Author comments on RC2 - egusphere-2025-636

MORITZ M. HEUER, HADYSA MOHAJERANI, MARKUS C. CASPER

In this manuscript, Heuer et al. present a multi-step calibration approach and test 12 parametrizations of pedotransfer function for plausibility with regard to spatial patterns of simulated process dominance.

I like the approach of the authors and think that it is a valuable contribution to process-based model calibration. I have only a few comments.

We would like to thank the reviewer for their helpful and valuable comments! We are glad the reviewer likes our approach and believes it to be a valuable contribution to process-based model calibration. We are certain the proposed changes improve the manuscript's clarity.

It is not always clear whether the authors are analysing evapotranspiration (ETa) or evaporation. It is important to clarify this and to make clear that the comparison between MODIS data and simulation results is valid. See for example L. 10 "Evapotranspiration parameters based on MODIS evaporation data". See also L. 178-180.

Thank you for pointing out these ambiguations. The MODIS data we refer to are evapotranspiration time-series and not just evaporation measurements. We will change the mentioned text lines within the manuscript to evapotranspiration in order to increase clarity.

The selection of the hydrograph efficiency metrics is not convincing. NSE and R2 are conceptually very similar. There is no advantage in using R2 in addition to NSE. Table 5 shows that a lower NSE results in a lower R2 and vice versa. So, R2 is not needed. I suggest adding a complementary metric e.g. PBIAS to capture another part of the hydrological cycle. Another option could be to use the three components of the KGE separately (as KGE-beta represents the same hydrological behaviour as PBIAS).

We thank the reviewer for this proposal. We will replace the R^2 metric with the *PBIAS* to include a complementary metric to the *NSE* and *KGE* metrics. This results in the following reworked efficiency metrics table:

PTF	MHd	HDHO	NSE_{cal}	KGE_{cal}	$PBIAS_{cal}$	NSE_{val}	KGE_{val}	$PBIAS_{val}$
1	3.19	0.9	0.86	0.89	-5.31	0.78	0.86	2.64
2	4.88	0.7	0.66	0.82	-3.79	0.59	0.79	1.28
3	3.59	0.8	0.72	0.84	-6.56	0.63	0.81	-0.15
4	3.36	0.5	0.67	0.79	-9.30	0.45	0.69	-5.25
5	3.68	0.7	0.72	0.83	-8.37	0.55	0.75	-3.16
6	6.60	0.6	0.49	0.73	-4.46	0.34	0.70	5.01
7	5.04	0.8	0.60	0.81	-0.04	0.49	0.77	8.67
8	2.76	0.8	0.87	0.89	-5.52	0.78	0.87	3.16
9	6.68	0.8	0.54	0.61	-5.38	0.49	0.60	-3.24
10	2.01	0.6	0.59	0.78	-4.94	0.47	0.75	3.16
11	5.36	0.8	0.55	0.76	-3.65	0.43	0.74	5.39
12	5.84	0.9	0.70	0.83	-5.00	0.64	0.80	0.64

Together with the changes in the table, we change the methodology and result parts where the used efficiency metrics are described and evaluated.

We changed lines 349 to 351 to:

> The third metric included was the PBIAS. This metric is a measurement of the average tendency of the simulated data to be larger or smaller than their observed counterparts (GUPTA et al., 1999).

We changed lines 428 to 433 to:

> The model shows high values for the efficiency measures for both the calibration and the validation period. Between calibration and validation, there is only a slight decrease in the NSE from 0.87 to 0.78, while the KGE decreases only minimally from 0.89 to 0.87. Values for the PBIAS slightly improve from -5.52% for the calibration period to 3.16% for the validation period. Efficiency measures for the split-sample test of other PTFs (Table 5) show a large value range. For example, PTF 1 also shows relatively high values for the NSE and KGE. However, PTFs 4, 6, 9, 10 and 11 show low values. All other PTFs show values in between. The PBIAS shows values of around -5.00% for most PTFs for the calibration period, while the values for the validation period are between -5 and 5% for all PTFs except PTF 7.

We changed lines 516 to 517 to:

> This is also supported by high values of efficiency measures such as NSE (0.78), KGE (0.87) and a low PBIAS (3.16%) for the validation period in the split-sample test.

L.371-374: This part needs to be reformulated. Three sentences mention that PTF 9 and 10 are not valid. I suggest to make one clear statement.

> [...] This was achieved by using a single layer aquifer with a thickness of 1 m and lateral hydraulic conductivities of $3E-5 \cdot m \cdot s^{-1}$. In step 3, the CDC height could also be adapted to the course of the gauging station curve for almost all PTFs except PTF 9 and 10. The corresponding calibrated values for d_r range from 6 for PTF 4 up to 60 for PTF 2. In the front part of the curve, the simulations almost exclusively run below the reference curve of the gauging station.

Table 5: Use consistently two digit numbers. For PTF1 and KGEval three digit numbers are shown.

We thank the reviewer for pointing this out! We corrected the values' decimal points to be the same for all entries.

References

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of hydrologic engineering*, *4*(2), 135–143.