approach is that it easily allows evaluating the model for various subsets of the data, e.g. only for high or low flows. However, in some cases, particularly when parameters have a linear influence on the fraction of rainfall generating runoff and the precipitation estimates do not have random errors but a systematic bias, the Monte Carlo approach may lead to wrong conclusions. Heistermann and Kneis (2011) give the following example: Assume a very simple linear runoff model $Q = \psi \cdot P$ where Q represents the runoff, ψ the runoff coefficient with values between 0 and 1, and P the precipitation. Monte Carlo simulations with uniform sampling over the runoff coefficient are performed for the true precipitation data set and a second precipitation data set characterised by a constant bias. In the next step the root mean squared error (RMSE) between simulated and true discharge, which can be calculated from the true precipitation and the true value for the runoff coefficient ψ_{true}, is evaluated for each simulation. It can be shown (see Heistermann and Kneis (2011)) that in a system with $\psi_{true} < 0.7$ ($\psi_{true} > 0.7$) the mean RMSE of a precipitation data set with a negative (positive) bias is lower than for the true precipitation data set so that the biased precipitation data set would be classified as the better one. While very obvious ill-posed settings may be avoided by careful analysis of the model, less obvious cases may not always be avoided from the outset. "

Monte Carlo calibration of model parameters using different data sources is a good way of detecting systematic errors and to estimate the parameter sensitivity if the right objective functions are used. If goodness-of-fit (for example Nash-Sutcliffe) is plotted against a measure of the water balance, systematic errors can be detected.

Reply: For evaluating areal precipitation estimates by hydrological modelling, Monte Carlo simulations and calibrating the model to each precipitation data set are two possible approaches, which both have advantages and disadvantages. In this study, the model calibration approach was chosen as it is less susceptible to errors of false rankings and because optimisation approaches usually achieve better model performances with the same number of simulations than a Monte Carlo approach.

Also in an optimisation approach one could use the model without a precipitation bias factor, optimise the model for example against Nash-Sutcliffe and detect systematic errors by analysing the bias in the discharge. However, there are several advantages of using a calibrated precipitation bias factor for analysing a systematic over- or underestimation of the precipitation data set: i) If we used a model without precipitation bias factor and one compares two precipitation data sets of which one has a systematic bias and the other has no bias, one would be able to detect this from the bias in the discharge (given that the precipitation bias is larger than what can be compensated by the model parameters). However, as the biased precipitation data set will lead to simulations with a lower Nash-Sutcliffe value, it is now less straightforward to detect which of the precipitation data sets has a better performance with respect to the temporal dynamics. ii) If a precipitation estimate has a large bias and one wants to evaluate its performance with respect to the temporal dynamics, it is not advisable to directly use it as input to a hydrological model, as the whole system might be in a different state and thus function differently. iii) The calibrated precipitation bias factor allows to directly quantify the bias of the precipitation estimate. In the model without precipitation bias factor, one would not be able to directly infer the bias of the precipitation data set from the bias in the simulated discharge, as due to non-linearities of the system the bias in the simulated discharge is often different from the bias in the simulated streamflow.

We agree that this point was not sufficiently discussed in the original manuscript and improved this section in the introduction.

4. The inclusion of a "precipitation bias factor" is often used in operational hydrological forecasting, and it is a way to overcome underprediction of precipitation. However, it also affects the water balance, and if you have other parameters that affect the balance, such as evaporisation parameters, the parameter values are not independent, and changing one set of parameters will affect others. You argue in 2.4.1 the benefits of including such a parameter, but I believe that this has too much of an impact on the parameter values and the whole hydrological model exercise. This is apparent in Fig 13, where the parameters that are used to calculate evapotranspiration are directly affecting the bias factor. This means that you are removing water from the model, either by changing the ET parameters of the bias coefficient. There are other ways of investigating the bias of the data sets, for example by a simple scatter plot between NS and mean bias, as I suggested earlier.

Reply: Please see the reply to major comment 3, for the question why we chose to evaluate the bias of a precipitation data set from the calibrated precipitation bias factor instead of from the bias in the simulated discharge. Here we address the concern that the precipitation bias factor has a too large impact on the water balance and other model parameters (and thus on the whole modelling exercise). Possible interactions of the precipitation bias factor and other model parameters are considered by the calibration framework described in section 3.3.4, which takes account of parameter uncertainties. For our case it turned out that changes in the precipitation bias factor could be compensated by other parameters only to a very small extent, as shown by the small uncertainty range of the calibrated precipitation bias factor (Fig. 8 and Fig. 10) and the low correlations of the precipitation bias factor to other parameters (see section 4.2.1).

Uncertainties resulting particularly from uncertainties in the evapotranspiration module are addressed in the sensitivity analysis in section 3.3.5. We added more explanation to the motivation behind this sensitivity analysis. As uncertainties in inputs to the evapotranspiration module would likely have an influence on the water balance (and thus the precipitation bias factor), we directed this uncertainty analysis towards inputs to the evapotranspiration module. The factors for selected inputs which were introduced for this sensitivity analysis were not intended as calibration parameters. Instead we used the available data from literature, maps or observed or modelled climate data as best guess, and the order of magnitude of possible effects on the precipitation bias factor was analysed through this sensitivity analysis.

5. Why are you using a runoff model to look at precipitation characteristics as linear trends, such as in the WRF data? This should be obvious through a check for linear trends in for example annual means. ERA40 has problems with trends over certain regions, and this could be such an example. I would suggest adding this analysis to the precipitation evaluation.

Reply: The main objective for repeating the analysis for different time periods was not to identify linear trends but to evaluate the robustness of the precipitation factor and the ranking of the objective function with respect to the selected time period. A precipitation bias factor which deviates from one but shows little variation between different time periods is still somewhat unsatisfactory but could, for further modelling exercises, be handled in a relatively straightforward way. In contrast, if the precipitation bias factor varies between time periods, the overall uncertainty range is much larger than the uncertainty range indicated by the parameter uncertainty for one period.

We agree that the problem with the trend in the downscaled ERA-40 data can also be detected by analysing and comparing the trend in the downscaled ERA-40 and in the other data sets, particularly as the other data are derived directly from station data and it is known that ERA-40 data have problems with trends. However, if one compares two precipitation data sets where one would not trust one of them more than the other (like for example MLR-all versus MLR-ind), one cannot use one of them as a reference. Then we need the comparison between simulated and observed discharge to find out for which of the two precipitation data sets the precipitation bias varies over time. Besides, the different precipitation data sets could also show similar trends which are however not consistent with the trend in the discharge data. This can only be detected using a simulation approach. For clarification, we also added this point in the manuscript.

6. In your conclusion you state that evaluation of the bias factor adds an additional performance measure to the analysis, but I do not agree. The misrepresentation of physical features as well as linear trends in data should be easily detected with precipitation statistics and metrics. The added value of using a runoff model would be useful to distinguish more in detail between datasets, for *example how high flows; low flows are modelled et cetera. The conclusion that the deficit is due to a linear trend in the precipitation and not changes in the catchment is trivial, since there are no such evidence in the other data.*

Reply: By using the precipitation bias factor and the objective function value one can distinguish between the performance with respect to the overall bias and the temporal dynamics so that both of these indicators can be seen as complementary measures for the performance of the precipitation data set. If the model includes a precipitation bias factor, it needs to be considered for the evaluation - otherwise two precipitation estimates could be ranked equal if they performed equally well with respect to the temporal dynamics even though their overall bias could be very different. For the question why we chose to include a precipitation bias factor (instead of for example inferring the bias of the precipitation data set from the bias in the simulated discharge), please refer to the reply to major comment 3.

We agree that differing trends in the precipitation data can also be detected by comparing linear trends of the precipitation data itself, but this was not the primary objective of the analysis of the precipitation factor for different time periods. Due to non-linearities of the system, a hydrological model is needed if there are differing trends in two precipitation data sets (and it is not known from the outset that one of the precipitation estimates has problems with representing trends) and one wants to find out which precipitation estimate, or possibly both, are inconsistent with the trend in the streamflow data. Please also refer to the reply to major comment 5.

Specific comments

1. P 10720, L6. You say that the study has a second aim, but I would argue that this is not an aim to test it for hydrological modelling purposes, that is an application of the method.

Reply: Please also see the reply to major comment 1. We now revised the abstract in order to emphasize that the two different aspects of the paper are directed towards one overall aim. Both parts have novel aspects and they are strongly linked, as, due to the difficulties with cross validation in data sparse regions where the locations of the gauges are not representative for the catchment, the evaluation approach using hydrological modelling (aspect two) is required for the evaluation of the precipitation estimates (aspect one).

2. P10723, L10. I disagree that discharge measurements are inflicted with smaller measurement errors than precipitation. It might be true for flows close to mean flows, but for low flows and especially high flows, the uncertainties can be huge.

Reply: We agree that particularly for high flows streamflow measurements are also very uncertain. Still, in many cases, especially if a relatively large fraction of the precipitation is solid (like in our study area), precipitation measurement errors are likely to be larger than streamflow measurement errors. The sentence was therefore changed to “Under average flow conditions, discharge measurements are also usually afflicted with smaller measurement errors than precipitation measurements, especially if they contain a large fraction of snow measurements.”

3. P10726, Section 2.2.1. Why was ERA40 used? There are newer products out, for example ERA-Interim. Was it because of the available data for precipitation? And what was the data period for the data used?

Reply: Thanks for the hint to add the data period, which was missing here. The period 1959-1990 was chosen due to the availability of the precipitation data, after 1990 many precipitation stations were closed.

4. P10727, L1-9. Please describe the method to correct for undercatch by Yang el at.

Reply: We added a sentence explaining how the equations in Yang et al. are derived and directly refer to the equation numbers.

5. P10727. Please describe the interpolation method in 2.2.3 with equations. I found it difficult to understand the method from the text itself.

Reply: We now added equations which formally describe the interpolation method.

6. P10728. Please describe the description in section 2.2.4 with equations

Reply: The equations in section 3.1.5 (old 2.2.4) are the same as in section 3.1.4 (old 2.2.3) , only M_i and M_j now refer to the values from the regression equations instead to the monthly accumulated downscaled reanalysis data. We therefore refer to the equations in 3.1.4 (old 2.2.3).

7. P10728. Merge section 2.2.4 and 2.2.5

Reply: The description of the IDW method has been extended so that this section now is not as short as in the original manuscript. We prefer to keep sections 3.1.3 – 3.1.5 (2.2.3 - 2.2.5) separate so that each section describes one interpolation approach. The order of the methods has been changed – the IDW method is now described first, as this method is also applied within the other two approaches.

8. P10728. Please spell out the acronyms GTS, GHCN, NCDC andFAO.

Reply: This has been corrected in the manuscript.

9. P10734. Please mention the used time periods for calibration earlier in the text

Reply: The time period is now also mentioned in sections 2.2.1 and 2.2.2.

10. P10734 eq 1. As mentioned earlier, the used objective function is suboptimal, since it mixes two different aspects of the modelled discharge, and you therefore loose information of your system.

Reply: The bias is included in the objective function in order to force the bias to be low, as a larger bias in the simulated discharge would complicate the interpretation of the precipitation bias factor. In this study, a possible bias in the simulated discharge due to for example a bias in the precipitation data set was intended to be compensated by the precipitation bias factor, which is then used to evaluate the bias in the precipitation data set (please see the reply to major comment 3 for the reasons why the bias is analysed using the precipitation bias factor instead using the bias in the simulated discharge).

11. P10735. How much does the glaciers affect the water balance?

Reply: The impact of glacier melt on the overall water balance is relatively small. Based on previous model results for the studied catchments the contribution of glacier melt on the annual time scale is estimated to be less than 3 % of the total inputs to the system (from rainfall, snow melt and glacier melt).

12. P10736, L7. “at the time” should be “at a time”

Reply: Is now corrected.

13. P10737, L5-10. Changing just one parameter at a time is hardly a sensitivity test. And as discussed earlier and what is apparent from Fig. 13, the parameter values of the ET are highly correlated with the precipitation bias factor.

Reply: We acknowledge that changing one parameter at a time is only a very simple sensitivity test. The approach was chosen because the expenses for a full sensitivity analysis would be very high and was not doable in this study. However, the simple sensitivity analysis is still very useful, as it provides an estimate of the effect of uncertainties in each of the selected inputs on the precipitation bias factor. This way the order of magnitude of uncertainties in the precipitation bias factor which can be expected to result from uncertainties in these inputs can be estimated (i.e. are these uncertainties are in the order of 5 %, 50 % or rather 100 %?). These uncertainties can then be set into relation to the differences in the precipitation bias factors of different precipitation estimates in order to assess whether differences in the precipitation bias factors are larger than the effect of the uncertainties in inputs to the evapotranspiration module. In this study, differences in the precipitation bias factors of different precipitation estimates were relatively large (for example for the subcatchments Ak-Tash, Cholma and Gulcha these differences were often larger than 0.5 corresponding to a 50 % difference in the precipitation estimates). Uncertainties in the precipitation bias factor resulting from uncertainties in the evapotranspiration module were estimated to be in the order of 0.1.

14. P10739. L5. Why was such a coarse resolution selected for the WRF model? Could it not be run with a higher resolution?

Reply: In principle it would be possible to nest another simulation with a resolution of for example 4 km within the existing 12 km simulation and it would be very interesting to analyse whether this would solve the problem of the overestimated precipitation in the catchment Gulcha, which was assumed to be caused by the relatively coarse resolution and the poor representation of the orography in this region. However, this was unfortunately not possible within this study. This said, even though a resolution of 12 km is not satisfactory in orographically complex regions, it is already a relatively high resolution compared to many other applications of regional climate models for long term simulations (e.g. in the ENSEMBLES project a resolution of 25 km was applied for downscaling ERA-40 data over Europe; van der Linden and Mitchell (2009)), and only recently finer resolutions became computationally feasible (Suklitsch et al., 2011).

15. P10740. L1-4. Please make a reference to Fig. 7 already here when it is first discussed.

Reply: Thanks, this has been done.

16. P10741. L4. As seen in fig. 9, a precipitation bias factor of 0.6 is physically not reasonable.

Reply: A precipitation factor of 0.6 (indicating overestimation of 40 %) is similarly inacceptable as a precipitation factor of for example 1.5 and higher as can be seen in Fig. 10 for the IDW precipitation data in the catchments Ak-Tash, Cholma and Gulcha. As, except for the APHRODITE data set, the precipitation data are undercatch corrected (section 3.1.2) we do not expect a systematic underestimation of precipitation; and overestimation of precipitation for example is a typical problem of RCMs. The information on the precipitation bias factor is then used for the evaluation of the precipitation data sets. For example a precipitation factor of 0.6 indicates a strong overestimation of precipitation so that for the catchment Gulcha the precipitation data set “WRFadj-all” has to be judged as poor. The uncertainties in the areal precipitation estimates are much larger than one would expect in orographically less complex regions with a more dense precipitation network, which results partly in considerable deviations of the precipitation bias factors from one.

17. P10743, L7. Something is missing “the best, the best 20”

Reply: We added “1” so that it now reads “the best 1, the best 20”.

18. P10743, L13. As you mention here, the model structure might have a major impact on your results, especially the chosen scheme for evapotranspiration.

Reply: We agree that the chosen model structure may influence the results and that particular the calculation of evapotranspiration plays a major role. A solution to this problem would be to repeat this analysis for an ensemble of hydrological models, but such an ensemble approach was not within the scope of this study. Evapotranspiration will be affected by the model structure and by uncertainties in the model inputs. As we use a relatively physically based approach for the calculation of evapotranspiration and the uncertainties in the model inputs are expected to be high in this region, we focused our sensitivity analysis on the model inputs. This can of course not replace an ensemble approach, but it at least gives some estimates in which order of magnitude changes in the inputs to the evapotranspiration module would influence our results.

19. I would suggest to merge figure 1 and 2, since they are very similar.

Reply: This has been done.

20. Please put Figure captions (a, b c etc) in the figure instead of the name of the method. The text is also very small and difficult to read.

Reply: The font size of figures 5, 6, 7, 10 and 11 has been increased. The name of the method is inserted in the figure, as we find that this simplifies the orientation for the reader, who otherwise always would have to look back and forth between the figure and the legend.

References

Gourley, J. J., and Vieux, B. E.: A method for evaluating the accuracy of quantitative precipitation estimates from a hydrologic modeling perspective, J. Hydrometeorol., 6, 115-133, 10.1175/jhm408.1, 2005.

Heistermann, M., and Kneis, D.: Benchmarking quantitative precipitation estimation by conceptual rainfall-runoff modeling, Water Resour. Res., 47, 10.1029/2010wr009153, 2011.

Suklitsch, M., Gobiet, A., Truhetz, H., Awan, N. K., Gottel, H., and Jacob, D.: Error characteristics of high resolution regional climate models over the Alpine area, Climate Dynamics, 37, 377-390, 10.1007/s00382-010-0848-5, 2011.

van der Linden, P., and Mitchell, J. F. B.: ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project, 160pp, 2009.