S1 ERA5-Land soil moisture evaluation

We evaluated ERA5-Land soil moisture against independent in situ observations from the International Soil Moisture Network (ISMN; Dorigo et al., 2011, 2021) to assess magnitude biases and agreement in wetness ordering. Consistent offsets or scaling differences are less consequential for our tree-based machine learning model and catchment-specific percentile inputs than non-systematic errors or biases that distort temporal variations and the ordering of wet and dry conditions.

We evaluated the upper three ERA5-Land soil layers: 0–7, 7–28, and 28–100 cm. We used ISMN sensors for calibration from three networks: CW3E (United States; Ralph et al., 2022), XMS-CAT (Spain; see Acknowledgements), and TERENO (Germany; Zacharias et al., 2011). For each network, we selected the stations based on record length, continuity, and overlap with Global Runoff Data Centre (GRDC) catchments where possible. The deepest ERA5-Land layer, 100–289 cm, was excluded because too few sensors met the depth and data-quality criteria. We averaged daily sensor measurements within each station and layer, then equally averaged available stations. ERA5-Land means equally weighted unique mapped pixels.

We fitted a pooled linear correction for each layer using calibration data from the three networks. Within each network, the first two-thirds of the combined available dates across layers defined the calibration period, and subsequent dates defined the evaluation period. We evaluated the corrections during the subsequent periods within those networks and against independent OzNet observations from Australia (Smith et al., 2012), which were excluded from calibration. Table S1 summarizes station coverage, sensor depths, and record, calibration, and evaluation periods. Performance was assessed using Spearman rank correlation (ρ) to quantify agreement in wetness ordering, squared Pearson correlation (R^2), relative root mean squared error (100 × RMSE divided by the paired ISMN mean), and bias (model minus ISMN mean).

ERA5-Land generally overestimated soil moisture. Nevertheless, squared Pearson correlations of 0.59–0.96 indicated linear association between the reanalysis data and the in-situ observations, and Spearman correlations of 0.73–0.96 indicated strong agreement in daily wetness ordering across network–layer combinations (Fig. S1). The pooled correction reduced relative RMSE and absolute bias in most combinations, indicating a correctable systematic component alongside residual discrepancies.

As an additional surface-soil-moisture benchmark, we evaluated ERA5-Land surface soil moisture (0–7 cm) and Soil Moisture Active Passive (SMAP) Level 3 product SPL3SMP_E (approximately 0–5 cm depth at 9 km grid spacing; O'Neill et al., 2023) against 0–5 cm observations from Spain's REMEDHUS network (González-Zamora et al., 2019; Fig. S2). We retained good-quality SMAP retrievals and estimated missing morning (AM) or evening (PM) values from the opposite overpass using pixel-specific linear models fitted over the satellite record. Daily SMAP values averaged available AM and PM values, followed by averaging within stations and equally across stations. We restricted all series to common valid dates, fitted separate ERA5-Land and SMAP corrections on the first two-thirds of those dates, and evaluated them on the remainder. Correction reduced relative RMSE and absolute bias for both products. ERA5-Land showed higher Spearman correlation (ρ = 0.94 versus 0.90) and squared Pearson correlation (R^2 = 0.89 versus 0.82) with ISMN than SMAP.

Together, these findings and ERA5-Land's longer record, complete spatial coverage, multi-depth structure, and consistency with the precipitation field support its use in this study. In the main analysis, we used the original ERA5-Land values and treated the correction as a diagnostic experiment. This evaluation covered soil moisture only; previous evaluations of ERA5-Land precipitation are discussed in Sect. 2.1 of the main text.

Table S1. ISMN network and sensor coverage and periods used for ERA5-Land evaluation and the SMAP comparison. Station-days are the sum of days with valid daily soil-moisture means across stations for each layer over the retained record; each station contributes at most one station-day per date. Periods may contain gaps. OzNet was used only for independent evaluation.

Network	ERA5-Land layer	ISMN sensor depths (cm)	Stations	Station-days	Record period	Calibration period	Evaluation period
CW3E	1	5	5	13857	2017-06-29 to 2025-07-06	2017-06-29 to 2022-11-17	2022-11-18 to 2025-07-06
	2	10; 15; 20	5	13660	2017-06-29 to 2025-07-06	2017-06-29 to 2022-11-17	2022-11-18 to 2025-07-06
	3	50; 100	1	2714	2017-10-07 to 2025-03-24	2017-10-07 to 2022-11-17	2022-11-18 to 2025-03-24
TERENO	1	5	3	12283	2011-08-04 to 2025-04-11	2011-08-04 to 2020-06-06	2020-06-07 to 2025-04-11
	2	20	3	13112	2011-08-04 to 2025-06-29	2011-08-04 to 2020-06-06	2020-06-07 to 2025-06-29
	3	50	4	16763	2010-01-14 to 2025-06-29	2010-01-14 to 2020-06-06	2020-06-07 to 2025-06-29
XMS-CAT	1	5	6	14321	2017-03-30 to 2025-07-06	2017-03-30 to 2022-07-14	2022-07-15 to 2025-07-05
	2	10; 20	9	20890	2016-08-01 to 2025-07-06	2016-08-01 to 2022-07-14	2022-07-15 to 2025-07-05
	3	30; 40; 50; 75; 100	10	23293	2016-08-01 to 2025-07-06	2016-08-01 to 2022-07-14	2022-07-15 to 2025-07-05
OzNet	1	0–8	4	10194	2001-11-13 to 2008-12-31	—	2001-11-13 to 2008-12-31
	2	0–30	10	20967	2001-11-13 to 2008-12-31	—	2001-11-13 to 2008-12-31
	3	30–60; 60–90	8	17264	2001-11-13 to 2008-12-31	—	2001-11-13 to 2008-12-31
REMEDHUS	1	0–5	22	125581	2005-03-15 to 2025-01-01	2015-04-01 to 2020-12-27	2020-12-29 to 2023-12-30

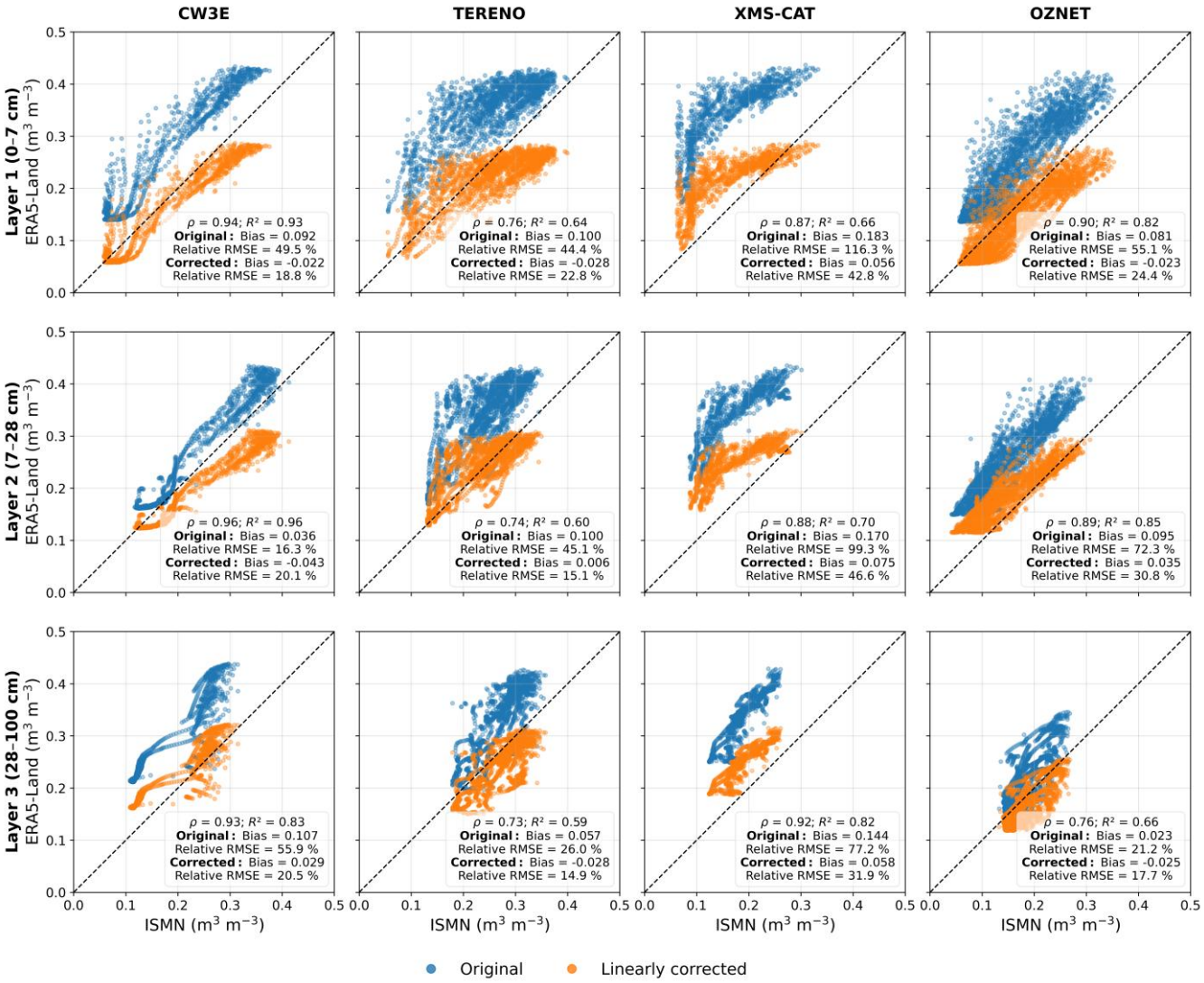


Figure S1: Evaluation of ERA5-Land soil moisture against ISMN observations for layers 1–3 (0–7, 7–28, and 28–100 cm) at CW3E, TERENO, XMS-CAT, and independent OzNet sites. Blue and orange points show original and pooled linearly corrected values during the evaluation periods in Table S1. Dashed lines indicate 1:1 agreement. Spearman ρ measures agreement in wetness ordering, and R^2 is the squared Pearson correlation. Both are reported once per panel because they are unchanged by the positive linear corrections. Relative RMSE is 100 × RMSE divided by the paired ISMN mean; bias is mean ERA5-Land minus mean ISMN soil moisture ($m^3\ m^{-3}$).

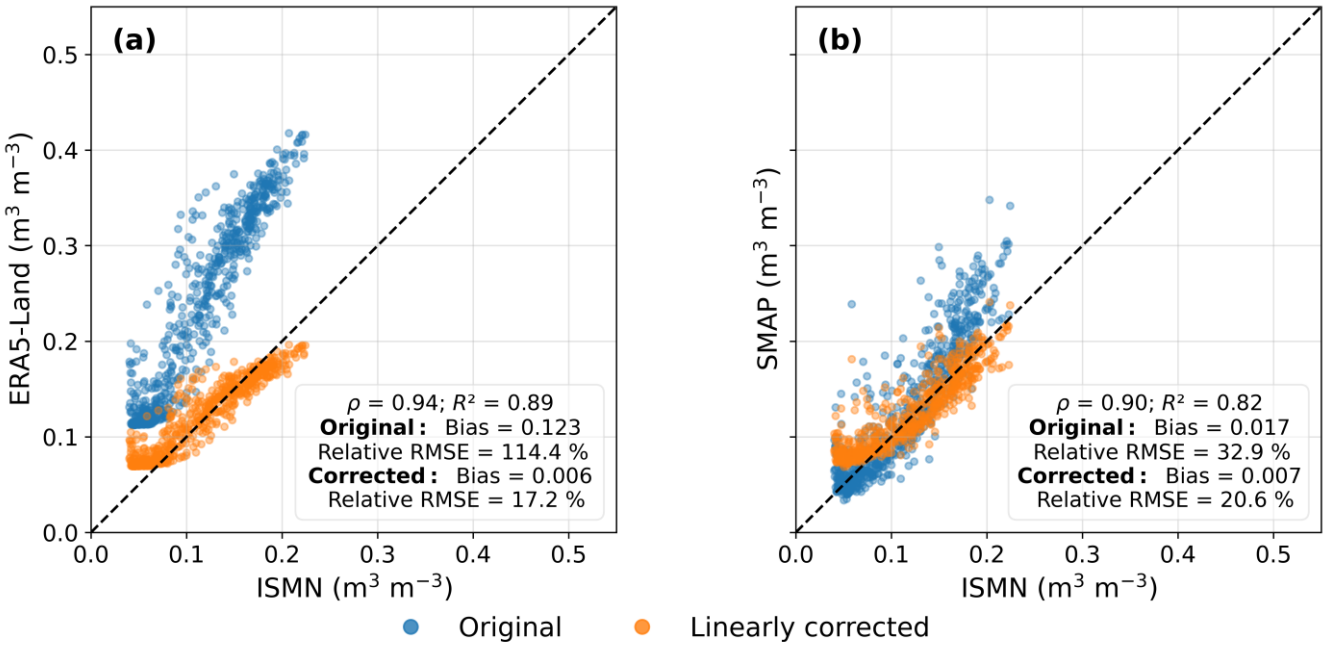


Figure S2: Surface soil moisture versus ISMN observations at REMEDHUS: (a) ERA5-Land layer 1 (0–7 cm) and (b) SMAP surface soil moisture (approximately 0–5 cm). Blue and orange points show original and linearly corrected values on common evaluation dates. Separate corrections were fitted during the calibration period in Table S1. Dashed lines indicate 1:1 agreement. Spearman ρ measures agreement in wetness ordering, and R^2 is the squared Pearson correlation. Both are reported once per panel because they are unchanged by the positive linear corrections. Relative RMSE is 100 × RMSE divided by the paired ISMN mean; bias is mean model minus mean ISMN soil moisture (m³ m⁻³).

S2 Predictor selection and lag-time estimation

To support the selection of precipitation and soil-moisture predictor timing, we estimated characteristic catchment lag times for the study catchments. The objective was to choose a predictor configuration that captures the short-term hydrological response relevant to annual-maximum streamflow events, while retaining a consistent event-based framework across a large and diverse sample of catchments.

Catchment lag time (t_L) was estimated using the empirical equation of Watt and Chow (1985), which was developed for natural catchments with limited lake or swamp storage:

$$t_L = 0.000326 \left(\frac{L}{\sqrt{S}} \right)^{0.79} \qquad \text{(S1)}$$

t_L is the lag time (h), L is the channel length (m), and S is the dimensionless channel slope (m m⁻¹). Because channel length was unavailable, it was estimated using Hack's law (Hack, 1957):

$$L = 1.4A^{0.6} \qquad \text{(S2)}$$

A is the catchment area (km²) and L is the estimated channel length (km). The resulting length was converted from kilometers to meters before applying Eq. S1. The channel slope values, provided in dm km⁻¹, were used as a proxy for S and converted to m m⁻¹ by multiplying by 10^{-4}.

The resulting t_L values were calculated for all 1,374 catchments in the initial model sample and converted from hours to days. The distribution of estimated lag times is shown in Fig. S3. The median lag time is 0.67 days, and the mean lag time is 0.83 days, indicating that most catchments in the study sample have short characteristic response times.

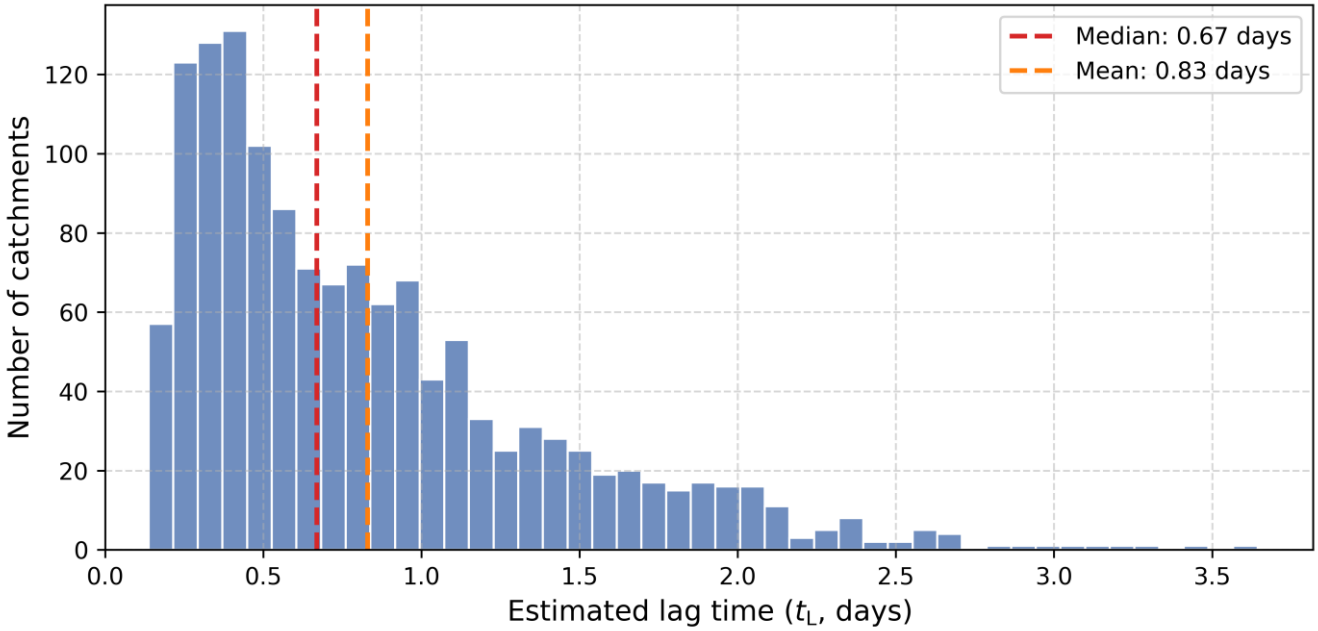


Figure S3: Distribution of empirical catchment lag times (t_L) estimated for the 1,374 catchments used for model training and evaluation, using Hack's law and the Watt–Chow equation. Dashed vertical lines indicate the median and mean lag times.

A key challenge in predictor selection is balancing three requirements: capturing the relevant catchment response time, focusing on the immediate drivers of flood generation, and maintaining a large multi-continental catchment sample. Longer precipitation lookback windows would capture catchments with slower responses but move the analysis away from the immediate event-scale forcing targeted in this study. To reduce this trade-off, the main analysis was restricted to catchments smaller than 5,000 km², limiting the inclusion of very large catchments whose slower hydrological responses may be less well represented by short near-event predictor windows.

The estimated lag-time stratification supports the use of a short precipitation window. As shown in Fig. S4a, 969 catchments (70.5 %) have estimated lag times of ≤ 1 day, 337 (24.5 %) have lag times of >1 to 2 days, and 62 (4.5 %) have lag times of >2 to 3 days. These intervals correspond to the assigned precipitation-day categories t, $t-1$, and $t-2$, respectively. Together, they account for 99.6 % of the catchments and fall within the selected 3-day precipitation window from $t-2$ to t. Only 6 catchments (0.4 %) fall outside this window. The relationship between lag category and catchment size is evident in Fig. S4b: catchments with longer estimated lag times tend to have larger drainage areas, with median areas increasing from 472 km² for the same-day category to 4,295 km² for catchments outside the 3-day precipitation window.

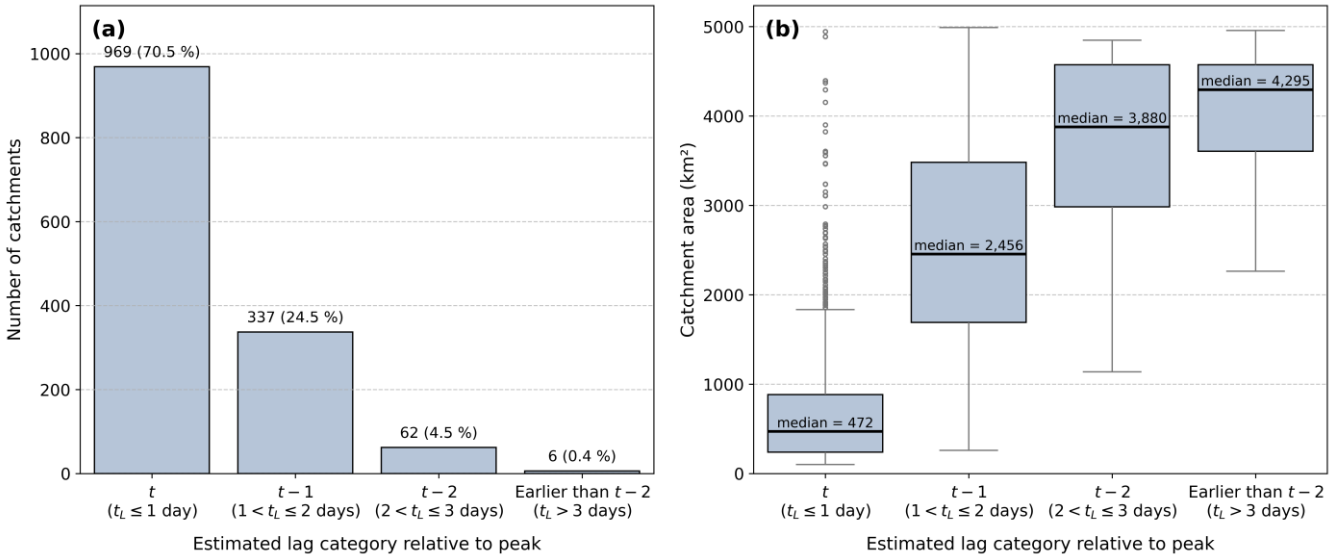


Figure S4: Catchment stratification by estimated temporal lag category. (a) Number and percentage of catchments assigned to each lag category. (b) Distribution of catchment area within each lag category, with median values indicated. The first three categories are assigned to precipitation days *t*, *t* − 1, and *t* − 2, respectively, and fall within the selected 3-day precipitation window (from *t* − 2 to *t*) and together account for 99.6 % of the study catchments.

To evaluate the robustness of this predictor selection, we performed a controlled predictor-sensitivity analysis. We compared four configurations while holding the selected LightGBM hyperparameters, cross-validation procedure, evaluation sample, and evaluation framework fixed, as described in Sect. 2.3 of the main text. The four configurations were as follows: (i) precipitation accumulated over t and $t-1$ with the four soil-moisture layers at $t-2$; (ii) precipitation accumulated over $t-1$ and $t-2$ with the four soil-moisture layers at $t-3$; (iii) the selected configuration, precipitation accumulated over $t-2$ to t with the four soil-moisture layers at $t-3$; and (iv) the same 3-day precipitation window with mean soil moisture percentiles across the four layers at $t-3$.

Predictive performance was very similar among the three configurations retaining the four individual soil-moisture layers (Table S2), with only small differences in daily RMSE, Nash–Sutcliffe efficiency (NSE), bias, mean absolute error (MAE), and median catchment-wise NSE. Replacing the individual soil-moisture layers with their mean, however, resulted in higher RMSE and MAE but lower NSE and median catchment-wise NSE, indicating that retaining vertical soil-moisture information was beneficial. Bias remained close to zero in all configurations.

The distribution of Shapley additive explanations (SHAP) contributions among the individual soil-moisture layers was more sensitive to antecedent timing. When soil moisture at $t-2$ was used, layers 2 and 3 had comparable event-level prominence, with the largest absolute SHAP value among the soil-moisture layers in 42.3 % and 43.0 % of events, respectively. In contrast, for both configurations using soil moisture at $t-3$, layer 3 was the strongest-contributing soil layer in approximately 73 % of events. Thus, the relative prominence of individual soil-moisture depths depends partly on the timing at which antecedent wetness is represented and should not be interpreted as evidence of a uniquely controlling physical depth.

Table S2. Sensitivity of model performance to predictor configuration based on the full daily evaluation sample. RMSE, NSE, and bias are reported as mean ± standard deviation across the 25 validation-fold model fits. MAE is calculated from the full daily out-of-fold predictions, and median catchment-wise NSE is calculated across the 1,374 catchments from their daily out-of-fold prediction series. RMSE, MAE, and bias are expressed in percentile units. P denotes accumulated precipitation percentile, and SM denotes four separate layer-specific soil-moisture percentiles, and mean SM denotes their mean.

Predictor configuration	RMSE	NSE	Bias	MAE	Median catchment-wise NSE (n = 1,374)
P (t, $t-1$), SM ($t-2$)	0.209 ± 0.001	0.576 ± 0.003	0.002 ± 0.002	0.157	0.600
P ($t-1$, $t-2$), SM ($t-3$)	0.209 ± 0.001	0.577 ± 0.003	0.002 ± 0.002	0.157	0.596
Selected model: P (t, $t-1$, $t-2$), SM ($t-3$)	0.209 ± 0.001	0.575 ± 0.003	0.002 ± 0.002	0.157	0.594
P (t, $t-1$, $t-2$), mean SM ($t-3$)	0.227 ± 0.001	0.500 ± 0.004	0.000 ± 0.002	0.179	0.549

Despite this redistribution of the contributions among the individual soil-moisture layers, the grouped precipitation–soil-moisture result was qualitatively robust. To isolate the effect of predictor configuration, the dominance index D was evaluated for the same 32,930 annual-maximum events retained for the main SHAP-based interpretation. The main-analysis event-selection criteria were not independently re-applied to each alternative configuration; instead, the annual-maximum event sample was held fixed across all four models (Fig. S5). The magnitude of D showed some sensitivity to predictor timing, particularly for the configuration using soil moisture at $t-2$. Nevertheless, mean and median D remained negative in all configurations, and events with $D < 0$ continued to constitute the majority. Although the relative magnitude of precipitation and soil-moisture contributions depends partly on predictor timing, the magnitude of the net grouped soil-moisture contribution exceeds the absolute precipitation contribution for most events across all tested configurations. Here, D compares the absolute precipitation contribution with the magnitude of the net contribution obtained by summing the signed SHAP values across the four soil-moisture layers; this differs from the individual-predictor ranking, in which the predictor with the largest absolute SHAP value is identified for each event.

Overall, the lag-time analysis and predictor-sensitivity tests support the selected configuration of 3-day accumulated precipitation from $t-2$ to t and four-layer antecedent soil moisture at $t-3$. The selected precipitation window covers the estimated response timing of 99.6 % of the study catchments. Soil moisture at $t-3$ precedes the entire precipitation window, thereby avoiding temporal overlap between the antecedent-state and event-precipitation predictors and reducing the possibility that soil moisture already reflects precipitation included in the event forcing. Predictive performance was comparable among

the layer-specific timing configurations, whereas averaging soil moisture across layers increased RMSE and MAE and reduced NSE and median catchment-wise NSE. Although the allocation of contributions among individual soil layers and the magnitude of *D* showed some sensitivity to predictor timing, the broader precipitation–soil-moisture contribution pattern remained qualitatively consistent. The selected setup therefore provides a practical compromise between physical consistency, event-scale interpretation, preservation of vertical soil-moisture information, sample size, and predictive performance.

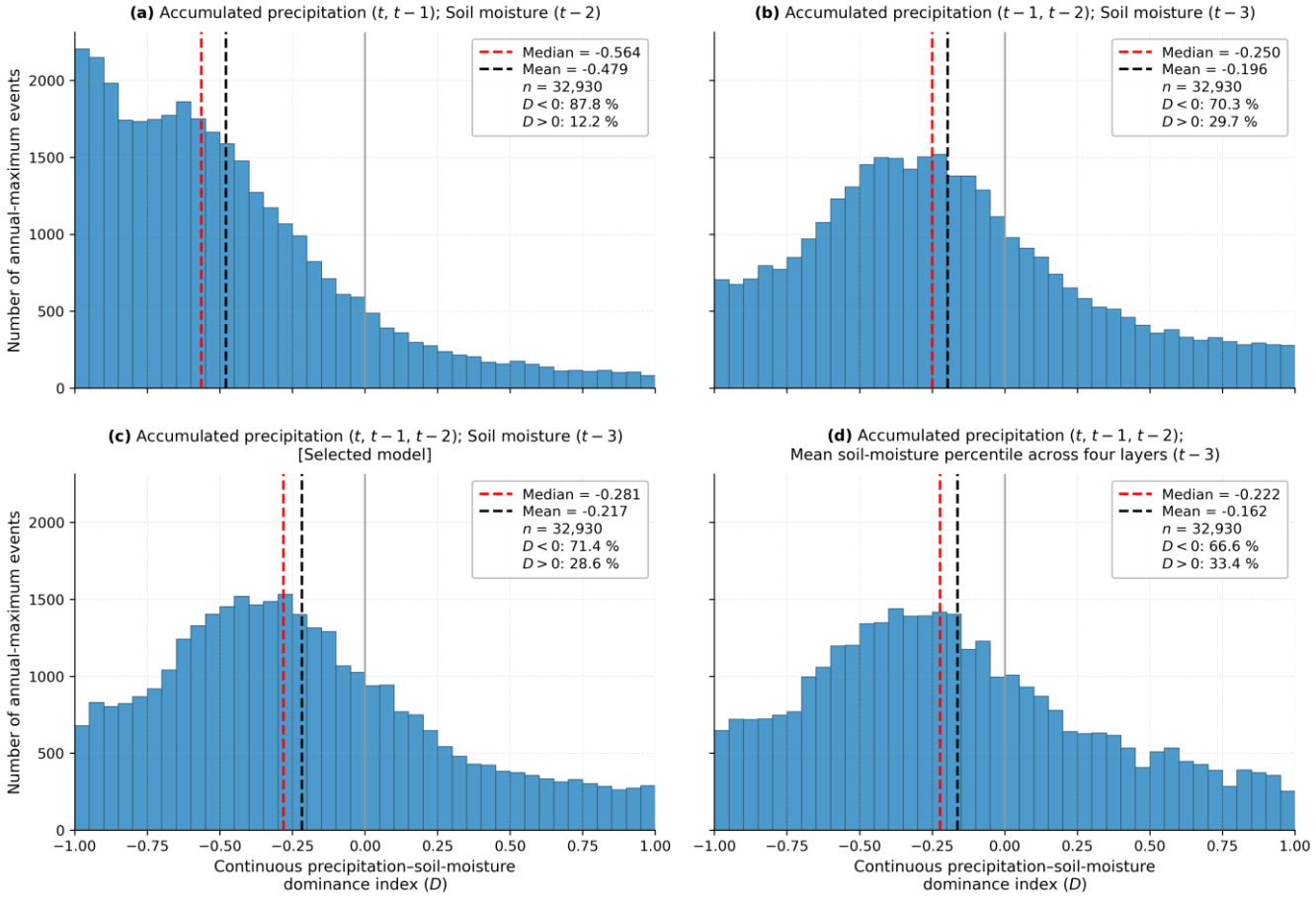


Figure S5: Sensitivity of the precipitation–soil-moisture dominance index *D* to predictor configuration. All configurations use the same 32,930 annual-maximum events retained for the main SHAP-based interpretation; event-selection criteria were not reapplied to alternative models, isolating sensitivity to predictor choice. Configurations comprise accumulated precipitation over (a) $t-1$ to t, with soil moisture at $t-2$; (b) $t-2$ to $t-1$, with soil moisture at $t-3$; (c) $t-2$ to t, with soil moisture at $t-3$ (selected model); and (d) the same precipitation window as (c), with soil-moisture percentiles averaged across four layers at $t-3$. Panels (a)–(c) retain four separate soil-moisture-layer predictors. All predictors are expressed as catchment-specific percentiles. Red and black dashed lines mark median and mean *D*, respectively; gray lines mark $D = 0$. Legends report event counts and percentages with soil-moisture dominance ($D < 0$) and precipitation dominance ($D > 0$).

S3 Model configuration and contribution sensitivity

This section provides additional information on model configuration and sensitivity analyses supporting the SHAP-based contribution results. Table S3 lists the candidate LightGBM hyperparameter values explored during model tuning. Figure S6

shows the contribution results obtained using a stricter event-level prediction-error threshold of |predicted − observed| ≤ 0.2, compared with the 0.3 threshold used in the main analysis.

Table S3. Candidate hyperparameter values explored for the LightGBM model.

Hyperparameter	Meaning	Candidate values
learning_rate	Step size of each boosting iteration	[0.01, 0.05, 0.1]
max_depth	Maximum depth of individual decision trees	[10, 11, 12]
num_leaves	Maximum number of leaves per tree	[511, 1023]
min_child_samples	Minimum number of training samples in a leaf node	[500, 1000, 2000]

As a sensitivity test, the SHAP-based contribution analysis was repeated using a stricter event-level prediction-error threshold of |predicted − observed| ≤ 0.2, instead of the 0.3 threshold used in the main analysis. This reduced the retained event sample from 32,930 to 25,693 well-predicted annual-maximum events. The individual-predictor contribution pattern remained stable: precipitation was the strongest individual predictor in 54.7 % of retained events, while one of the four soil-moisture layers was the strongest individual predictor in the remaining 45.3 %. Layer 3 was the strongest individual predictor in 40.5 % of all retained events. These results indicate that the individual-predictor contribution pattern is not sensitive to the selected event-error threshold. We retained the 0.3 absolute-error threshold in the main analysis because it preserved a larger event sample while maintaining robust contribution patterns.

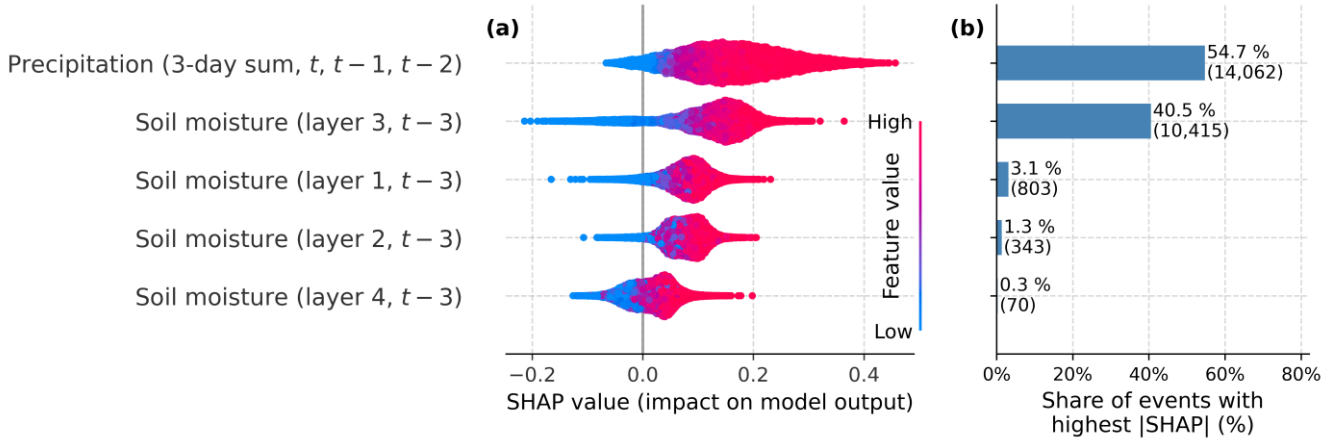


Figure S6: Sensitivity of individual-predictor SHAP contributions to a stricter event-level prediction-error threshold of |predicted − observed| ≤ 0.2. (a) SHAP summary plot for retained annual-maximum events. Predictors are ordered by mean absolute SHAP value. Each point represents one event; the x-axis shows the SHAP value, indicating the predictor's contribution to the predicted streamflow percentile, and color indicates the predictor percentile. (b) Share of retained annual-maximum events for which each predictor has the largest absolute SHAP value. The upper bar shows events for which precipitation was the strongest individual predictor; the remaining bars show events for which a soil-moisture layer was the strongest individual predictor, separated by layer. Labels show percentages and event counts. Soil-moisture layers 1–4 correspond to depths of 0–7, 7–28, 28–100, and 100–289 cm, respectively.

S4 Additional analyses

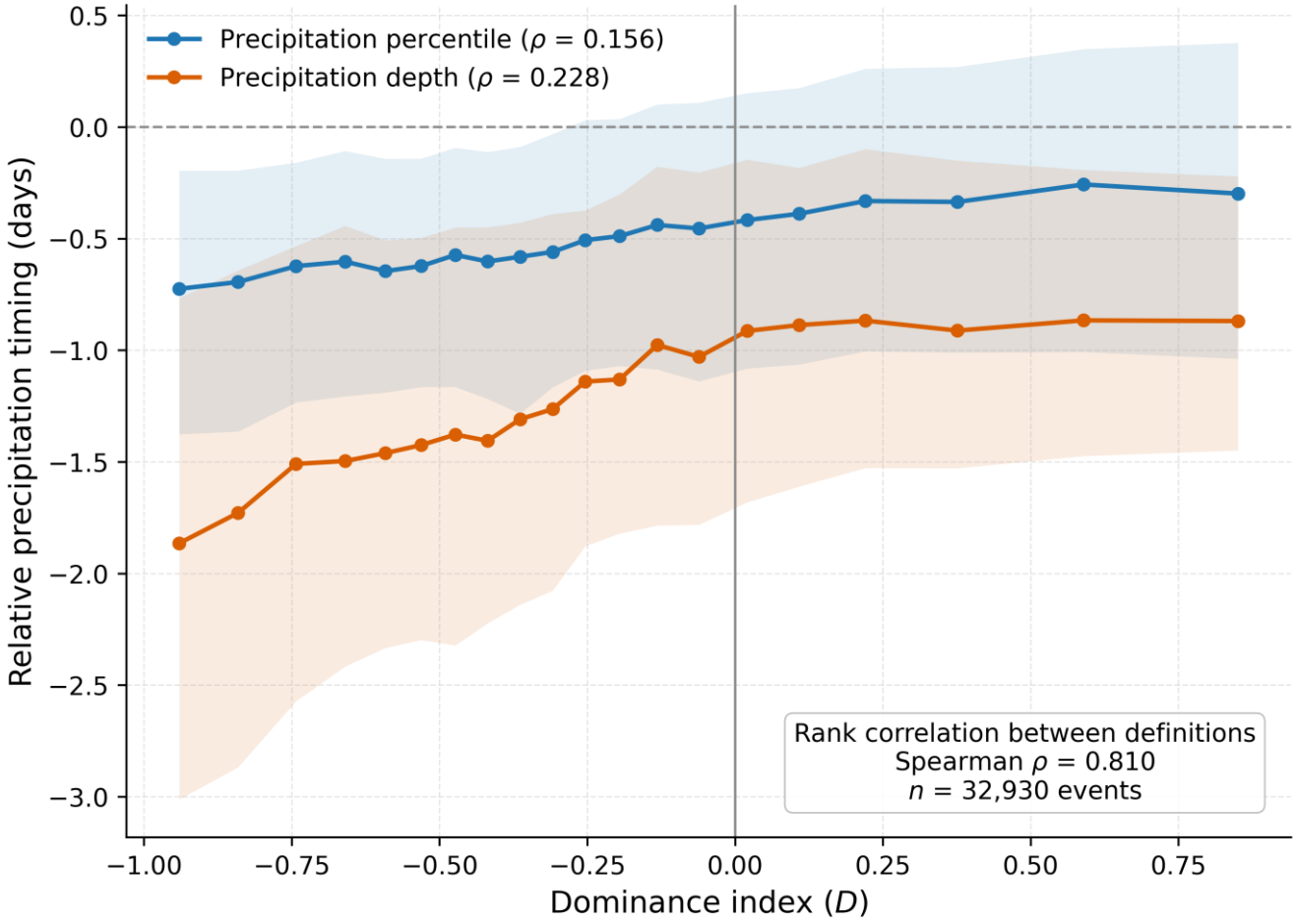


Figure S7: Sensitivity of relative precipitation timing to precipitation weighting. The precipitation center of mass relative to the annual-maximum streamflow peak was calculated over the 15-day window from $t - 7$ to $t + 7$, using either daily catchment-specific precipitation percentiles (blue) or daily precipitation depth at mm d^{-1} (orange) as weights. Negative values indicate precipitation concentrated before the streamflow peak. Events were divided into 20 equal-count bins according to the dominance index (D); points and lines show the median timing within each bin, and shaded bands show the corresponding interquartile ranges. The vertical gray line marks $D = 0$, and the horizontal dashed line marks precipitation centered on the peak day. Spearman's ρ values in the legend describe the association between D and each timing definition across the 32,930 individual events. The annotation reports the event-level Spearman correlation between the two timing definitions.

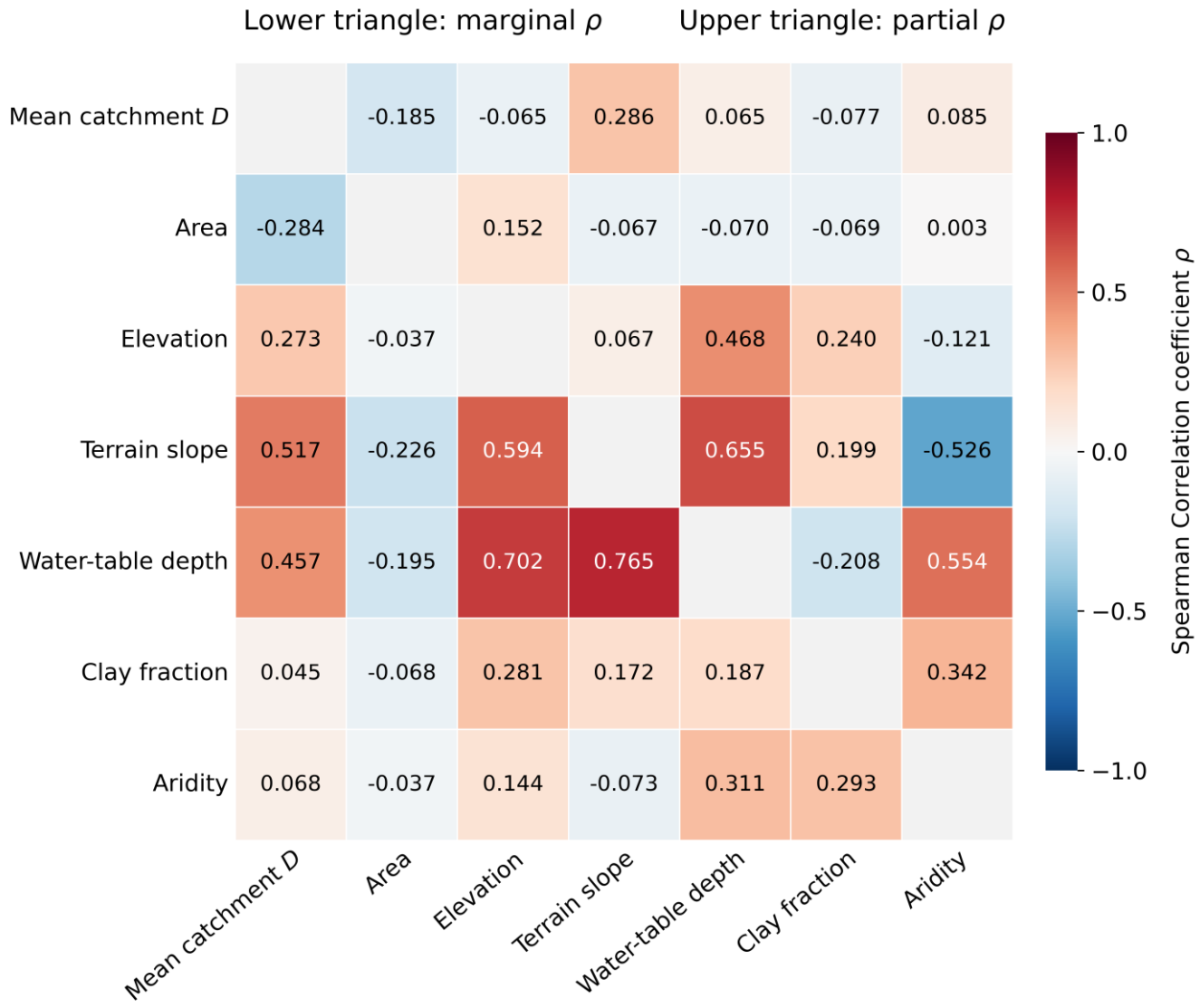


Figure S8: Marginal and partial Spearman correlations (ρ) among the catchment-mean dominance index (D) and six continuous catchment attributes. The lower triangle shows pairwise marginal Spearman correlations, whereas the upper triangle shows partial Spearman correlations after controlling for all remaining variables. Coefficients were calculated across the same 1,199 catchments with at least five retained annual-maximum events. Red and blue indicate positive and negative correlations, respectively.

Acknowledgements

We acknowledge the work of Agnès Lladós and Lola Boquera as well as the XMS-CAT network team in support of the ISMN for providing the soil-moisture observations used in Supplement S1.

References

Dorigo, W., van Oevelen, P., Wagner, W., Drusch, M., Mecklenburg, S., Robock, A., and Jackson, T.: A New International Network for in Situ Soil Moisture Data, Eos Trans. AGU, 92, 141–142, https://doi.org/10.1029/2011EO170001, 2011.

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., Preimesberger, W., Xaver, A., Annor, F., Ardö, J., Baldocchi, D., Bitelli, M., Blöschl, G., Bogena, H., Brocca, L., Calvet, J.-C., Camarero, J. J., Capello, G., Choi, M., Cosh, M. C., van de Giesen, N., Hajdu, I., Ikonen, J., Jensen, K. H., Kanniah, K. D., de Kat, I., Kirchengast, G., Kumar Rai, P., Kyrouac, J., Larson, K., Liu, S., Loew, A., Moghaddam, M., Martínez Fernández, J., Mattar Bader, C., Morbidelli, R., Musial, J. P., Osenga, E., Palecki, M. A., Pellarin, T., Petropoulos, G. P., Pfeil, I., Powers, J., Robock, A., Rüdiger, C., Rummel, U., Strobel, M., Su, Z., Sullivan, R., Tagesson, T., Varlagin, A., Vreugdenhil, M., Walker, J., Wen, J., Wenger, F., Wigneron, J. P., Woods, M., Yang, K., Zeng, Y., Zhang, X., Zreda, M., Dietrich, S., Gruber, A., van Oevelen, P., Wagner, W., Scipal, K., Drusch, M., and Sabia, R.: The International Soil Moisture Network: serving Earth system science for over a decade, Hydrol. Earth Syst. Sci., 25, 5749–5804, https://doi.org/10.5194/hess-25-5749-2021, 2021.

González-Zamora, Á., Sánchez, N., Pablos, M., and Martínez-Fernández, J.: CCI soil moisture assessment with SMOS soil moisture and in situ data under different environmental conditions and spatial scales in Spain, Remote Sens. Environ., 225, 469–482, https://doi.org/10.1016/j.rse.2018.02.010, 2019.

Hack, J. T.: Studies of longitudinal stream profiles in Virginia and Maryland, U.S. Geological Survey Professional Paper 294-B, U.S. Geological Survey, 45–97, https://doi.org/10.3133/pp294B, 1957.

O'Neill, P. E., Chan, S., Njoku, E. G., Jackson, T., Bindlish, R., Chaubell, J., and Colliander, A.: SMAP Enhanced L3 Radiometer Global and Polar Grid Daily 9 km EASE-Grid Soil Moisture (SPL3SMP_E, Version 6), NASA National Snow and Ice Data Center Distributed Active Archive Center [data set], https://doi.org/10.5067/M20OXIZHY3RJ, 2023.

Ralph, F. M., Wilson, A. M., Alden, D., Ellis, C. J., Cooper, A., Paulsson, K., Kawzenuk, B. K., and Yao, P.: Center for Western Weather and Water Extremes (CW3E) Surface Meteorological Observational Dataset, UC San Diego Library Digital Collections [data set], https://doi.org/10.6075/J0SX6DDH, 2022.

Smith, A. B., Walker, J. P., Western, A. W., Young, R. I., Ellett, K. M., Pipunic, R. C., Grayson, R. B., Siriwardena, L., Chiew, F. H. S., and Richter, H.: The Murrumbidgee soil moisture monitoring network data set, Water Resour. Res., 48, W07701, https://doi.org/10.1029/2012WR011976, 2012.

Watt, W. E. and Chow, K. C. A.: A general expression for basin lag time, Can. J. Civ. Eng., 12, 294–300, https://doi.org/10.1139/l85-031, 1985.

Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens, O., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Helle, G., Kiese, R., Kunstmann, H., Klotz, S., Munch, J. C., Papen, H., Priesack, E., Schmid, H. P., Steinbrecher, R., Rosenbaum, U., Teutsch, G., and Vereecken, H.: A Network of Terrestrial Environmental Observatories in Germany, Vadose Zone J., 10, 955–973, https://doi.org/10.2136/vzj2010.0139, 2011.