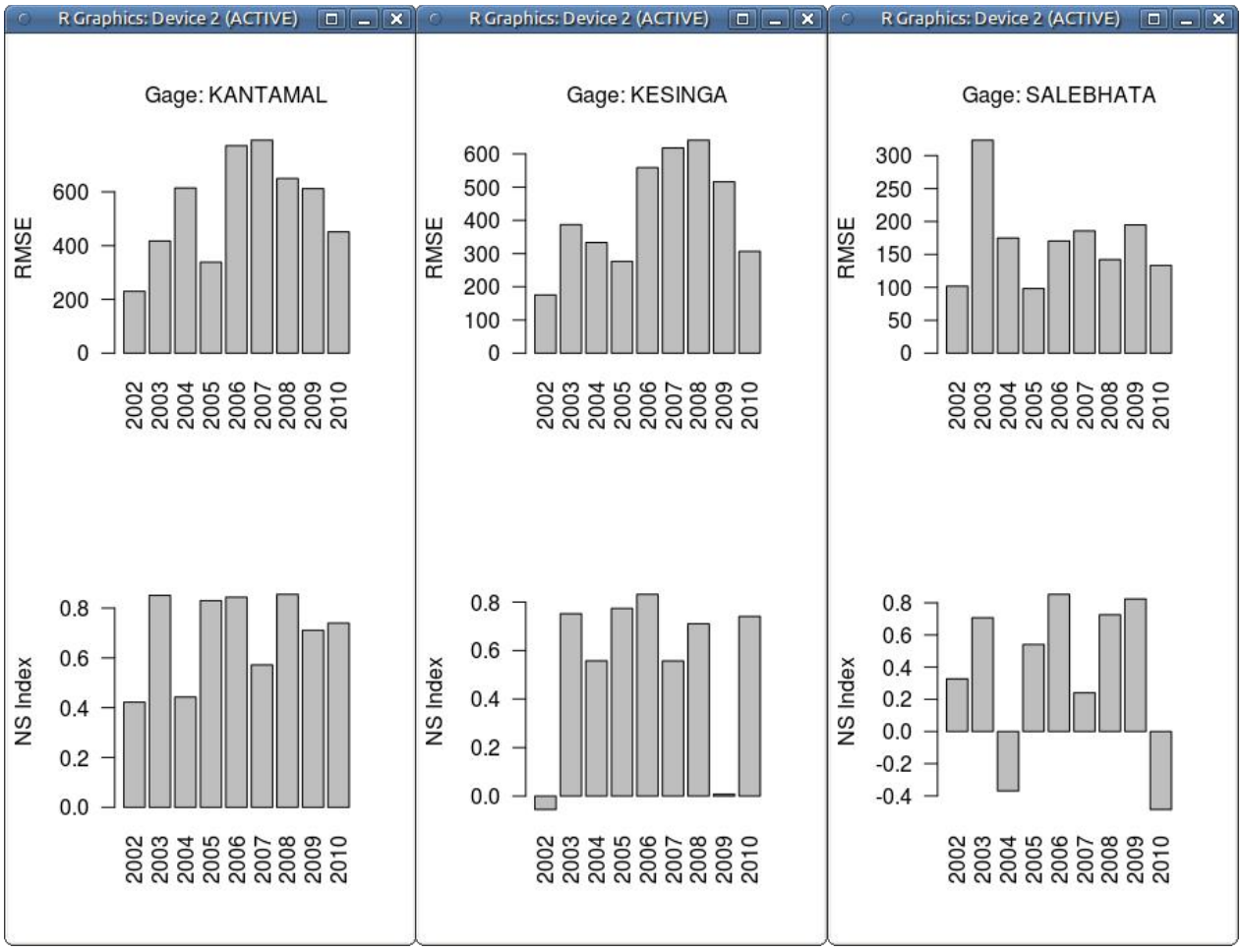
We thank reviewer #1 for the comments on our reply to his/her original review. Again, our reply can be found below, including a copy of the reviewer's comments (gray boxes). Suggested revisions to the manuscript are printed in green color.

Reviewer comment 1

Based on your argument, NS index includes the variation of observations, so the interannual variability dominates the sensitivity of the validation results. Can you use RMSE or MSE as indices? The interannual variability is not reflected in them. You can use Normalized RMSE or MSE so that each year is comparable.

Yes, in the mentioned context, use of RMSE (or MSE) may be advantageous. Use of the normalized variants, NMSE or NRMSE, however, is not a cure for the problem because scaling with the range of observations (in NMSE or NRMSE) is conceptually similar to scaling with the variance (in NS Index).

We inserted some graphics of annual RMSE (top) and NS Indices (bottom) below for three gages (columns). Rainfall forcing was G24, i. e. daily rain gage data.



Look, for example, at the results for Salebhata in the rightmost column. For 2010, the NS Index is the worst of all years but the RMSE is below its median. In 2009, both NS Index and RMSE show the 2nd highest values in spite of the fact that both indices differ in their orientation (optimum NS Index: 1, opt. RMSE: 0). Thus, if these indices are applied to the given data one should probably report both values.

In our opinion, a better solution to this problem would be a more clever split-sampling that balances the variance in the two sub-samples. Then, there is a good chance to obtain consistent results for different indices of the goodness-of-fit, including RMSE and NS Index. A possible (and tested) split-sampling approach was mentioned in our reply to comment #4 of the original review (lines 106-115).

In the presented study, we circumvented the issue by choosing a different approach to validation. We did not consider temporal sub-samples but we tested the model's transferability in space. Therefore, with respect to Section 3.3.6 of the manuscript (model validation), we don't see the need for further changes going beyond our suggestions in reply to the original review (see lines 119-127 of our first reply at http://www.hydrol-earth-syst-sci-discuss.net/11/C683/2014/hessd-11-C683-2014-supplement.pdf).

It you insist on an (additional) validation using temporal sub-sampling, we suggest to report the goodness-of-fit for the year 2010. This year was not considered in the calibration data set. As indicated by the figures above, the model performance for this year is comparable to the performance in the 'calibration years', with one exception (NS Index at Salebhata).

Reviewer comment 2

The low NS values in 2004, 2009 and 2010 can be attributed to rainfall overestimation in TRMM. I suggest that you put this into conclusions because it is an important part of the performance of TRMM.

The current version of the conclusions section (line 15) already contains the following statement:

"At the same time, the remote sensing data frequently overestimate rainfall amounts observed at the ground."

We suggest to extend this as follows:

"At the same time, the remote sensing data frequently overestimate rainfall amounts observed at the ground. The latter fact is responsible for a relevant number of flaws in the hydrological simulations."

In our opinion, this general statement is appropriate for the conclusions. We already suggested to add a more detailed statement after line 7 of page 1182 (see lines 186-188 of our first reply at http://www.hydrol-earth-syst-sci-discuss.net/11/C683/2014/hessd-11-C683-2014-supplement.pdf).