https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

Open Access Earth System Science Data Discussions

SEAS5-BCSD: A bias-corrected and downscaled global seasonal forecast reference dataset for 1981–2024

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

[1]Karlsruhe Institute of Technology, Campus Alpin, Institute of Meteorology and Climate Research IMKIFU, 82467 Garmisch-Partenkirchen, Germany
[2]University of Augsburg, Institute of Geography, 86159 Augsburg, Germany
[3]University of Augsburg, Center for Climate Resilience, 86159 Augsburg, Germany

Correspondence: Jan N. Weber (jan.weber@kit.edu)

Abstract.

Seasonal forecasts offer valuable information on upcoming conditions for the water, energy, and agricultural sectors. However, applications of raw data from global seasonal forecasts are limited, as they can show substantial biases and temporal drifts. In this study, we present a bias-corrected and downscaled global seasonal forecast reference dataset for precipitation and 2-m temperature for 1981 till 2024, provided at monthly resolution. We achieve this with the Bias Correction and Spatial Disaggregation (BCSD) method, combining ECMWF SEAS5 seasonal forecasts with ERA5 reanalysis data. The resulting post-processed product is a spatially refined and improved dataset for a wide range of seasonal applications in the water, energy and agricultural sectors. Unlike existing products, the dataset provides bias-corrected forecasts for all SEAS5 ensemble members over the full hindcast period (1981-2016) and even beyond (till 2024). The dataset spans all global land areas at 0.25° spatial resolution with a forecast lead time of up to seven months. It comprises 25 ensemble members for the period 1981-2016 and 51 ensemble members for 2017-2024. To assess probabilistic forecast quality, we conduct a comprehensive performance evaluation, using the Brier Skill Score (BSS) and the Continuous Ranked Probability Skill Score (CRPSS). The BCSD-corrected temperature forecasts outperform climatology across nearly all regions and lead times, with highest skill in flat and warm regions. Precipitation skill is highest in the tropics and humid regions. Semi-arid areas show solid skill during the rainy season but reduced performance in dry months. This skillful global seasonal forecast reference dataset can now be explored by the community for subsequent forecast evaluation, drought prediction studies, and water resource management applications. The BCSD-corrected seasonal forecast dataset is publicly available as NetCDF data under a Creative Commons Attribution 4.0 International License (CC BY 4.0) at the World Data Center for Climate (WDCC; DOI: 10.26050/WDCC/SEAS5-BCSD, Weber et al., 2026).

1 Introduction

Droughts, prolonged heat waves, heavy rainfall, and large-scale floods: recent years have shown that improved adaptation to hydrometeorological extremes needs to be achieved in many regions worldwide. These extreme events occur with increasing frequency and intensity due to climate change (Di Capua and Rahmstorf, 2023). Among them, droughts pose a particularly se-

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

vere threat to society due to their extensive spatial scale and prolonged duration compared to floods (United Nations Convention to Combat Desertification, 2022).

Knowledge of hydrometeorological wet/dry or hot/cold anomalies months ahead allows improved preparedness for stakeholders, such as water managers and large-scale farmers, to mitigate the impacts of extreme periods. While floods are often associated with short warning times of hours to days due to their high spatial complexity (Cloke and Pappenberger, 2009), droughts and heatwaves need to be predicted well over a month in advance due to their large-scale and long-term nature. Extended predictability allows stakeholders to implement effective countermeasures, which can be scaled according to the predicted severity of the event (Portele et al., 2021).

Seasonal forecasts provide valuable insights into upcoming conditions, offering a forecast horizon of up to twelve months. The drivers of predictability in seasonal forecasting systems are primarily twofold. The initial atmospheric state at the beginning of the simulation plays a key role in the first few weeks of the forecast, as it contains essential information about circulation patterns. For longer forecast durations, predictability is largely governed by teleconnections, which are primarily influenced by sea surface temperature (SST) variations (Materia et al., 2014). Despite these sources of skill, the inherently chaotic nature of the atmosphere limits the accuracy of forecasts several months in advance (e.g., Vitart et al., 2017). Multiple national and international meteorological services, including the German Weather Service (DWD), the UK Met Office, the Australian Bureau of Meteorology, and the European Centre for Medium-Range Weather Forecasts (ECMWF), operate seasonal forecasting systems (e.g., Fröhlich et al., 2021; Johnson et al., 2019; Hudson et al., 2017; MacLachlan et al., 2014). However, the raw output of these models often lacks realism in terms of magnitude and variability (Yin et al., 2023, Figure 1), making it unsuitable for direct application in downstream models such as hydrological or agricultural simulations. To address this, a range of bias correction methods have been developed and applied to seasonal forecast datasets, resulting in a combination of reference and model data. Bias-corrected seasonal forecasting datasets are available at both regional and global scales, but most provide only a limited number of ensemble members (Beck et al., 2022, e.g.,). Extreme climatic conditions, such as severe large-scale droughts, are inherently difficult to capture due to their rare occurrence. In many cases, only a small fraction of ensemble members represent such extremes. Therefore, robust probabilistic assessment of climate risks requires access to the full ensemble distribution. We provide bias-corrected forecasts for all 25 ensemble members of the high-resolution (36 km) SEAS5 system, covering the period from 1981 to 2016 and all 51 ensemble members from 2017 onward. Rather than serving as a competing product, this reference dataset is intended to complement existing products and to support a broader range of applications and analyses. While numerous studies have evaluated forecast skill at regional scales (e.g., Crespi et al., 2021; Wang et al., 2019), comprehensive global analysis of bias corrected seasonal forecasts is still sparse. This study aims to close this gap by presenting a bias-corrected global seasonal forecast dataset based on ECMWF's SEAS5 system, designed to serve as a reference dataset. In addition, we provide a detailed global assessment of the dataset's strengths and limitations using a suite of established forecast evaluation metrics.

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

Open Access Earth System Science Data Discussions

2 Methods

2.1 Data

Our dataset is based on the high-resolution version of the SEAS5 Seasonal Forecasting System of the European Centre for Medium-Range Weather Forecasts (ECMWF). SEAS5 provides global coverage at approximately 36 km resolution, with daily timesteps and 91 atmospheric layers (Johnson et al., 2019). The forecast horizon lies at 215 days. SEAS5 is released on the 5th day of each month, minimizing the gap between the start of a new month and the availability of a new forecast.

SEAS5 offers a wide range of variables, of which we focus on total precipitation (tp) and mean 2-m air temperature (t2m). A key advantage of SEAS5 is the availability of historical reforecasts for 1981–2016, which allows the construction of cumulative distribution functions (CDFs) for statistical bias correction using an extensive data pool. This enhances the detection of extreme events by expanding the range of possible outcomes (this point will be discussed in detail later in the context of the BCSD method). For the analysis in this article, we use the 1981–2016 hindcast period with cross validation. SEAS5 reforecasts during this period are produced with 25 ensemble members, whereas the operational forecast system generates 51 members for real-time forecasts; including the operational period in the statistical analysis would therefore introduce an inhomogeneity in ensemble size (Johnson et al., 2019). Reducing the operational ensemble to 25 members to match the hindcast set is not advisable due to the chaotic nature of the stochastic perturbations in SEAS5 and the potential loss of valuable spread information inherent to the system. Using only the consistent hindcast ensemble ensures that skill metrics reflect the forecast system's performance and not changes in ensemble sampling characteristics.

As a reference product, we selected ECMWF's ERA5 reanalysis. ERA5 covers the period from 1940 to three days before the current date, fully encompassing both the reference and validation periods. It has a spatial resolution of $0.25° \times 0.25°$ and an hourly temporal resolution. ERA5 has been proven to be one of the most reliable global reanalysis datasets, covering all required variables for our study (Hersbach and co authors, 2020). It has been extensively validated for selected variables (e.g., Lavers et al., 2022), for global applications (Nogueira, 2020) regional studies (Bandhauer et al., 2021), etc. That being said, we are well aware that ERA5 has its' limitations (see Appendix A1). Nevertheless, being one of the most commonly used global references for precipitation and temperature, we are convinced, that ERA5 is an appropriate and representative reference for evaluating the forecast skill of the SEAS5-BCSD dataset. To ensure comparability with SEAS5, ERA5 data are aggregated to daily sums or means, respectively. We refrained from using ERA5-Land due to its delayed availability, typically two to three months after the release of the standard ERA5 dataset. Additionally, ERA5 offers the same set of ground-level variables as SEAS5, facilitating direct comparison.

2.2 Bias correction and spatial disaggregation

Several studies (e.g., Weber et al., 2023; Lorenz et al., 2021; Portele et al., 2021) highlight the limitations of uncorrected SEAS5 data and emphasize the necessity of bias correction. Lorenz et al. (2021) demonstrated significant performance improvements for semi-arid regions when applying a statistical bias correction. Building on this, Portele et al. (2021) demonstrated across seven drought-prone regions worldwide that using normalized seasonal forecasts can yield substantial economic benefits,

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

including a real-world case in which approximately 16 million USD in potential losses at a hydropower reservoir in Sudan could have been avoided. Using a similar statistical correction approach, Weber et al. (2023) identified forecast skill over Germany, a region with a more complex seasonal climate (reference needed) and thus more challenging predictability than semi-arid regions. These findings motivated us to develop a global bias correction approach. The primary challenge is to design a computationally efficient method that provides bias-corrected forecasts within hours of downloading the raw SEAS5 data, ensuring the timely availability of the valuable first lead month.

For statistical bias correction of original and raw SEAS5, we here employ the Bias Correction and Spatial Disaggregation (BCSD, Wood et al., 2004) method to generate the new dataset SEAS5-BCSD. The resulting dataset is therefore not a re-distribution of SEAS5 model output but a statistically post-processed reference product that combines information from both SEAS5 forecasts and ERA5 reanalysis data. BCSD consists of two main components: (i) a bias correction step that applies quantile-quantile mapping to harmonize the forecast with the reference dataset and (ii) a spatial downscaling step to refine the resolution of the bias-corrected data.

We modified the sequence of Bias Correction (BC) and Spatial Disaggregation (SD) following the methodology outlined in Lorenz et al. (2021) for avoiding numerical issues. The SD process consists of a linear interpolation of the relative differences between SEAS5 and a coarse-grid ERA5 climatology. Bias correction is subsequently applied on each forecast day using Empirical Quantile Mapping (EQM), which is well suited for precipitation extremes (Golian and Murphy, 2022). This method constructs a cumulative distribution function (CDF) for each pixel separately for both the reference dataset and the seasonal forecast over the reference period. Each daily CDF is based on all corresponding days from the 36-year reference period and hence an empirical distribution following Boé et al. (2007). To smooth the CDF, a fixed-width time window of 15 days centered on the target date is utilized, incorporating values from preceding and succeeding days. This approach helps to capture more extreme events. Consequently, the CDF becomes better resolved and extended at both its upper and lower ends. However, if the time window is too large, it may incorporate data from the wrong season, introducing an own bias when correcting the target date. Overall, each daily ERA5 CDF is constructed from 1116 data points per grid cell. For SEAS5, the inclusion of ensemble members expands the dataset to a total of 27,900 values per forecast day. Due to seasonal forecast drifts varying by issue month, SEAS5 CDFs are computed individually for each issue months 215 days.

Applying bias correction on a global scale presents computational challenges, particularly concerning processing time. To accelerate calculations, the CDFs are precomputed and stored rather than generated dynamically. While this significantly enhances computational efficiency, it necessitates substantial storage capacity. To manage storage requirements, the CDFs are downsampled to 200 data points, reducing the total storage footprint significantly.

For bias correction of new forecasts, each forecasted value is first mapped onto the SEAS5 CDF to determine its probability. This probability is then used to obtain the corrected value by mapping it onto the ERA5 CDF. However, if a forecasted value exceeds the historical maximum or falls below the historical minimum, the CDF approach alone is insufficient due to the absence of a corresponding probability. Several methods exist to address this issue, including fitting a statistical distribution such as the Generalized Pareto Distribution (Volosciuk et al., 2017), applying linear extrapolation (Gonzalez-Aparicio and Hidalgo, 2011), or implementing a hybrid approach that includes clipping extreme values (Holthuijzen et al., 2022). In this study, precipitation

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

extremes are handled using a linear extrapolation/scaling approach, while temperature extremes are corrected with an additive delta approach.

Additionally, special handling is required for cases where precipitation values reach 0 mm in either the ERA5 or SEAS5 CDFs. If the dry-day probability in ERA5 exceeds that in SEAS5, all forecasted values below the ERA5 probability threshold are set to 0 mm. Conversely, if the SEAS5 dry-day probability surpasses that of ERA5, for each new forecast of 0 mm a random sample is drawn from the ERA5 CDF below the probability value, ensuring consistency in dry-day frequency between datasets.

Accurate daily forecasts over extended lead times exceeding two hundred days are not realistic due to inherent limitations in predictability (Krishnamurthy, 2019). Therefore, forecasts were aggregated to a monthly resolution to enhance robustness and interpretability. In the following, the aggregated time steps are referred to as "Lead Month" 0 through 7, with the initialization month denoted as the "issue month". Lead Month 0 comprises forecast days 1 to 30 (depending on the respective calendar month), representing one month into the future. Although the bias correction method is applied globally across all grid cells, oceanic regions and Antarctica were subsequently masked out due to their limited relevance for most land-based hydrological and agricultural applications.

2.3 Continuous Ranked Probability Skill Score (CRPSS)

To estimate forecast skill, we employ various performance metrics to capture a plethora of characteristics of the bias-corrected forecasts. The evaluation of forecast accuracy can be conducted using the Continuous Ranked Probability Score (CRPS), following Hersbach (2000). For each pixel and each time step, the CRPS quantifies the deviation of forecasted values from the observed values, in this case, the ERA5 reanalysis dataset. Due to the probabilistic nature of SEAS5 forecasts, a simple coefficient of determination (R^2) is insufficient, as it does not account for ensemble members. Instead, the CRPS is used, defined as:

$$\mathrm{CRPS}(F, x) = \int_{-\infty}^{\infty} \left(F(y) - H(y - x)\right)^2 dy, \tag{1}$$

where $H(y-x)$ is the Heaviside step function, $F(y)$ represents the cumulative distribution function (CDF) of the probabilistic forecast values, and x is the observed value. The CRPS thus quantifies the area between the Heaviside function, which transitions from 0 to 1 at the observed value, and the forecast CDF. This computation is performed individually for each pixel and each time step. The CRPS values range from 0 to ∞, where a value of 0 indicates a perfect forecast.

To facilitate the comparison of CRPS values across different forecasting approaches, the Continuous Ranked Probability Skill Score (CRPSS) is utilized:

$$\mathrm{CRPSS} = 1 - \frac{\mathrm{CRPS}_{\mathrm{fcst}}}{\mathrm{CRPS}_{\mathrm{ref}}}. \tag{2}$$

Here, $\mathrm{CRPS}_{\mathrm{fcst}}$ is the CRPS of the forecast under evaluation, while $\mathrm{CRPS}_{\mathrm{ref}}$ refers to the CRPS of a reference dataset, such as a climatology derived from a reanalysis dataset like ERA5 or an alternative forecast (e.g., uncorrected data). The CRPSS

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

ranges from 1 to $-\infty$, where a value of 1 represents a perfect forecast, 0 signifies performance equivalent to the reference, and negative values indicate performance worse than climatology. However, in a probabilistic forecasting system like SEAS5, achieving a CRPSS of 1 is not possible, as ensemble spread is a desirable feature. Maintaining the SEAS5 ensemble spread results in a maximum achievable CRPSS of approximately 0.6–0.8 for precipitation, depending on the region.

2.4 Brier Skill Score (BSS)

In addition, the Brier score is applied to the data. The score, introduced by Brier (1950), can be focused on a specific event, e.g., a forecast of precipitation lower than the 33rd percentile. Therefore, the Brier Score is useful for the evaluation of commonly used probability-based forecast maps such as tercile maps and is also suitable for assessing the performance of extreme forecasts. We calculate the Brier Score following Brier (1950) as

$$\mathrm{BS} = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2 \tag{3}$$

where N is the number of observations, f_t is the forecast at time t, indicating whether the event is forecasted to occur ($f_t = 1$) or not ($f_t = 0$). o_t is the corresponding observation, in our case derived from ERA5 data, similarly categorized as either 0 or 1. Analogous to the CRPSS, the Brier Skill Score (BSS) can be calculated to compare the Brier Score of two forecasts directly with each other:

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS_{fcst}}}{\mathrm{BS_{ref}}}. \tag{4}$$

where $\mathrm{BS_{fcst}}$ is the Brier Score of the forecast under evaluation, and $\mathrm{BS_{ref}}$ is the Brier Score of the reference forecast. The resulting values are interpreted in the same way as the CRPSS, for positive values the forecast outperforms the reference, while negative values indicate that the skill favors the reference. Grid cells with 0 mm of precipitation as the lowest threshold are excluded from the analysis. For a forecasted value of exactly 0 mm, it would be ambiguous whether this value should be classified into the driest or the second-driest category, thus introducing potential misclassification in threshold-based verification.

2.5 Receiver Operating Characteristics (ROC) Curve

The ROC curve is a diagnostic tool used to assess a forecasting system's ability to discriminate between event occurrence and non-occurrence by plotting the true positive rate against the false positive rate across varying probability thresholds (Marzban, 2004). The true positive rate (TPR) and false positive rate (FPR) are defined as:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{5}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}, \tag{6}$$

where TP, FP, FN, and TN denote true positives, false positives, false negatives, and true negatives, respectively. By varying the probability threshold used to define an event, a continuous ROC curve is obtained that summarizes the forecast's discrimination performance. For the SEAS5–BCSD forecasts, 26 probability levels are considered, corresponding to the 25 ensemble members

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

plus the case of non-fulfillment. The curve thus reflects how often an event (e.g., dry or wet condition) is correctly forecast relative to how often it is falsely predicted. An ideal forecasting system would be located in the upper-left corner of the ROC plot, while a system with no skill, equivalent to climatology or random guessing, would lie along the 1:1 diagonal (dashed line) (Tóth et al., 2003). Note that the ROC curve evaluates discrimination ability only and does not convey information on bias or event frequency.

3 Results

The following section provides a global overview of skill scores across variables, lead times, and correction stages, before delving into more detailed analyses of spatial, climatic and seasonal patterns.

3.1 Global overview of performance

In the first lead month (approx. 30–60 days after initialization) the forecast already exhibits substantial bias for both variables on a global scale when using uncorrected forecasts (Figure 1). As illustrated in the left panel of Figure 1, the most pronounced loss of skill for precipitation occurs in arid and semi-arid regions. In terms of absolute precipitation amounts, however, the largest biases are found in tropical regions, with missing precipitation over the Amazon basin and an overestimation in the Congo basin. For temperature, most land areas exhibit a cold bias, which is particularly pronounced over the Sahara and in mountainous regions. For both variables, distinct wave-like patterns are visible, especially along major mountain ranges. These patterns may originate from deficiencies in the representation of orographic gravity wave drag or, more systematically, from spectral parameterizations in the original model output that are transferred to the Gaussian grid during post-processing. The presence of these biases results in negative global skill scores for the uncorrected forecasts (Table A1). With increasing lead time, systematic drift becomes even more pronounced (Table A1), further degrading forecast performance relative to the climatology. This highlights the urgent need for bias correction to adjust forecasted values toward more realistic levels.

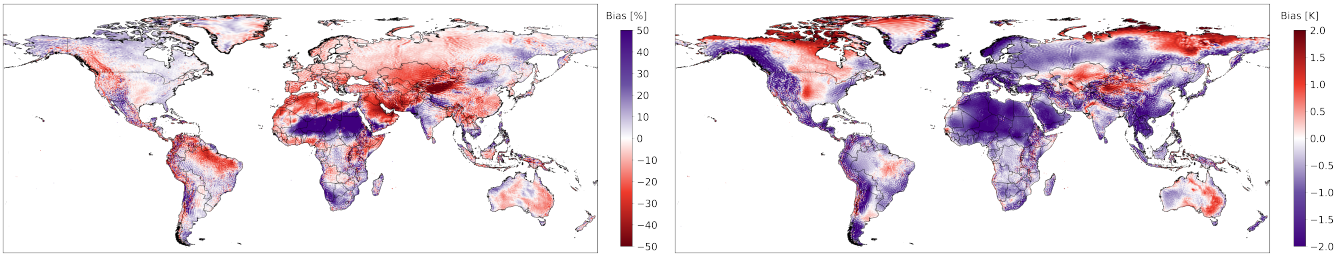


Figure 1. Bias of uncorrected SEAS5 (“raw”) forecasts Lead Month 1 relative to ERA5 for precipitation (left) and temperature (right). Precipitation bias is shown as relative deviation (%), and temperature bias as absolute deviation (K).

To quantify these skill levels more systematically across space and lead time, we next examine the CRPSS values aggregated by lead month and region. Figure 2 provides a detailed global assessment of the BCSD forecasts using the CRPSS. Across

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

nearly all land regions, BCSD outperforms the uncorrected forecasts by a large margin, with particularly strong gains observed for temperature. For both variables, Lead Month 0 exhibits high skill, with values exceeding 0.15 globally for precipitation and 0.25 for temperature in many regions. Starting from Lead Month 1, a noticeable decline in skill is observed. In Lead Month 1 to 3, CRPSS values drop to lower levels globally. Nevertheless, 69.7 % of all grid cells still show skillful precipitation forecasts relative to the climatology, and 97.4 % for temperature. The CRPSS difference between BCSD-corrected and uncorrected forecasts (not shown) remains relatively stable across the subsequent lead months, indicating that all forecast horizons benefit similarly from the BCSD correction. The best performing regions are concentrated around the equator, while areas with lower skill include Central Russia for both variables in Central Russia, and the Sahara for precipitation. In Lead Months 4 to 6, the spatial pattern remains similar. As expected, skill decreases with increasing lead time, though many regions still retain positive skill values.

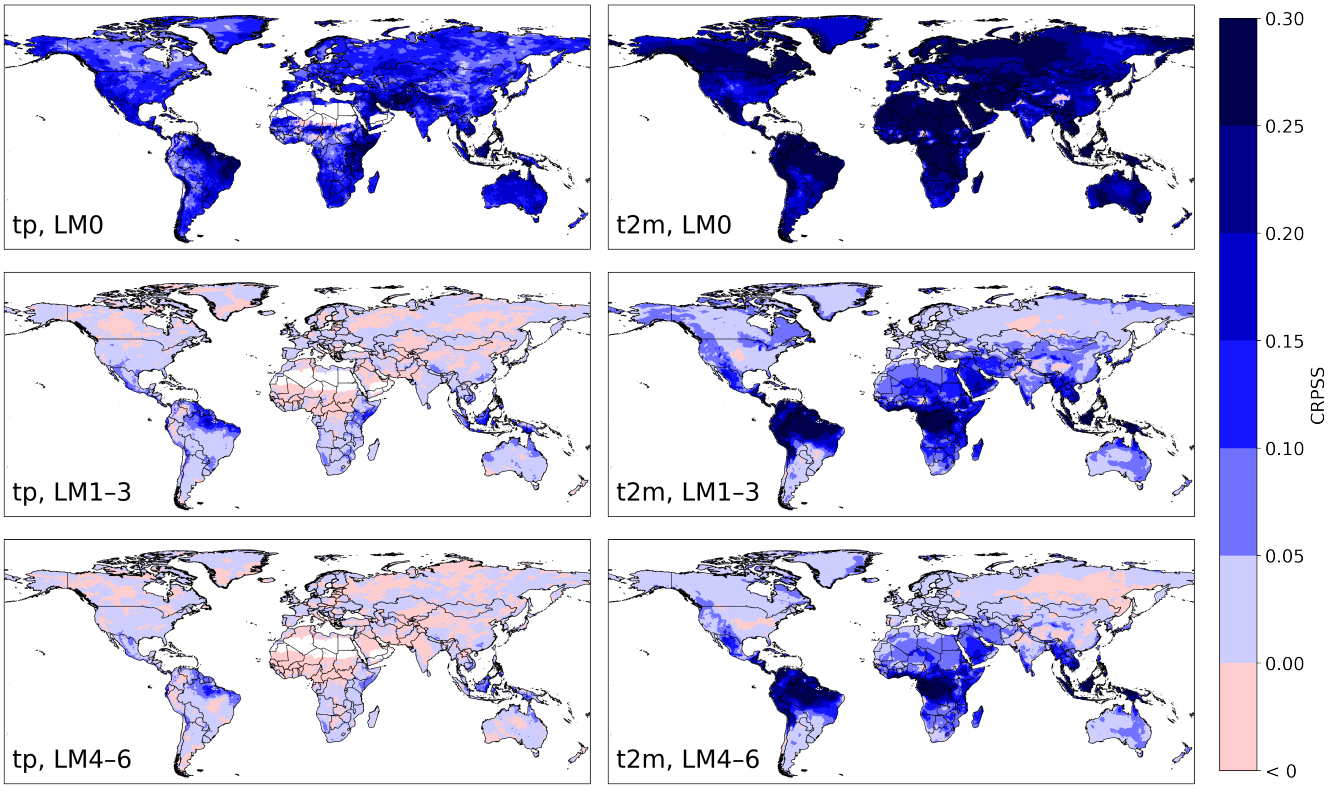


Figure 2. SEAS5-BCSD, Continuous Ranked Probability Skill Score relative to climatology for different lead months (LMs), shown for precipitation (left) and temperature (right). Lead Month 0 (top), Lead Months 1–3 (middle), and Lead Months 4–6 (bottom) are displayed. Blue colors indicate higher skill than climatology, with darker shades representing better performance. Dry grid cells are excluded.

Since skill in average conditions does not always translate to skill in extremes, we complement the CRPSS analysis with a categorical assessment based on probabilistic thresholds. Accordingly, Figure 3 provides a more detailed evaluation of Brier

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

Skill Score (BSS) performance across several drought and wetness thresholds. The figure includes the widely used tercile thresholds (top row), quintile thresholds (third row), and the 10th and 90th percentiles (second row), the latter serving as proxies for more extreme events. For the categories in-between, the middle tercile and quintile categories is shown in the bottom row. These "normal conditions" forecasts exhibit far less skill than the dry/wet categories with only a relatively weak BSS values evident even in Lead Month 0. In contrast, forecasts for dry and wet categories demonstrate significantly higher skill. Evaluating the percentage of land area with positive BSS values, the tercile approach achieves the highest coverage with 99.1 % (dry) and 99.4 % (wet), followed closely by the quintile approach with 96.8 % and 98.5 %, respectively. In terms of the magnitude of skill, the tercile-based forecasts also show higher mean BSS values overall. The main regional difference between the tercile and quintile approaches are concentrated in North-East Asia, where quintile skill is somewhat reduced. For the more extreme categories, skill is not symmetrically distributed between the wet and dry ends of the spectrum: while wet extremes are predicted with similar skill to the corresponding tercile and quintile categories, forecasts for dry extremes show reduced skill in several regions, particularly compared to their wet counterparts. For temperature (Figure A1), overall skill is higher, both in terms of the fraction of positively skilled grid cells and the magnitude of the BSS values. Only a few regions in the Himalayas display reduced skill in the tercile categories, along with negative skill in the quintile and extreme categories. In contrast to precipitation, the forecast performance for hot and cold anomalies is nearly symmetrical. However, as with precipitation, categories representing “near-normal” conditions show substantially reduced skill.

Since the forecasts in Lead Month 0 for abnormal dry conditions (below the 33rd percentile) perform best, a closer inspection of subsequent lead times is warranted. As expected, the spatial extent of regions with positive skill decreases with increasing lead time (Fig. 4). In most parts of the world, this decline appears gradual. A notable exception is Central Europe, where skill drops substantially in Lead Month 1 before improving again in later months. Regions exhibiting strong skill in LM0 (cf. Fig. 3), such as northern Brazil, Mexico, Indonesia, and southern Africa, generally maintain positive skill through LM6. In contrast, other areas like Australia and India show a marked reduction in predictive skill at longer lead times. High-latitude regions, including boreal Canada and Siberia, exhibit only weak signals throughout. When turning to above average temperature forecasts (see Figure A2), a stronger and more spatially consistent skill is evident, even at longer lead times such as LM6. The equatorial region again shows particularly strong performance, other than that no obvious spatial pattern arise. Overall, Brier Skill Score values for temperature substantially exceed those of precipitation, especially for lead times further in the future, underlining the comparatively higher predictability of temperature on seasonal scales.

3.2 Statistical consistency and predictive characteristics

To complement the regional analysis, a statistical evaluation of the dataset is provided. An essential requirement for the BCSD method is that it does not shift the distribution of the values into unrealistic territory. In particular, during the reference period, the statistical characteristics of the bias-corrected SEAS5 values should align closely with that of ERA5 due to the quantile mapping approach. Figure 5 illustrates this by showing the distribution of BCSD-corrected SEAS5 monthly precipitation values from 1981 to 2016. The bin thresholds are defined by ERA5 percentiles (5th, 10th, 15th, ..., 95th), resulting in bins that each contain 5 % of the ERA5 data. The SEAS5 data would also be uniformly distributed across these bins, although perfect

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

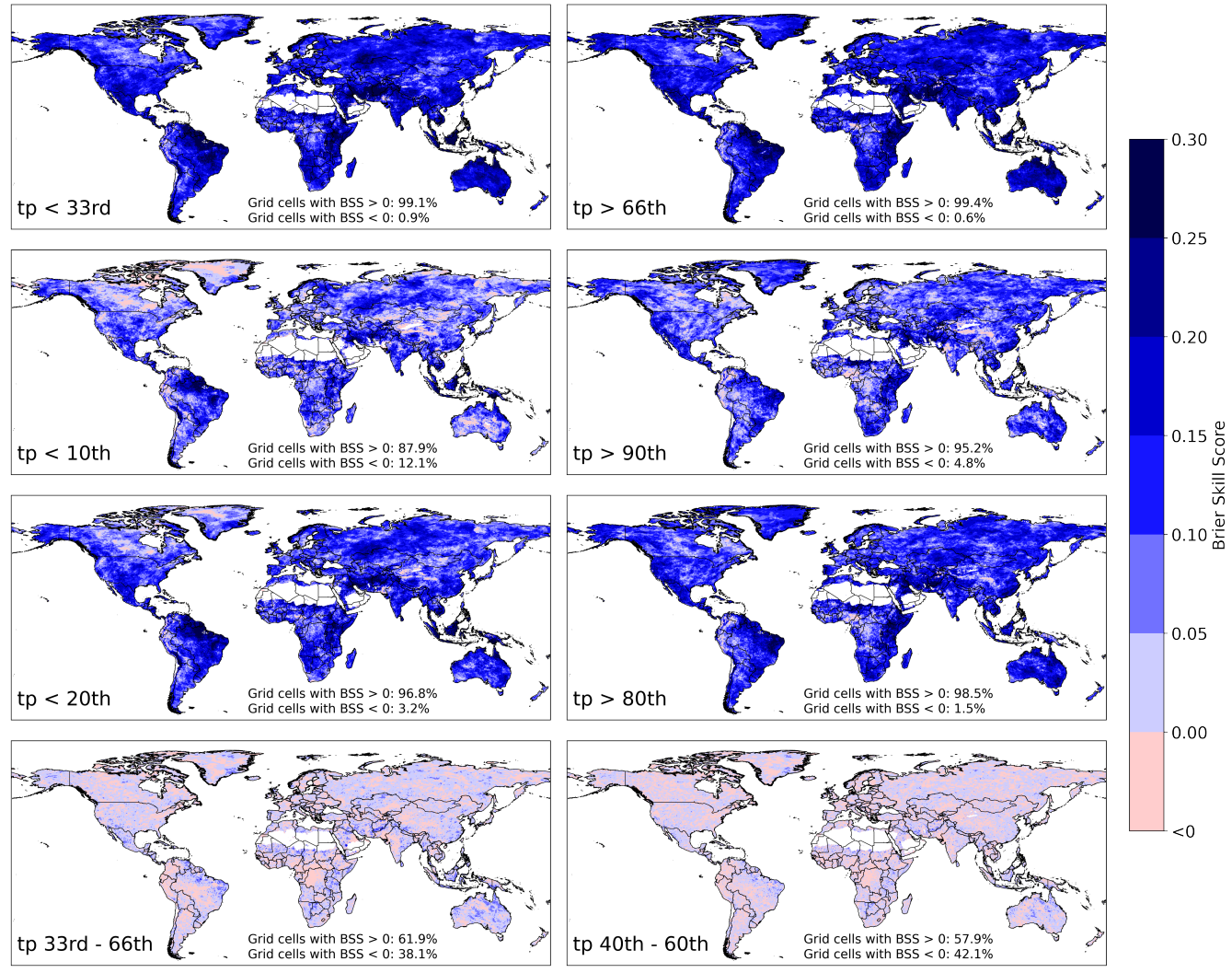


Figure 3. SEAS5-BCSD, Brier Skill Score (BSS) relative to climatology for precipitation at Lead Month 0. Different percentile thresholds are shown, indicated in the lower-left corner of each panel. Blue colors indicate higher skill than climatology, with darker shades representing better performance. Dry grid cells are excluded.

agreement is not expected because the bias correction is applied at the daily scale, while the analysis is based on monthly means. The histogram confirms a strong agreement between ERA5 and SEAS5, particularly the precipitation values above 0.5 mm/day are all in perfect proportion to the expected 5 %. In the very dry categories, a slight underrepresentation of the driest two bins is noted, with an overrepresentation in the adjacent (still dry) bins. This discrepancy becomes more pronounced in later lead months but remains minor in Lead Month 0. Additionally, a marginal underrepresentation is noted in the wettest bins, though this effect is negligible.

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

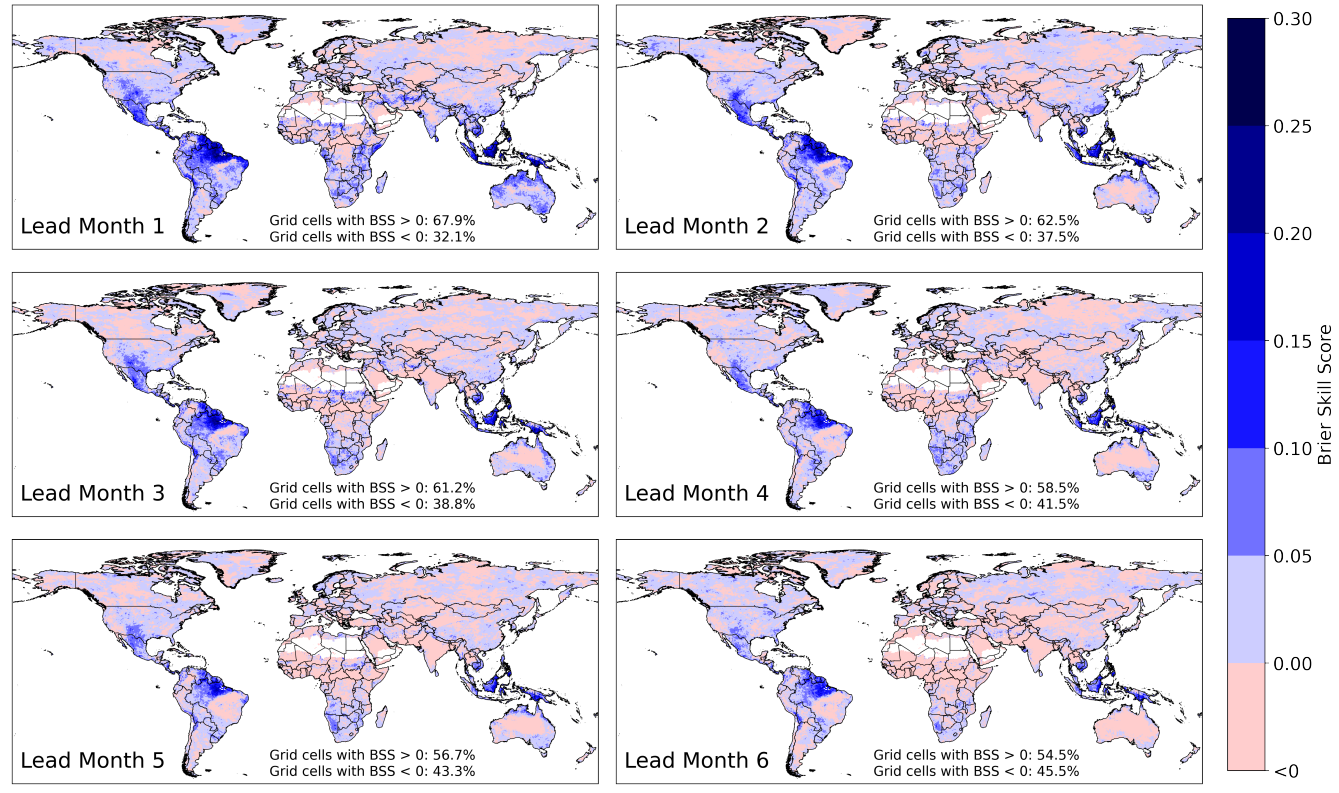


Figure 4. SEAS5-BCSD, Brier Skill Score (BSS) relative to climatology for precipitation, focusing on dry conditions below the 33rd percentile for Lead Months 1–6. Blue colors indicate higher skill than climatology, with darker shades representing better performance. Dry grid cells are excluded.

While the histogram confirms the consistency of the corrected forecast distribution, it does not reflect the forecasting system's ability to correctly anticipate event occurrence. To address this, we turn to the Receiver Operating Characteristics (ROC) analysis. Figure 6 shows values left of the 1:1 diagonal for all variables, lead months and percentile thresholds, the area under the ROC curve (AUC) exceeds 0.5, indicating positive skill relative to a climatological forecast. In Lead Month 0, skill levels are high across both temperature and precipitation. In Lead Month 1, forecast performance decreases, particularly for precipitation, but remains skillful. These findings are consistent with earlier assessments in Figure 4 and Figure 2. Temperature forecasts maintain robust discriminatory power, with ROC curves for Lead Month 1 still comparable to the precipitation performance in Lead Month 0. Among percentile categories, only marginal differences are observed; with the only exception being a slightly better performance of the dry categories in Lead Month 0 of precipitation, suggesting improved detectability of dry conditions at shorter lead times when the classification threshold is low.

Building on this evaluation, we investigate whether internal ensemble agreement correlates with forecast skill, providing further insights into forecast confidence. Figure 7 explores the role of ensemble agreement, a component already implicitly

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

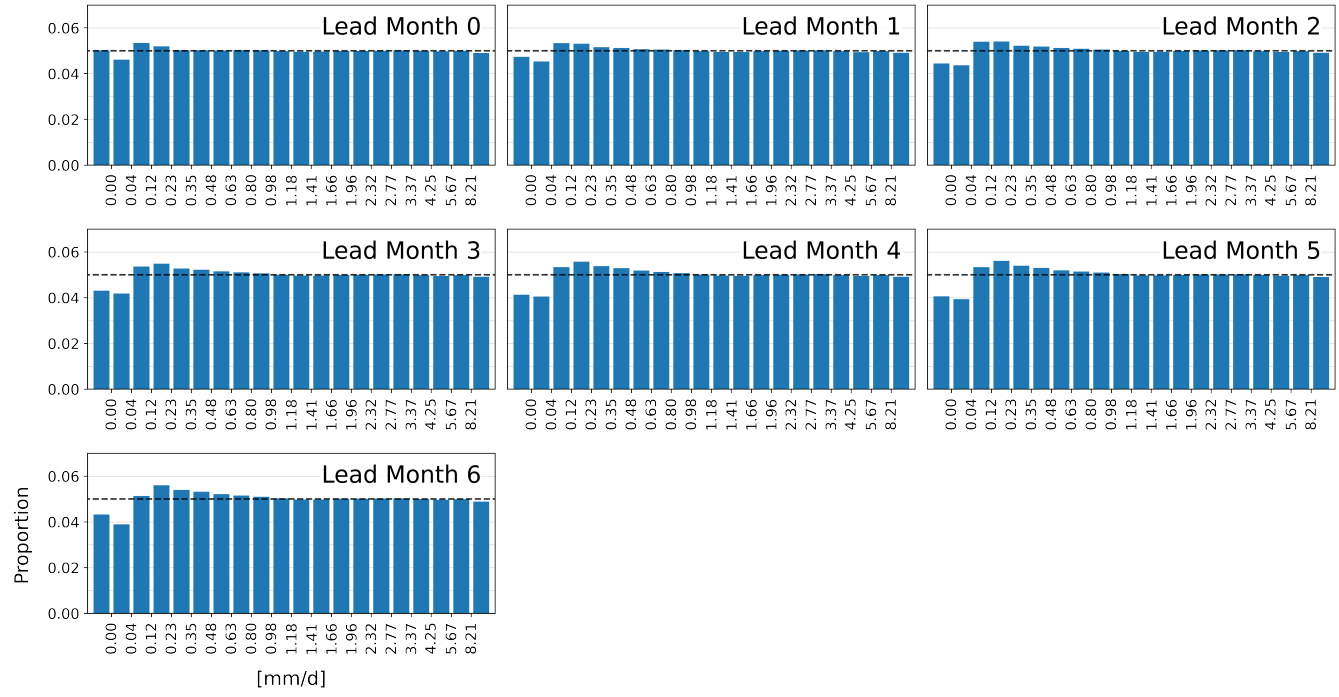


Figure 5. Histogram of BCSD-corrected SEAS5 monthly precipitation values for the period 1981–2016. Bin thresholds are defined by ERA5 percentiles (5th to 95th), such that each bin contains 5 % of the ERA5 data. A uniform distribution of SEAS5 values across bins would indicate perfect agreement with the ERA5 reference distribution.

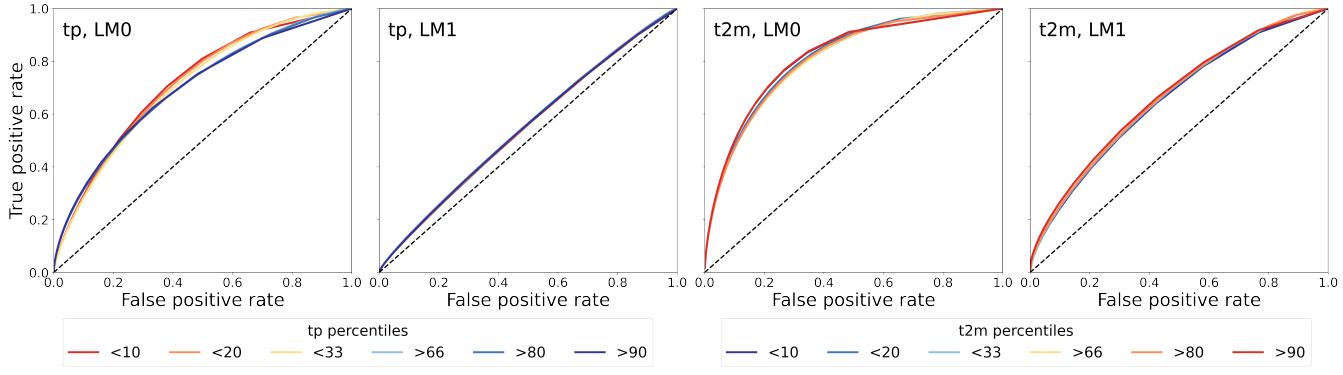


Figure 6. SEAS5-BCSD, Receiver Operating Characteristic (ROC) curves for precipitation (left) and mean 2-m temperature (right) for Lead Months 0 and 1. Percentile thresholds are used to distinguish between abnormal and strong dry/wet conditions. The dashed line indicates random chance.

considered in CRPSS and BSS, in greater detail. The graph illustrates how forecast skill varies across different percentile thresholds as a function of ensemble agreement, i.e., the proportion of ensemble members falling into a given category. This

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

proportion can be interpreted as a measure of internal certainty within the SEAS5 forecast system. For the 20th and 33rd percentiles (representing dry conditions), forecast skill improves nearly linearly with increasing ensemble agreement above a share of 33 %. This suggests that the greater the confidence of the ensemble in predicting dry conditions, the better the forecast performance. For more extreme dryness (10th percentile), a similar trend is observed up to an ensemble share of 20 %, although at generally lower skill levels. Beyond this point, the curve plateaus and then falls off, though the number of such cases becomes sparse, limiting interpretability. In contrast, for wet conditions, the relationship between the wet, strong wet and extreme wet categories is near identical. Nevertheless, across all percentiles and categories, forecast skill generally remains in the positive range for Lead Month 0, with the exception of extreme droughts. In summary, for most threshold categories, a higher degree of ensemble agreement correlates with better forecast performance, highlighting the added value of ensemble certainty in probabilistic seasonal predictions.

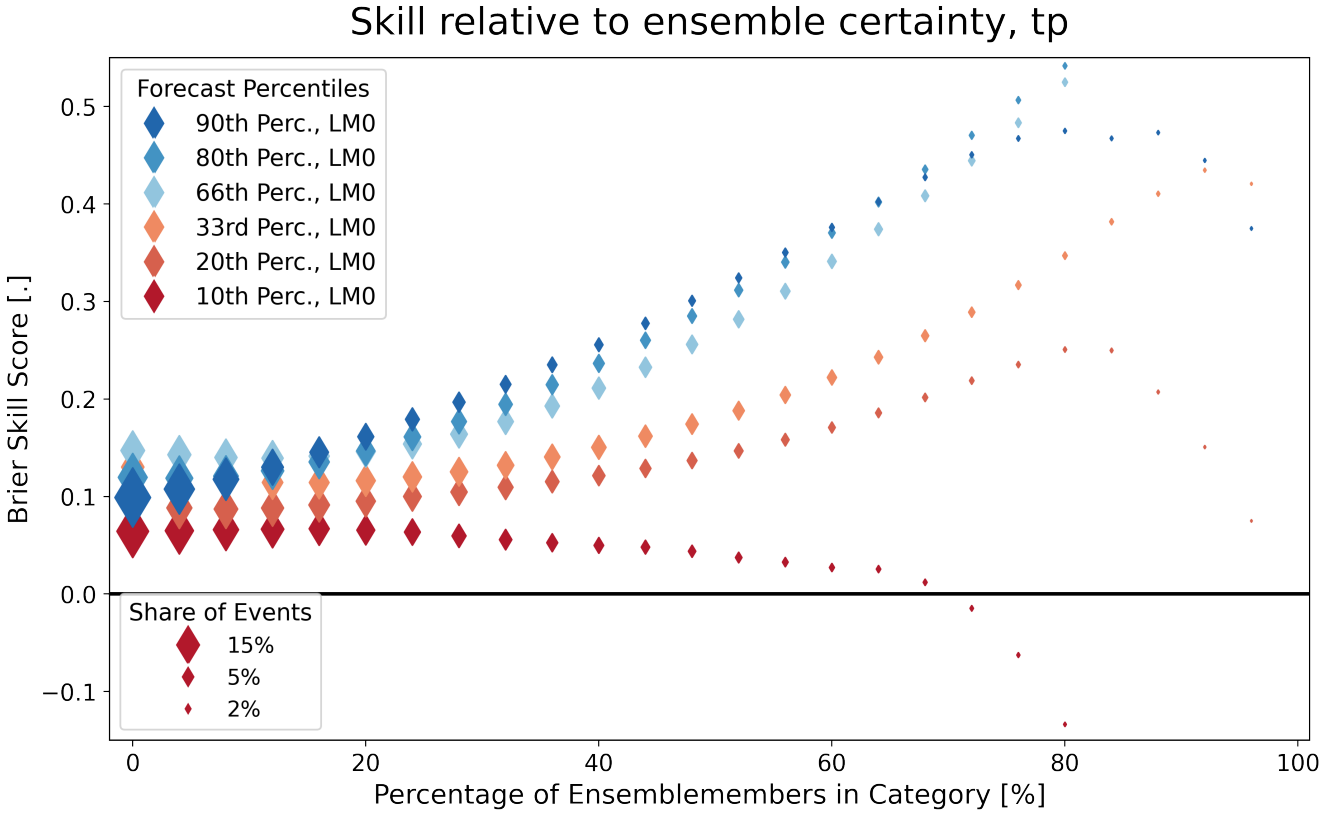


Figure 7. SEAS5-BCSD, Brier Skill Score (BSS) relative to climatology as a function of the number of ensemble members within a category, shown for precipitation at Lead Month 0. Three drought/wet thresholds are depicted. The size of the rectangles denotes the share of events.

We further assess how forecasts persist and evolve as new initialization dates approach the target month, a key aspect for users requiring early, reliable warnings. While Figure 7 focused on selected percentile thresholds, it is equally important to assess

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

the full distribution of ensemble members. Figure 8 illustrates how SEAS5 forecasts are distributed across tercile categories, conditional on the ERA5 classification. For example, the top-left panel includes only those grid cells where ERA5 indicates wetter-than-normal conditions. The bars then show the share of the forecasted classes for these grid cells with lead times from seven (leftmost bar per plot) to one month ahead. A purely climatological forecast would distribute 33.3 % of cases into each tercile. Thus, if the bars with the darker hue exceed 33.3 %, SEAS5 outperforms climatology. A clear trend emerges: the closer the forecast issue date is to the target month, the more accurate the tercile classification. From Lead Month 6 to Lead Month 1, this improvement is gradual, followed by a significant jump in performance at Lead Month 0. With regard to precipitation, dry conditions are most reliably predicted. When dry conditions occur, SEAS5 forecast already indicated them three months in advance with a 19 % higher probability than climatology. Conversely, "normal" conditions are more difficult to capture and are not predicted more accurately than climatology at any lead time. As observed in previous figures, temperature forecasts show higher skill than precipitation, except in the "normal" tercile, which remains below climatological skill. However, when ERA5 indicates unusually warm conditions, the BCSD-corrected forecasts captured this correctly in half of all cases in Lead Month 2, 50 % better than the one-in-three expectation from climatology. Figure 8 can also be interpreted from the opposite perspective by examining the rate of incorrect, reversed forecasts, i.e., instances where the predicted tercile was opposite to the observed category. This provides insight into the potential risks for users relying on categorical forecasts. For example, when ERA5 indicated dry conditions, 29 % of forecasts predicted wet conditions two months prior, dropping to just 16 % at Lead Month 0. Similarly, for hot months, only a fourth of the grid cells predicted cold conditions half a year ahead, and this number decreased to 10 % one month in advance.

The longer a condition is consistently forecasted across subsequent issue months, the higher the forecast skill becomes. For drought conditions, the tercile hit rate for Lead Month 0 reaches 49.6 %, corresponding to 17.7 million correctly predicted grid cells out of 35.8 million grid cells that showed drought condition in LM0. If the same month was already predicted as dry one issue month earlier, this rate increases slightly to 50.2 % (8.2 million of 16.3 million grid cells). Including predictions from Lead Months 2 to 6 from preceding issue months leads to a steady improvement in skill, rising to 50.8 % (4.7 of 9.3 million), 51.4 % (3.1 of 6.0 million), 52.0 % (2.2 of 4.2 million), 52.6 % (1.6 of 3.1 million), and 53.3 % (1.2 of 2.3 million), respectively. This effect is even more pronounced for mean temperature, with more cases of consistent forecasting as well. Starting from a hit rate of 61.3 % (22.2 of 36.1 million) for Lead Month 0, the rate increases markedly with each earlier forecast: 64.9 % (13.1 of 20.2 million) for Lead Month 1, 66.4 % (9.5 of 14.2 million), 67.7 % (7.5 of 11.1 million), 68.9 % (6.2 of 9.0 million), 70.0 % (5.3 of 7.6 million), and finally 70.8 % (4.6 of 6.5 million) when considering forecasts from Lead Month 6.

3.3 Forecast skill in further context

The preceding analyses and figures showed that the skill of precipitation forecasts has clear limitations on a global scale. The global CRPSS (cf. Fig.2) indicates relatively high performance in equatorial regions and poor performance in subtropical, arid zones. To investigate this relationship more systematically, each grid cell was assigned to a specific climate zone according to the Köppen–Geiger Climate Classification (Beck et al., 2018). Figure9 illustrates the typical pattern: high skill in Lead Month 0 followed by a stagnation or gradual decline in subsequent lead months. Among the different climate zones, the tropical

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

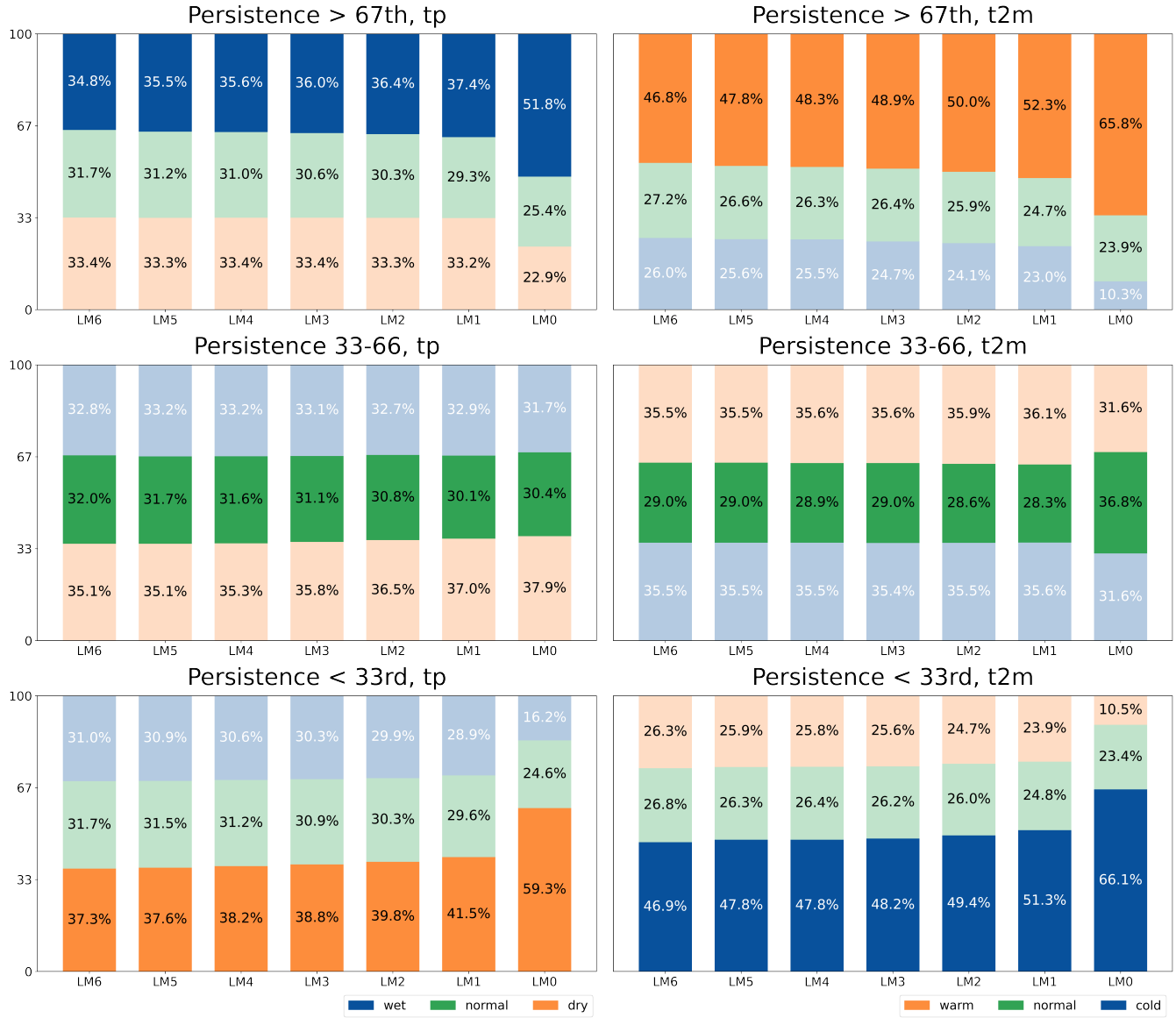
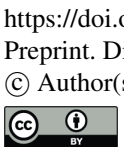


Figure 8. SEAS5-BCSD, Tercile persistence for precipitation (left) and temperature (right). For each ERA5-classified tercile, the corresponding SEAS5 tercile classifications are tracked across the seven lead months. Values represent percentage distributions aggregated over all grid cells and the reference period. Darker shading indicates the correct (persistent) category.

regions clearly stand out with the highest forecast skill for precipitation. Most months in these regions maintain positive skill even at Lead Month 6. In contrast, the neighboring arid zones are already outperformed by climatology from Lead Month 2 onward. At this lead time, forecasts for temperate and polar regions still retain a slight advantage over climatology, though their skill converges toward zero at longer lead times. Cold regions perform similarly to arid ones in terms of CRPSS, albeit with a

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026



Open Access Earth System Science Data Discussions

narrower spread. To quantify the benefit of the BCSD correction across climate zones, we also computed the CRPSS relative to the raw SEAS5 forecasts (not shown). The tropics again show the strongest improvement, with a median CRPSS gain of 0.28, followed by temperate zones (0.15), arid (0.11, with substantial spread), polar (0.07), and cold regions (0.06). Differences between lead months are minor, suggesting that tropical precipitation forecasts benefit most consistently and substantially from the BCSD approach.

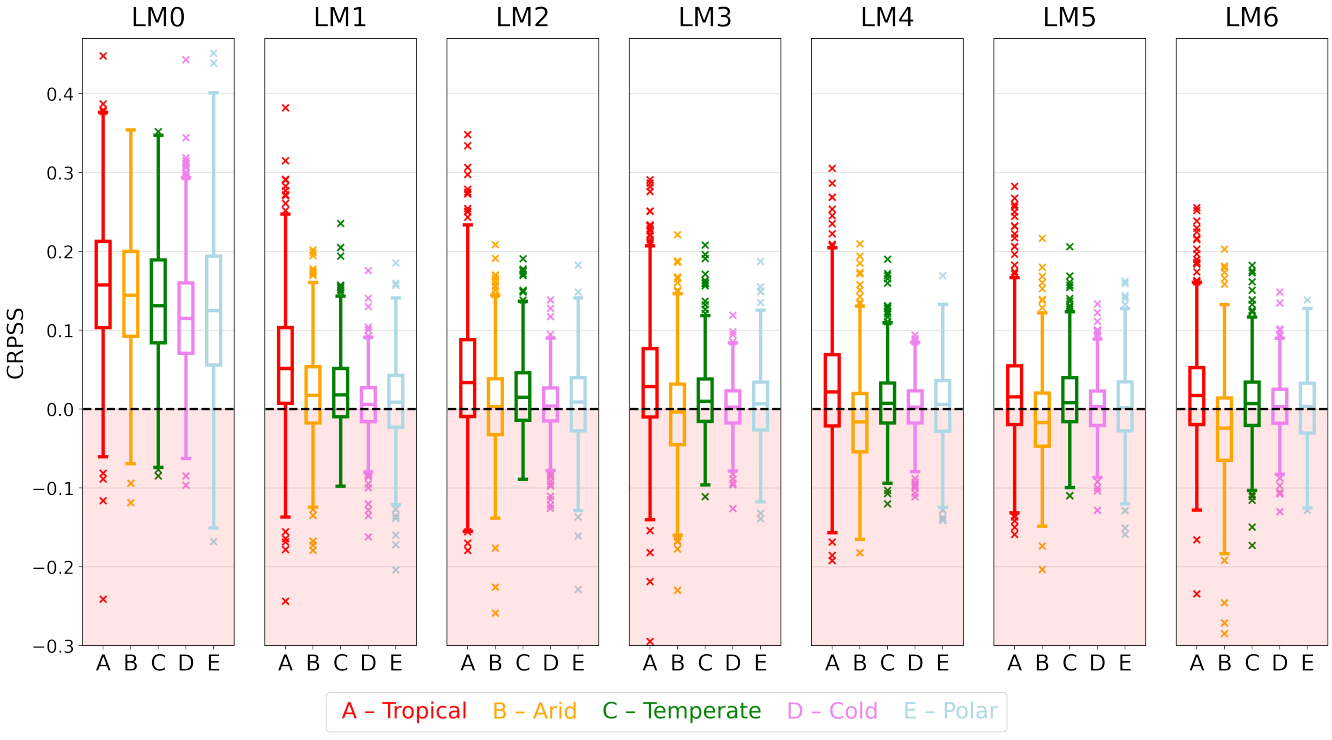


Figure 9. SEAS5-BCSD, Spatially aggregated Continuous Ranked Probability Skill Score (CRPSS) relative to climatology for precipitation, shown for Lead Months 0–6 over the period 1981–2016. Results are grouped by major climate classes according to the Köppen–Geiger classification.

Figure 2 reveals a performance bias toward hot or cold conditions in Lead Months 1 to 6, but in Lead Month 0, this is not apparent. One potential factor influencing temperature forecast skill is elevation. Raw SEAS5 forecasts have strong biases in mountain areas (Figure 1, right panel), which the BCSD should reduce. By isolating elevation effects, more nuanced insights may emerge for Lead Month 0. Seasonal forecasting in mountainous regions is particularly challenging (Uttarwar et al., 2025). To assess the potential influence of elevation on temperature forecast skill, Figure 10 shows the CRPSS for different elevation-related metrics for Lead Month 0. Elevation data are derived from the 30 m-resolution TanDEM-X Global DEM (Gonzalez et al., 2020), aggregated to match the 0.25° ERA5 grid. On the left, the mean elevation per grid cell is shown. A clear trend emerges: for Lead Month 0, CRPSS values are highest in lower-lying regions and tend to decrease gradually with increasing

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

elevation. To investigate whether this trend is more related to absolute peak elevation or general plateau height, CRPSS is also shown for the maximum and minimum elevations. Maximum elevation displays a pronounced stratification, suggesting that the presence of low-lying terrain is more conducive to higher forecast skill than regions with high peaks. In high plateaus or mountain ranges, where all grid cells lie above 3000 m, the temperature forecast skill is somewhat reduced, as seen in the third panel from the left. Additionally, terrain complexity, calculated as the mean absolute difference from the mean elevation, also correlates with skill, though less strongly than mean elevation.

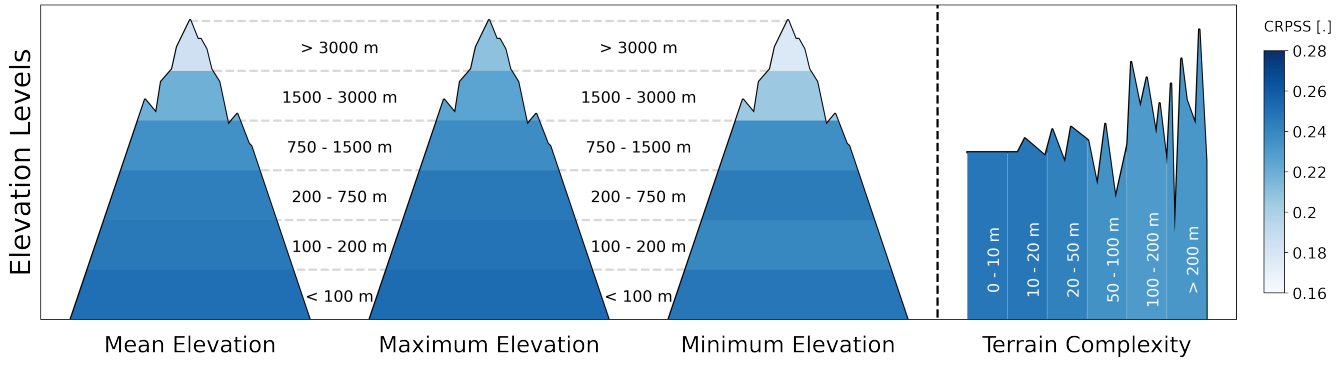


Figure 10. SEAS5-BCSD, CRPSS of temperature for Lead Month 0, stratified by elevation metrics. Mean elevation refers to the average value of the high-resolution DEM within each grid cell, maximum elevation denotes the highest point, and minimum elevation the lowest. Terrain complexity is defined as the mean absolute deviation of all high-resolution DEM values from the grid cell mean.

Given the apparent relationship between forecast skill and annual precipitation totals (see Figure A3), a more detailed investigation of this dependency is warranted. A clear positive relationship can be observed: the greater the rainfall amount, the higher the forecast skill. As the CRPSS is a normalized score comparing forecasts to climatology, this trend cannot be attributed to larger absolute errors at high precipitation values, as is the case for metrics like CRPS or mean absolute error. This relationship aids in interpreting the spatial skill patterns shown in Figure 4. Regions with higher annual precipitation generally correspond to areas of improved forecast skill. Since Figure 4 displays skill for drought forecasts, already arid regions inherently exhibit lower Brier Skill Scores, as drought conditions are climatologically common and thus harder to distinguish from the baseline. This relationship is particularly evident in Central Australia, where low precipitation levels coincide with low skill values in the BCSD-corrected forecasts. The highest forecast performance is found in regions receiving more than 2000 mm yr^{-1}, beyond which the performance curve flattens. At the other end of the spectrum, regions with low precipitation amounts show a clear decline in skill (see Figure A3). For Lead Month 6, the CRPSS values begin to fall consistently below zero at around 330 mm yr^{-1} (not shown). Forecasts closer to the target month maintain positive skill down to lower precipitation thresholds, with the transition points at approximately 10, 90, 195, 205, 305, and 235 mm yr^{-1} for Lead Months 0 through 5, respectively. Below these thresholds, forecast performance declines rapidly. An interesting pattern emerges for Lead Month 0: in arid (but not hyper-arid) regions with less than 250 mm yr^{-1}, forecast skill slightly increases as precipitation decreases,

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

peaking around 30 mm yr^{-1}. This may reflect the relatively good performance of the system in predicting drought conditions in these dry environments.

In the vicinity of 300 mm yr^{-1}, where forecast skill becomes consistently positive, albeit at a low level, many semi-arid regions of the Earth are located, typically characterized by a pronounced rainy–dry season cycle. These two seasons often differ markedly in their observed and forecasted rainfall amounts. To assess their impact separately, Figure 11 presents a seasonal analysis based on the Brier Skill Score (BSS). Only regions with a strong seasonality were included, defined as those where the two driest months contribute less than 20 % of the rainfall of the two wettest months. Very arid areas with annual totals below 250 mm yr^{-1} were excluded. The upper graph is subdivided into three horizontal sections corresponding to the forecast skill for mild (top row), moderate (middle), and severe droughts (bottom). For Lead Month 0, all categories show positive skill in both seasons. While the rainy season maintains positive skill throughout the entire lead time range, forecast performance during the dry season quickly deteriorates, with BSS values turning negative after one month. In contrast, seasonal variation in the mid-latitudes is primarily temperature-driven. Accordingly, Figure 11 also shows BSS values for summer and winter in regions where the three warmest months are at least 10 °C warmer than the three coldest. In these areas, seasonal differences in forecast skill are less pronounced. For hot and moderate warm seasons (again, relative and not absolute), winter exhibits slightly better performance. However, in the case of mild seasons (bottom row), the differences between winter and summer largely vanish.

4 Discussion

Our findings stress the variability in seasonal forecast performance across different climate regimes and conditions. Particularly noteworthy is the limited forecast skill in arid regions, which warrants closer examination in the following discussion.

4.1 Underperformance in arid regions/Climate-dependend performance

The analysis revealed that BCSD forecasts tend to be outperformed by climatology in arid regions. This is supported by regional studies, such as Zamora et al. (2021), who found a weak Brier Skill Score for extremely dry predictions in the US using a different seasonal forecasting product. Thrasher et al. (2012) also discussed the limitations of BCSD performance over extreme climates. One likely reason for this skill deficit is the low signal-to-noise ratio in dry regions: the absolute differences between dry and very dry conditions are small, making them more difficult to predict accurately (Hao et al., 2018). Even minor biases or shifts in forecast distribution can therefore result in a loss of skill relative to climatology. To assess whether the observed skill reduction originates from the bias correction method itself or from limitations in the SEAS5 model (which was developed after the study by Thrasher et al. (2012)) Figure 5 provides further insight. It shows that the lowest precipitation values are slightly underrepresented, while marginally wetter (but still dry) categories are correspondingly overrepresented. Before the bias correction, SEAS5 forecasts contain more days without precipitation over land than ERA5, typical for climate prediction systems due to the drizzle effect (Boé et al., 2007). The BCSD process appears to reverse this pattern. Wetter categories are represented very well in contrast. This discrepancy may point to limitations in how dry-day probabilities (DDP)

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

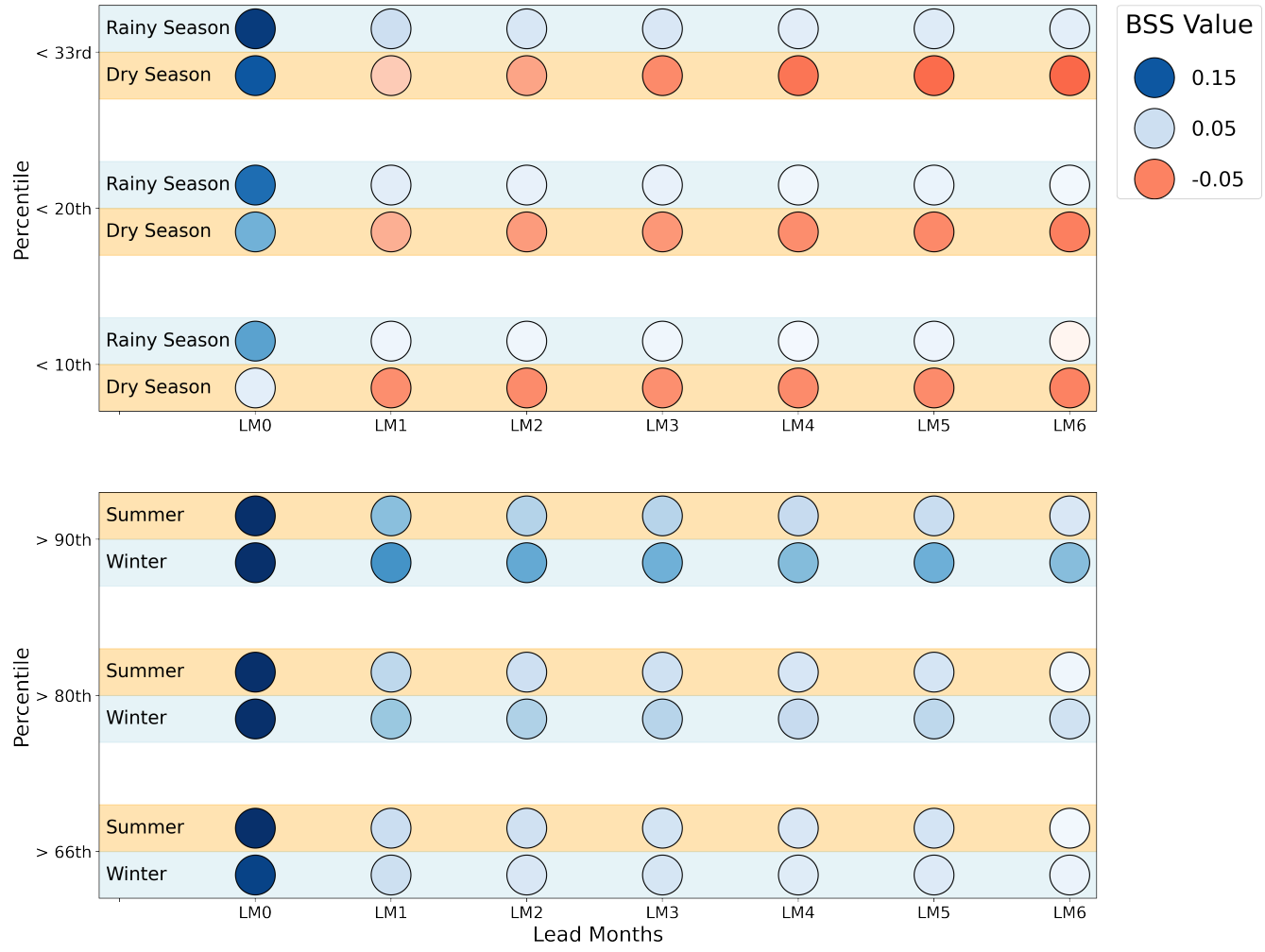


Figure 11. Brier Skill Score (BSS) for different seasons. Precipitation is separated into rainy and dry seasons (top panel), temperature into summer and winter seasons (bottom panel). The color intensity is proportional to the BSS value.

are handled by the system. In the context of climate projections, Rajczak et al. (2016) demonstrated that Quantile Mapping can significantly improve the representation of dry spell lengths in comparison to overly wet raw projections, thereby also enhancing the accuracy of DDP estimates. This is supported by Themeßl et al. (2012), although for RCMs not for seasonal forecasting systems. On the other hand, Maraun (2013) reported that Quantile Mapping struggles with correctly adjusting drizzle precipitation, highlighting limitations in handling low-intensity events. As described in Section 2.2, if the DDP of SEAS5 is larger than of the reference, random sampling is applied to draw values from the respective cumulative distribution function (CDF). If the fixed number of quantile bins (200) is insufficiently granular, this can lead to mismatches. Since SEAS5 generally exhibits slightly more dry days, this issue is more frequent than in the reverse case, where values are forced to zero

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

in the correction. A further contributing factor could be the temporal smoothing introduced by the 31-day moving window used in the correction process. In very dry months, this may result in contamination by wetter days from neighboring seasons, shifting values into higher bins (e.g., from the lowest to the third-lowest category) based on very small differences (border at 1.2 mm/month). Additionally, monthly averaging may amplify the influence of individual forecasted precipitation events. This could result in SEAS5 displaying non-zero values even in months with extremely low actual rainfall, potentially skewing distributions further.

4.2 Computational efficiency

Timely availability of bias-corrected forecasts is crucial, particularly since forecast skill is as shown highest during the first lead month. As SEAS5 forecasts are released on the 5th of each month at 12 UTC, our goal was to produce BCSD-corrected outputs for six variables on the same day. To meet this target, we needed to make substantial efforts to reduce computational demand. The primary improvement in computational efficiency stems from storing pre-computed cumulative distribution functions (CDFs), rather than generating them on-the-fly, as done in previous studies (e.g., Lorenz et al., 2021). To reduce processing time during the CDF preparation stage, we approximated the empirical CDFs using a reduced number of quantile points: 200 instead of the full set of available values (1116 for ERA5 and 27,900 for SEAS5). This linear interpolation approach is widely used in the literature (e.g., Themeßl et al., 2012; Boé et al., 2007) and reflects a compromise between computational efficiency and statistical accuracy. Although higher-resolution CDFs (e.g., 500 or full data) may enhance correction performance (Gudmundsson et al., 2012), test runs demonstrated only marginal differences in forecast skill between 200, 500, and the full dataset. In contrast, using only 100 points led to noticeable degradation in performance. Therefore, 200 quantile points were selected as a balanced solution. This reduction also resulted in significantly lower storage requirements, with the pre-computed CDFs occupying substantially less than a fully downscaled uncorrected SEAS5 archive needed for on-the-fly corrections. We employed an equally spaced selection of quantiles. While Gergel et al. (2024) emphasize the importance of carefully resolving the tails of the distribution, arguing for denser sampling at extremes, an evenly spaced selection offers robustness for ensemble-based systems where most members fall within the central range of the distribution (Cannon et al., 2015). In our view, both strategies are defensible depending on the application. Further efficiency was gained by parallelizing the correction across ensemble members. This approach not only distributed the computational load but also avoided redundant file access, reducing overall runtime by approximately 10 %. Combined, these optimizations enabled the complete bias correction of all six target variables within roughly twelve hours, making same-day release of bias-corrected forecasts feasible, contingent on data availability and download times.

5 Data Availability

The global Bias-Corrected and Spatially Disaggregated (BCSD) forecast dataset is available on the World Data Center for Climate (WDCC) via https://www.wdc-climate.de/ui/entry?acronym=SEAS5-BCSD (DOI: 10.26050/WDCC/SEAS5-BCSD, Weber et al. (2026)). The dataset includes all issue dates from 1981 to 2024. Forecasts are provided as monthly aggregates on

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

a 0.25° x 0.25° grid with a land mask applied, while maintaining the full ensemble dimension. Separate files are provided for each variable; total precipitation (tp) and 2 m temperature (t2m) are currently available. For the 1981–2016 period, individual yearly files are approximately 500 MB in size, and about 1 GB thereafter, corresponding to roughly 28 GB per variable for the complete dataset.

In addition, the BCSD forecasts are distributed via the Karlsruhe Institute of Technology (KIT) – Campus Alpin THREDDS Data Server. Unlike the products archived in WDCC, the operational forecasts are exclusively available through the THREDDS service. These data are published with a delay of about one to two days following the official ECMWF seasonal forecast release (on the fifth of each month). Access to the operational products can be requested by contacting christof.lorenz@kit.edu or jan.weber@kit.edu under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

6 Code Availability

The BCSD processing code used to generate the daily SEAS5-BCSD dataset (prior to monthly aggregation) is available at Zenodo (Lorenz et al., 2025). The repository provides the scripts used for bias correction and spatial disaggregation of SEAS5 seasonal forecasts. The code is openly accessible under the license specified in the repository. The version used for this study corresponds to the state of the repository as of November 2025.

7 Conclusion

Raw SEAS5 forecasts exhibit only limited skill, particularly at longer lead times. Their performance becomes increasingly weak beyond the first month. Our presented global BCSD correction significantly improves forecast quality, providing improved predictions for most regions several months ahead. The corrected temperature forecasts are especially skillful. Across nearly all land regions, they outperform climatology even at a lead time of seven months. Forecasts perform best in flat, low-lying regions, while performance declines in high-altitude regions and complex terrains, although even there, skill remains positive. Seasonal differences in temperature forecast skill (e.g., between warm and cold seasons) are minor. For precipitation, skill is more limited. While bias correction brings substantial improvements in tropical regions, gains are smaller in cold-continental areas. Regions with high annual rainfall generally benefit from good forecast performance. In semi-arid zones, skill is solid during the rainy season but declines markedly in dry months. In hyper-arid regions, forecasts are generally unreliable. This may stem from how dry-day probabilities are handled in the bias correction process. We anticipate that drought indicators incorporating temperature, such as the Standardized Precipitation Evapotranspiration Index (SPEI; Vicente-Serrano et al., 2010), will further improve drought detection, particularly in hot and dry regimes where evaporation plays a key role. Forecasts of abnormally wet or dry conditions generally exhibit higher skill than those predicting near-normal conditions. Moreover, the forecast reliability increases with ensemble agreement: The greater the number of ensemble members indicating a given abnormal category, the higher the likelihood that this category will be observed. This reflects a consistent relationship between forecast probability and observed frequency. Further improvements could be achieved by refining technical aspects such as the

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

moving window size or the number of quantile bins used for the cumulative distribution functions (CDFs), or by adopting a more robust reference dataset than ERA5. However, within the context of a computationally efficient global system, the current results are already highly satisfactory.

Potential applications of the SEAS5-BCSD dataset are broad. The dataset can strengthen regional assessments of seasonal forecast performance and enable anticipatory risk management in climate-sensitive sectors such as water resource management (e.g., reservoir operations), agriculture, and inland waterway transport. The availability of the complete ensemble set is particularly advantageous for identifying low-probability, high-impact scenarios and supports more robust probabilistic risk assessments. Since both temperature and precipitation are provided, compound events such as concurrent heat and drought stress can be anticipated at an early stage. In addition, the dataset provides a reference framework for evaluating future forecasting products generated using similar post-processing approaches. As both the code and data are publicly available, the analyses are fully reproducible. The dataset's ensemble structure enables a comprehensive assessment of forecast reliability and uncertainty, incorporating both ensemble spread and probabilistic information. This transparency strengthens confidence in seasonal forecasts and, consequently, in early warning systems based on them. Such trust is essential for risk-based decision-making and represents a key step toward actionable climate services. Another important application is the analysis of forecast performance for large-scale drought events and related climate extremes, potentially warning millions of inhabitants months in advance of an emerging drought or failing rainy season. In the context of climate change, which is expected to increase the frequency and intensity of extreme events and challenge assumptions of stationarity, reliable seasonal forecasts become an essential component of climate adaptation strategies. The availability of bias-corrected ensemble members over the full hindcast period enables consistent retrospective analyses and skill assessments across multiple decades, thereby increasing confidence in current forecasts. From a modeling perspective, the dataset provides a global test bed with high-resolution, bias-corrected temperature and precipitation fields that can serve as input for hydrological, agricultural, and impact modeling studies.

In summary, the BCSD-corrected SEAS5 forecasts can be applied across most non-arid regions worldwide, with skillful predictions available up to seven months in advance. Together, these characteristics establish a foundation for future research and operational climate services. Owing to improvements in the processing workflow, the corrected data are typically released on the sixth day of each month, facilitating timely application.

Appendix A: Performance evaluation metrics

A1 ERA5 data set limitations

Despite its widespread use as a reference, ERA5 is not without limitations, particularly in data-sparse regions and for variables like precipitation, where reanalysis uncertainties can propagate into both forecast evaluation and bias correction. For Ethiopia, Ahmed et al. (2024) found that CHIRPS generally outperformed ERA5, especially in the Ethiopian Highlands, where both datasets overestimated rainfall and underestimated the observed high rainfall variabilities. However, ERA5 is able to reasonably capture spatial patterns and rainfall distribution. On a global scale, ERA5 also exhibits a slight wet bias, but performs well in extratropical regions, particularly for precipitation (Lavers et al., 2022). Comparisons with satellite-based products such as

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

IMERG over China suggest that ERA5 performs less well in tropical regions, but tends to outperform satellite datasets in continental and cold temperate areas (Xu et al., 2022). Despite these known limitations, ERA5 was chosen as the reference for this study due to its overall adequate performance, full global coverage and near real-time availability with only a short latency of five days. Additionally, ERA5 offers consistent data across multiple variables, facilitating a seamless transition between precipitation and temperature analyses and bias correction approaches.

We acknowledge that trend calculation in ERA5 is not recommended. However, since the BCSD method also relies on this reference dataset, the comparison remains valid. The same trend effects should be reproduced when using different datasets as input for the CDFs.

Table A1. CRPSS of raw SEAS5 forecasts against climatology, averaged per continents.

Region	LM0	LM1	LM2	LM3	LM4	LM5	LM6
Africa	-0,221	-0,203	-0,167	-0,150	-0,148	-0,148	-0,156
Asia	0,015	-0,107	-0,113	-0,113	-0,117	-0,114	-0,117
Australia	0,140	0,013	-0,004	-0,003	-0,012	-0,001	-0,028
South America	-0,088	-0,176	-0,186	-0,189	-0,197	-0,203	-0,209
Europe	0,071	-0,046	-0,051	-0,048	-0,052	-0,050	-0,053
North America	0,062	-0,041	-0,046	-0,050	-0,047	-0,049	-0,047

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

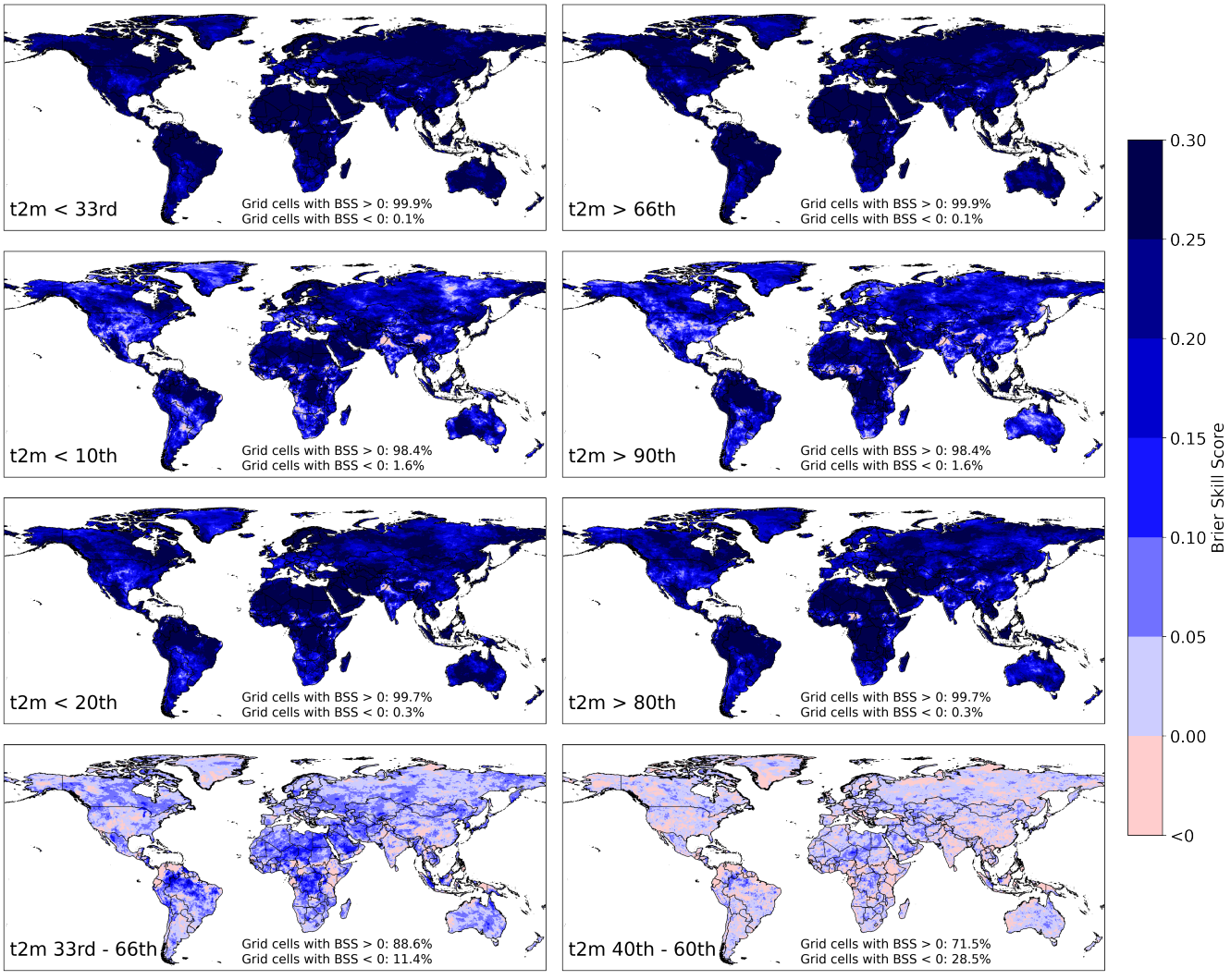


Figure A1. SEAS5-BCSD, Brier Skill Score (BSS) relative to climatology for 2m-temperature at Lead Month 0. Different percentile thresholds are shown, indicated in the lower-left corner of each panel. Blue colors indicate higher skill than climatology, with darker shades representing better performance.

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

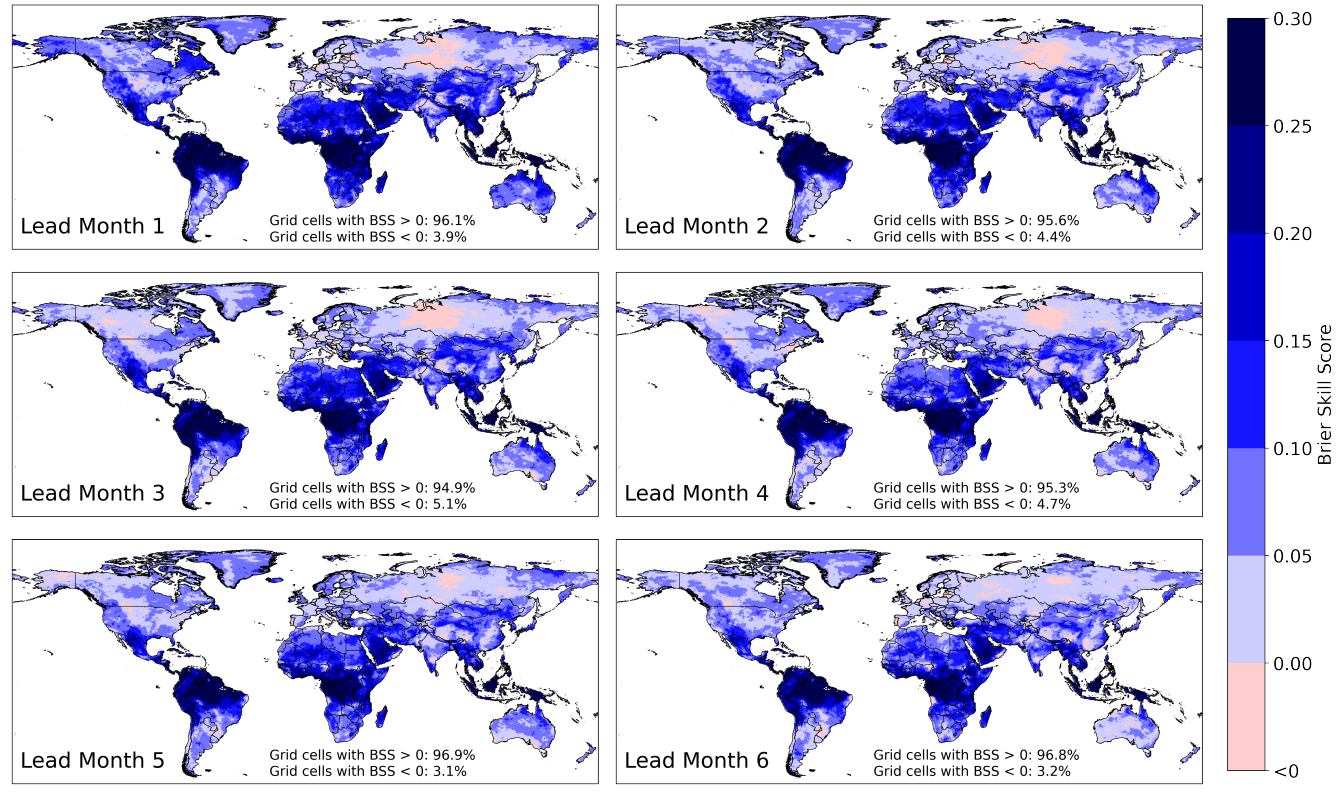


Figure A2. SEAS5-BCSD, Brier Skill Score (BSS) relative to climatology for 2m-temperature, focusing on dry conditions below the 33rd percentile for Lead Months 1–6. Blue colors indicate higher skill than climatology, with darker shades representing better performance.

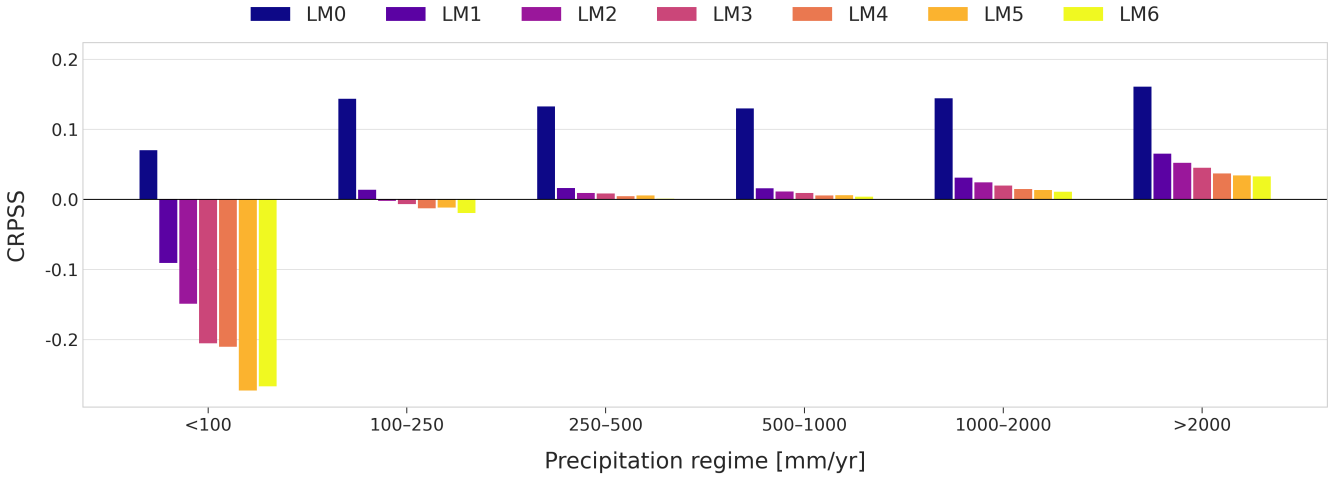


Figure A3. SEAS5-BCSD, yearly mean precipitation sum compared to the Continuous Ranked Probability Skill Score (CRPSS) relative to climatology. The area is grouped by precipitation thresholds into 6 bins, each showing seven lead months.

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

Author contribution

Christof Lorenz, Tanja C. Schober, Rebecca Wiegels, and Jan N. Weber developed the BCSD code. Jan N. Weber implemented and applied the methodology at the global scale and performed the analyses. Jan N. Weber prepared the manuscript with contributions from all co-authors.

Competing interests

Co-author Christof Lorenz is a member of the editorial board of ESSD.

Acknowledgements

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). We acknowledge the ECMWF for providing ERA5 and SEAS5 data. Artificial intelligence–based language tools were used to support English language editing; the scientific content and interpretations remain the sole responsibility of the authors.

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

References

Ahmed, J. S., Buizza, R., Dell'Acqua, M., Demissie, T., and Pè, M. E.: Evaluation of ERA5 and CHIRPS rainfall estimates against observations across Ethiopia, Meteorology and Atmospheric Physics, 136, https://doi.org/10.1038/sdata.2018.214, 2024.

Bandhauer, M., Isotta, F., Lakatos, M., Lussana, C., Båserud, L., Izsák, B., Szentes, O., Tveito, O. E., and Frei, C.: Evaluation of daily precipitation analyses in E-OBS (v19.0e) and ERA5 by comparison to regional high-resolution datasets in European regions, International Journal of Climatology, 42, 727–747, https://doi.org/10.1002/joc.7269, 2021.

Beck, H., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, Nature Scientific Data, 5, https://doi.org/10.1038/sdata.2018.214, 2018.

Beck, H., van Dijk, A. I. J. M., Larraondo, P. R., McVicar, T. R., Pan, M., Dutra, E., and Miralles, D. G.: MSWX: Global 3-Hourly 0.1° Bias-Corrected Meteorological Data Including Near-Real-Time Updates and Forecast Ensembles, Bulletin of the American Meteorological Society, 103, https://doi.org/10.1175/BAMS-D-21-0145.1, 2022.

Boé, J., Terray, L., Habets, F., and Martin, E.: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies, International Journal of Climatology, 27, 1643–1655, https://doi.org/10.1002/joc.1602, 2007.

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

Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?, Journal of Climate, 28, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1, 2015.

Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, Journal of Hydrology, 375, 613–626, https://doi.org/10.1016/j.jhydrol.2009.06.005, 2009.

Crespi, A., Petitta, M., Marson, P., Viel, C., and Grigis, L.: Verification and Bias Adjustment of ECMWF SEAS5 Seasonal Forecasts over Europe for Climate Service Applications, Climate, 9, https://doi.org/10.3390/cli9120181, 2021.

Di Capua, G. and Rahmstorf, S.: Extreme weather in a changing climate, Environmental Research Letters, 18, https://doi.org/10.1088/1748-9326/acfb23, 2023.

Fröhlich, K., Dobrynin, M., Isensee, K., Gessner, C., Paxian, A., Pohlmann, H., Haak, H., Brune, S., Früh, B., and Baehr, J.: The German Climate Forecast System: GCFS, Journal of Advances in Modeling Earth Systems, 13, https://doi.org/10.1029/2020MS002101, 2021.

Gergel, D. R., Malevich, S. B., McCusker, K. E., Tenezakis, E., Delgado, M. T., Fish, M. A., and Kopp, R. E.: Global Downscaled Projections for Climate Impacts Research (GDPCIR): preserving quantile trends for modeling future climate impacts, Geoscientific Model Development, 17, 191–227, https://doi.org/10.5194/gmd-17-191-2024, 2024.

Golian, S. and Murphy, C.: Evaluating Bias-Correction Methods for Seasonal Dynamical Precipitation Forecasts, Journal of Hydrometeorology, 23, 1350–1363, https://doi.org/10.1175/JHM-D-22-0049.1, 2022.

Gonzalez, C., Bachmann, M., Jose-Luis, B.-B., Rizzoli, P., and Zink, M.: A Fully Automatic Algorithm for Editing the TanDEM-X Global DEM, Remote Sensing, 12, https://doi.org/10.3390/rs12233961, 2020.

Gonzalez-Aparicio, I. and Hidalgo, J.: Dynamically based future daily and seasonal temperature scenarios analysis for the northern Iberian Peninsula, International Journal of Climatology, 32, 1825–1833, https://doi.org/10.1002/joc.2397, 2011.

Gudmundsson, L., Bremnes, J., Haugen, J., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, Hydrology and Earth System Sciences, 16, 3383–3390, https://doi.org/10.5194/hess-16-3383-2012, 2012.

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

Hao, Z., Singh, V. P., and Xia, Y.: Seasonal Drought Prediction: Advances, Challenges, and Future Prospects, Reviews of Geophysics, 56, 108–141, https://doi.org/10.1002/2016RG000549, 2018.

Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather and Forecasting, 15, 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Hersbach, H. and co authors: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Holthuijzen, M., Beckage, B., Clemins, P. J., Higdon, D., and Winter, J. M.: Robust bias-correction of precipitation extremes using a novel hybrid empirical quantile-mapping method, Theoretical and Applied Climatology, 149, 863–882, https://doi.org/10.1007/s00704-022-04035-2, 2022.

Hudson, D., Alves, O., Hendon, H. H., Lim, E.-P., Liu, G., Luo, J.-J., MacLachlan, C., Marshall, A. G., Shi, L., Wang, G., Wedd, R., Young, G., Zhao, M., and Zhou, X.: ACCESS-S1 The new Bureau of Meteorology multi-week to seasonal prediction system, Journal of Southern Hemisphere Earthy System Science, 67, 132–159, https://doi.org/10.1071/ES17009, 2017.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B. M.: SEAS5: the new ECMWF seasonal forecast system, Geoscientific Model Development, 12, 1087–1117, https://doi.org/10.5194/gmd-12-1087-2019, 2019.

Krishnamurthy, V.: Predictability of Weather and Climate, Earth and Space Science, 6, 1043–1056, https://doi.org/10.1029/2019EA000586, 2019.

Lavers, D. A., Simmons, A., Vamborg, F., and Rodwell, M. J.: An evaluation of ERA5 precipitation for climate monitoring, Quarterly Journal of the Royal Meteorological Society, 148, 3152–3165, https://doi.org/10.1002/qj.4351, 2022.

Lorenz, C., Portele, T. C., Laux, P., and Kunstmann, H.: Bias-corrected and spatially disaggregated seasonal forecasts: a long-term reference forecast product for the water sector in semi-arid regions, Earth System Science Data, 13, 1106–1114, https://doi.org/10.5194/essd-2020-177, 2021.

Lorenz, C., Wiegels, R., Chwala, C., Fersch, B., Weber, J. N., Sawadogo, W., Portele, T. C., and Borkenhagen, C.: PyCast-S2S: A Python Framework for Subseasonal-to-Seasonal Forecast Post-Processing, https://doi.org/10.5281/ZENODO.16926092, software, Version 0.8.0, 2025.

MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A., Gordon, M., Vellinga, M., Williams, A., Comer, R., Camp, J., Xavier, P., and Madec, G.: Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system, Quarterly Journal ot the Royal Meteorological Society, 141, 1072–1084, https://doi.org/10.1002/qj.2396, 2014.

Maraun, D.: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, Journal of Climate, 26, 2137–2143, https://doi.org/10.1175/JCLI-D-12-00821.1, 2013.

Marzban, C.: The ROC Curve and the Area under It as Performance Measures, Weather and Forecasting, 19, 1106–1114, https://doi.org/10.1175/825.1, 2004.

Materia, S., Borrelli, A., Bellucci, A., Alessandri, A., Di Pietro, P., Athanasiadis, P., Navarra, A., and Gualdi, S.: Impact of Atmosphere and Land Surface Initial Conditions on Seasonal Forecasts of Global Surface Temperature, Journal of Climate, 27, 9253–9271, https://doi.org/10.1175/JCLI-D-14-00163.1, 2014.

Nogueira, M.: Inter-comparison of ERA-5, ERA-interim and GPCP rainfall over the last 40 years: Process-based analysis of systematic and random differences, Journal of Hydrology, 583, https://doi.org/10.1016/j.jhydrol.2020.124632, 2020.

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

Portele, T. C., Lorenz, C., Dibrani, B., Laux, P., Bliefernicht, J., and Kunstmann, H.: Seasonal forecasts offer economic benefit for hydrological decision making in semi-arid regions, Nature Sicentific Reports, 11, https://doi.org/10.1038/s41598-021-89564-y, 2021.

Rajczak, J., Kotlarski, S., and Schär, C.: Does Quantile Mapping of Simulated Precipitation Correct for Biases in Transition Probabilities and Spell Lengths?, Journal of Climate, 29, 1605–1615, https://doi.org/10.1175/JCLI-D-15-0162.1, 2016.

Themeßl, M. J., Gobiet, A., and Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, Climatic Change, 112, 449–468, https://doi.org/10.1007/s10584-011-0224-4, 2012.

Thrasher, B., Maurer, E., McKellar, C., and Duffy, P.: Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping, Hydrology and Earth System Sciences, 16, 3309—3314, https://doi.org/10.5194/hess-16-3309-2012, 2012.

Tóth, Z., Talagrand, O., Candille, G., and Zhu, Y.: Probability and Ensemble Forecasts, in: Forecast Verification: A Practitioner's Guide in Atmospheric Science, edited by Jolliffe, I. T. and Stephenson, D. B., pp. 137–163, John Wiley & Sons, Chichester, UK, ISBN 0-471-49759-2, 2003.

United Nations Convention to Combat Desertification: Drought in Numbers 2022, https://www.unccd.int/sites/default/files/2022-05/Drought%20in%20Numbers.pdf, accessed: 2025-03-31, 2022.

Uttarwar, S. B., Napoli, A., Avesani, D., and Majone, B.: Elevation-driven biases in seasonal weather forecasts: Insights from the Alpine region, Physics and Chemistry of the Earth, 139, https://doi.org/10.1016/j.pce.2025.103957, 2025.

Vicente-Serrano, S. M., Begueria, S., and López-Moreno, J. I.: A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index, Journal of Climate, 23, 1696–1718, https://doi.org/10.1175/2009JCLI2909.1, 2010.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H.-S., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., Manoussakis, M., Mastrangelo, D., MacLachlan, C., McLean, P., Minami, A., Mladek, R., Nakazawa, T., Najm, S., Nie, Y., Rixen, M., Robertson, A. W., Ruti, P., Sun, C., Takaya, Y., Tolstykh, M., Venuti, F., Waliser, D., Woolnough, S., Wu, T., Won, D.-J., Xiao, H., Zaripov, R., , and Zhang, L.: The Subseasonal to Seasonal (S2S) Prediction Project Database, Bulletin of the American Meteorological Society, 98, 163–173, https://doi.org/10.1175/BAMS-D-16-0017.1, 2017.

Volosciuk, C., Maraun, D., Vrac, M., and Widmann, M.: A combined statistical bias correction and stochastic downscaling method for precipitation, Hydrology and Earth System Sciences, 21, 1693–1719, https://doi.org/10.5194/hess-21-1693-2017, 2017.

Wang, Q., Shao, Y., Song, Y., Schepen, A., Robertson, D. E., Ryu, D., and Pappenberger, F.: An evaluation of ECMWF SEAS5 seasonal climate forecasts for Australia using a new forecast calibration algorithm, Environmental Modelling Software, 122, https://doi.org/10.1016/j.envsoft.2019.104550, 2019.

Weber, J. N., Lorenz, C., Portele, T. C., and Kunstmann, H.: Seasonal forecasts for Germany: enhancing the predictive capability of global SEAS5 ensemble forecasts using bias correction, Hydrologie und Wasserbewirtschaftung, 67, 90–109, https://doi.org/10.5675/HyWa_2023.2_2, 2023.

Weber, J. N., Lorenz, C., Schober, T. C., and Kunstmann, H.: Global Bias-corrected Seasonal Forecast (SEAS5-BCSD) Archive 1981 To 2024 For Precipitation And Temperature, https://doi.org/10.26050/WDCC/SEAS5-BCSD, dataset, version 1.0, 2026.

Wood, A. W., Leung, L., Sridhar, V., and Lettenmaier, D.: Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs, Climatic Change, 62, 189–216, https://doi.org/10.1023/B:CLIM.0000013685.99609.9e, 2004.

Xu, J., Ma, Z., Yan, S., and Peng, J.: Do ERA5 and ERA5-land precipitation estimates outperform satellite-based precipitation products? A comprehensive comparison between state-of-the-art model-based and satellite-based precipitation products over mainland China, Journal of Hydrology, 605, https://doi.org/10.1016/j.jhydrol.2021.127353, 2022.

https://doi.org/10.5194/essd-2026-166
Preprint. Discussion started: 15 April 2026

Yin, G., Yoshikane, T., Kaneko, R., and Yoshimura, K.: Improving Global Subseasonal to Seasonal Precipitation Forecasts Using a Support Vector Machine-Based Method, Journal of Geophysical Research: Atmospheres, 128, https://doi.org/10.1029/2023JD038929, 2023.

625 Zamora, R. A., Zaitchik, B. F., Rodell, M., Getirana, A., Kumar, S., Arsenault, K., and Gutmann, E.: Contribution of Meteorological Downscaling to Skill and Precision of Seasonal Drought Forecasts, Journal of Hydrometeorology, 22, 2009–2031, https://doi.org/10.1175/JHM-D-20-0259.1, 2021.