Supplement of Hydrol. Earth Syst. Sci., 30, 2817–2835, 2026
https://doi.org/10.5194/hess-30-2817-2026-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

Reconstructed soil moisture droughts in Belgium reveal 2011–2020 was the driest decade since 1970

Katoria Lekarkar et al.

Correspondence to: Katoria Lekarkar (katoria.lesaalon.lekarkar@vub.be)

S1 Drought-day classification skill

To explicitly assess whether the model reproduces drought conditions observed by the in situ measurements, we performed a daily drought-day classification using a percentile threshold. For each station, drought days in the observations were defined as days on which the standardized soil moisture anomaly fell below the 20th percentile of the observed anomaly distribution over the overlapping period,

$$\tau = P_{20}\left(z_t^{\text{obs}}\right), \tag{S1}$$

where z_t^{obs} denotes the observed daily standardized anomaly and $P_{20}(\cdot)$ is the empirical 20th percentile computed using all overlapping days with valid observations. Observed and model drought-day indicators were then defined as

$$D_t^{\text{obs}} = \mathbb{I}\left(z_t^{\text{obs}} \leq \tau\right), \qquad D_t^{\text{mod}} = \mathbb{I}\left(z_t^{\text{mod}} \leq \tau\right), \tag{S2}$$

where z_t^{mod} is the model anomaly $\mathbb{I}(\cdot)$ equals 1 when the condition is true (drought day) and 0 otherwise (non-drought day).

We quantified drought-day classification skill using the standard contingency-table counts: hits (H, true positives), misses (M, false negatives), false alarms (F, false positives), and correct negatives (C, true negatives).

Hits (H) are days on which both the model and the observations indicate drought ($D_t^{\text{mod}} = 1$ and $D_t^{\text{obs}} = 1$). Misses (M) are observed drought days that the model fails to flag as drought ($D_t^{\text{mod}} = 0$ and $D_t^{\text{obs}} = 1$). False alarms (F) are days flagged as drought by the model but not by the observations ($D_t^{\text{mod}} = 1$ and $D_t^{\text{obs}} = 0$). Correct negatives (C) are days on which both the model and the observations indicate non-drought conditions ($D_t^{\text{mod}} = 0$ and $D_t^{\text{obs}} = 0$). These counts were obtained by summing the number of days satisfying each condition over the evaluation period.

$$H = \sum_t \mathbb{I}\left(D_t^{\text{mod}} = 1 \wedge D_t^{\text{obs}} = 1\right), \quad M = \sum_t \mathbb{I}\left(D_t^{\text{mod}} = 0 \wedge D_t^{\text{obs}} = 1\right), \tag{S3}$$

$$F = \sum_t \mathbb{I}\left(D_t^{\text{mod}} = 1 \wedge D_t^{\text{obs}} = 0\right), \quad C = \sum_t \mathbb{I}\left(D_t^{\text{mod}} = 0 \wedge D_t^{\text{obs}} = 0\right), \tag{S4}$$

from which we computed;

(i) the hit rate/recall

$$\text{Recall} = \frac{H}{H+M} \tag{S5}$$

(ii) the false alarm rate,

$$\text{FAR} = \frac{F}{F+C} \tag{S6}$$

and (iii) the F_1 score,

$$\text{F}_1 = \frac{2H}{2H+F+M} \tag{S7}$$

Hit rate/recall measures the fraction of observed drought days correctly detected by the model, FAR measures the fraction of observed non-drought days incorrectly flagged as drought, and the F_1 score summarizes overall drought-day classification skill by balancing misses and false alarms.

Using a daily drought-day classification based on the 20th-percentile threshold, the model shows high skill in reproducing observed drought conditions across the 21 stations. Averaged over stations, the hit rate (recall) was 0.74, indicating that the model detected 74% of observed drought days, while the false alarm rate was only 0.05, implying that about 5% of observed non-drought days were incorrectly flagged as drought. The mean F_1 score (summarizing the balance between misses and false alarms) was 75%.

S2 Model Performance Evaluation: Streamflow

We quantified the skill of the model to reproduce observed discharge using the Nash–Sutcliffe Efficiency (NSE; Nash and Sutcliffe (1970)), which is defined as

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{T} \left(Q_t^{\mathrm{obs}} - Q_t^{\mathrm{sim}}\right)^2}{\sum_{t=1}^{T} \left(Q_t^{\mathrm{obs}} - \overline{Q}^{\mathrm{obs}}\right)^2} \tag{S8}$$

where Q_t^{obs} and Q_t^{sim} are observed and simulated discharge at time step t, $\overline{Q}^{\mathrm{obs}}$ is the mean observed discharge over the evaluation period, and T is the number of time steps. NSE ranges from $-\infty$ to 1, where NSE $= 1$ indicates perfect agreement, NSE ≈ 0 indicates skill comparable to using the observed mean discharge as a predictor, and NSE < 0 indicates performance worse than this mean-flow benchmark. Following common practice in hydrological model evaluation, NSE values ≥ 0.5 are generally considered satisfactory for streamflow simulations while NSE ≥ 0.75 is judged as very good model performance (Moriasi et al., 2007).

S2.1 Model performance for daily discharge

The skill of the model in reproducing observed daily discharge is summarised in Figure S1. To ensure a robust assessment, we retained only gauging stations with at least 10 years of observations and excluded stations with peak discharge ≤ 10 $\mathrm{m^3\,s^{-1}}$. Overall, the model reproduced daily discharge well. During calibration, the mean NSE across stations was 0.62, and 80% of stations achieved NSE ≥ 0.5. Validation results were comparable, with a mean NSE of 0.63 and 83% of stations exceeding 0.5, indicating good temporal transferability (Klemeš, 1986).

Spatial patterns in NSE (Figure S1) show broadly consistent skill across the domain. The highest NSE values (NSE ≥ 0.75) occurred predominantly in larger basins, where drainage areas can be delineated more reliably. Lower NSE values were more common in smaller basins, and especially in the low-relief northern part of the domain where catchment boundaries are less

distinct. In some cases, anthropogenic modification of rivers, such as canalisation, diversions and diking, which is common in the northern lowlands and which is not implemented in the model, explained poor model performance at some gauging stations. Notwithstanding these few cases, the results demonstrate that the model provides a reliable, spatially consistent basis for assessing soil moisture dynamics over the country.

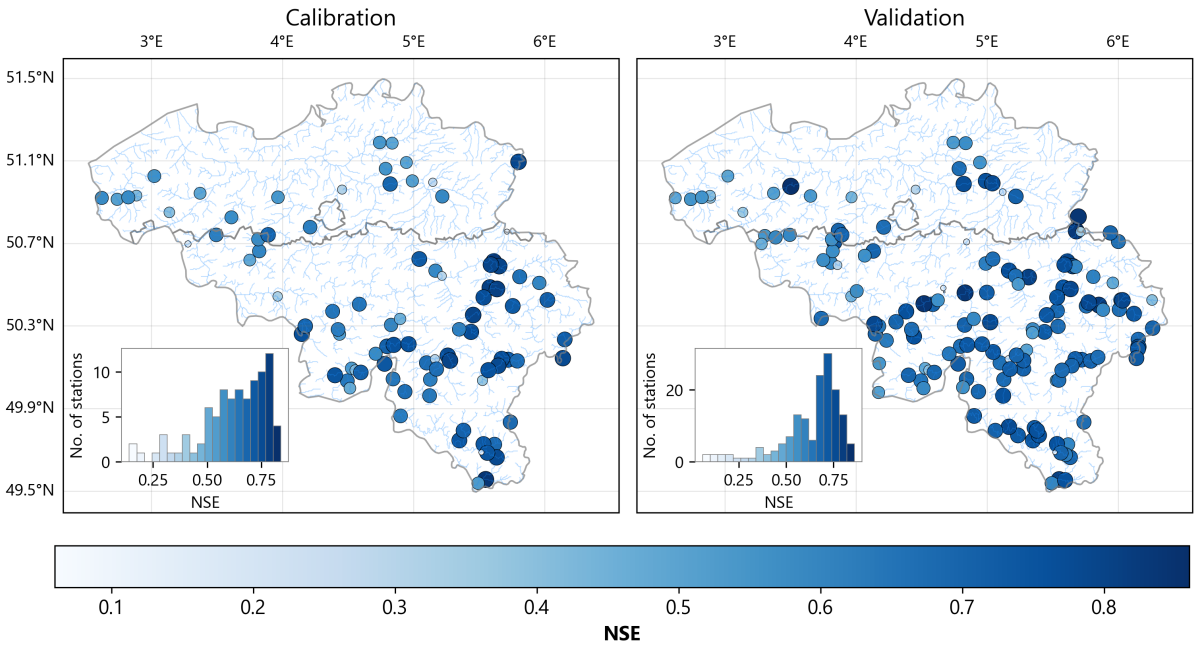

Figure S1: Model performance at gauging stations across Belgium during calibration and validation periods. The colour intensity and size of each circle are proportional to the NSE value. The inset histograms show the distribution of NSE values across all stations for each period.

S3 Comparison between SMI and SPEI droughts

When analysing drought recovery, SPEI-1 reacts fastest and shows an earlier termination of droughts. Although the exact pattern of recovery is event-specific, drought recovery follows the same general order: short-term water balance anomalies (SPEI-1) normalize first followed by seasonal water balance anomalies (SPEI-3), before soil moisture conditions emerge out of drought. This pattern is most evident during the drought events of 1975-1977 and 2018-2019 (Figure S2). During the 1975–1977 drought event, all the indices show that the drought-affected area peaked by August 1976. According to the evolution of SPEI-1, the drought had virtually terminated by around November 1976. Yet, by this time almost half of the domain area was still under SPEI-3 drought, while SMI shows closer to 90% of the domain was still under drought. By the time the SPEI-3 drought terminated in January 1977, more than one-third of the domain was still under SMI drought, which took until February 1977 to terminate. A similar sequence of recovery was observed during the 2018-2019 drought. The 2016–2017 drought was interrupted

by intermediate wet conditions during March and April 2017, which led to partial drought recovery and consequently a smaller margin between SPEI-3 and SMI recoveries. However, the wet spell did not split the event because the month-to-month overlap in the drought area still exceeded the 640 km^2 merging threshold; thus, the drought remained as a single multi-temporal event.

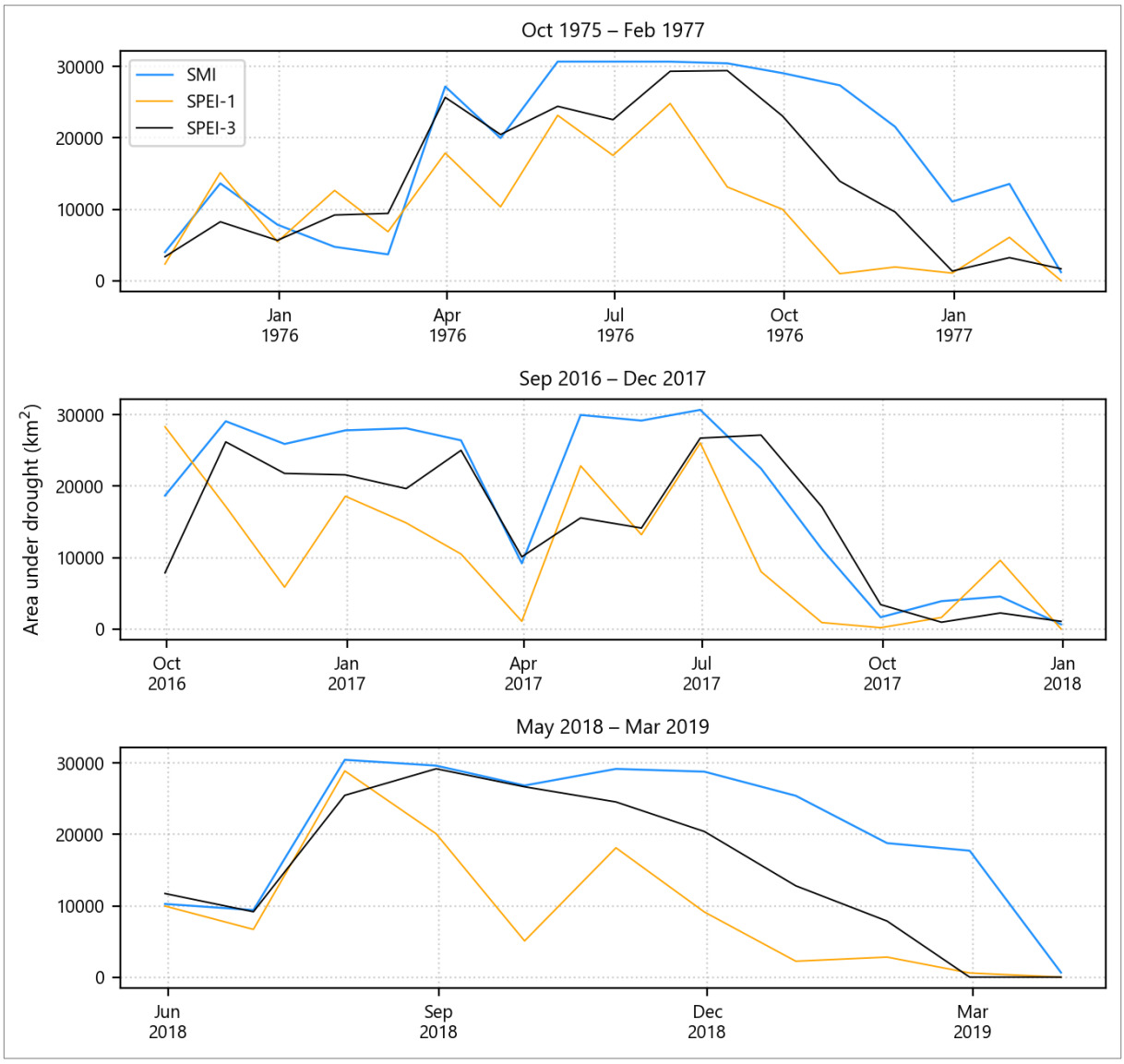

Figure S2: Evolution of the area affected by drought during the three biggest drought events, represented using the three indices.

S3.1 Spatial patterns of drought persistence and cumulative exposure.

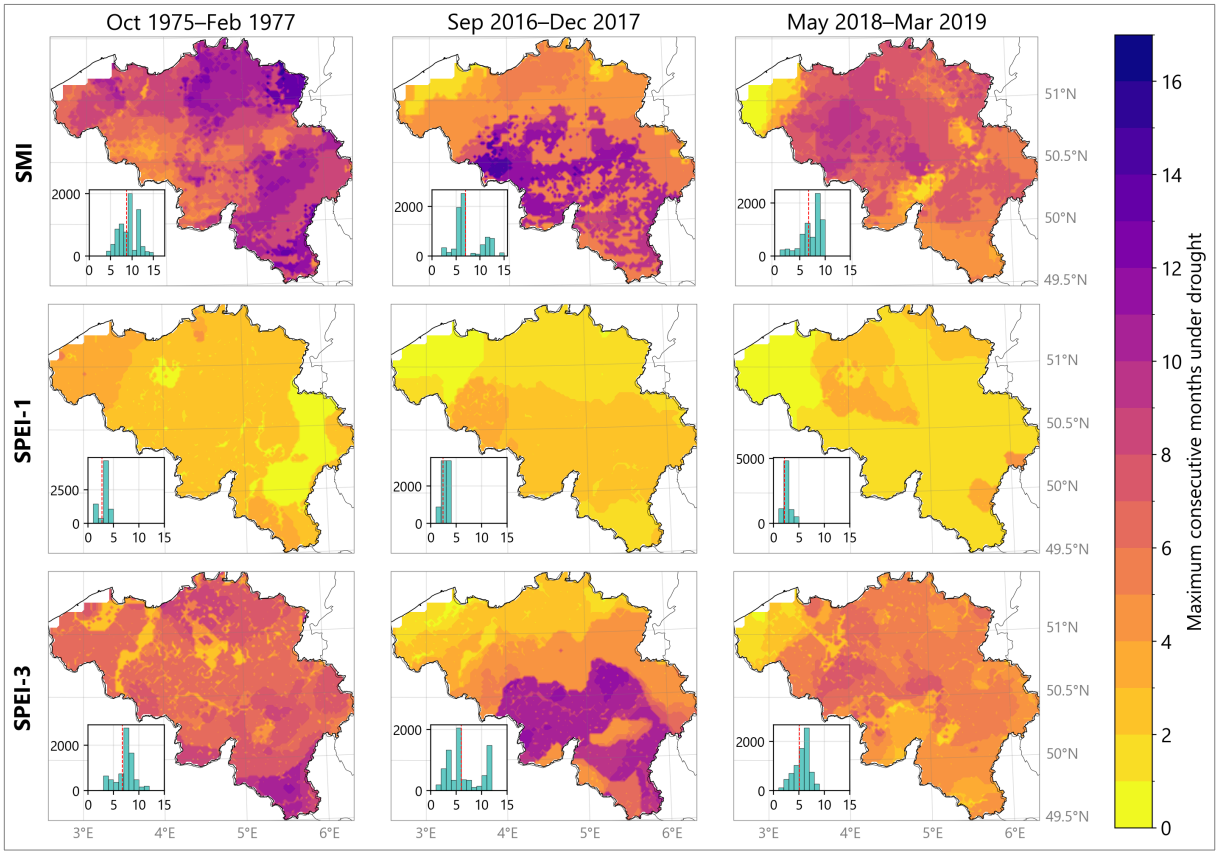

Figure S3: Pixel-wise drought persistence for the three largest droughts during the period of analysis for SMI, SPEI-1 and SPEI-3. Persistence is computed as the longest consecutive sequence of months where at least moderate conditions exist in each gridcell. The inset histogram shows the distribution per-pixel drought persistence. The vertical red line shows the mean persistence.

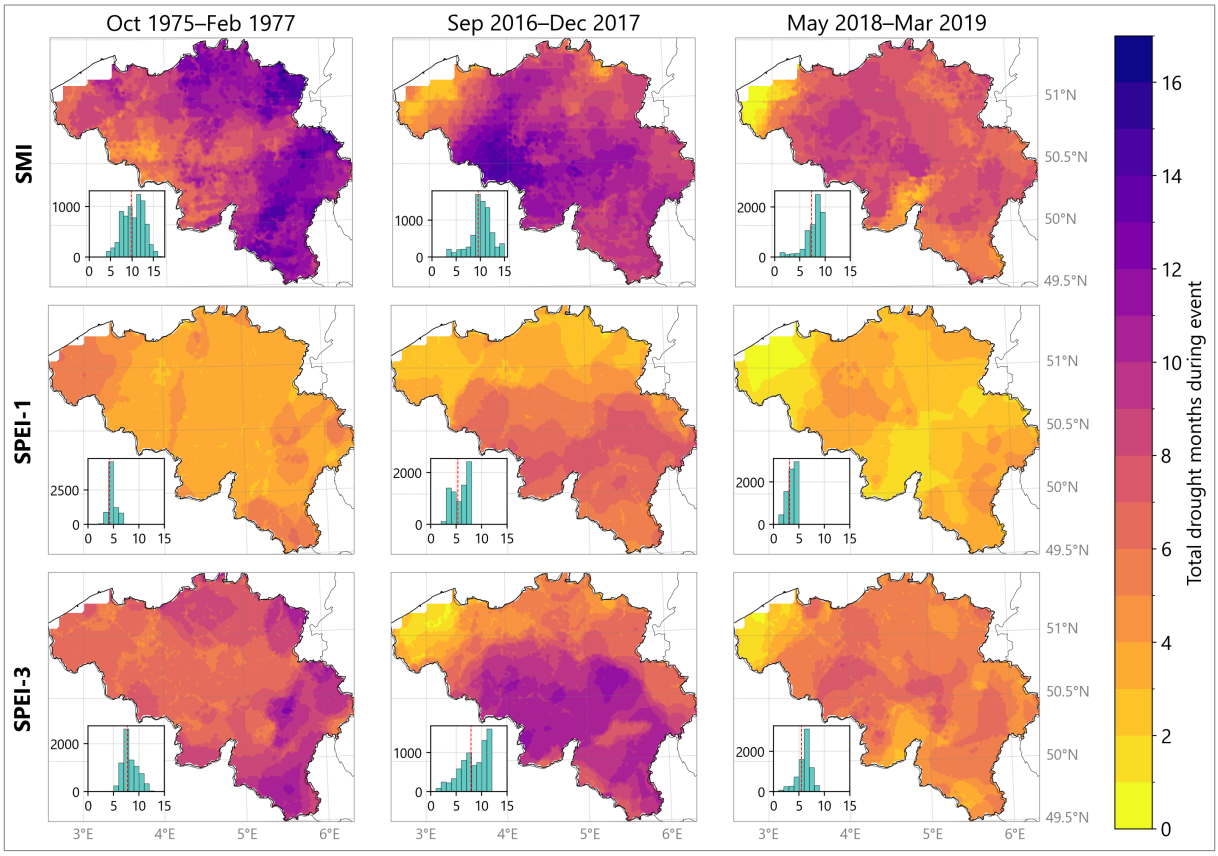

Figure S4: Pixel-wise cumulative drought exposure for the three largest droughts during the period of analysis for SMI, SPEI-1 and SPEI-3. Exposure is computed as the total number of months, not necessarily consecutive, under at least moderate drought drought in each grid cell. The inset histograms are similar to Figure S3 but for cumulative exposure.

References

Klemeš, V.: Operational testing of hydrological simulation models, Hydrological sciences journal, 31, 13–24, 1986.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Transactions of the ASABE, 50, 885–900, 2007.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles, Journal of hydrology, 10, 282–290, 1970.