Response to review comments of Remko C. Nijzink on the manuscript "Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products"
Patricia López López, Edwin H. Sutanudjaja, Jaap Schellekens, Geert Sterk and Marc F. P. Bierkens

The authors would like to thank Remko C. Nijzink for his time and constructive and valuable comments on the manuscript. His suggestions will help us to improve the quality of our manuscript. We have included detailed responses to his comments and suggestions in the supplementary .pdf file. We have also included a modified version of the original manuscript. Please note the supplement to this comment and the modified manuscript, with modifications in blue.

General comments:

Comment 1:

My most important point considers the calibration. It consists of 81 runs with three different values for the calibrated prefactors. I fully understand that running a distributed model has a high computation cost, but this number of runs seems rather low to me. There is a big risk of undersampling, leading to results that can just be a mere coincidence. For example, the most optimal value of f_w *may actually be 0.889, whereas only 0.75, 1 and 1.25 are explored in the study. Once again, I understand the burden of computational efforts, but at least the authors may want to reflect on this limitation in their discussion. In addition, it is mentioned that, except for these four prefactors, the remaining parameters were kept fixed. How many parameters are kept fixed and to what extend is the model already directed towards a certain solution by the choice of fixing certain parameters? For example, according to Figure 2 PCRGLOB-WB uses an interception routine. If the maximum interception capacity is kept fixed, it will probably influence the results for GLEAM versus the modelled evaporation.*

Answer:

We understand the reviewer’s comment about the number of runs carried out for the study. We considered four different prefactors for calibration and we kept the remaining ones fixed (more information about these parameters can be found in Sutanudjaja et al., 2011). With this approach, we aimed to understand the state and uncertainty of model parameters, improving our knowledge about the importance or influence of each model parameter for each calibration scenario in this Moroccan basin. We agree with the reviewer that more runs (e.g. f_w = {0.75, 0.8, 0.85, … , 1.25}) would be needed to precisely estimate the optimal parameters set. However, as the reviewer has pointed out, running a large-scale hydrological model, such as PCR-GLOBWB, requires high computational cost in terms of computer power and time. Therefore, we could not perform more runs at this stage. For future work, the range of prefactors values and then, the number of runs will be increased. Nevertheless, we will acknowledge this study limitation in section 5. Discussion and conclusions as follows:

P16L11: “… is selected. For these combinations, and due to computational limitations, only four prefactors were considered leading to 81 model runs per precipitation product. Using more prefactor values and therefore, more runs may improve the estimation of the optimal parameters set for each calibration scenario. A suggestion …"

Comment 2:

In addition to this, the step-wise calibration consists out of first calibrating on evaporation with GLEAM, and, in a second step, on soil moisture. I just wonder how much influence this order in calibration influences the results, especially as soil moisture strongly influences evaporation. Did you consider a step-wise calibration with first ESA CCI soil moisture and then GLEAM in a second step?

Answer:

As the reviewer pointed out in his comment, the step-wise calibration scenario is based on both observations, evapotranspiration and soil moisture. In the first step, all prefactors are calibrated using GLEAM actual evapotranspiration and in the second step, those prefactors that have been identified are held constant and the remaining ones are calibrated according to ESA CCI soil moisture. Results showed that GLEAM actual evapotranspiration can be used to calibrate only the reference potential evapotranspiration (f_e), whereas ESA CCI soil moisture allows the identification of the minimum soil water capacity (f_w) and the saturated hydraulic conductivities of the soil layers (f_k). Therefore, in the step-wise calibration scenario firstly f_e is identified and kept fixed, and secondly, f_w and f_k are identified. We agree with the reviewer that the calibration based on evapotranspiration also influences prefactors different to f_e (and the calibration based on soil moisture also influences prefactors different to f_w and f_k), but the impact is less significant and does not allow the identification of those prefactors. If we consider a step-wise calibration scenario using first ESA CCI soil moisture and GLEAM actual evapotranspiration in a second step, f_w and f_k would be first identified and maintain constant and f_e would be identified later. This means that, for this particular study, a change in the order in which the observations are considered in the step-wise calibration scenario would not imply different results, because each observation allows the identification of different model prefactors.

Comment 3:

I also wonder what the reasoning is behind the choice to compare the ESA CCI surface soil moisture with the soil moisture of the first three soil layers of PCR-GLOBWB. As mentioned by the authors (page 8, line 30) the ESA CCI soil moisture only represents the first 0.5-2cm, so wouldn't it make more sense to just compare with only the very first soil layer (first 5cm according to page 5, line 14) in PCR-GLOBWB? In this way, all parameters affecting the soil moisture in all the three layers will react, which can also be noted from the results for fk , but one could wonder whether this is for the right reason.

Answer:

According to P7L27-28: "ESA CCI surface soil moisture observations were compared to simulated soil moisture with the first of the three vertical soil layers in PCR-GLOBWB." To clarify this and following the reviewer's suggestion, we will include a note as follows:

P7L23-24: "… ESA CCI surface soil moisture observations were compared to simulated soil moisture of the first of the three vertical soil layers in PCR-GLOBWB (top 5 cm of soil). …"

Comment 4:

Often, a comparison is made between the reference scenario S0 and the scenarios S2- S4. Nevertheless, S0 is merely an uncalibrated model and especially for Ait Ouchene (Figure 6), the model performances are rather poor. Therefore, not much is needed to achieve improvements in this case. Isn't it much more interesting to focus more on comparing S1 with S2-S4? In other words, how close can we get to a calibration on streamflow with help of GLEAM and ESA CCI? It would be interesting to see if differences occur in Figures 8-10 for S1 and S2-4. Ideally, there would be no difference, but I expect that this will not be the case.

Answer:

We agree with the reviewer and we will replace Figures 8, 9 and 10 with new figures (Figures 7 and 8). Figure 7 will show comparisons between estimated and observed evapotranspiration, soil moisture and discharge including calibration scenarios S0 (reference scenario) and S4 (step-wise calibration). Similarly, to Figure 7, Figure 8, will show comparisons between estimated and observed evapotranspiration, soil moisture and discharge including calibration scenarios S1 (in-situ discharge calibration) and S4 (step-wise calibration). These two figures should improve the presentation of the calibration results. At the same time, section 4, results will be restructure in two subsections: 4.1. Calibration results and 4.2. Validation results. Section 4.1. Calibration results will be adjusted accordingly (please see modifications in blue in the manuscript).

Detailed comments (P: page, L: line or lines):

Comment 5:

P8.L2-3. I don't know if these specific stations were used for MSWEP, but as MSWEP used station data as input (also remarked by the authors on P7.L19-20), isn't it logical that MSWEP provided a better fit to the station data?

Answer:

MSWEP, WFDEI and EI precipitation values were interpolated to two weather station locations to calculate various performance metrics. These weather stations were not used to generate any of the global precipitation products, ensuring an independent validation. Furthermore, this precipitation evaluation was carried out only for two rainfall stations that were found inside the Oum Er Rbia basin. These measurements were considered too scarce in number and spatially sparse to cover the entire basin and therefore to extract firm conclusions about which products outperforms the other ones.

Comment 6:

P10.L15-17. Why model at a daily basis and only compare on monthly values? What is the temporal resolution of the data (discharge, GLEAM and ESA CCI)?

Answer:

As the reviewer has mentioned, PCR-GLOBWB runs at a daily temporal resolution and the meteorological data, the GLEAM actual evapotranspiration and the ESA CCI soil moisture observations are daily too. Preliminary analysis and comparison of the calibration scenarios were made at a daily temporal resolution initially. Daily and monthly results were similar. Therefore, for practical reasons and to simplify the manuscript, we decided to include results relative only to the monthly temporal resolution. Moreover, for water resources applications, monthly estimates can be sufficient in this particular area.

Comment 7:

P12.L10. I can see that fw shows a clear pattern, but I don't see this clearly for fe.

Answer:

We agree with the reviewer that f_w can be well identified from Figures 5a and Figure 5b. The pattern for f_e identification, although is not as clear as f_w, is also visible, especially in Figure 5b when WFDEI and MSWEP precipitation are used. Higher values of f_e, in particular f_e = 1.25, result in higher KGE values.

We will improve Figure 5 in different ways: we will use different colours and dot shapes to indicate different values of f_e, we will modify the horizontal axis of each scatterplot limiting the tick marks and numbers to the values of the used calibration prefactors and we will change the label of y-axis to indicate when KGE values are based on discharge, actual evapotranspiration and surface soil moisture using subscripts KGE_q, KGE_{evap} and KGE_{sm}. Moreover, we will modify the figure analysis as follows:

P11L26-30: "… Figures 5a and 5b (calibration scenario S1) are similar. From these figures, f_e (1st column) and f_w (4th column) are well identified by discharge calibration at both gauging stations when forced with any of the three precipitation products. f_e = 1.25 and f_w = 1.25 lead to the highest KGE_q values. However, it is not possible to identify the best prefactors of f_j (2nd column) and f_k (3rd column). There are no clear and distinct maximum values in the scatterplots of these figures, hence f_j = 0 and f_k = 0 are used. …"

Comment 8:

P13.L16-17. These numbers refer to the WFDEI-case

Answer:

We will correct these KGE values as follows:

P13L20-25: "… From Figure 6 (5th column), calibration using GLEAM evapotranspiration and ESA CCI soil moisture leads to further improvements than when these observations are separately used. For example, when MSWEP precipitation is used to model discharge at Mechra Eddahk station, KGE_q varies between 0.703, 0.693, 0.613 and 0.573 for calibration scenarios S1, S4, S2 and S3, respectively (KGE_q = 0.561 for the reference scenario S0). At Ait Ouchene station (see Figure 2 in the Supplementary Information), KGE_q varies between 0.520, 0.342, 0.331 and 0.271 for calibration scenarios S1, S4, S2 and S3, respectively (KGE_q = 0.542 for the reference scenario S0). …"

Comment 9:

P14.L9-11. It may as well be model structural deficiencies as wrong parameterizations. It is a bit easy to blame the input data directly, especially as it happens for two out of three input *products. It must be noted as well that even though EI has the peaks in 2002 right, it also underestimates the peaks in 1996 and 1997.*

Answer:

Indeed, these differences may be also related with model structural deficiencies. We will modify the manuscript according to the reviewer's suggestion:

P14L13-17: "… scenarios. From Figure 7c, the step-wise calibrated run (S4) reproduces the monthly observed discharge well, except some simulated extreme peaks which were not observed, e.g. January and June in 2002 and some which were not simulated properly, e. g. January and May in 1996 and 1997. This lack of fit may be due to errors in the precipitation data, because higher discharge differences are shown when WFDEI and MSWEP products are used in comparison to EI product. Other possible reasons may be related with model structural deficiencies. When …"

Comment 10:

Throughout the manuscript, the terms KGE, NSE etc. are used and sometimes refer to a case with evaporation and sometimes to cases with soil moisture or discharge. For clarity, it might be good to add a subscript (e.g. KGE_E, KGE_{SM} *etc.).*

Answer:

We believe that the reviewer's suggestion will facilitate the understanding of the manuscript, especially in section 4. Results. Therefore, we will modify the manuscript indicating with a subscript when the performance metrics are calculated for precipitation (precip), evapotranspiration (evap), soil moisture (sm) and discharge (q). Moreover, we will include the following sentence:

P10L8-10: "… is 0.

When the performance metrics were calculated between simulated and observed soil moisture estimates, the subscript sm was added to the metric, i.e. NSE_{sm}, KGE_{sm}, $RMSE_{sm}$, MAE_{sm}, r_{sm} and $PBias_{sm}$. Similarly, when comparing actual evapotranspiration estimates, precipitation and discharge, the added subscripts were evap, precip and q, respectively. …"

References:

- Sutanudjaja, E. H., Van Beek, L. P. H., De Jong, S. M., van Geer, F. C., & Bierkens, M. F. P. (2011). Large-scale groundwater modeling using global datasets: a test case for the Rhine-Meuse basin. Hydrology and Earth System Sciences, 15(9), 2913.

Additional modifications in figures and figures to be included

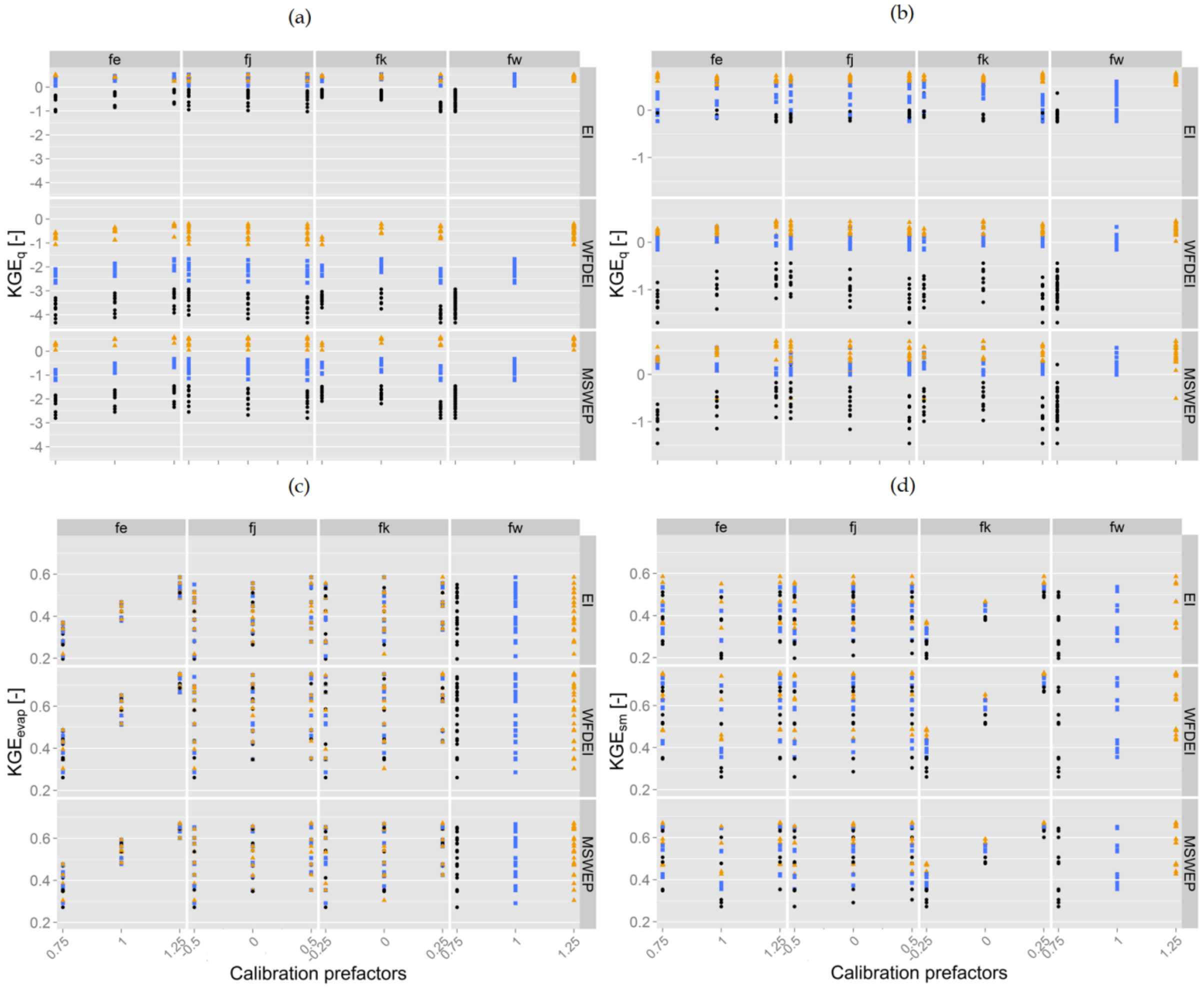

Figure 5. *Scatterplots of discharge performance indicator KGE based on the monthly observations versus prefactors f_e, f_j, f_k and f_w for the calibration scenarios S1 ((a) Ait Ouchene (b) Mechra Eddahk), S2 (c) and S3 (d). In each sub-figure, columns indicate the different calibrated prefactors and rows indicate the three global precipitation products used as model forcing. Different colours and dot shapes indicate different f_w values.*

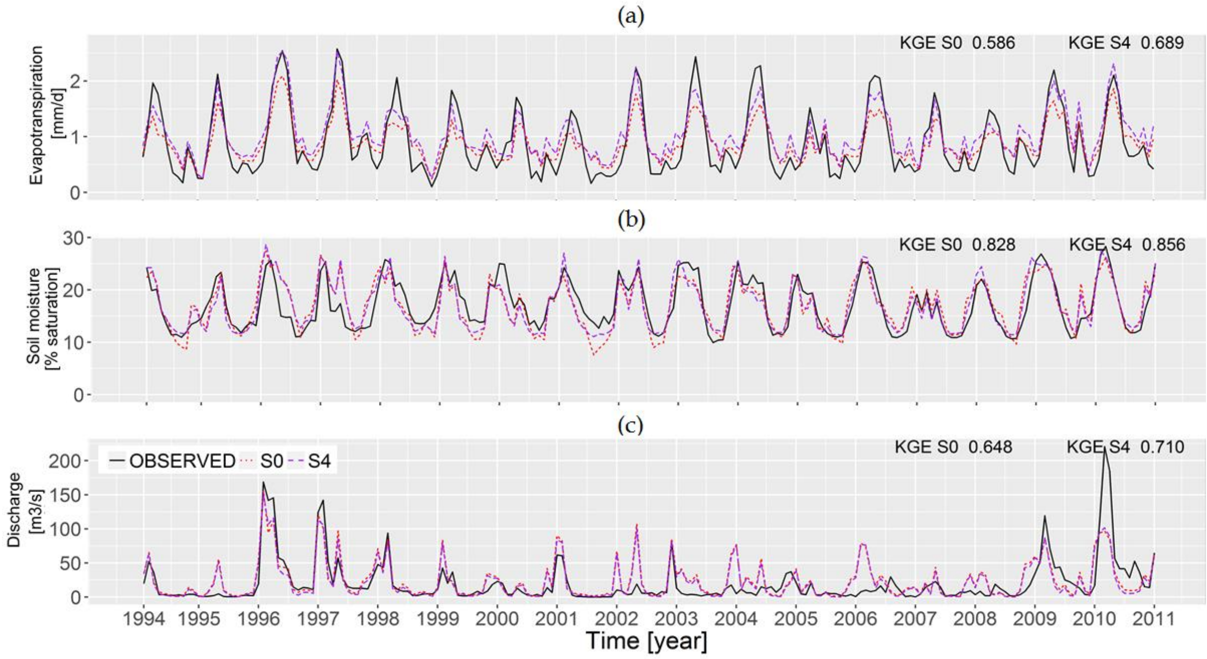

Figure 7. *(a) Monthly GLEAM actual evapotranspiration (black) and estimated actual evapotranspiration (red and purple) time series. (b) Monthly ESA CCI soil moisture (black) and estimated soil moisture (red and purple) time series. (c) Monthly observed discharge (black) and estimated discharge (red and purple) time series. The red dashed lines represent estimates from calibration scenario S0 (reference scenario). The purple dashed lines represent the calibrated time series from calibration scenario S4 which are taken from the runs that yield the best simulations. Estimated time series over the entire Oum Er Rbia basin for the validation time period obtained with MSWEP precipitation are shown.*

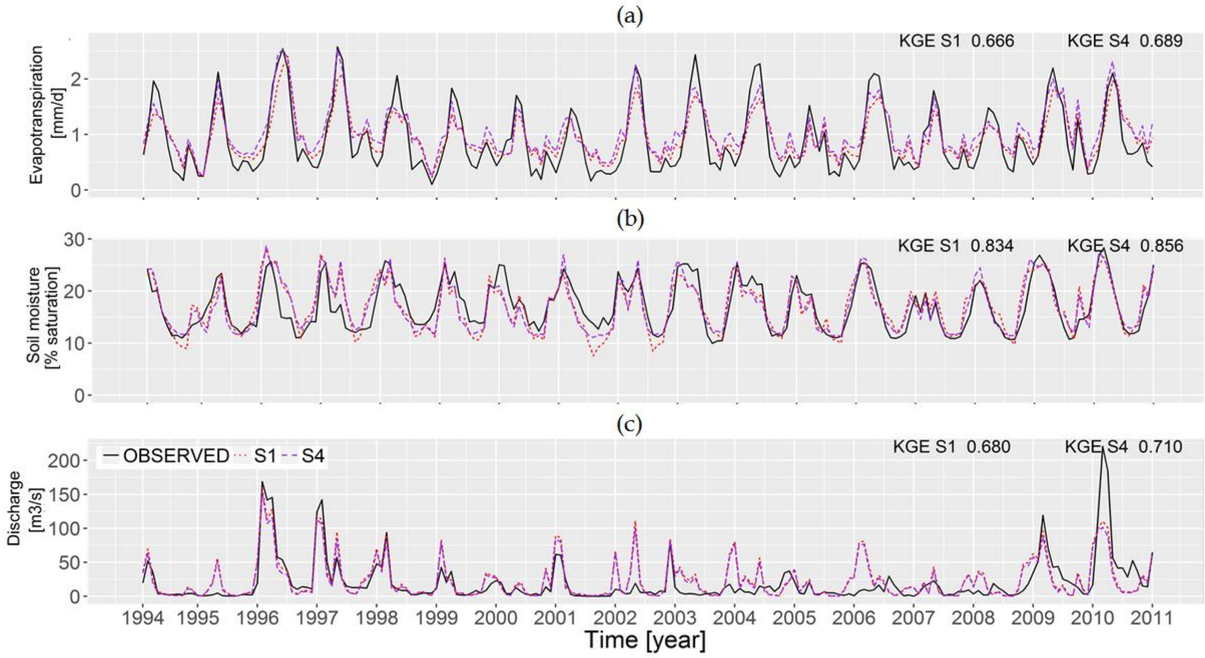

Figure 8. *(a) Monthly GLEAM actual evapotranspiration (black) and estimated actual evapotranspiration (red and purple) time series. (b) Monthly ESA CCI soil moisture (black) and estimated soil moisture (red and purple) time series. (c) Monthly observed discharge (black) and estimated discharge (red and purple) time series. The red dashed lines represent estimates from calibration scenario S1. The purple dashed lines represent the calibrated time series from calibration scenario S4 which are taken from the runs that yield the best simulations. Estimated time series over the entire Oum Er Rbia basin for the validation time period obtained with MSWEP precipitation are shown.*