We are grateful to reviewer #1 for his/her helpful comments on the discussion paper. Our reply can be found below. For convenience, we included a copy of the reviewer's comments (gray boxes). Text designated for inclusion in the revised manuscript is printed in green color.

Reviewer comment 1

Page 1174, Line 6-12. Why did you apply a bias-correction to 3B42RT data and use it for hydrological simulation? A very important purpose of carrying out rainfall evaluation study is to provide valuable sights to people about the suitability of satellite rainfall in the study area. Real-time 3B42RT data might be a good source for operational discharge prediction, so testing the rainfall prediction capability of original 3B42RT data seems to be more important than that of corrected 3B42RT. I suggest that you do hydrological simulations with the original 3B42RT data as forcing.

We clearly see the potential value of the 3B42RT product for operational hydrological forecasting. Hence, we also share the reviewer's opinion that this product's real-time character should be preserved in the hydrological verification. We can, therefore, follow the reviewer's thoughts when he asks for an analysis of the raw 3B42RT product.

We argue, however, that the most valuable verification is one that investigates the 'best-possible' product available for real time use. In our opinion, this is what our bias-corrected 3B42RT variant (abbreviated SRC in the manuscript) represents, because

1. the bias-correction leads to a significant improvement in rainfall estimates (and hydrological model outputs).
2. our variant of bias-correction is well applicable in operational systems (in any study basin and for all times).

Unfortunately, the second point did not become clear in the manuscript. We suggest to rephrase the paragraph in question (page 1174, lines 6 - 12 in the discussion paper) as follows:

"The real time product (3B42-RT) showed a moderate negative bias when compared with the gage-adjusted 3B42 product. We corrected for this deficit by multiplying the former product with an adjustment factor reflecting the deviation between the two satellite data sets. In order not to introduce artificial skill, factors were derived independently for two sub-sets of the 11 yr time series. The factor determined on the first half of the time series was applied to the second half and vice versa. Note that this correction does not make explicit use of recent rain gage data since the adjustment factor is derived from historic satellite data alone. Hence, the real-time character of the 3B42-RT product, i. e. its operational applicability, is fully preserved."

Consequently, we propose to abstain from a separate analysis of the raw 3B42 data since the results are likely to be 'worse than necessary'.

Reviewer comment 2

Page 1178, Line 4. 3.1.1 should be 3.1.3.

Yes, the reference is incorrect. Needs to be corrected to "The rainfall data sets introduced in Sects. 3.1.1 - 3.1.3 form [...]"

Reviewer comment 3

Page 1178, Line 6. How do you decide the parameters for inverse-distance interpolation? Do you do testing work to choose the best ones (e.g. cross validation) since the parameters can have effect on the results?

For the presented study, we re-used the optimal IDW settings identified in an earlier radar-QPE benchmark experiment (hourly rainfall data, different basin) by means of cross-validation (XV). With respect to the TRMM data, the used settings appear plausible. They guarantee that, for each target point, weighted information from (at most) the 4 nearest grid cells is used, with strong preference for the nearest cell.

In response to the comment, we carried out a XV experiment for the daily Mahanadi rain gauge data. In particular we performed a leave-one-out XV and computed the RMSE from the observed and estimated rain rates (average over stations and time steps). The IDE power was varied between 1 and 4. The number of neighbors (which equals the number of search sectors) was varied between 1 and 8. The underlying software (see Chapter 3 of http://echse.bitbucket.org/downloads/documentation/echse_tools_doc.pdf) accesses the same geo-statistical routines as the hydrological model used in our study.

The XV experiment revealed that the MSE is highest for the nearest-neighbor method (Thiessen polygons) and gradually decreases as the number of neighbors increases. Values greater than 4 result in marginal improvements only. Regarding the power parameter, the smallest MSE was obtained for a value of 1. The error for power 2 is only little larger, but it increases significantly towards a power of 4. These results were consistent for different thresholds of minimum rainfall (0, 5, 10 mm/day) and several tested sub-sets of rain gages. We conclude that the adopted IDW parameters (4 sectors, power of 2) are not the best possible choice <u>in terms of XV</u> but rather close to the optimum.

We know from other experiments (described in Heistermann, M., Kneis, D. (2011): Benchmarking quantitative precipitation estimation by conceptual rainfall-runoff modelling, Water Resources Research, 47, W06514) that the best-performing regionalization method according to XV (point scale) does not necessarily perform best in runoff simulation (basin scale). In the presented case, the hydrological simulations were found to be practically insensitive with respect to an optimization of the IDW parameters. In particular, decreasing the power from 2 to 1 and increasing the number of neighbors from 4 to 8 (as suggested by XV) did not result in a consistent increase in the Nash-Sutcliffe Index and/or decrease in the percentage bias. In all cases, changes in the absolute value of the Nash-Sutcliffe were <= 0.01 (sign differs for gages!) and changes in the percentage Bias were <= 2% (also with variable sign).

We suggest to append the following sentence to the paragraph in question (page 1178, line 7): "Suitability of the interpolation parameters was verified experimentally."

Reviewer comment 4

Page 1180, Line 3-5. 'The split-sample approach was found to be very sensitive to the choice of the particular time periods used for calibration and validation'. Is this because of the model itself of the rainfall inputs? If it is the former, the model may be not physically robust.

The same hydrological model engine was applied to a number of other basins (mostly in Germany) using rather high-quality rainfall and stream flow data. In these applications we didn't see indications for a general lack of physical robustness.

We suppose that the mentioned sensitivity of validation results with respect to the time period is largely connected to the strong inter-annual variability of the Monsoon (cf. Fig. 5 in the manuscript, for example). Since the Nash-Sutcliffe Index is defined as 1-MSE(obs, sim)/VAR(obs), it is quite sensitive to the amplitude of the observations in the analyzed time period, i. e. the intensity of the Monsoon. This becomes obvious when considering a fictive multi-year period with constant MSE. In such a case, the lowest annual values of the Nash-Sutcliffe Index would correspond to the driest years.

We elaborated on the problem a bit further and computed the Nash-Sutcliffe Index at one of the gages (Kantamal) for differently designed split samples. We found that the reported sensitivity problem actually vanishes for more cleverly constructed split-samples. If each of the two sub-samples contains a subset of years, we get rather different values of the Nash-Sutcliffe Index (i. e. the reported sensitivity problem occurs). In accordance with expectation (see above), we get a worse result for the sub-sample containing dryer years. However, if we split the series by filtering for day numbers <=15 (first sub-sample) and day numbers >15 (second sub-sample), we obtain a very similar goodness-of-fit for the two sub-samples. Note, however, that this way of constructing split samples also has limitations. For example, splitting the series into sub-samples on the basis of odd and even days would hardly make sense because of the series' auto-correlation.

We suggest to replace the first paragraph of Section 3.3.6 by the following text:

“Typically, only some part of an observed hydrograph is used for model calibration while the other part is reserved for validation. In this case study, however, the split-sample approach was found to be very sensitive to the choice of the time periods used for calibration and validation, respectively. In particular, the problem occurred if each of the two sub-samples consisted of a different sub-set of years. In such a case, the value of the Nash-Sutcliffe Index is largely controlled by the intensity of the Monsoon in the selected years (recall the denominator in Equation 4). We circumvented this problem by adopting an alternative strategy of model validation where the calibrated parameter sets are exchanged between neighboring catchments. This strategy analyzes the parameter’s transferability in space rather than time.”

Reviewer comment 5

Table 6 and Table 7. The calibration results are only shown for three catchments. How do you deal with the two catchments in terms of calibration?

This information was indeed missing in the manuscript. The gages Tikarpara and Mundali are located downstream of the Hirakud Dam and the flow rates are largely controlled by the release from that reservoir. Moreover, the information on release rates was available to us in the form of daily averages only. In view of these facts (low signal-to-noise ratio) we decided not to calibrate the model for the two catchments.

We suggest to add the following text as the very end of Section 3.3.5:

“We abstained from calibrating the model for the catchments of Tikarpara and Mundali since stream flow at these two gages is largely controlled by operation of the Hirakud Dam. The parameters for the two catchments were set to the average of the respective calibrated values for Kesinga, Kantamal, and Salebhata.”

Reviewer comment 6

Is rainfall data from the whole time period (i.e. 2000-2010) used for calibration? For a specific catchment, what calibrated parameters (i.e. from which other catchment) are used to drive the hydrological model for validation? I think you need to give more words on the calibration and validation so that my understanding of the results can be more reasonable.

To answer the first question, we propose to insert the paragraph below at the end of Section 3.3.5 (before the text added in response to comment 5):

"The model was calibrated on observed stream flow data from 2002-01-01 to 2009-12-31. Model runs were initialized about 1.75 years in advance (2000-03-01) but the outputs for this 'warm-up' period were discarded. The estimated initial state for 2000-03-01 was generated in a long-term simulation using recycled meteorological data of the years 2001-2010 as forcings."

The second question is precisely answered in Tables 6 and 7. An instruction how to read these tables is also given at page 1180, lines 11-14. We suggest not to repeat the information in the text.

Reviewer comments 7 + 8

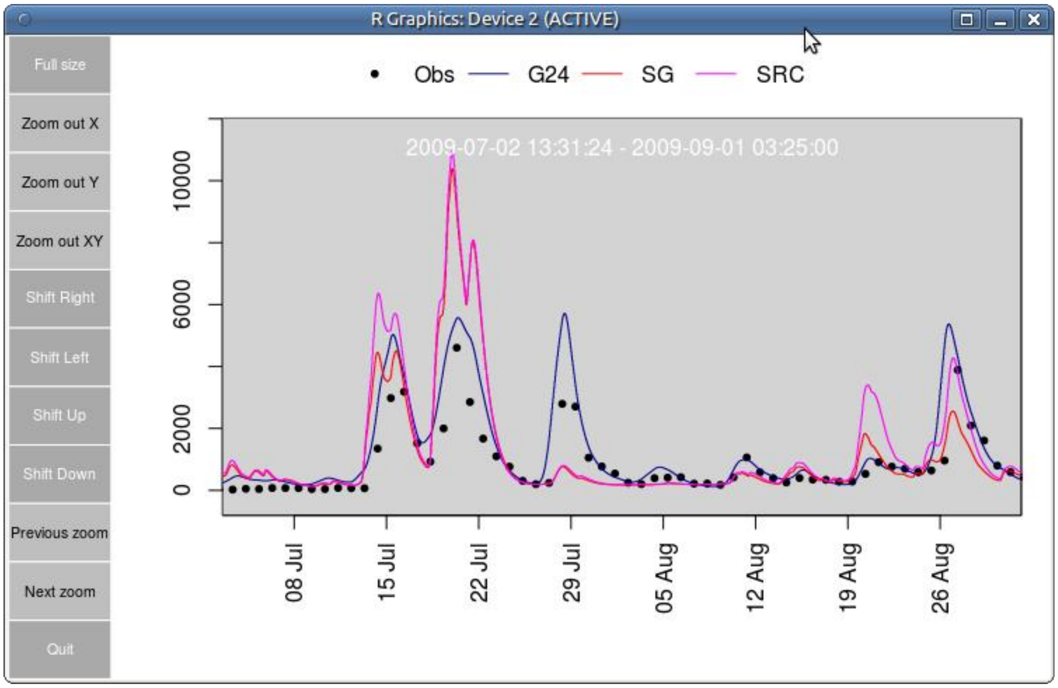
Page 1182, Line 10. 'bias is' may be 'bias was'.
Page 1182, Line 12. 'but it's' may be 'but its'.

Will be corrected.

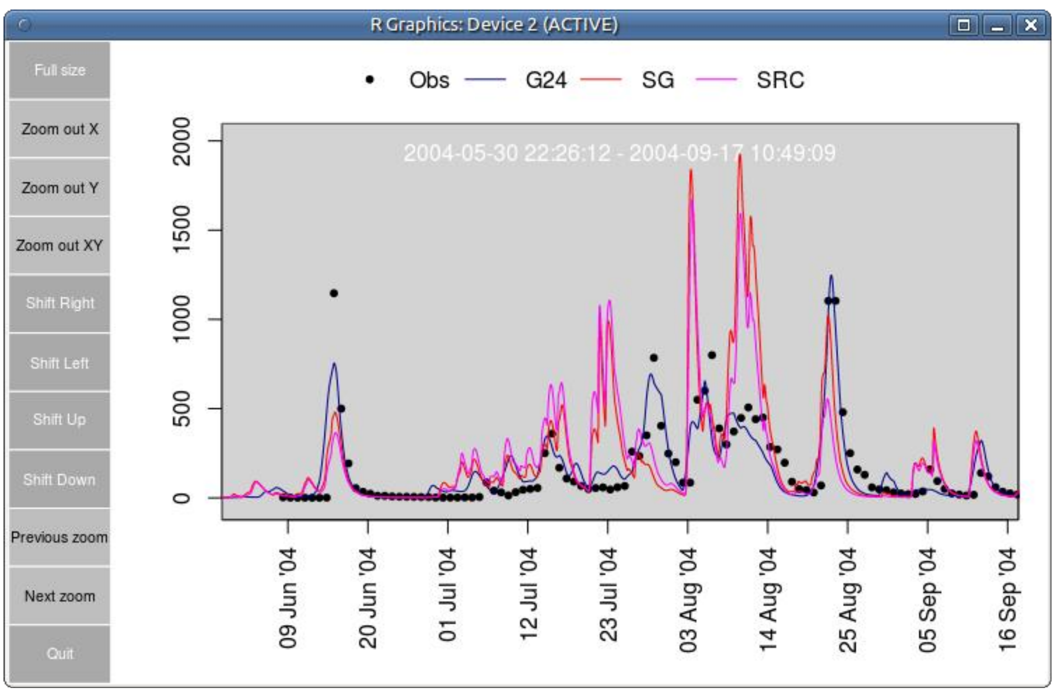
Reviewer comments 9

Fig. 4. Why do NS indices show very large annual variations (even extending to below or near zero) in Salebhata and Kesinga catchment for hydrological modeling with TRMM rainfall as forcing?

The years with the low values of the Nash-Sutcliffe Index are 2004 (both gages) as well as 2009 (Kesinga) and 2010 (Salebhata). In all these cases, the bad fit it mainly caused by a massive over-estimation of precipitation in the satellite data. This is exemplified by the two graphics below.



Stream flow (m3/s) at Kesinga, 2009



Stream flow (m3/s) at Salebhata, 2004

We suggest to add the following sentence after line 7 of page 1182:

”The very low Nash-Sutcliffe Indices obtained in some years for SG and SRC input (gages Kesinga and Salebhata) are largely due to occasions of severe rainfall overestimation by the two satellite-based products.”

Reviewer comments 10

TRMM rainfall-forced hydrological model outperforms in Tikarpara and Mundali than in the other three catchments in terms of NS indices (Fig. 4). Does the reservoir in the upstream play a role in this result because the other three ones are independent of reservoirs?

Yes. This fact is already mentioned at page 1181, lines 17-20. In response to the comment, we suggest to add another sentence for clarification at page 1182, line 7:

"Note again that the high Nash-Sutcliffe Indices obtained at Tikarpara and Mundali for *all* precipitation estimates are mainly due to the insertion of known reservoir release rates in the model (cf. Section 3.3.4)."