Supporting information for "Leveraging pre-storm soil moisture estimates for enhanced land surface model calibration in ungauged hydrologic basins" (Crow, Reichle and Dong).

1. Noah-MP parameterizations

Table S.1. Parameterizations for the 16-member Noah-MP runoff ensemble.

Configuration number	Noah-MP runoff parameterization	Key runoff parameter value	Key ET parameter value
1	SIM GW	$f = 2.0$ [m^{-1}]	$\lambda = 5.0$ [-]
3	SIM GW	$f = 4.0$ [m^{-1}]	$\lambda = 5.0$ [-]
4	SIM GW	$f = 8.0$ [m^{-1}]	$\lambda = 5.0$ [-]
5	SIM TOP	$f = 3.0$ [m^{-1}]	$\lambda = 5.0$ [-]
6	SIM TOP	$f = 6.0$ [m^{-1}]	$\lambda = 5.0$ [-]
7	SIM TOP	$f = 12.0$ [m^{-1}]	$\lambda = 5.0$ [-]
8	SIM TOP	$f = 24.0$ [m^{-1}]	$\lambda = 5.0$ [-]
9	SIM TOP	$f = 48.0$ [m^{-1}]	$\lambda = 5.0$ [-]
10	FD	$REFKDT = 0.5$ [-]	$\lambda = 5.0$ [-]
11	FD	$REFKDT = 1.0$ [-]	$\lambda = 5.0$ [-]
12	FD	$REFKDT = 3.0$ [-]	$\lambda = 5.0$ [-]
13	FD	$REFKDT =5.0$ [-]	$\lambda = 5.0$ [-]
14	BATS	$q = 2$ [-]	$\lambda = 5.0$ [-]
15	BATS	$q = 4$ [-]	$\lambda = 5.0$ [-]
16	BATS	$q = 8$ [-]	$\lambda = 5.0$ [-]

Table S.2. Parameterizations for the 25-member Noah-MP ET ensemble.

Configuration number	Noah-MP runoff parameterization	Key runoff parameter value	Key ET parameter value
1	SIM GW	$f = 2.0$ [m^{-1}]	$\lambda = 1.0$ [-]
2	SIM GW	$f = 2.0$ [m^{-1}]	$\lambda = 2.0$ [-]
3	SIM GW	$f = 2.0$ [m^{-1}]	$\lambda = 3.0$ [-]
4	SIM GW	$f = 2.0$ [m^{-1}]	$\lambda = 5.0$ [-]
5	SIM GW	$f = 2.0$ [m^{-1}]	$\lambda = 10.0$ [-]
6	SIM TOP	$f = 6.0$ [m^{-1}]	$\lambda = 1.0$ [-]
7	SIM TOP	$f = 6.0$ [m^{-1}]	$\lambda = 2.0$ [-]
8	SIM TOP	$f = 6.0$ [m^{-1}]	$\lambda = 3.0$ [-]
9	SIM TOP	$f = 6.0$ [m^{-1}]	$\lambda = 5.0$ [-]
10	SIM TOP	$f = 6.0$ [m^{-1}]	$\lambda = 10.0$ [-]
11	FD	$REFKDT = 1.0$ [-]	$\lambda = 1.0$ [-]
12	FD	$REFKDT = 1.0$ [-]	$\lambda = 2.0$ [-]
13	FD	$REFKDT =1.0$ [-]	$\lambda = 3.0$ [-]
14	FD	$REFKDT =1.0$ [-]	$\lambda = 5.0$ [-]
15	FD	$REFKDT =1.0$ [-]	$\lambda = 10.0$ [-]
16	FD	$REFKDT = 3.0$ [-]	$\lambda = 1.0$ [-]
17	FD	$REFKDT = 3.0$ [-]	$\lambda = 2.0$ [-]
18	FD	$REFKDT =3.0$ [-]	$\lambda = 3.0$ [-]
19	FD	$REFKDT =3.0$ [-]	$\lambda = 5.0$ [-]
20	FD	$REFKDT =3.0$ [-]	$\lambda = 10.0$ [-]
21	BATS	$q = 8$ [-]	$\lambda = 1.0$ [-]
22	BATS	$q = 8$ [-]	$\lambda = 2.0$ [-]
23	BATS	$q = 8$ [-]	$\lambda = 3.0$ [-]
24	BATS	$q = 8$ [-]	$\lambda = 5.0$ [-]
25	BATS	$q = 8$ [-]	$\lambda = 10.0$ [-]

2. Inter-day storm events

As discussed in the main text (Section 3.1), there is a potential for inter-day storm events to introduce a spurious high bias into sampled SRCS values that could conceivably overestimate the fraction of study basins where the PC assumption holds (Figure 2). Figure S.1 regenerates Figure 2 except now masking all storm events where the preceding day has an accumulation greater than $P_{min}/5$ (2 mm d^{-1}). In this way, it explicitly removes all storms characterized by significant inter-day rainfall. Note that, contrary to expectations, the percentage of basins where $SRCS_{PC}$ values fall within the 95% confidence interval of sampled SRCS *increases* from 61% in Figure 2 to 68% in Figure S.1. Therefore, there is no indication that inter-day storm events are spuriously increasing values of sampled SRCS and thus over-stating the validity of the PC assumption. Also note that the number of qualifying basins (containing more than 50 frozen-soil-free and snow-free events) declines from 617 in Figure 2 to 572 in Figure S.1 due to the more stringent masking of qualifying storm events.

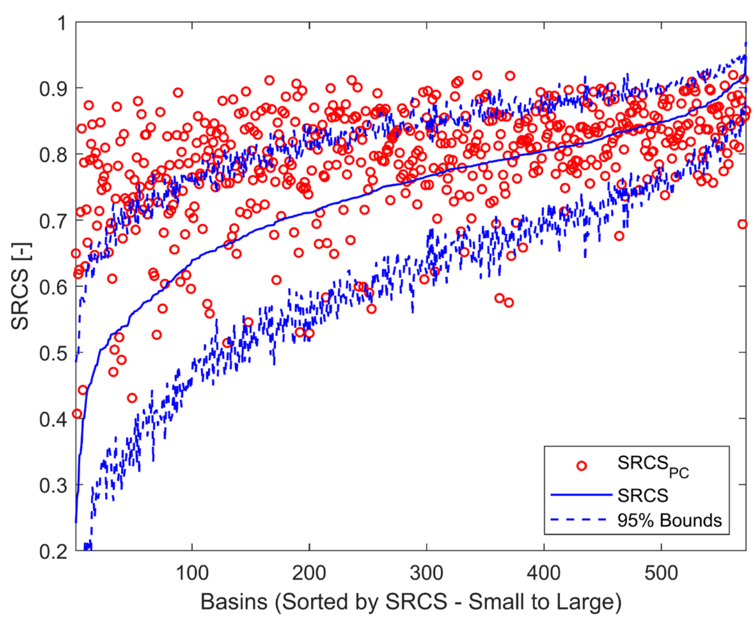

Figure S.1. As in Figure 2 except that all storm events in which the preceding day (i.e., last pre-storm day) has a rainfall accumulation greater than $P_{min}/5$ (2 mm d^{-1}) are masked.

3. Training and testing periods

Figure 5 in the main text is based on calibrated results using the entire available historical period for training. Here, we regenerate Figure 5 for the alternative cases of using either the first (Figure S.2) or second half (Figure S.3) of our analysis period for training and the respective other (non-overlapping) half of the period for testing.

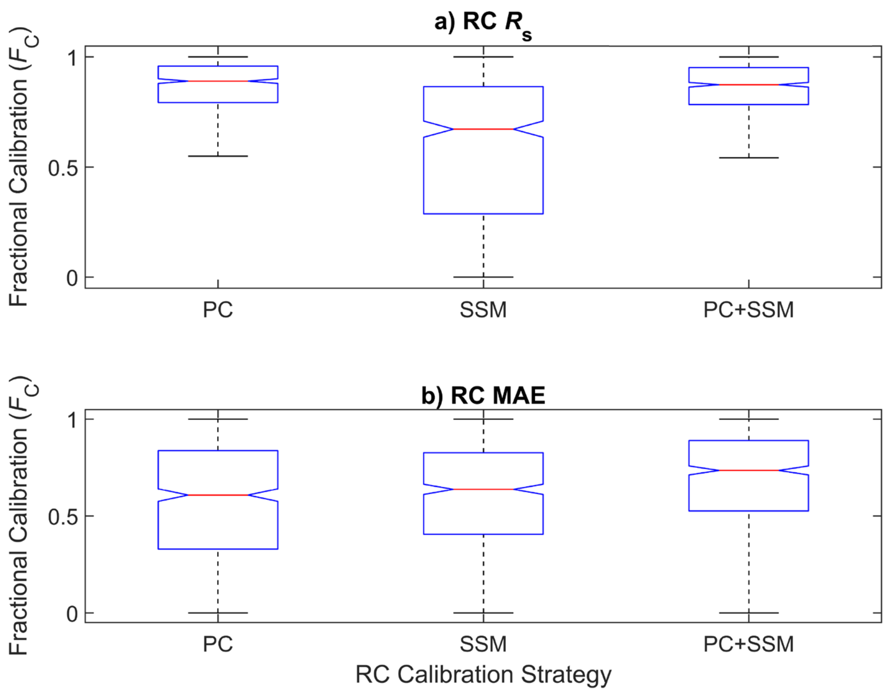

Figure S.2. As in Figure 5 except based on training during the first half of our analysis period (1 April 2015 to 14 December 2017) and testing during the second half of the period (15 December 2017 to 30 August 2020).

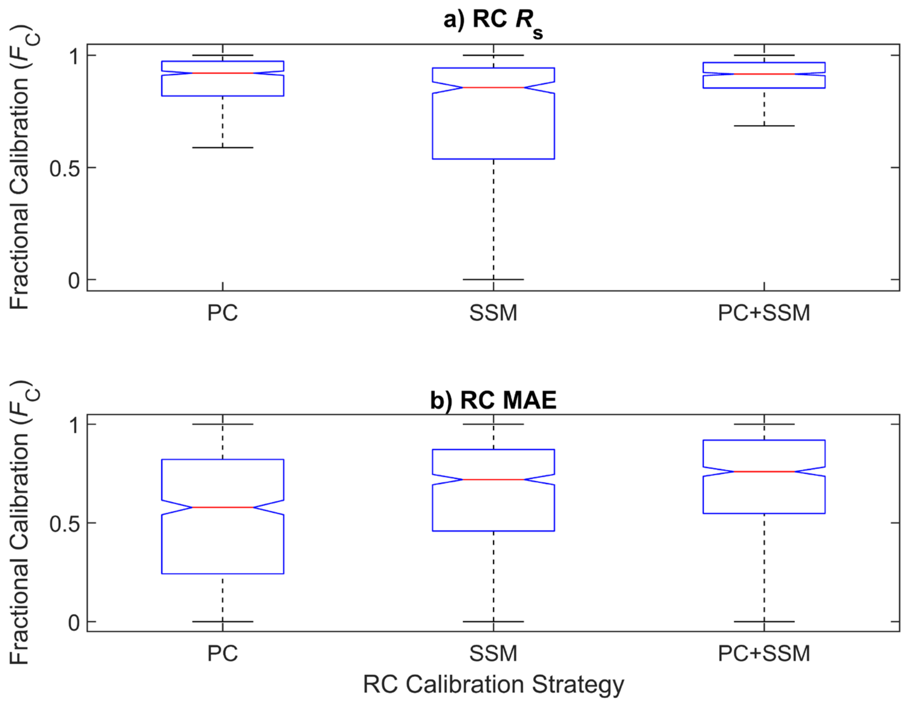

Figure S.3. As in Figure 5 except based on training during the second half of our analysis period (15 December 2017 to 30 August 2020) and testing during the first half of the period (1 April 2015 to 14 December 2017).

4. Role of bias in MAE results

The combination of very good PC+SSM calibration results for the Spearman rank correlation (R_s) metric and relatively poor corresponding results for the MAE metric in Figure 5 suggests that the PC+SSM calibration strategy has difficulty identifying basin-specific RC biases in calibrated Noah-MP results. This suggestion is confirmed in Figure S.4 – which illustrates a strong relationship between PC+SSM calibration MAE F_C results and RC bias. Note how low F_C values are strongly associated with non-zero RC bias values.

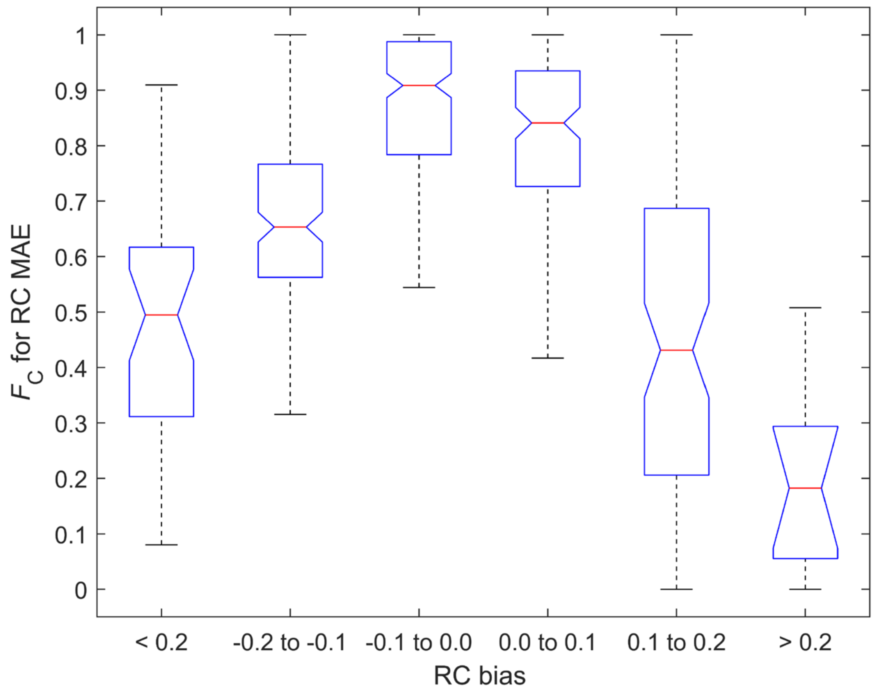

Figure S.4. Box-and-whisker plots describing the basin-wise distribution of Noah-MP F_C results binned by RC MAE results after PC+SSM calibration (Figure 6b).