Predictability of large-scale extreme droughts from global bias-corrected seasonal forecasts

Jan N. Weber, Christof Lorenz, Tanja C. Schober, Hannes Dehn, Rebecca Wiegels, Harald Kunstmann

Angaben zur Veröffentlichung / Publication details:

Weber, Jan N., Christof Lorenz, Tanja C. Schober, Hannes Dehn, Rebecca Wiegels, and Harald Kunstmann. 2026. "Predictability of large-scale extreme droughts from global bias-corrected seasonal forecasts." *Frontiers in Water* 8: 1911791. https://doi.org/10.3389/frwa.2026.1911791.

frontiers in Water

TYPE Original Research
PUBLISHED 10 August 2026
DOI 10.3389/frwa.2026.1911791

OPEN ACCESS

EDITED BY
Alexandra Gemitzi,
Democritus University of Thrace, Greece

REVIEWED BY
Qibo Xu,
University of California, Los Angeles (UCLA), United States
Armand Tchinda Feudjio,
The University of Bamenda, Cameroon

*CORRESPONDENCE
Jan N. Weber
✉ jan.weber@kit.edu

RECEIVED 17 June 2026
REVISED 19 July 2026
ACCEPTED 22 July 2026
PUBLISHED 10 August 2026

CITATION
Weber JN, Lorenz C, Schober TC, Dehn H, Wiegels R and Kunstmann H (2026) Predictability of large-scale extreme droughts from global bias-corrected seasonal forecasts. *Front. Water* 8:1911791. doi: 10.3389/frwa.2026.1911791

Predictability of large-scale extreme droughts from global bias-corrected seasonal forecasts

Jan N. Weber[1,2]*, Christof Lorenz[1], Tanja C. Schober[1], Hannes Dehn[1,2], Rebecca Wiegels[1] and Harald Kunstmann[1,2,3]

[1]Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research Atmospheric Environmental Research (IMKIFU), Garmisch-Partenkirchen, Germany, [2]Institute of Geography, University of Augsburg, Augsburg, Germany, [3]Center for Climate Resilience, University of Augsburg, Augsburg, Germany

Extreme droughts are increasing in severity and frequency due to climate change, highlighting the need for reliable early-warning information at seasonal time scales. Bias-corrected seasonal forecasts offer a promising basis for anticipating drought conditions several months in advance, yet their predictive skill and potential operational value for the most severe droughts at the global scale remains insufficiently quantified. Here, we assess the performance of seasonal drought forecasts derived from the SEAS5-BCSD dataset using the Standardized Precipitation Evapotranspiration Index (SPEI). Thirty-six of the most extreme drought events between 1981 and 2024 are identified from ERA5 reanalysis, selecting two events per continent and accumulation period (SPEI-1, SPEI-3, and SPEI-6), and assessing forecast performance using probabilistic, spatial, and impact-oriented verification metrics, including the Continuous Ranked Probability Skill Score (CRPSS), the Fractions Skill Score (FSS), and the Potential Economic Value (PEV). The results demonstrate that SEAS5-BCSD drought forecasts outperform climatology for nearly all analyzed events. Across all verification metrics, forecast skill is generally highest for SPEI-1 events and for the selected African droughts, whereas the selected North American events exhibit comparatively lower skill. While the exact location and intensity of the most severe drought cores remain difficult to predict, useful probabilistic information is often preserved within the ensemble distribution. Maximum PEV values frequently exceed 0.7, indicating substantial potential value for decision-making under suitable cost-loss assumptions. These findings highlight the benefits of probabilistic, ensemble-based interpretation of seasonal drought forecasts. Rather than relying solely on the ensemble mean, considering the full forecast distribution enables more risk-informed drought preparedness by accounting for both the likelihood and the potential severity of extreme events.

KEYWORDS

bias correction and spatial disaggregation, decision support, drought, SEAS5, seasonal forecast, SPEI

1 Introduction

Devastating large-scale droughts are becoming increasingly frequent and severe as a consequence of climate change (Vicente-Serrano et al., 2022; OECD, 2025). In terms of overall impact, they rank among the most destructive natural hazards worldwide, often exceeding floods and storms regarding the number of people affected, as well as associated agricultural and economic losses (CRED and UNDRR, 2020; FAO and WMO, 2018). Therefore, reliable information on the spatial extent, duration, and severity of drought events is essential for effective preparedness and mitigation strategies (Graham et al., 2011; Hao et al., 2018; Bandyopadhyay et al., 2020). A comprehensive and up-to-date global drought database is currently lacking. Existing efforts, such as the European Commission's drought database based on the Standardized Precipitation Index (SPI) (European Commission, J. R. C.)), provide valuable information on severe drought events across multiple accumulation periods. However, these datasets typically end around 2016 and therefore do not capture the marked increase in drought frequency and severity observed over the past decade (Vicente-Serrano et al., 2022). This limitation restricts the ability to assess recent extreme events and to place them into a consistent global context.

In this context, seasonal forecasts, issued by (inter-)national meteorological services, enable the estimation of drought occurrence probabilities up to 12 months in advance, thereby supporting early drought risk assessment. A prerequisite for the meaningful application of such forecasts, however, is a robust evaluation of their predictive skill (Doblas-Reyes et al., 2013). Previous studies have investigated the performance of seasonal drought forecasting systems primarily at regional scales, for example in the Mediterranean (Arnone et al., 2020), Latin America (Carrão et al., 2018), and Spain (Madrigal et al., 2018). These analyses generally reported modest but positive skill relative to climatology, largely confined to the first three lead months with the first lead month being clearly superior to the latter ones (Carrão et al., 2018; Vorobevskii et al., 2024; Weber et al., 2023). In contrast, comprehensive global-scale assessments remain comparatively scarce. Existing global investigations often focus on climatological or ensemble-mean behavior (e.g., Hao et al., 2014; Beck et al., 2022) or specific aspects of drought predictability such as onset timing (Yuan and Wood, 2013; Chevuturi et al., 2021), rather than systematically evaluating extreme events across regions and lead times. Notably, Weber et al. (2026b) found that SEAS5-BCSD forecasts exhibit higher skill for moderate drought conditions than for near-normal states, suggesting that forecast skill may vary substantially with drought severity and motivating further analysis of more extreme events. The choice of drought indicator also influences reported forecast skill. Studies relying primarily on precipitation-based metrics such as the Standardized Precipitation Index (SPI) have often found limited significant skill (Arnone et al., 2020; Carrão et al., 2018). Incorporating evapotranspiration, either through hydrological modeling or through the Standardized Precipitation Evapotranspiration Index (SPEI), has yielded more encouraging results (Vorobevskii et al., 2024; Brands et al., 2025). For Europe, Shyrokaya et al. (2025) reported slightly improved performance of SPEI compared to SPI, for semi-arid regions on multiple continents, Portele et al. (2021) showed the same. On a global scale, Vicente-Serrano et al. (2023) demonstrated that SPEI provides a robust drought hazard indicator across diverse climatic regions for monitoring, with reduced skill in some areas primarily attributable to uncertainties in precipitation estimates within reanalysis datasets. Overall, these findings indicate that multivariate drought indicators such as SPEI may offer advantages for global seasonal drought predictability assessments. Despite these advances, three important knowledge gaps remain. First, most global studies focus on climatological skill or ensemble means rather than individual extreme drought events. Second, few studies systematically compare forecast performance across continents and lead times using a consistent methodology. Third, the operational value of probabilistic drought forecasts has rarely been quantified. This study addresses these gaps by evaluating the predictability of 36 of the most severe historical drought events worldwide using probabilistic verification metrics, spatial verification, and an assessment of Potential Economic Value (PEV).

To address this gap, we adopt a forecast-only evaluation strategy. Unlike seamless drought outlook approaches that combine monitoring and forecasting information (e.g., Dutra et al., 2014), this framework enables a direct attribution of skill to the seasonal prediction system itself and thus provides a conservative estimate of drought predictability. Because raw seasonal forecast output commonly exhibits systematic biases and model drift relative to reference datasets such as ERA5 (Yin et al., 2023; Weber et al., 2026b), we employ a bias-corrected seasonal forecast dataset (SEAS5-BCSD). This dataset is derived from the ECMWF SEAS5 system using the Bias Correction and Spatial Disaggregation (BCSD) method (Weber et al., 2026a), ensuring improved consistency with reference climatology. As a drought indicator, we focus on the Standardized Precipitation Evapotranspiration Index (SPEI). To obtain a comprehensive assessment of forecast skill under extreme conditions, we analyze two of the strongest drought events per continent (excluding Antarctica) and aggregation time scale (1, 3, and 6 months) of the last 44 years, resulting in a total of 36 drought cases. These events form the basis for a systematic evaluation of the reliability and limitations of global bias-corrected seasonal extreme drought forecasts, addressing how well the most extreme droughts in recent history can be predicted using such forecasts and whether systematic strengths and weaknesses can be identified.

2 Methods

2.1 Data

Drought detection and forecast verification are based on the ERA5 reanalysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 offers global coverage, temporal continuity from 1940 to present, and consistent representation across variables (Hersbach et al., 2020), making it a suitable reference for this study. The dataset also serves as the climatological ensemble, with a leave-one-out approach (Arlot and Celisse, 2010) applied for 1981–2016. From 2017 onwards, the full ensemble of reference years (1981–2016) is used as climatological ensemble for drought occurrence.

As forecast input, the bias-corrected, monthly aggregated SEAS5-BCSD dataset (Weber et al., 2026a) is employed. In this global dataset, the ECMWF SEAS5 forecasts (Johnson et al., 2019) were corrected toward ERA5 using the Bias Correction and Spatial Disaggregation (BCSD) method (Wood et al., 2002; Lorenz et al., 2021; Weber et al., 2026b), with the cumulative distribution functions based on the 1981–2016 reference period, corresponding to the SEAS5 hindcast period. The dataset has a spatial resolution of $0.25^\circ \times 0.25^\circ$ and has demonstrated improved skill compared to raw SEAS5 forecasts, particularly in tropical regions and at longer lead times (Weber et al., 2026b). SEAS5-BCSD uses linear extrapolation (scaling) for precipitation and an additive delta approach for temperature records, widely used approaches to capture the extent of extremes (Boé et al., 2007; Gonzalez-Aparicio and Hidalgo, 2011; Holthuijzen et al., 2022). The availability of data from January 1981 to the present enables the assessment of both recent and historical drought events. The SEAS5 variables used for calculating the drought hazard indicator are total precipitation and mean temperature at 2 m.

While Portele et al. (2021) demonstrated that seasonal drought skill can be obtained without explicit bias correction due to the normalization inherent in SPEI, we employ a bias-corrected dataset to ensure a consistent representation of the forecast distribution, particularly in its tails. Raw seasonal forecasts may misrepresent extreme precipitation and dry-day frequencies, which can affect the standardized drought index calculation. Particular attention is given to the treatment of unprecedented extreme values that lie outside the calibration range of the empirical CDF. Within the BCSD framework, multiplicative delta scaling is applied to precipitation and additive delta scaling to temperature, ensuring physically consistent extrapolation beyond the reference distribution. This approach is more robust than a simple *post-hoc* standardization of the derived water balance, as the bias correction is applied directly to both variables prior to drought index calculation. The handling of new extremes is increasingly important in the context of climate change, given the rising frequency of record-breaking temperature events.

Although our evaluation is conducted at monthly resolution, forecasts are generated at daily time steps, and unrealistic daily extremes can propagate into monthly aggregates. Applying BCSD at daily scale improves the representation of precipitation intensity and dry-day occurrence, particularly in complex terrain where raw model output may deviate substantially from reanalysis data (Lorenz et al., 2021). This provides a more physically consistent basis for drought classification.

2.2 Data processing

Drought conditions in forecast and reference data are quantified using the Standardized Precipitation Evapotranspiration Index (SPEI), computed for both reanalysis and forecasts for all months since 1981. Unlike the Standardized Precipitation Index (SPI; McKee et al., 1993), the SPEI incorporates an estimate for evaporation, providing a more complete representation of the water balance. The SPEI was derived following the method of Vicente-Serrano et al. (2010), using precipitation and mean temperature as input variables. To capture droughts at different accumulation timescales, SPEI-1, SPEI-3, and SPEI-6 were calculated, corresponding to 1-, 3-, and 6-month windows. This roughly corresponds to the temporal extent of a meteorological, an agricultural and a hydrological drought (Wilhite and Glantz, 1985). Areas with extremely low climatological precipitation, defined as regions receiving less than 0.01 mm.day^{-1} at the 33rd percentile of monthly precipitation, were masked out. In such hyper-arid regions, SPEI values are undefined or numerically unstable due to the near absence of precipitation (Vicente-Serrano et al., 2010). Consequently, skill metrics are evaluated only over precipitation-relevant regions, avoiding artificial skill artifacts in permanently dry areas. For continental-scale means, the fitting of the curve was applied after computing the aggregated water balance. The normalization reference period was set to 1981–2016, with 1980 included to initialize SPEI-3 and SPEI-6, ensuring consistency with the reference period used in the BCSD procedure. While some studies (e.g., Dutra et al., 2014; Sutanto et al., 2020) incorporate monitoring-period information into the forecasting window to construct seamless SPI-3/-6/-12 predictions, we deliberately restrict the evaluation to forecast-only lead times. Specifically, we consider Lead Month 0 (LM0) for SPEI-1, LM0-2 for SPEI-3, and LM0-5 for SPEI-6, ensuring that the assessed skill is exclusively derived from the seasonal forecast and not influenced by observed monitoring data.

2.3 Identification of the most severe droughts

The objective is to identify the two most severe drought events per continent, separately for SPEI-1, SPEI-3, and SPEI-6. The selection of two events per continent and aggregation scale represents a compromise between focusing on the most extreme droughts and ensuring a minimal sample size for a meaningful comparison. Restricting the analysis to a single event would risk overemphasizing case-specific characteristics, while including a larger number of events would dilute the focus on truly exceptional drought conditions. Although using a fixed number of events per continent means that some globally more severe droughts are not included, this strategy ensures a balanced representation of all continents and avoids biases arising from differences in continent size or drought-affected area. Figure 1 describes the process visually. Ranking droughts objectively is inherently challenging, as severity can be defined in terms of spatial extent, duration, intensity, or socio-economic vulnerability. In this study, the focus is limited to atmospheric factors alone, excluding regional vulnerability. Drought duration is represented by different accumulation periods, intensity by varying SPEI thresholds (Malekian et al., 2026), and spatial extent by the use of spatial skill scores (Skok and Roberts, 2016).

The analysis covers the six inhabited continents: Africa, Asia, Australia, Europe, North America, and South America; Antarctica is not considered (see Figure 2). Greenland, located on the North American tectonic plate (Bird, 2003), is included in North America. Asia and Europe are separated along the Ural mountain range, while Oceania is considered together with Australia. We note that, under this delineation, European drought events are more likely to occur within the Russian landmass, as large-scale drought patterns cover a smaller contiguous land area in Western and Central Europe due to the region's fragmented coastline and substantial maritime influence. For each continent and aggregation

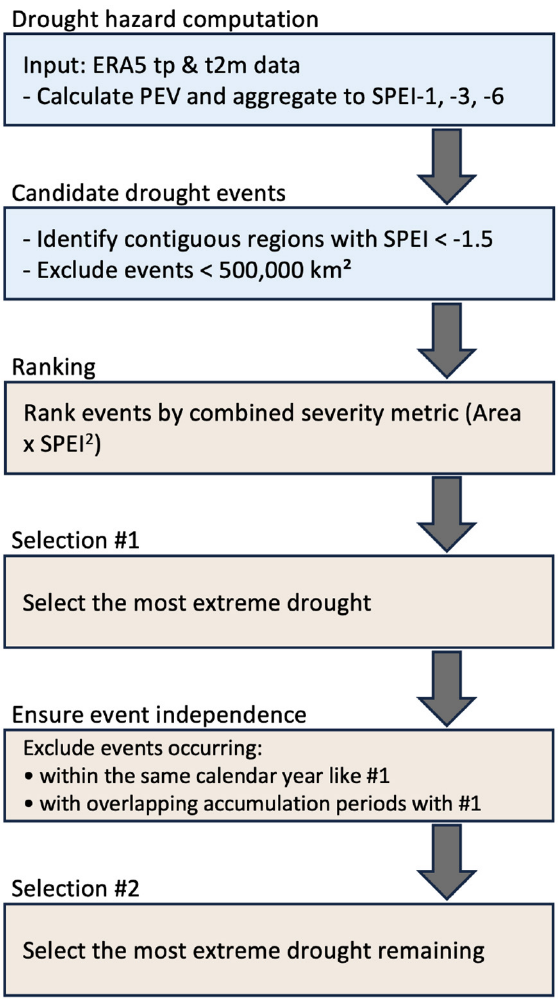

FIGURE 1
Flowchart illustrating the selection process for identifying two drought events per continent and aggregation period, resulting in a total of 36 droughts. Blue boxes indicate steps executed once, while sand-colored boxes represent steps repeated for each continent and aggregation period.

scale, the 5 months with the largest area affected by SPEI $<$ -1.5 during 1981–2024, representing severe drought conditions (Vicente-Serrano et al., 2010) are selected based on ERA5 data. More restrictive thresholds such as -2 primarily capture localized extreme cores and may therefore underestimate the full spatial extent of major drought events. In addition, 5 months with the lowest SPEI values are identified to capture drought intensity. For this purpose, grid cells with SPEI above -1.5 are masked. Although a minimum affected-area criterion was initially considered, it proved unnecessary because all high-ranking candidate events already exceeded the intended threshold. The final event selection was therefore unaffected by this criterion. The event selection deliberately emphasizes large-scale droughts because seasonal prediction systems are expected to exhibit skill primarily for broad-scale hydroclimatic anomalies rather than localized events. This procedure yields up to ten candidate months per continent and aggregation scale, covering both the most intense and the most spatially extensive droughts. Among the remaining candidates, the strongest drought is identified using an objective drought ranking metric that combines drought duration, affected area, and intensity (Equation 1):

$$R = D\sqrt{A, I^2}, \tag{1}$$

where D denotes the accumulation period (1, 3, or 6 months for SPEI-1, SPEI-3, and SPEI-6, respectively), A is the affected area with SPEI < -1.5, and I is the mean SPEI within the affected area. Larger values of R therefore correspond to droughts that are affecting larger areas, and exhibit greater intensity.

To ensure temporal diversity, events occurring within the same calendar year and an overlapping accumulation period are subsequently removed before identifying the second drought using the same procedure. The final set of events is summarized in Table A1 in Appendix. For clarity, the SPEI < -1.5 threshold is used solely to identify the month associated with the most severe drought conditions; all subsequent analyses for the respective month are conducted using the full range of SPEI values spatially, not restricted to areas below the threshold. The selected events correspond to the two highest-ranked droughts per continent according to the objective Drought Event Index, which combines affected area, intensity and duration.

Notably, 32 of the 36 identified droughts occurred after the reference period ending in 2016. This is consistent with Vicente-Serrano et al. (2022), who reported an increase in agricultural droughts driven by higher atmospheric evaporative demand especially in the last years, and Nwayor and Robeson (2024), who showed decreasing SPEI values in a warming world. Furthermore, this provides a largely independent evaluation period, as the vast majority of these drought months were not included in the CDF calculations used for the BCSD.

With 36 drought events identified over a 44-year period, not all large-scale droughts are represented in this analysis. The relatively limited sample size is an inherent consequence of focusing exclusively on the most extreme, large-scale events, rather than including moderate or short-lived droughts as in, for example, Richardson et al. (2020). The events were selected objectively according to predefined ranking criteria rather than randomly sampled, ensuring that the analysis focuses on the most societally relevant large-scale droughts. While this selection reduces the number of cases and limits (temporal) diversity, it allows a targeted assessment of forecast performance under the most challenging and societally relevant conditions. At the same time, the sample size exceeds that of purely event-based case studies (e.g., Suárez-Almiñana et al., 2022), providing a more robust basis for identifying systematic patterns in seasonal drought predictability.

2.4 Evaluation metrics

To evaluate continental-scale drought forecast performance, three complementary indicators are employed. The Continuous Ranked Probability Skill Score (CRPSS) assesses the agreement between the forecast ensemble and the reanalysis by comparing their cumulative distribution functions (Equations 2, 3). It is derived from the Continuous Ranked Probability Score (CRPS,

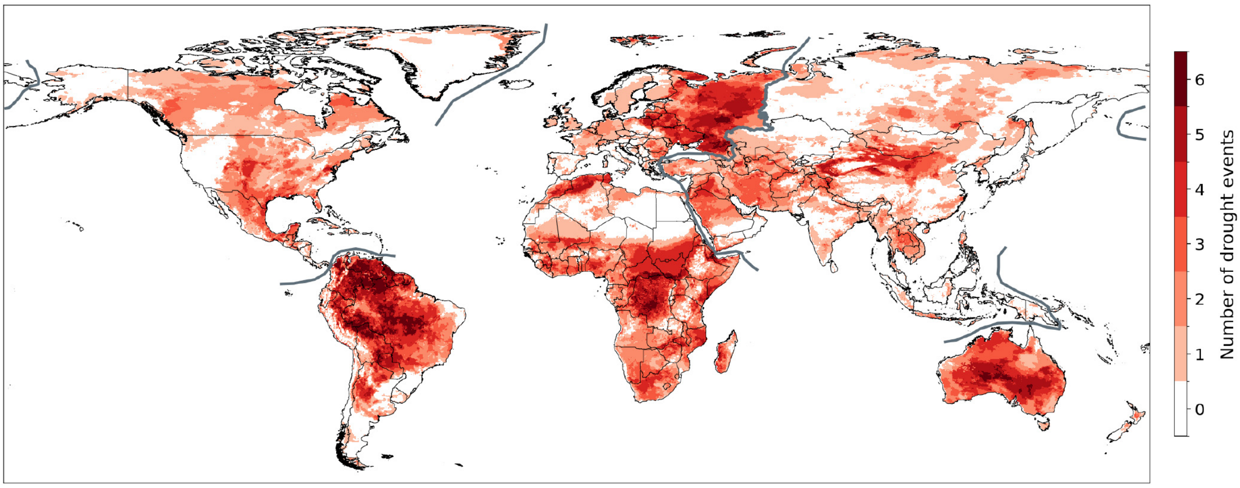

FIGURE 2
Map of the 36 selected drought events based on ERA5 reanalysis data. Areas with SPEI < −1.5 are shaded in red; darker shading indicates regions where multiple drought extents overlap. The borders between continents are marked with thick gray lines.

Hersbach, 2000), which measures the area between the forecast ensemble distribution and the Heaviside step function representing the reanalysis.

$$\mathrm{CRPS}(F,x) = \int_{-\infty}^{\infty} \left(F(y) - H(y-x)\right)^2 dy, \tag{2}$$

The CRPSS is then calculated as

$$\mathrm{CRPSS} = 1 - \frac{\sum \mathrm{CRPS}_{\mathrm{fcst}}}{\sum \mathrm{CRPS}_{\mathrm{ref}}}. \tag{3}$$

with ref being, e.g., the climatology. The metric therefore considers the full spectrum of values, from dry to wet conditions, as well as the distance of each ensemble member from the reanalysis, on a grid cell basis. For spatial and/or temporal aggregates, the mean SPEI is calculated first and the CRPSS is then computed for this aggregated series. A value of 1 indicates a perfect forecast, 0 performance equivalent to climatology, and negative values performance worse than climatology. Due to ensemble spread, a perfect score of 1 is generally not attainable in probabilistic forecasting systems (Hersbach, 2000).

While CRPSS captures overall distributional performance, complementary metrics focus on categorical drought events. For these, thresholds such as SPEI < -1 or SPEI < -1.5 are applied, classifying both forecasts and reanalysis into binary outcomes (event or non-event) at each grid cell. Based on these classifications, several event-based metrics are derived and explained in the following.

The Hit Rate (HR) measures the proportion of correctly predicted drought events among all observed drought events (Owens and Hewsom, 2018). For climatology, assuming a normal distribution, a Hit Rate of approximately 15.9% is expected for SPEI < -1, and about 2.3% for SPEI < -2, corresponding to the lower-tail probabilities of a standard normal distribution. An increase beyond these baseline values for SEAS5-BCSD indicates improved event detection. The False Positive Rate (FPR) is defined as the fraction of forecasted drought events that do not occur against the total number of actual negative events (Equation 4):

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}, \tag{4}$$

ideally, HR exceeds FPR.

To relate the Hit Rate (HR) and False Positive Rate (FPR), Receiver Operating Characteristic (ROC) curves are calculated. The ROC curve expresses the trade-off between HR and FPR as a function of the probability threshold, defined here by the fraction of ensemble members forecasting the binary drought event (threshold exceeded or not exceeded; Marzban, 2004). A forecast with no skill relative to random chance yields identical HR and FPR values for all thresholds and therefore lies along the diagonal of the ROC space, corresponding to an Area Under the Curve (AUC) of 0.5 (Tóth et al., 2003). Increasing forecast skill shifts the ROC curve toward the upper-left corner of the diagram, resulting in larger AUC values, with a theoretical maximum of 1 for a perfect forecast.

To account for both misses and false alarms, the Brier Skill Score (BSS, Brier, 1950) is employed (Equation 5). The BSS is based on the Brier Score (BS), defined as

$$\mathrm{BS} = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2, \tag{5}$$

where N is the number of observations, f_t is the probabilistic forecast at time t, and o_t is the corresponding binary outcome derived from ERA5. The BSS is then calculated analogously to the CRPSS, with 1 representing perfect forecasts (theoretical only for ensembles Weigel et al., 2007), 0 as good as the reference, and negative values worse than the reference. While the CRPSS can be viewed as a continuous analog of the BSS (Hersbach, 2000), the BSS provides sharper insights into the ability of forecasts to capture specific extreme events, which are the main focus of this study.

Precipitation forecasts are prone to displacement errors: while the total precipitation over a large region may be well represented, grid-cell-level comparisons like the preceding indicators can yield poor scores due to spatial shifts of only a few kilometers, resulting in so-called double-penalty effects (Mittermaier et al., 2011). To account for such spatial pattern errors, the Fractions Skill Score (FSS) is employed (Equation 6). This neighborhood verification method, widely used for high-resolution short-term precipitation forecasts (Necker et al., 2024), is also applicable to longer-term predictions. Conceptually, the FSS is a variation of the Brier Skill Score that compares fractional coverage of events within neighborhoods of increasing size. Following Mittermaier et al. (2011), it is defined as

$$\mathrm{FSS} = 1 - \frac{\mathrm{FBS}}{\mathrm{FBS}_{\mathrm{worst}}}, \tag{6}$$

where FBS denotes the Fractions Brier Score (Equation 7) and $\mathrm{FBS}_{\mathrm{worst}}$ its maximum possible value (all precipitation misplaced). The FBS is composed as follows:

$$\mathrm{FBS} = \frac{1}{N}\sum_{i=1}^{N}\left(f_i - o_i\right)^2, \tag{7}$$

where N denotes the total number of grid cells, f_i is the fraction of neighboring grid cells around location i exceeding the threshold in the forecast, and o_i is the corresponding fraction in the observations.

An FSS of 1 corresponds to a perfect forecast (unattainable for ensemble systems), while values above 0.5 are typically considered indicative of useful skill, provided that the domain is large compared to the typical spatial error (Skok and Roberts, 2016), which is the case for this analysis due to the continent-wide domain.

Beyond purely statistical verification metrics, we assessed the practical usefulness of the forecasts using the Potential Economic Value (PEV) (Equation 8), which quantifies the economic benefit of a probabilistic forecast within the classical cost-loss decision framework (Murphy, 1977; Richardson, 2000). PEV compares the expected expense of decisions based on the forecast against decisions based solely on climatology. For each drought threshold (SPEI $< -1, -1.5$, and -2), binary drought occurrence was derived from the observations, ensemble forecasts, and climatological reference. The PEV was then evaluated for forecast probability thresholds (p_{th}) between 0.05 and 0.95 and cost-loss ratios (c/l) (Equation 9) between 0.01 and 0.99. Following Richardson (2000), the PEV is defined as

If $\frac{c}{l} < p$, the Potential Economic Value is given by

$$\mathrm{PEV} = \frac{\frac{C}{L} - p - \frac{C}{L}(1-p)F + p\left(1 - \frac{C}{L}\right)H}{\frac{C}{L}(1-p)}, \tag{8}$$

whereas for $\frac{C}{L} \geq p$

$$\mathrm{PEV} = \frac{-\frac{C}{L}(1-p)F + p\left(1 - \frac{C}{L}\right)H}{p\left(1 - \frac{C}{L}\right)}. \tag{9}$$

Here, H denotes the Hit Rate, F the False Positive Rate, p the climatological probability of the event, and C/L the assumed cost-loss ratio. PEV values larger than zero indicate that using the forecast is economically preferable to relying on climatology alone, whereas values below zero imply that climatological information yields a lower expected cost. A value of one represents a perfect forecast.

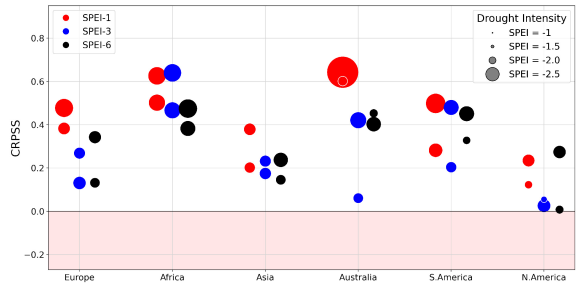

FIGURE 3
Mean Continuous Ranked Probability Skill Score of the selected droughts across continents and lead times. The size of the circles indicate the drought severity.

3 Results

3.1 Overall drought forecast performance

We first assess the overall forecast performance for the selected extreme drought events across continents and accumulation periods. The CRPSS overview (Figure 3) indicates positive skill for all selected drought events. Among the indices, SPEI-1 attains slightly higher CRPSS values than SPEI-3 and SPEI-6, consistent with the expected accumulation of forecast errors (Lorenz et al., 2021; Weber et al., 2026b) at longer accumulation periods. The difference between SPEI-3 and SPEI-6 is negligible. Across continents, the African droughts exhibited the highest median CRPSS (Table A2 in Appendix), with all six droughts reaching CRPSS values above 0.38. Asia and North America exhibit lower values, generally below those of Africa, while Australia and South America maintain consistently high skill except for isolated cases. Europe performs intermediately with values between 0.13 and 0.43. Overall, the selected tropical drought events tend to exhibit higher forecast skill than those at higher latitudes, although this pattern should be interpreted with caution given the limited number of events. Regarding drought intensity, Figure 3 reveals a clear systematic trend; a tendency for the most severe droughts to exhibit higher CRPSS values compared to milder events. This pattern is supported by a correlation analysis, which indicates a statistically significant ($p \approx 3.03 \times 10^{-6}$) negative relationship between CRPSS and SPEI values. Given that SPEI is negative during drought conditions, this implies that more extreme droughts are forecasted more accurately than less severe droughts (still SPEI < -1.5) within the relatively small sample of thirty-six events.

Because CRPSS evaluates the entire forecast distribution, forecast errors associated with normal and wet conditions also contribute to the score. To isolate forecast performance for drought events, we therefore additionally employ the Brier Skill Score (BSS), which is calculated specifically for drought occurrence. To distinguish the performance of different drought categories,

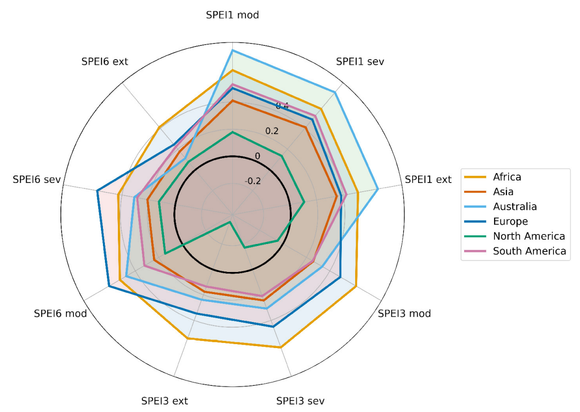

FIGURE 4
Radar plot of Brier Skill Score (BSS) across drought severity and accumulation periods for each continent. Each axis corresponds to a specific combination of severity and accumulation period. The black circle represents a BSS of 0 (no skill). Values farther from the center indicate higher skill. Lines closer to the outer edge of the plot indicate better performance of the forecast relative to climatology.

thresholds are defined at SPEI < -1 (moderate droughts), < -1.5 (severe droughts), and < -2 (extreme droughts) on grid-cell level, following Vicente-Serrano et al. (2010). This provides a more complete picture of forecast performance, especially for the most extreme areas.

The radar plot (Figure 4) summarizes the BSS across drought categories, lead times, and continents. Ideally, lines should extend far from the center, with the black ring marking climatology. For SPEI-1, all continents achieve positive skill, North America is the closest to zero. Moderate drought areas are detected most reliably, while severe and extreme areas show weaker skill, consistent with the correlation analysis shown in Figure 3. This effect is minor at the short lead time of 1 month but becomes most pronounced at 6 months. At a lead time of 3 months (SPEI-3), skill decreases further; both North American forecasts fall below climatology, resulting in negative BSS. SPEI-6 produces a similar picture, with progressively lower skill as drought severity increases. Overall, two robust patterns emerge: (i) forecast skill declines with increasing drought severity category, and (ii) longer lead times reduce skill. Among continents, African droughts are consistently best predicted (always above 0.32 BSS), while North American droughts are the most difficult.

In addition to climatology, the drought detection performance of SEAS5-BCSD was evaluated against two additional baseline forecast approaches: a persistence forecast based on the observed SPEI value of the preceding month, and a linear regression model using the Niño3.4 index (NOAA, 2026). The median Brier Skill Score (BSS) values are predominantly positive across the evaluated drought thresholds, indicating that SEAS5-BCSD generally outperforms both benchmark approaches (Table 1). The highest skill is found for mild drought conditions, whereas for more severe drought thresholds the linear regression occasionally exhibits slightly higher skill. Forecast performance is highest for short-duration droughts (SPEI-1), while forecasts for long-duration droughts (SPEI-6) generally outperform those for intermediate durations (SPEI-3). Overall, both benchmark approaches provide meaningful reference forecasts, while SEAS5-BCSD generally demonstrates added value beyond these simple baseline methods.

Since the SPEI is standardized, the climatological frequency of drought categories (e.g., SPEI < -2) is fixed, preventing an artificial inflation of extreme drought forecast frequencies. To assess whether the extreme cores of the 36 extreme droughts were at the correct spatial position, the Hit Rate (HR) and False Positive Rate (FPR) were calculated. For the short-term droughts (SPEI-1), HR is substantially larger than FPR in the mean, 0.630 to 0.279 for moderate, 0.565 to 0.226 for strong and 0.497 to 0.182 for extreme drought areas. For SPEI-3 and SPEI-6, both the Hit Rate and False Positive Rate decrease compared to SPEI-1, but the separation between the two remains similar. Mean Hit Rates range from 0.48 to 0.32 compared with False Positive Rates between 0.30 and 0.15, indicating that the forecasts continue to detect substantially more drought pixels than they falsely predict. Under a climatological forecast, hit rates and false positive rates are equal and correspond to the climatological frequency of drought events ($\approx$ 16%, 7%, and 2% for SPEI $<$ -1, -1.5, and -2, respectively). The substantially higher hit rates obtained here, combined with consistently larger HR than FPR, therefore indicate clear discriminatory skill relative to climatology. The relatively high number of false positives compared to climatology is primarily due to spatial displacement of forecast drought patterns. The more severe the drought category, the higher the HR/FPR ratio, indicating fewer false positives per correctly detected event. Of the 36 analyzed droughts, only one event (Aus 09/2013) exhibits a higher False Positive Rate than Hit Rate. This isolated case contrasts with the overall behavior of the dataset and is therefore interpreted as an event-specific exception rather than evidence of a systematic forecast characteristic.

To provide a clearer overview of the HRFPR relationship, Figure 5 shows the ROC curves for extreme drought events (SPEI < -2). All analyzed droughts exhibit AUC values greater than 0.5, indicating positive skill relative to a random forecast. In panel (a), short-duration drought forecasts (SPEI-1) show particularly high predictability, with all AUC values exceeding 0.8. For longer accumulation periods (SPEI-3 and SPEI-6; panels (b) and (c)), forecast skill remains positive for all events, although AUC values are slightly reduced compared to SPEI-1. Reduced skill is evident for Australia 09/2013 and North America 02/2017. No consistent continent-wide pattern emerges across the analyzed events.

However, a known limitation of the Brier Skill Score (BSS) and hit/false positive rates is their double penalization of spatially displaced precipitation patterns (Mittermaier et al., 2011). To account for such displacement errors, the Fractions Skill Score (FSS) is calculated, which evaluates spatial agreement within increasing neighborhood sizes. The window size is defined in grid cells; thus, a window size of one allows for a tolerance of one grid cell ($\approx$ 25 km) in any direction, two corresponds to $\approx$ 50 km, and so forth. The FSS results indicate that most strong drought forecasts (SPEI < -1.5) exceed the 0.5 benchmark commonly interpreted as useful skill (Figure 6, Skok and Roberts, 2016). SPEI-1 (left panel) achieves the highest overall performance, with a mean FSS of 0.742 across window sizes and a mean maximum of 0.804 across events. Only one case (North America 03/2012) remains below 0.5 for window sizes $\leq$ 75 km, suggesting that lower BSS values in these cases primarily stem from small-scale

TABLE 1 Median Brier Skill Score (BSS) of SEAS5-BCSD relative to three benchmark forecasts across all lead months for mild, strong and extreme drought thresholds. Values in parentheses indicate the minimum and maximum BSS across the evaluated lead months. Positive BSS values indicate that SEAS5-BCSD outperforms the respective benchmark forecast.

Threshold	Persistence	Linear Niño3.4 regression	Climatology
SPEI-1 < −1	0.55 (0.39–0.79)	0.63 (0.13–0.87)	0.51 (0.04–0.82)
SPEI-3 < −1	0.50 (0.18–0.73)	0.47 (−0.01–0.81)	0.30 (−0.08–0.73)
SPEI-6 < −1	0.55 (0.07–0.76)	0.59 (0.12–0.67)	0.44 (−0.05–0.67)
SPEI-1 < −1.5	0.49 (0.37–0.75)	0.56 (−0.04–0.78)	0.49 (−0.09–0.75)
SPEI-3 < −1.5	0.43 (0.22–0.71)	0.35 (−0.34–0.74)	0.25 (−0.32–0.70)
SPEI-6 < −1.5	0.49 (0.27–0.72)	0.42 (0.03–0.65)	0.33 (−0.06–0.62)
SPEI-1 < −2	0.41 (0.04–0.68)	0.43 (−0.16–0.67)	0.41 (−0.16–0.66)
SPEI-3 < −2	0.39 (0.05–0.71)	0.22 (−0.78–0.62)	0.17 (−0.74–0.61)
SPEI-6 < −2	0.37 (−0.02–0.70)	0.27 (−0.06–0.52)	0.24 (−0.09–0.50)

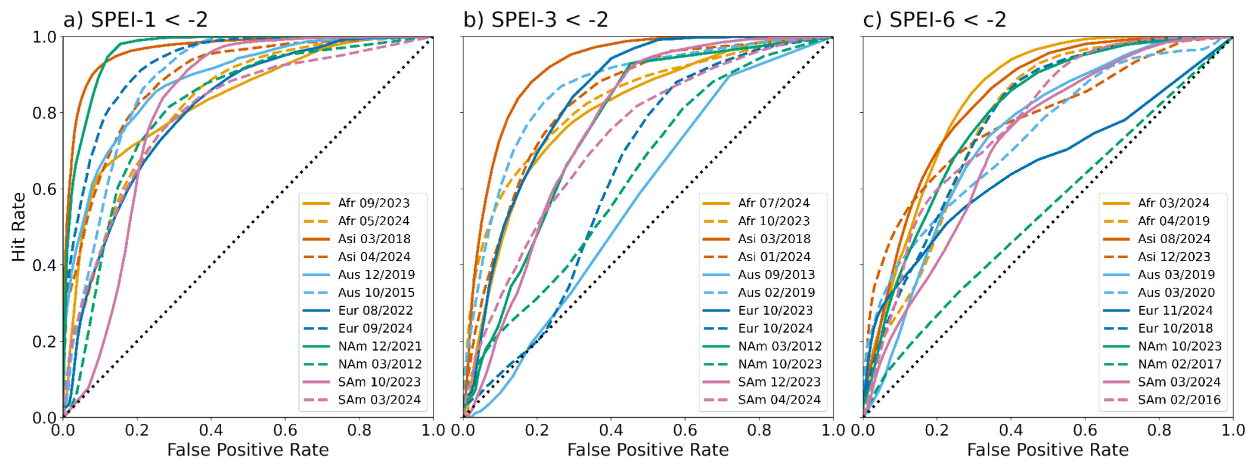

FIGURE 5
ROC curves for all analyzed drought events by continent, evaluated for extreme drought conditions (SPEI < −2). **(a–c)** Show results for SPEI-1, SPEI-3, and SPEI-6, respectively.

spatial displacement rather than a complete failure to capture the event. The continental differences identified in the CRPSS and BSS are also reflected in the FSS, with tropical regions generally exhibiting higher spatial coherence. SPEI-3 (middle panel) displays the largest spread in performance, ranging from the best overall forecast (Africa 07/2024) to the weakest (Australia 09/2013), yet 10 of 12 cases exceed the 0.5 skill threshold. SPEI-6 (right panel), by contrast, performs consistently well for SPEI -1.5, with all but one event (North America 02/2017) exceeding a FSS value of 0.5. Furthermore, nine drought events surpass the benchmark already at window sizes ≤ 75 km, indicating high spatial agreement on small-scale deviations. The increase of FSS with larger neighborhood sizes suggests that the spatial pattern of droughts is captured reasonably well, even if the precise location deviates by several grid cells. This indicates a scale-dependent predictability that improves when forecasts are interpreted over regional rather than local domains.

For the selected events, areas of extreme drought generally exhibited lower spatial skill than severe drought areas. For SPEI-1, the median maximum FSS was 0.745 (95% bootstrap CI: 0.660–0.888), compared to 0.813 (0.738–0.905) for severe drought areas (Figure A1 in Appendix). At longer lead times, the median maximum FSS decreased further, reaching 0.621 (0.534–0.793) for SPEI-3 and 0.685 (0.545–0.793) for SPEI-6. Although the confidence intervals overlap substantially, these results suggest that the detection of the most intense drought cores is more challenging than that of larger severe drought areas. The comparatively lower FSS for extreme droughts likely reflects both the limited sample size and the inherent difficulty of reproducing the spatial coherence of rare, high-intensity anomalies. In particular, the spread of the precipitation ensemble can lead to under- or overestimation of the spatial extent and intensity of the most severe drought conditions. Nevertheless, a substantial proportion of drought events still exceeded the commonly used FSS benchmark of 0.5 at the native grid resolution (30 of 36 SPEI-1 events, 17 of 36 SPEI-3 events, and 25 of 36 SPEI-6 events), indicating useful spatial forecast skill even before neighborhood smoothing is applied.

To complement the conventional verification metrics, the Potential Economic Value (PEV) was evaluated for all selected drought events (Figure 7). Across all drought severities and aggregation periods, the forecasts exhibited positive maximum PEV values, indicating economic benefit relative to climatological decision making. Bootstrap confidence intervals likewise remained

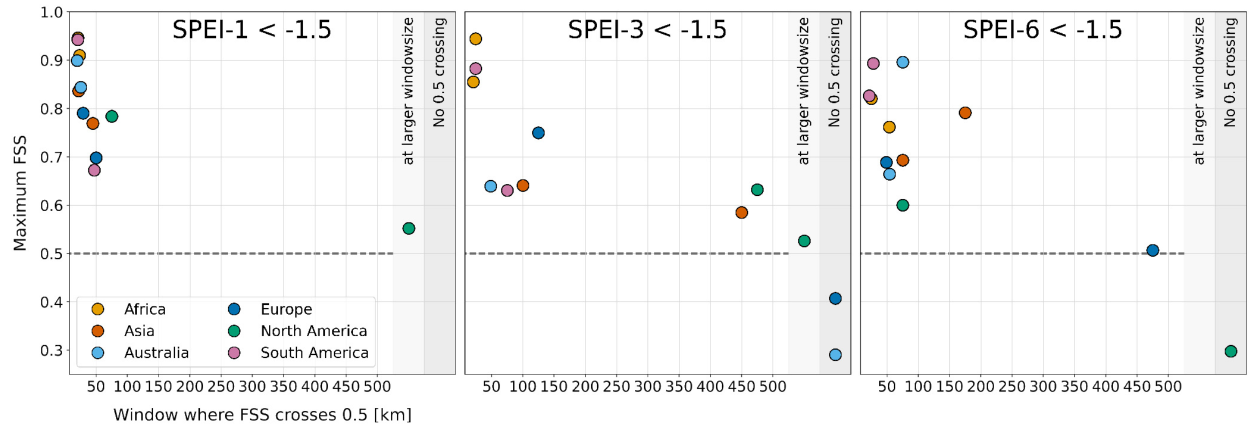

FIGURE 6
Fractions Skill Score (FSS) of strong drought occurrence (SPEI < −1.5) for SPEI-1 (left), SPEI-3 (center), and SPEI-6 (right). Colors indicate the different continents. The dashed horizontal line marks the 0.5 skill benchmark. Points located within the shaded areas either do not exceed the 0.5 threshold or only reach it at spatial window sizes greater than 20 grid cells.

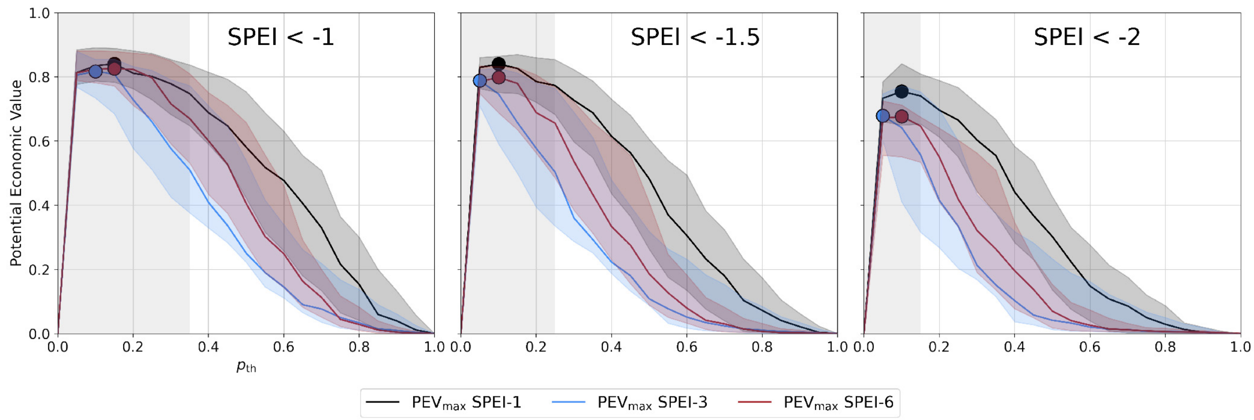

FIGURE 7
Potential Economic Value (PEV) as a function of the forecast probability threshold p_{th} for mild (left), severe (middle), and extreme (right) droughts. Curves show the maximum PEV over all tested cost-loss ratios for SPEI-1, SPEI-3, and SPEI-6. Markers indicate the probability threshold corresponding to the maximum PEV. Shading denotes the 95% bootstrap confidence interval (1,000 resamples). Grey areas indicate forecast probability thresholds for which at least 75% of drought events achieved hit rates exceeding 50%.

above zero, suggesting that this result is robust with respect to event sampling. As expected from the low climatological frequency of severe droughts, the optimal cost-loss ratios were generally small, with maximum PEV typically obtained for cost-loss ratios around 0.05 (not shown). For SPEI < -1.5, positive PEV values persisted up to cost-loss ratios of approximately 0.3, while for SPEI < -1 they remained positive up to roughly 0.5. Mild droughts (SPEI < -1) yielded the highest economic value, although even forecasts for extreme droughts (SPEI < -2) reached PEV values exceeding 0.7. Consistent with the CRPSS and Brier Skill Score analyses, forecasts initialized 1 month in advance generally achieved the highest PEV. Six-month aggregation periods tended to outperform the 3-month aggregation, although the differences were not statistically significant. Hatched regions indicate forecast probability thresholds for which at least 75% of drought events achieved hit rates exceeding 50%, highlighting a broad range of probability thresholds with consistently high event detection.

3.2 Case studies of individual drought events

To present the forecast performance in greater detail, we focus on three representative droughts rather than discussing every case individually. Based on the CRPSS, we select one well-predicted event for SPEI-1, one rather poorly predicted event for SPEI-3, and another well-predicted event for SPEI-6. Specifically, these are Africa 05/2024, North America 10/2023, and Australia 03/2020 (see Table A1 in Appendix).

Africa 05/2024 represents the drought with the largest impact area in Africa among those analyzed (see Table A1 in Appendix). Negative SPEI values covered most of Africa (Figure 8), with extensive regions falling below an SPEI of –2. The most severely affected areas include Chad and the Central African Republic, where values below –4 are detected, as well as large parts of the Democratic Republic of the Congo. The northern Sahel zone is

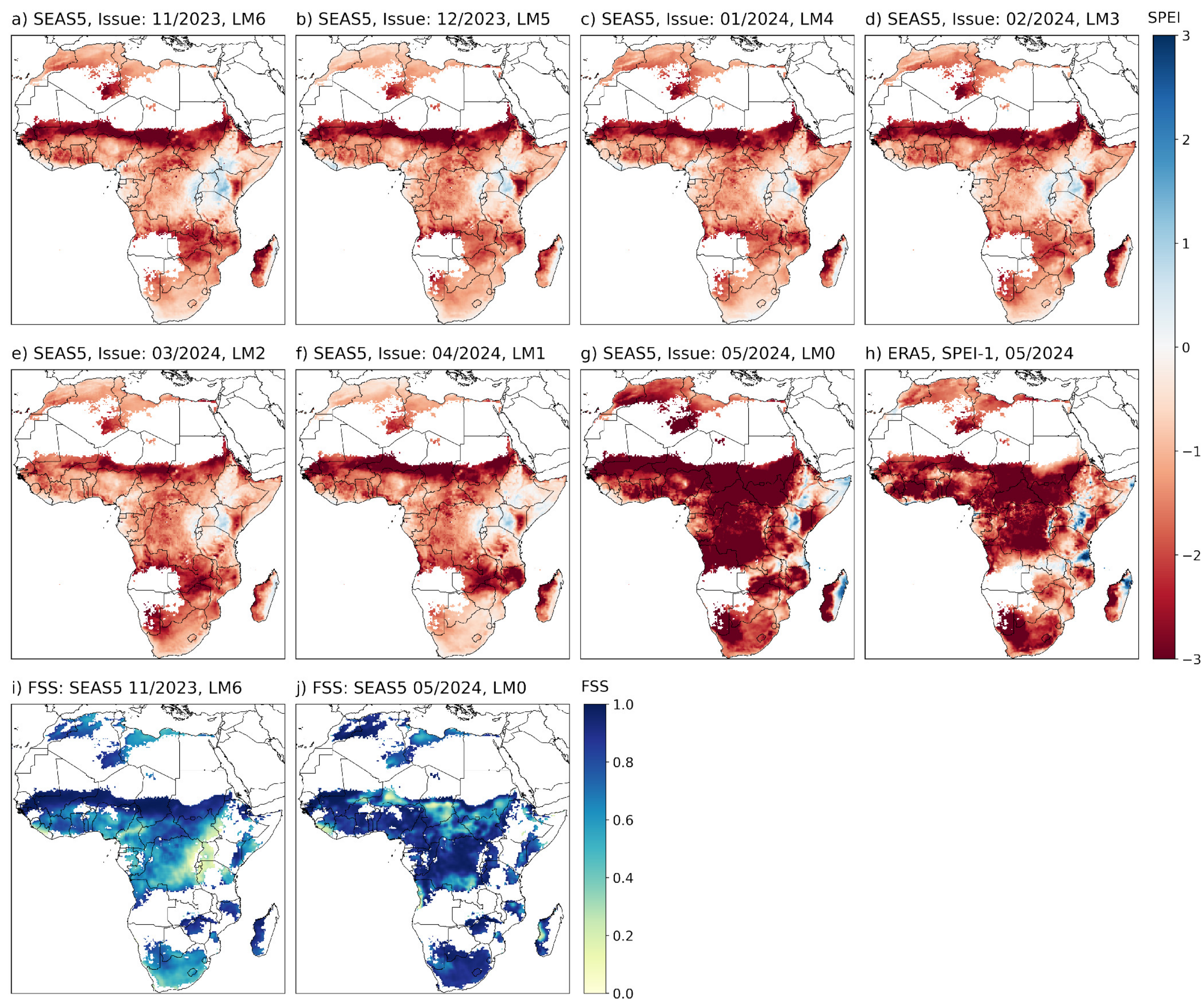

FIGURE 8
(a–g) SEAS5-BCSD forecasts of SPEI-1 valid for May 2024 from different issue months (lead month abbreviated as LM). **(h)** Reference SPEI-1 from ERA5 for May 2024. Extreme arid regions in the target month are masked. **(i)** Fractions Skill Score (FSS) of the SEAS5 SPEI-1 forecast (issue month 11/2023) for strong drought conditions (SPEI < −1.5), computed using a 100 km window. **(j)** FSS of SEAS5 SPEI-1 (issue month 05/2024).

found as a hotspot of drought. These ERA5-based findings are consistent with NOAA (2024). Figures 8a–g displays the SPEI-1 forecasts for this month from all seven available lead times. The dominance of red shading, indicating low SPEI values, across all forecast issue months suggests that SEAS5-BCSD successfully captured the drought tendency up to 7 months in advance. The corrected forecasts exhibit a spatially consistent dry signal from lead months six to one, with positive SPEI anomalies confined mainly to around Uganda, while the northern Sahel and the Zambezi River Basin show the strongest dryness. The mean SPEI across Africa fluctuates between −1.32 and −1.43 during this period, indicating severe drought conditions over large parts of the continent. At lead month 0, the average SPEI drops sharply to −2.74, with intensified dryness particularly over the Central African Republic, the Congo River Basin and southern Namibia. In contrast, localized wetter conditions intensify over the Horn of Africa and eastern Madagascar. Comparison with ERA5 data reveals a high degree of spatial agreement, with minor deviations in the western Sahel, where the forecasts indicate record-dry conditions, while observations show a still strong but less extreme drought. To quantify these qualitative findings, Figure 8j displays the FSS for Lead Month 0, using the strong drought threshold and a window size of 100 km. FSS values are shown only for grid cells where SPEI < -1.5 occurred in ERA5. At Lead Month 0, the majority of affected areas exhibit FSS values well above the commonly used 0.5 threshold. Especially the drought hot spots in South Africa and the Congo basin are captured with values ranging consistently above 0.9. The non-drought areas of eastern Africa show 0. As an example of longer lead times, the forecast issued in November 2023 is shown in Figure 8i. Here, 83.0% of all grid cells exceed the 0.5 threshold, compared to 92.4% at Lead Month 0. The lowest values occur in eastern Africa, where a wet anomaly was correctly predicted but displaced by more than the prescribed window size. The highest FSS values are found in the northern Sahel, where values of 0.9 are consistently reached.

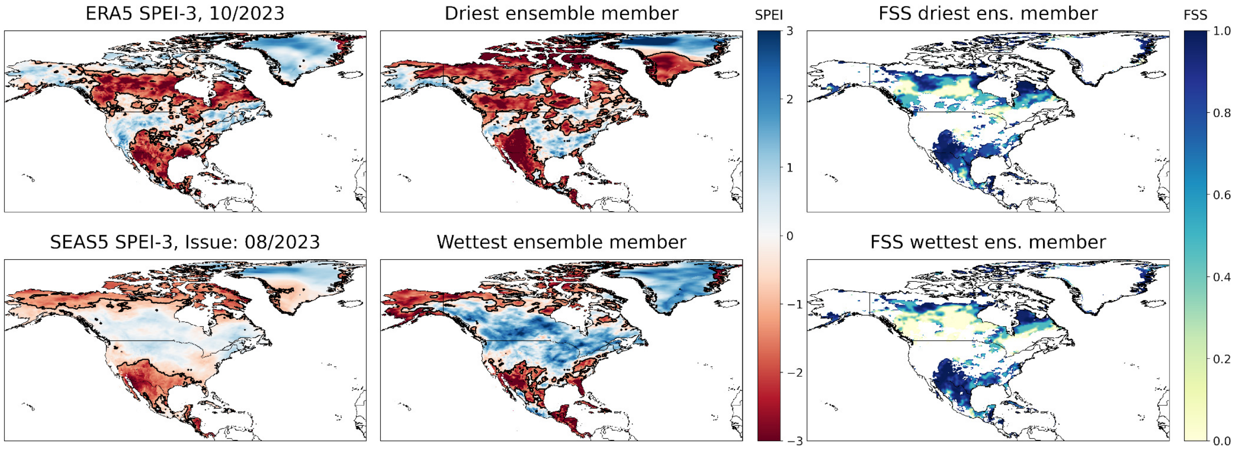

FIGURE 9
SPEI-3 for October 2023 over North America. The left column shows ERA5 reanalysis **(Top)** and the SEAS5-BCSD ensemble mean aggregated over lead months 0-2 **(Bottom)**. The middle column displays SPEI-3 for the driest and wettest ensemble members. Contours indicate the moderate drought threshold (SPEI = −1). The right column presents the Fractions Skill Score (FSS) for the respective ensemble member, calculated using a 100 km neighborhood window.

Overall, the seasonal forecasts perform very well, capturing the extent of the event several months in advance, although the drought intensity was slightly underestimated prior to lead month 0, which in turn over-corrected this bias. The spatial location of the drought is reproduced accurately, even up to 7 months ahead.

To further investigate the limits of drought forecasting skill, a comparatively weak-performing event is examined (Figure 9). The 10/2023 North American drought, responsible for a very strong forest fire season in Canada (Couillard et al., 2026), exhibits the lowest CRPSS among the analyzed SPEI-3 droughts and the second lowest performance overall, albeit still featuring positive CRPSS values and thus outperforming climatology (Figure 3). A direct comparison of the ensemble mean forecast (upper right panel Figure 9) with the ERA5 reference SPEI-3 field (upper left panel Figure 9) reveals a bipolar drought structure, characterized by two concurrent drought centers. In the reference data, a major drought affects large parts of Mexico and the adjacent southern United States, while a second drought extends across central Canada. The forecast reproduces the southern drought with high fidelity, capturing its intensity, spatial extent, and location well, with the exception of Baja California. Over Canada, SEAS5-BCSD also indicates drought conditions but underestimates their intensity and shifts the affected region too far north. This highlights a key limitation of seasonal forecasting systems, where spatial displacement of large-scale atmospheric patterns can result in substantial regional errors in precipitation anomalies.

To assess ensemble spread, the driest and wettest ensemble members (based on domain-averaged SPEI) are shown in the lower panels of Figure 9. For the Mexican drought, even the wettest ensemble member consistently indicates extreme drought conditions, demonstrating a robust signal across the ensemble for this particular case. The spatial structure of the southern drought is well-preserved in both extreme members, with the wettest member even capturing abnormally dry conditions along the southeastern U.S. coast more clearly than the driest member. The driest member identifies drought conditions across North America but exhibits a slight southward shift and introduces an additional drought signal over northern Canada and southern Greenland that is absent in ERA5. Conversely, the wettest member shows better agreement over Greenland but displaces the Canadian drought northward.

To quantify this qualitative assessment, the Fractions Skill Score (FSS) is analyzed for this event (Figure 9, right column). For both the driest and wettest ensemble members, nearly all regions south of the Canadian border exhibit positive skill, with local FSS values exceeding 0.9 in many areas. In northern Quebec and parts of the Northwest Territories, both ensemble extremes suggest that the drought pattern is slightly displaced. The displacement is smaller for the driest ensemble member (top right) than for the wettest. The shift results in lower FSS values along the southern edge of the impact area. As the neighborhood window size increases (not shown), FSS values rise, reflecting the close proximity of the forecasted drought signal despite its spatial offset. Overall, the driest ensemble member achieves a mean FSS of 0.649, while the wettest member attains a slightly lower but still positive mean of 0.569. For the stronger drought threshold (SPEI < -1.5), the 0.5 FSS benchmark is exceeded at a neighborhood scale of approximately 475 km (see Figure 6). A maximum FSS value of 0.632 (0.617 for extreme droughts, Figure A1 in Appendix) is reached, indicating useful spatial skill at larger scales. These results support the earlier interpretation that reduced BSS values primarily stem from large-scale spatial displacement rather than a failure to capture the overall drought signal.

Overall, this spatial analysis demonstrates that although the aggregate skill for this event is limited, the ensemble contains valuable predictive information. The consistent detection of drought signals, despite spatial displacement and intensity errors, suggests that ensemble-based interpretation can still provide meaningful guidance for large-scale drought monitoring.

As a final case, we analyze a SPEI-6 drought event. The Australia 03/2020 drought, spanning from October 2019 to March 2020, was selected because the 2019–20 season was the second hottest on record (1940–2024) and fell into approximately the

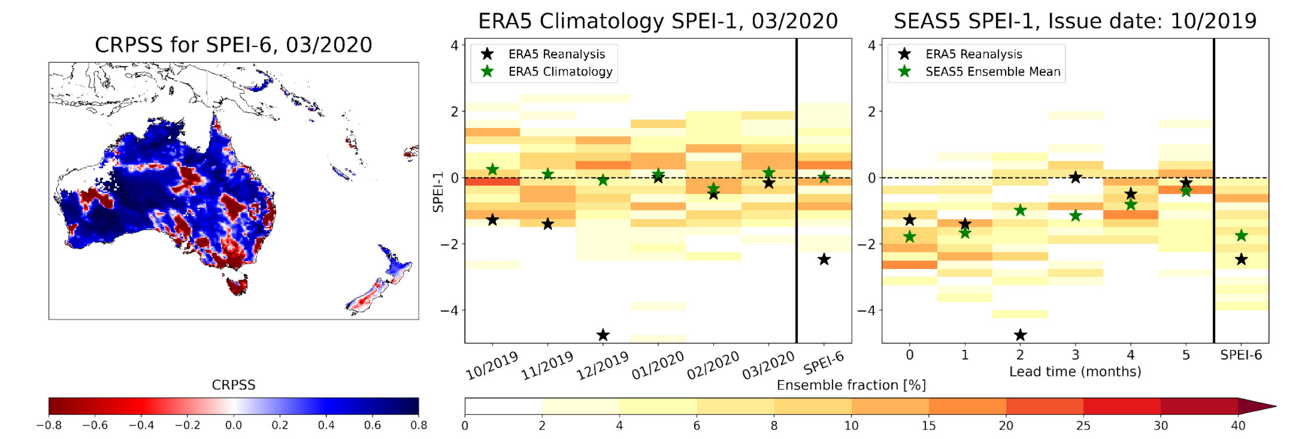

FIGURE 10
Ensemble fractions for each SPEI-1 category during the 2019/20 Australian drought for ERA5 climatology **(Middle)** and SEAS5-BCSD **(Right)**. Observed SPEI-1 values are shown as black stars and the ensemble mean as green stars. The rightmost column in both panels shows the ensemble fraction for SPEI-6 over the full period. The Continuous Ranked Probability Skill Score (CRPSS) relative to climatology for the corresponding SPEI-6 event is displayed on the **(Left)**.

15th precipitation percentile for that time of year. These 6 months correspond to the climatologically wettest season averaged over Australia, meaning that the concurrence of exceptional heat and strong rainfall deficits produced a pronounced drought. The preceding year (2018–19) featured weak El Niño conditions (Rayner et al., 2003), which typically increase predictability of the Australian rainy season (Johnson et al., 2019; Ma et al., 2023). By contrast, 2019–20 exhibited neutral to weak El Niño conditions, limiting the extent to which SEAS5-BCSD could rely on teleconnection signals to anticipate dry anomalies. Therefore, skill arising independently of ENSO forcing should be apparent.

Figure 10 displays the CRPSS for this event in the left panel. Predominantly dark blue shading indicates strong positive skill of the bias-corrected forecast relative to climatology. The main impact region of the drought, located in central Australia near the Gibson Desert, is captured very well. Compared to the preceding North American drought case, the spatial extent is represented more accurately, with no major displacement errors. The continental mean SPEI-6 is also reproduced well, with a forecast value of -1.21 compared to -1.13 in the reanalysis (see rightmost column of the right panel).

Given the long lead time of more than 150 days, reduced forecast skill is expected compared to the SPEI-1 case (see Figures 3–6). To examine the drought evolution in more detail, the SPEI-6 was decomposed into six consecutive SPEI-1 values. Although this changes the effective weighting of individual months compared to a combined SPEI-6 calculation, the trade-off is justified: the decomposition allows a clearer assessment of the ensemble spread and month-to-month signal accumulation. Note that January and February contribute the largest fraction of annual precipitation in Australia, and therefore exert the greatest influence on the resulting SPEI-6 values on the right.

For climatology, the ensemble fraction reproduces the last 3 months reasonably well, as these months show only slightly dry conditions and the climatology exhibits a broadly centered distribution (middle panel of Figure 10). In contrast, during months with very low observed SPEI-1 values, climatology fails to capture the severity of the drought, particularly in December, when record-low values occurred.

SEAS5-BCSD, on the other hand, exhibits a pronounced dry tendency in the first two lead months, producing an ensemble mean close to the observed value together with a reduced ensemble spread (right panel of Figure 10). Although the extreme December drought is underestimated, several ensemble members reach values near -4, suggesting that the system (given sufficient perturbation) can in principle reproduce events of such magnitude. The last three, more moderate, months are predicted to be near-normal but slightly dry, again with a noticeably narrower ensemble spread than climatology, resulting in a more precise forecast.

Across all six lead months, only very few SEAS5-BCSD ensemble members produce positive SPEI values. These occur mainly in Lead Months 3 and 5. Since the observed conditions during these months were close to neutral (SPEI $\approx$ 0), the limited presence of wet outliers indicates generally appropriate model behavior. When focusing on SPEI-6, shown as the rightmost column in the right panel, no ensemble member exhibits values above zero, consistent with the reanalysis data. By contrast, the ERA5 climatology is centered around zero and therefore remains far from the observed conditions. Consequently, SEAS5 clearly outperforms the climatological reference.

4 Discussion

4.1 Signal versus skill: misplacement, damping, and lead-time dependence

The discrepancy between the consistently positive CRPSS and the BSS occasionally falling below climatology arises primarily from differences in what these scores emphasize. The CRPSS accounts for both dry and wet anomalies, whereas the BSS focuses solely on drought events, making it more sensitive to errors in intensity and spatial placement. Consequently, reduced BSS values do not necessarily indicate an absence of predictive signal, but rather diminished reliability in the representation of drought magnitude and spatial coherence. At longer lead times, decreases in CRPSS or even negative BSS values should therefore not be interpreted as a complete loss of predictability. For several drought events, the Fraction Skill Score (FSS) still exceeds the 0.5 threshold when evaluated at larger spatial windows (see Figure 9), indicating that the signal persists but is spatially displaced. Similarly, the Hit Rate consistently exceeding the False Positive Rate (with only one exception), and both being substantially higher than climatological values, further supports the presence of a genuine drought signal that is affected by displacement, intensity damping, or temporal smoothing rather than being entirely absent. Finally, the ability of ensemble tails to capture extreme conditions even when the ensemble mean remains moderate highlights that valuable predictive information can persist beyond what is reflected in mean-based or event-based skill scores. Taken together, these results suggest that the reduction of probabilistic skill at longer lead times is primarily driven by growing uncertainty in spatial positioning and intensity, rather than by a lack of drought predictability.

4.2 Continental differences in drought predictability

When comparing continental-scale performance, the evaluated African drought events generally exhibit the highest forecast skill, whereas the North American events tend to show lower skill (Figures 3, 4 and Table A2 in Appendix). However, the bootstrap confidence intervals indicate that continental differences should be interpreted with caution given the limited sample size. The predictability of extreme droughts is strongly controlled by the dominant drivers of hydroclimatic variability (Williams et al., 2023). In regions where large-scale modes such as sea surface temperature (SST) anomalies and teleconnections exert a strong influence (Dai and Wigley, 2000), seasonal drought predictability tends to be enhanced. The selected African drought events provide an example of this behavior: the strong coupling between tropical climate variability, SSTs, and precipitation (Nicholson, 2017), combined with comparatively lower synoptic-scale variability relative to mid-latitudes (Trenberth and Stepaniak, 2003), results in robust and spatially coherent forecast skill (Tanessong et al., 2024; Tchinda et al., 2026). While predictability varies within Africa (Deman et al., 2022), the dominance of large-scale tropical forcing across much of the continent enhances average skill at continental scale. Similarly, South America, which is strongly influenced by SST anomalies as well, exhibits relatively high skill, particularly at longer lead times (Figure 6, right panel). Europe, by comparison, is less directly influenced by SSTs and primarily affected by the North Atlantic Oscillation (NAO; Pozo.Vázquez et al., 2001), resulting in generally lower seasonal forecast skill. In contrast, North America is characterized by more dynamic atmospheric variability and a weaker dominance of large-scale teleconnection forcing (Alexander et al., 2002), which limits seasonal predictability. An additional challenge arises from the large fraction of North American land area located at high latitudes, where forecast skill is generally reduced (Weber et al., 2026b). This reduction may partly reflect observational limitations in polar regions, where the sparse *in situ* observing network hampers both reanalysis quality and forecast initialization (Smith et al., 2019). Consistent with the signal-versus-skill distinction discussed above, reduced scores over North America largely reflect spatial displacement and damped intensity rather than a complete loss of predictability at longer lead times (Figure 9). This interpretation is supported by Fraction Skill Scores exceeding 0.5 at larger spatial windows (Figure 6), indicating that forecasts capture the correct large-scale drought patterns despite local mismatches. Finally, continental-scale conclusions are based on only six extreme drought events per continent. Although bootstrap confidence intervals indicate that the reported median skill estimates are reasonably robust for the selected event set, the limited sample size and event-focused selection warrant caution when generalizing these findings to continental-scale drought predictability.

4.3 Limitations

A central limitation of this analysis is its dependence on ERA5 as reference data. While ERA5 is known to exhibit regional precipitation biases, for example in parts of East Africa (Ahmed et al., 2024), it reliably captures large-scale spatial patterns and rainfall distributions. Owing to its global coverage, internal consistency, and seamless availability of all required variables for drought indicator calculation (Hersbach et al., 2020), ERA5 represents a pragmatic and widely accepted reference for global-scale drought forecast evaluation.

It should furthermore be noted that ERA5 is used both as the reference dataset for the bias-correction procedure and for forecast verification. Consequently, the evaluation quantifies the ability of SEAS5-BCSD to reproduce the ERA5 representation of historical droughts rather than independent *in-situ* observations, as discussed in Weber et al. (2026b). Although the use of a different observational or reanalysis dataset would inevitably alter the absolute skill estimates, ERA5 provides a consistent reference for both the post-processing and verification framework adopted in this study. A different observational or reanalysis dataset for verification would introduce an additional source of uncertainty arising from differences between reference datasets, thereby complicating the interpretation of forecast performance. Using the same reference dataset for post-processing and verification therefore provides methodological consistency, but does not guarantee positive forecast skill or improvements relative to the chosen benchmark forecasts, since the BCSD adjusted precipitation and temperature distributions rather than optimizing forecast skill metrics such as CRPSS, BSS, or ROC.

The reference climatology period (1981–2016) differs from the WMO standard normal period (1991–2020; WMO, 2021). This choice was necessitated by the increase in SEAS5 ensemble size in 2017, for which no robust statistical harmonization approach exists (Manzanas et al., 2022). While absolute climatological values may differ, the internally consistent use of a single reference period ensures the comparability of all skill metrics within this study. Extremely low SPEI values partly reflect the use of this fixed reference climatology, so unprecedented recent droughts naturally produce increasingly negative standardized anomalies. Such values should therefore be interpreted as indicators of exceptionally rare conditions relative to the reference period rather than precise estimates of return period. Furthermore, uncertainties in ERA5 precipitation and evapotranspiration, particularly in data-sparse regions, may influence the exact magnitude of these extreme values. Additionally, the findings presented here are specific to SPEI-based drought forecasts. Different drought indices, such as SPI or PDSI, emphasize different physical processes and may therefore exhibit different levels of predictability.

Another limitation concerns the attribution of drought forecast skill to individual drivers. In particular, the relative contribution of temperature versus precipitation to SPEI-based skill is not explicitly quantified. Previous studies indicate that rising temperatures increasingly amplify drought severity in SPEI relative to SPI under global warming (Nwayor and Robeson, 2024). Given the high skill of BCSD-corrected temperature forecasts (Weber et al., 2026b), temperature likely plays a substantial role in the overall drought predictability, but its exact contribution warrants targeted follow-up analysis. For multi-month drought indices (SPEI-3 and SPEI-6), the influence of Lead Month 0 and subseasonal predictability in the first forecast weeks may affect overall skill. While the Australian case study (Figure 10) suggests that meaningful predictive information persists beyond the first month, a systematic assessment across events remains an avenue for future research. Compared to previous studies reporting higher skill for seasonal SPEI forecasts (e.g., Dutra et al., 2014), the scores presented here are lower. This difference arises from the strict restriction to purely forecasted months, whereas seamless approaches incorporate monitoring data into longer accumulation periods, inflating skill by construction. Although more conservative, the present approach provides an uncontaminated assessment of true forecast performance, rendering positive skill particularly meaningful.

Finally, the analysis was not stratified by regional rainy and dry seasons. At continental scale, such a separation is difficult to define consistently due to (1) differences in rainy season type (bimodal versus unimodal), (2) shift of season by teleconnections like ENSO (Goswami and Xavier, 2005) and (3) regions (e.g., northern Europe) without a clearly defined rainy season. Moreover, a stratification would substantially reduce sample size, limiting statistical robustness. Additionally, the SPEI calculation already accounts for the intra-seasonal differences between the year by standardization by climatology on a monthly (aggregation) scale.

4.4 Seasonal drought forecasts in decision support

Despite these limitations, the analysis demonstrates that under strong dry conditions, SEAS5-BCSD forecasts contain useful probabilistic drought signals several months in advance, even when spatial placement or intensity remains uncertain. Figure 7 demonstrates the resulting possible economic implications using the Potential Economic Value (PEV). A key strength of the system lies in its ensemble-based formulation, which provides probabilistic information that is particularly valuable for assessing the likelihood of extreme drought events. Deterministic skill scores alone underestimate the practical value of ensemble forecasts for early warning applications (Chevuturi et al., 2021). Unlocking the full potential of ensemble predictions therefore requires a comprehensive analysis of the ensemble distribution, rather than reliance on simple ensemble-mean approaches (e.g., Palmer et al., 2004). Even when the ensemble mean underestimates drought severity, individual ensemble members can reproduce extreme conditions, as illustrated for the Australian case (Figure 10). Likewise, events with weak aggregate skill may still contain useful information within the ensemble structure, particularly in the behavior of extreme members (Figure 9). These examples illustrate the potential value of examining the full ensemble distribution, but they do not constitute a systematic assessment of the operational value of ensemble-tail information or its associated false-positive characteristics. Furthermore, stakeholders should also monitor drought signals in regions adjacent to their area of interest, as slight spatial displacement may occur, particularly at longer lead times. Additional predictive insight can be gained by examining the temporal consistency of forecasts across successive issue months: drought signals that persist across lead times tend to be more reliable (Figure 8), in agreement with findings by Weber et al. (2026b). Another qualitative indicator of forecast robustness is the coherence of the drought signal across ensemble members, for example when only a few wet outliers contrast an otherwise consistent dry pattern, or when spatial drought structures are reproduced consistently across members, as seen for the Mexican drought in Figure 9. Such ensemble-based interpretation is consistent with WMO guidance on climate forecast use (Graham et al., 2011) and with operational drought outlook practices in regional climate outlook forums, such as the Greater Horn of Africa Climate Outlook Forum (GHACOF).

Based on these findings, several practical recommendations for the interpretation of seasonal drought forecasts can be derived:

- Consider not only the ensemble mean but also the ensemble distribution when assessing forecast uncertainty.
- Inspect the full ensemble distribution, including the most extreme members, as these may provide additional context for low-probability, high-impact scenarios. But treat signals in individual extreme ensemble members as complementary information rather than stand-alone evidence for decision making.
- Evaluate the magnitude of ensemble spread as an indicator of forecast uncertainty and confidence.
- Assess drought signals in neighboring regions to account for potential spatial displacement.
- Compare forecasts across multiple lead times, as temporally consistent signals tend to be more reliable.

More reliable seasonal ensemble forecasts therefore enable more nuanced and risk-aware drought responses when interpreted appropriately. For example, Portele et al. (2021) demonstrate substantial economic benefits for water reservoir management when seasonal forecasts are incorporated into decision-making. Crucially, users should be sensitized not to focus solely on the most likely outcome, but also to consider low-probability, high-impact scenarios represented in the tails of the ensemble distribution, which are inherently difficult to capture using deterministic metrics alone (Graham et al., 2011). A key next step is the improved integration of such ensemble-based information into existing decision support frameworks and operational drought monitoring systems (Bruno Soares et al., 2018; Suárez-Almiñana et al., 2022).

5 Conclusion

This study demonstrates that, for the selected catalog of 36 large-scale historical drought events, global bias-corrected seasonal forecasts from SEAS5 provide useful probabilistic information for anticipating large-scale, extreme droughts across all continents several months in advance. For nearly all analyzed events, forecast skill exceeds climatology, often substantially, with the selected African droughts generally noting the best forecast performance. While short-lived droughts and forecasts issued close to onset exhibit the highest skill, meaningful predictive information persists even at lead times of up to 6 months.

Predictability is highest for the spatial extent of moderate drought conditions, whereas the precise location and intensity of the most extreme drought cores remain more challenging. Importantly, reduced deterministic or event-based skill scores do not imply an absence of predictive signal. Even in cases of lower overall performance, the ensemble distribution may retain useful predictive information that is not fully reflected by deterministic or event-based skill scores. The presented case studies suggest that this information can include low-probability, high-impact outcomes, as well as indicators of spatial displacement and elevated hit rates, although the operational value of ensemble-tail information warrants further dedicated investigation. These findings further suggest that analyses relying solely on the ensemble mean or a limited subset of members risk discarding critical information relevant for extreme event preparedness.

Across the selected events, the forecast system exhibited useful skill on all continents, suggesting that SEAS5-BCSD can provide valuable probabilistic information for large-scale drought preparedness in a wide range of climatic settings. The results highlight the value of probabilistic, ensemble-based seasonal drought forecasts for early warning and risk-informed decision-making. When interpreted appropriately alongside local expertise and impact information, such forecasts can support well-coordinated drought preparedness strategies that enable the timely implementation of low-cost and no-regret measures months in advance, scaled to both the severity and the likelihood of the predicted event. The consistent performance across the analyzed events demonstrates the broad applicability of the evaluation framework across diverse climatic regions. Strengthening the integration of ensemble drought forecasts into operational planning frameworks therefore represents a key opportunity to enhance proactive drought risk management.

As noted in the introduction, a comprehensive and up-to-date global drought database is currently lacking. The identification of the 36 extreme drought events presented in this study therefore represents a first step toward such a resource. To facilitate reuse and further analysis, the event catalog is published alongside this study (Table A1 in Appendix), complemented by spatial delineations of the drought-affected areas (shape files; Weber (2026)). While the primary focus of this work lies on the evaluation of seasonal forecast skill, this additional dataset provides a consistent, observation-based reference for future studies on large-scale drought events and their predictability. This may support comparative studies, model benchmarking, and the evaluation of drought impacts across regions and time scales.

Based on the findings of this study, several directions for future research emerge. One promising approach would be to analyze the specific drought events investigated here using impact models, such as hydrological or agricultural models, in order to quantify the potential benefits of seasonal drought forecasts in retrospective case studies. By comparing model simulations that incorporate SEAS5-BCSD forecasts with simulations relying solely on climatological information, it would be possible to estimate how early drought warnings could have influenced management decisions and to quantify potential gains, for example in terms of crop yield, reservoir storage, hydropower production, or economic savings. From a modeling perspective, further research could also focus on identifying the large-scale drivers of the analyzed drought events. These may include teleconnections or characteristic synoptic-scale circulation patterns. Investigating whether SEAS5 successfully predicts these large-scale precursors would provide valuable insight into the sources of forecast skill. Such analyses could ultimately support targeted improvements in seasonal forecasting systems and help refine bias-correction approaches.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

JW: Writing – review & editing, Writing – original draft, Software, Conceptualization, Investigation, Formal analysis, Visualization, Validation, Methodology, Data curation. CL: Supervision, Conceptualization, Writing – review & editing, Data curation, Software, Funding acquisition, Project administration. TS: Writing – review & editing, Conceptualization, Supervision. HD: Writing – review & editing, Investigation. RW: Software, Writing – review & editing. HK: Supervision, Project administration, Conceptualization, Resources, Writing – review & editing, Funding acquisition.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the Development of an operational, multisectoral, global drought hazard prediction system (OUTLAST) project (grant no. 02WGR1642C) funded by the German Federal Ministry of Research, Technology and Space (BMFTR).

Acknowledgments

We acknowledge the ECMWF for providing ERA5 and SEAS5 data.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. Artificial intelligence-based language tools were used to support English language editing; the scientific content and interpretations remain the sole responsibility of the author(s).

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2026.1911791/full#supplementary-material

References

Ahmed, J. S., Buizza, R., Dell'Acqua, M., Demissie, T., and Pè, M. E. (2024). Evaluation of era5 and chirps rainfall estimates against observations across Ethiopia. *Meteorol. Atmos. Phys*. 136:17. doi: 10.1007/s00703-024-01008-0

Alexander, M. A., Bladé, I., Newman, M., Lanzante, J. R., Lau, N.-C., and Scott, J. D. (2002). The atmospheric bridge: the influence of enso teleconnections on air-sea interaction over the global oceans. *J. Climate* 15, 2205–2231. doi: 10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2

Arlot, S., and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statist. Surv*. 4, 40–79. doi: 10.1214/09-SS054

Arnone, E., Cucchi, M., Dal Gesso, S., Petitta, M., and Calmanti, S. (2020). Droughts prediction: a methodology based on climate seasonal forecasts. *Water Resour. Manag*. 34, 4313–4328. doi: 10.1007/s11269-020-02623-3

Bandyopadhyay, N., Bhuiyan, C., and Saha, A. K. (2020). Drought mitigation: critical analysis and proposal for a new drought policy with special reference to Gujarat (India). *Progress Disaster Sci*. 5:100049. doi: 10.1016/j.pdisas.2019.100049

Beck, H., van Dijk, A. I. J. M., Larraondo, P. R., McVicar, T. R., Pan, M., Dutra, E., et al. (2022). Mswx: global 3-hourly 0.1○ bias-corrected meteorological data including near-real-time updates and forecast ensembles. *Bull. Am. Meteorol. Soc*. 103, E710–E732. doi: 10.5194/egusphere-egu23-2246

Bird, P. (2003). An updated digital model of plate boundaries. *Geochem. Geophys. Geosyst*. 4:1027. doi: 10.1029/2001GC000252

Boé, J., Terray, L., Habets, F., and Martin, E. (2007). Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *Int. J. Climatol*. 27, 1643–1655. doi: 10.1002/joc.1602

Brands, S., Díez González-Pardo, J., Herrera, S., Bedia, J., Manzanas, R., Esteban, R.-G., et al. (2025). Seasonal drought predictions in the Mediterranean using the spei index: paving the way for their operational applicability in climate services. *Climate Serv*. 38:100555. doi: 10.1016/j.cliser.2025.100555

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Rev*. 78, 1–3. doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Bruno Soares, M., Daly, M., and Dessai, S. (2018). Assessing the value of seasonal climate forecasts for decision-making. *WIREs Climate Change* 9:e523. doi: 10.1002/wcc.523

Carrão, H., Naumann, G., Dutra, E., Lavaysse, C., and Barbosa, P. (2018). Seasonal drought forecasting for Latin America using the ecmwf s4 forecast system. *MDPI Climate* 6:48. doi: 10.3390/cli6020048

Chevuturi, A., Turner, A. G., Johnson, S., Weisheimer, A., Shonk, J. K., Stockdale, T. N., and Senan, R. (2021). Forecast skill of the Indian monsoon and its onset in the ecmwf seasonal forecasting system 5 (seas5). *Climate Dyn*. 56, 2941–2957. doi: 10.1007/s00382-020-05624-5

Couillard, P.-L., Laflamme, J., Frégeau, M., Cardona, L., Morneau, C., and Payette, S. (2026). Drought as the main climatic driver of major boreal fire seasons over the last century in Eastern Canada. *Environ. Res. Lett*. 21:054016. doi: 10.1088/1748-9326/ae49a0

CRED and UNDRR (2020). *The Human Cost of Disasters: An Overview of the Last 20 Years (2000–2019)*. Technical report, Centre for Research on the Epidemiology of Disasters (CRED), United Nations Office for Disaster Risk Reduction (UNDRR), Brussels.

Dai, A., and Wigley, T. (2000). Global patterns of enso-induced precipitation. *Geophys. Res. Lett*. 27, 1283–1286. doi: 10.1029/1999GL011140

Deman, V. M., Koppa, A., Waegeman, W., MacLeod, D. A., Bliss Singer, M., and Miralles, D. G. (2022). Seasonal prediction of horn of africa long rains using machine learning: the pitfalls of preselecting correlated predictors. *Front. Water* 4:1053020. doi: 10.3389/frwa.2022.1053020

Doblas-Reyes, F. J., Gardía-Serrano, J., Lienert, F., Biescas, A. P., and Rodrigues, L. R. L. (2013). Seasonal climate predictability and forecasting: status and prospects. *WIREs Climate Change* 4, 245–268. doi: 10.1002/wcc.217

Dutra, E., Pozzi, W., Wetterhall, F., Di Giuseppe, F., Magnusson, L., Naumann, G., et al. (2014). Global meteorological drought – part 2: seasonal forecasts. *Hydrol. Earth Syst. Sci*. 18, 2669–2678. doi: 10.5194/hess-18-2669-2014

European Commission, J. R. C. (2026). *Drought events Based on SPI 12 Version 1.0*. Dataset. European Commission.

FAO and WMO (2018). *2017 the Impact of Disasters and Crises on Agriculture and Food Security*. Technical Report, Food and Agriculture Organization of the United Nations, Rome.

Gonzalez-Aparicio, I., and Hidalgo, J. (2011). Dynamically based future daily and seasonal temperature scenarios analysis for the northern Iberian peninsula. *Int. J. Climatol*. 32, 1825–1833. doi: 10.1002/joc.2397

Goswami, B., and Xavier, P. K. (2005). Enso control on the south Asian monsoon through the length of the rainy season. *Geophys. Res. Lett*. 32:L18717. doi: 10.1029/2005GL023216

Graham, R., Yun, W.-T., Kim, J., Kumar, A., Jones, D., Bettio, L., et al. (2011). Long-range forecasting and the global framework for climate services. *Climate Res*. 47, 47–55. doi: 10.3354/cr00963

Hao, Z., AghaKouchak, A., Nakhjiri, N., and Farahmand, A. (2014). Global integrated drought monitoring and prediction system. *Sci. Data* 1:140001. doi: 10.1038/sdata.2014.1

Hao, Z., Singh, V. P., and Xia, Y. (2018). Seasonal drought prediction: advances, challenges, and future prospects. *Rev. Geophys*. 56, 108–141. doi: 10.1002/2016RG000549

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecast*. 15, 559–570. doi: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The era5 global reanalysis. *Q. J. R. Meteorol. Soc*. 146, 1999–2049. doi: 10.1002/qj.3803

Holthuijzen, M., Beckage, B., Clemins, P. J., Higdon, D., and Winter, J. M. (2022). Robust bias-correction of precipitation extremes using a novel hybrid empirical quantile-mapping method. *Theoret. Appl. Climatol*. 149, 863–882. doi: 10.1007/s00704-022-04035-2

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al. (2019). Seas5: the new ecmwf seasonal forecast system. *Geosci. Model Dev*. 12, 1087–1117. doi: 10.5194/gmd-12-1087-2019

Lorenz, C., Portele, T. C., Laux, P., and Kunstmann, H. (2021). Bias-corrected and spatially disaggregated seasonal forecasts: a long-term reference forecast product for the water sector in semi-arid regions. *Earth Syst. Sci. Data* 13, 1106–1114. doi: 10.5194/essd-13-2701-2021

Ma, Y., Sun, J., Dong, T., Yu, W., and Dong, W. (2023). More profound impact of cp enso on australian spring rainfall in recent decades. *Climate Dyn*. 60, 3065–3079. doi: 10.1007/s00382-022-06485-w

Madrigal, J., Solera, A., Suárez-Almiñana, S., Parades-Arquiola, J., Andreu, J., and Sánchet-Quispe, S. T. (2018). Skill assessment of a seasonal forecast model to predict drought events for water resource systems. *J. Hydrol*. 564, 574–587. doi: 10.1016/j.jhydrol.2018.07.046

Malekian, R., Sharifi, A., Garmdareh, S. E. H., Veysi, S., Farooque, A. A., Randhawa, G. S., et al. (2026). Era5-land-based drought assessment using spi and spei to support smart agriculture in Prince Edward Island, Canada. *Smart Agric. Technol*. 14:102077. doi: 10.1016/j.atech.2026.102077

Manzanas, R., Torralba, V., Lledó, L., and Bretonnière, P. (2022). On the reliability of global seasonal forecasts: sensitivity to ensemble size, hindcast length and region definition. *Geophys. Res. Lett*. 49:e2021GL094662. doi: 10.1029/2021GL094662

Marzban, C. (2004). The roc curve and the area under it as performance measures. *Weather Forecast*. 19, 1106–1114. doi: 10.1175/825.1

McKee, T. B., Doesken, N. J., and Kleist, J. (1993). "The relationship of drought frequency and duration to timescales," in *Eighth conference on applied climatology* (Boston, MA: American Meteorological Society (AMS)).

Mittermaier, M., Roberts, N., and Thompson, S. A. (2011). A long-term assessment of precipitation forecast skill using the fractions skill score. *Meteorol. Applic*. 20, 176–186. doi: 10.1002/met.296

Murphy, A. H. (1977). The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Monthly Weather Rev*. 105, 803–816. doi: 10.1175/1520-0493(1977)105<0803:TVOCCA>2.0.CO;2

Necker, T., Wolfgruber, L., Kugler, L., Weissmann, M., Dorninger, M., and Serafin, S. (2024). The fractions skill score for ensemble forecast verifcation. *Quat. J. R. Meteorol. Soc*. 150, 4457–4477. doi: 10.1002/qj.4824

Nicholson, S. E. (2017). Climate and climatic variability of rainfall over Eastern Africa. *Rev. Geophys*. 55, 590–635. doi: 10.1002/2016RG000544

NOAA (2024). *Monthly Global Drought Narrative for May 2024*. Silver Spring: NOAA.

NOAA (2026). *Niño anom 3.4 index using noaa ersst v6 from ncei*. Silver Spring: NOAA.

Nwayor, I. J., and Robeson, S. M. (2024). Exploring the relationship between spi and spei in a warming world. *Theoret. Appl. Climatol*. 155, 2559–2569. doi: 10.1007/s00704-023-04764-y

OECD (2025). *Global Drought Outlook. Trends, Impacts and Policies to Adapt to a Drier World*. Paris: OECD. doi: 10.1787/d492583a-en

Owens, R. G., and Hewsom, T. D. (2018). *Ecmwf Forecast User Guide*. Reading.

Palmer, T., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Délécluse, P., et al. (2004). Development of a european multimodel ensemble system for seasonal-to-interannual prediction (demeter). *Bull. Am. Meteorol. Soc*. 85, 853–872. doi: 10.1175/BAMS-85-6-853

Portele, T. C., Lorenz, C., Dibrani, B., Laux, P., Bliefernicht, J., and Kunstmann, H. (2021). Seasonal forecasts offer economic benefit for hydrological decision making in semi-arid regions. *Nat. Sci. Rep*. 11:10581. doi: 10.1038/s41598-021-89564-y

Pozo.Vázquez, D., Esteban-Parra, M., Rodrigo, F., and Castro-Díez, Y. (2001). A study of nao variability and its possible non-linear influences on European surface temperature. *Climate Dyn*. 17, 701–715. doi: 10.1007/s003820000137

Rayner, N., Parker, D., Horton, E., Folland, C., Alexander, L., Rowell, D., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res*. 108:4407. doi: 10.1029/2002JD002670

Richardson, D., Fowler, H. J., Kilsby, C. G., Neal, R., and Dankers, R. (2020). Improving sub-seasonal forecast skill of meteorological drought: a weather pattern approach. *Nat. Haz. Earth Syst. Sci*. 20, 107–124. doi: 10.5194/nhess-20-107-2020

Richardson, D. S. (2000). Skill and relative economic value of the ecmwf ensemble prediction system. *Q. J. R. Meteorol. Soc*. 145, 210–231. doi: 10.1002/qj.49712656313

Shyrokaya, A., Pappenberger, F., Messori, G., Pechlivanidis, I., Cloke, H., and Di Baldassarre, G. (2025). How good is my drought index? Evaluating predictability and ability to estimate impacts across europe. *Environ. Res. Lett*. 20:034051. doi: 10.1088/1748-9326/adb869

Skok, G., and Roberts, N. (2016). Analysis of fractions skill score properties for random precipitation fields and ecmwf forecasts. *Q. J. R. Meteorol. Soc*. 142, 2599–2610. doi: 10.1002/qj.2849

Smith, G. C., Allard, R., Babin, M., Bertino, L., Chevallier, M., Corlett, G., et al. (2019). Polar ocean observations: a critical gap in the observing system and its effect on environmental predictions from hours to a season. *Front. Mar. Sci*. 6:429. doi: 10.3389/fmars.2019.00429

Suárez-Almiñana, S., Andreu, J., Solera, A., and Madrigal, J. (2022). Integrating seasonal forecasts into real-time drought management: Júcar River Basin case study. *Int. J. Disaster Risk Reduct*. 70:102777. doi: 10.1016/j.ijdrr.2021.102777

Sutanto, S. J., van der Weert, M., Blauhut, V., and Van Lanen, H. A. J. (2020). Skill of large-scale seasonal drought impact forecasts. *Nat. Haz. Earth Syst. Sci*. 20, 1595–1608. doi: 10.5194/nhess-20-1595-2020

Tanessong, R. S., Fotso-Nguemo, T. C., Kaissassou, S., Guenang, G. M., Komkoua Mbienda, A. J., Djiotang Tchotchou, L. A., et al. (2024). Climate forecast skill and teleconnections on seasonal time scales over central africa based on the north american multi-model ensemble (nmme). *Meteorol. Atmos. Phys*. 136:19. doi: 10.1007/s00703-024-01018-y

Tchinda, A. F., Nonki, R. M., Dickmu, P. L., Mamadou, O., and Tanessong, S. R. (2026). Assessment of nmme and seas5 forecasts for West African Jjas rainfall. *Pure Appl. Geophys*. 183, 1457–1482. doi: 10.1007/s00024-025-03887-9

Tóth, Z., Talagrand, O., Candille, G., and Zhu, Y. (2003). "Probability and ensemble forecasts," in *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, eds. I. T. Jolliffe and D. B. Stephenson (Chichester: John Wiley & Sons), 137–163.

Trenberth, K. E., and Stepaniak, D. P. (2003). Covariability of components of poleward atmospheric energy transports on seasonal and interannual timescales. *J. Climate* 16, 3691–3705. doi: 10.1175/1520-0442(2003)016<3691:COCOPA>2.0.CO;2

Vicente-Serrano, S. M., Begueria, S., and López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *J. Clim*. 23, 1696–1718. doi: 10.1175/2009JCLI2909.1

Vicente-Serrano, S. M., Domínguez-Castro, F., Reig, F., Tomás-Burguera, M., Peña-Angulo, D., Latorre, B., et al. (2023). A global drought monitoring system and dataset based on era5 reanalysis: a focus on crop-growing regions. *Geosci. Data J*. 10, 505–518. doi: 10.1002/gdj3.178

Vicente-Serrano, S. M., Peña-Angulo, D., Beguería, S., Domínguez-Castro, F., Tomás-Burguera, M., Noguera, I., et al. (2022). Global drought trends and future projections. *Philos. Trans. R. Soc. A* 380:20210285. doi: 10.1098/rsta.2021.0285

Vorobevskii, I., Luong, T. T., and Kronenberg, R. (2024). Seasonal forecasting of local-scale soil moisture droughts with global brook90: a case study of the european drought of 2018. *Nat. Haz. Earth Syst. Sci*. 24, 681–697. doi: 10.5194/nhess-24-681-2024

Weber, J. N. (2026). *Extreme Continental Droughts Based on ERA5 SPEI (1981 - 2024)*. Dataset, version 1.0, Under Review. Karlsruhe: RADAR4KIT. Available online at: https://radar.kit.edu/radar/en/dataset/3yx9nft8bm6t9svh. doi: 10.35097/3yx9nft8bm6t9svh

Weber, J. N., Lorenz, C., Portele, T. C., and Kunstmann, H. (2023). Seasonal forecasts for germany: enhancing the predictive capability of global seas5 ensemble forecasts using bias correction. *Hydrol. Wasserbewirtschaftung* 67, 90–109. doi: 10.5194/egusphere-egu23-14589

Weber, J. N., Lorenz, C., Schober, T. C., and Kunstmann, H. (2026a). *Global Bias-Corrected Seasonal Forecast (seas5-bcsd) Archive 1981 to 2024 for Precipitation and Temperature*. Garmisch-Partenkirchen: Dataset, version 1.0.

Weber, J. N., Lorenz, C., Schober, T. C., Wiegels, R., and Kunstmann, H. (2026b). Seas5-bcsd: a bias-corrected and downscaled global seasonal forecast reference dataset for 1981–2024. *Earth Sys. Sci. Data*. 18, 5259–5282. doi: 10.5194/essd-18-5259-2026

Weigel, A. P., Liniger, M. A., and Appenzeller, C. (2007). The discrete brier and ranked probability skill scores. *Monthly Weather Rev*. 135, 118–124. doi: 10.1175/MWR3280.1

Wilhite, D. A., and Glantz, M. H. (1985). Understanding: the drought phenomenon: the role of definitions. *Water Int*. 10, 111–120. doi: 10.1080/02508068508686328

Williams, N., Scaife, A., and Screen, J. (2023). Underpredicted enso teleconnections in seasonal forecasts. *Geophys. Res. Lett*. 50:e2022GL101689. doi: 10.1029/2022GL101689

WMO (2021). *Updated 30-Year Reference Period Reflects Changing Climate*. Geneva: WMO.

Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P. (2002). Long-range experimental hydrologic forecasting for the Eastern United States. *J. Geophys. Res. Atmos*. 107, ACL 6-1-ACL 6-15. doi: 10.1029/2001JD000659

Yin, G., Yoshikane, T., Kaneko, R., and Yoshimura, K. (2023). Improving global subseasonal to seasonal precipitation forecasts using a support vector machine-based method. *J. Geophys. Res.: Atmos*. 128:17. doi: 10.1029/2023JD038929

Yuan, X., and Wood, E. F. (2013). Multimodel seasonal forecasting of global drought onset. *Geophys. Res. Lett*. 40, 4900–4905. doi: 10.1002/grl.50949