Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluation of ERA5-Land and ERA5-Drought Reanalysis Data for SPI/SPEI-Based Drought Monitoring in the Northern Korean Peninsula

YONG SIK HAM

hys12211126@163.com

China University of Geosciences (Beijing)

Wei He

China University of Geosciences (Beijing)

Kyong-Bok Sonu

China University of Geosciences (Beijing)

Research Article

Keywords: Standardized Precipitation Index (SPI), Standardized Precipitation Evapotranspiration Index (SPEI), ERA5-Land, Drought monitoring, Reanalysis evaluation, Democratic People's Republic of Korea (DPRK)

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11697162/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

In regions with sparse ground observation networks, reanalysis-based drought indices are important for understanding spatiotemporal characteristics and early monitoring. However, because reanalysis products differ in their processing and spatial resolution, their reliability may vary; therefore, validation against observations is necessary before operational use. This study evaluates SPI/SPEI derived from ERA5-Land (0.1°) and ERA5-Drought (0.25°) against 154 meteorological stations across the Northern Korean Peninsula (1981–2020). ERA5-Land SPI was calculated from precipitation, and SPEI using the Thornthwaite PET method, whereas ERA5-Drought used pre-computed products provided by ECMWF (Penman-Monteith PET). The evaluation at 1-, 3-, 6-, and 12-month scales used continuous metrics (NSE, R, RMSE, Bias) and categorical drought detection skills (POD, FAR, CSI, HSS) across four climate zones. Overall, ERA5-Land-based SPI/SPEI consistently outperformed ERA5-Drought. The latter showed severe degradation at longer timescales (e.g., NSE = − 0.19 for SPI-6) primarily due to extreme simulated outliers; clipping these improved SPI-6 NSE to + 0.41 and reduced Bias from 0.17 to 0.11, though ERA5-Land remained superior. Spatially, both performed best in the maritime East Coast Zone (ECZ) and worst in complex, mountainous inland zones (NIZ and MIZ). In the 2001 spring drought case study, both failed to capture long-term persistence, and ERA5-Drought exhibited severe phase shifts. These findings highlight that ERA5-Land-based SPI/SPEI is highly suitable for operational drought monitoring in the Northern Korean Peninsula, whereas ERA5-Drought should be used cautiously at 6–12 month scales.

1. Introduction

Drought is a climate-related disaster caused by precipitation deficits that affects water resources, agriculture, and ecosystems over large areas. Unlike other weather disasters, drought develops slowly, persists for extended periods, and expands gradually in space (Wilhite and Glantz, 1985; Wilhite 2000). In the Democratic People's Republic of Korea (DPRK), which lies in the East Asian monsoon region, most of the annual precipitation is concentrated in the summer (June–August), whereas the other seasons (spring, autumn, and winter) are relatively dry. Consequently, meteorological, agricultural, and hydrological droughts recur on seasonal and interannual time scales.

The Standardized Precipitation Index (SPI) and the Standardized Precipitation Evapotranspiration Index (SPEI) are widely used to quantify drought (McKee et al., 1993; Vicente-Serrano et al., 2010). The SPI relies solely on precipitation, whereas the SPEI incorporates the water balance between precipitation and potential evapotranspiration and therefore remains stable under warming conditions (Vicente-Serrano et al., 2010). Both indices are computed at multiple time scales (e.g., 1, 3, 6, and 12 months); conventionally, 1–3 month scales represent meteorological/agricultural drought, while 6–12 month (and longer) scales represent hydrological drought. However, the performance of these indices can vary significantly depending on the time scale, the dataset, and regional climatic characteristics (Aschale et al., 2026; Ceppi et al., 2025; Huang et al., 2026).

Drought monitoring requires reliable long-term data. In regions where observing networks are sparse, reanalysis products serve as an important surrogate for observations (Bodjrenou et al., 2025; Zhan et al., 2016). The ERA5 reanalysis and its high-resolution counterpart ERA5-Land provide long-term precipitation and temperature (Muñoz-Sabater et al., 2021), while ERA5-Drought directly provides global SPI/SPEI indices (Keune et al., 2025). Numerous studies have utilized these datasets for drought assessment across diverse regions, including Iran (Gorjizade and Hafshejani, 2026), Algeria (Ceppi et al., 2025), India (Harishnaika et al., 2026), Brazil (Fernandes et al., 2024), and Morocco (Choukri et al., 2026; Oukaddour et al., 2025).

However, reanalysis products are not free from systematic biases, particularly in precipitation, which can distort drought signals (Oukaddour et al., 2025; Wałęga et al., 2026). For example, Aschale et al. (2026) found that ERA5-Land shows limitations in capturing short-term variability and complex terrain effects, often leading to overestimations of drought frequency. Similarly, studies in India and the US have highlighted biases in monsoon precipitation and temperature from reanalysis products, which limit their ability to measure drought severity accurately (Shah and Mishra, 2014; Mo et al., 2011). Furthermore, Bodjrenou et al. (2025) noted that while reanalyses perform well at monthly and annual scales, their accuracy degrades significantly at daily and hourly scales. Therefore, validation against local station data is essential before applying reanalysis-based indices for regional drought assessment.

Several studies have evaluated reanalysis-based drought indices against observations. Chen et al. (2019) validated multiple reanalysis datasets across China, finding disparities in their ability to represent drought severity. Abu Arra et al. (2025) evaluated ERA5-Land and IMERG in Türkiye, while Wałęga et al. (2026) validated ERA5-Land for meteorological drought in Poland. Crucially, Huang et al. (2026) demonstrated that uncertainties in drought propagation assessments originate more from dataset choice than from index selection, emphasizing the critical need for dataset validation. Reanalysis-based drought indices have also been evaluated against station observations on the Korean Peninsula, but these efforts have focused on earlier reanalysis products and on the southern part of the peninsula. ERA5-Drought was released only in April 2025 and, to the best of our knowledge, has not yet been validated against station data in the northern Korean Peninsula. This study therefore contributes (i) the first station-based validation of ERA5-Drought for this region and (ii) a systematic examination of how extreme index outliers propagate into validation metrics across time scales and climate zones.

A critical but often overlooked issue is how precipitation biases in reanalysis propagate through the distribution-fitting step of SPI/SPEI calculation. Because these indices are obtained by fitting a parametric distribution to accumulated water deficits and transforming the cumulative probability to a standard normal variate, the derivation itself introduces additional uncertainties beyond those in the input variables (Keune et al., 2025). A few severe outliers in the reference period can distort the fitted tail and produce index values that exceed the range of approximately ± 3 (the SPI/SPEI convention; values beyond this range lie on the edge of the distribution and are best treated simply as extreme events rather than as exceptionally extreme; Keune et al., 2025). In a validation framework, these values inflate residual variance and can distort performance metrics even when the bulk of the time series is

reasonably reproduced. This propagation pathway—from input precipitation bias to extreme index outliers to degraded validation metrics—has not been systematically examined for ERA5-Drought in the DPRK context.

The objectives of this study are as follows. First, using observation-based SPI/SPEI at 154 stations across the DPRK for 1981–2020 as the reference, we quantitatively validate and compare SPI/SPEI derived from ERA5-Drought and ERA5-Land, hereafter referred to as SPI/SPEI-ERA5-Drought and SPI/SPEI-ERA5-Land, respectively, at time scales of 1, 3, 6, and 12 months. Second, we assess performance by time scale and climate zone, and evaluate the capability to detect dry events using categorical metrics. This study aims to provide a robust scientific basis for selecting the most suitable dataset and index for drought monitoring and early warning systems in the DPRK.

The paper is structured as follows: Section 2 introduces the study area, data, and methods; Section 3 presents the validation results; Section 4 discusses the findings; and Section 5 concludes the study.

2. Study Area and Methods

2.1 Study Area and Data

The study area covers the northern part of the Korean Peninsula. Figure 1 shows the spatial distribution of the meteorological stations used in this study and the delineated climate zones. Following the climate-zone classification proposed by Om et al. (2024) and Kim et al. (2025) for the northern Korean Peninsula, the stations were grouped into four zones: the North Inland Zone (NIZ), Mid Inland Zone (MIZ), West Coast Zone (WCZ), and East Coast Zone (ECZ).

The monthly precipitation and mean temperature data used in this study were obtained from the State Hydro-Meteorological Administration (SHMA) of the DPRK. In this study, we performed quality control (QC) and homogeneity testing on the raw station records ourselves, following the procedure described by Om et al. (2022) and Kim et al. (2025). Specifically, we conducted (i) a data-quality examination, in which discrepancies between related weather elements were identified using logical conditions and values outside the monthly and annual upper and lower thresholds derived from climatological statistics were flagged and removed; (ii) interpolation of missing values using linear fluctuation trends and relationships with neighboring stations showing the highest correlation over the same period; and (iii) an investigation of station relocations, in which the series were split at the relocation year and subjected to a t-test for the mean and a chi-squared test for the difference between distributions.

Of the 188 SHMA stations, 154 were retained; the remaining 34 were all excluded because of missing monthly records during 1981–2020. The 154 selected stations comprise 30 stations in the NIZ, 30 in the MIZ, 56 in the WCZ, and 38 in the ECZ. Station name, latitude, longitude, and elevation of the 154 selected stations are provided in Table S1. The 1980 record was used only as the spin-up year for the 12-month accumulation of SPI and SPEI, and all index calculations and validation were performed for 1981–2020 (480 months), the common period of the ERA5-Land and ERA5-Drought products.

Three datasets were used in this study: (1) in-situ monthly precipitation and mean temperature from the 154 stations over 1981–2020 (480 months); (2) ERA5-Land monthly precipitation and mean temperature at 0.1° resolution over the same period; and (3) SPI/SPEI directly provided by the ERA5-Drought product at 0.25° resolution (https://cds.climate.copernicus.eu/). For ERA5-Land, the total precipitation variable was used; because this variable is provided as a daily-average rate (m), monthly precipitation was obtained by multiplying the daily-average value by the number of days in each month and converting the unit from m to mm. ERA5-Drought is a dataset produced and distributed directly by ECMWF, and its SPI/SPEI products are computed from ERA5 atmospheric reanalysis precipitation (not ERA5-Land), using Penman-Monteith PET, a 1991–2020 reference period, a Stagge-style zero-precipitation correction, and Shapiro-Wilk quality flags (Keune et al., 2025).

Importantly, the DPRK contributes only a limited number of stations to the WMO Global Telecommunication System (GTS), whereas the present validation uses 154 SHMA stations. The remaining stations, which are not exchanged through the GTS, are almost certainly not assimilated into ERA5 or ERA5-Land. Moreover, the reanalysis grids (0.1° for ERA5-Land and 0.25° for ERA5-Drought) represent area-averaged states that are not tied to any single station. The validation performed here is therefore quasi-independent, providing an external test of the reanalysis-based drought indices rather than a self-consistent comparison.

2.2 Calculation of Drought Indices

2.2.1 SPI

The SPI expresses how much the precipitation over a given period deviates from normal (climatological) conditions on the scale of the standard normal distribution (McKee et al., 1993; WMO, 2012). First, the accumulated precipitation over k months is computed:

$$P_k(t) = \sum_{i=0}^{k-1} P(t-i)$$

where P(t − i) is the precipitation in the i-th month preceding and including the current month; the k-month accumulation is a backward accumulation ending at the current month. The probability distribution of the accumulated precipitation is estimated by fitting a gamma distribution to the monthly series of the 1991–2020 reference period using maximum likelihood estimation (MLE) (Guttman, 1999; Lloyd-Hughes and Saunders, 2002; WMO, 2012). The gamma fit is performed separately for each accumulation scale k and for each calendar month; that is, January values are fitted to January data, February values to February data, and so on, and this procedure is repeated independently for each k (WMO, 2012). The probability density function of the gamma distribution is:

$$f(x) = \frac{x^{\alpha-1}\exp(-x/\beta)}{\beta^{\alpha}\,\Gamma(\alpha)}, \quad x > 0$$

where α, β, and Γ denote the shape parameter, the scale parameter, and the gamma function, respectively. Because zero-precipitation months exist in the precipitation series, the cumulative probability accounting for the zero-precipitation fraction p_0 in the reference period is defined as (Guttman, 1999):

$$G(x) = p_0 + (1 - p_0) F_\Gamma\ (x; \alpha, \beta), \quad x > 0,\ G(0) = p_0$$

Finally, the SPI is obtained by transforming the cumulative probability G using the inverse of the standard normal cumulative distribution function Φ^{-1}:

$$SPI = \Phi^{-1}(G(P_k))$$

2.2.2 SPEI

The SPEI is based on the water balance between precipitation and potential evapotranspiration (Vicente-Serrano et al., 2010). The potential evapotranspiration (PET) required for the SPEI is estimated using the Thornthwaite (1948) method, which requires only the monthly mean temperature and latitude. This method was chosen because it can be applied identically to both the observation and ERA5-Land datasets, thereby ensuring methodological consistency in the SPEI calculation. The monthly heat index is first computed as

$$I_m = \left(\frac{T_m}{5}\right)^{1.514}, \quad T_m > 0$$

$$I_m = 0, \quad T_m \le 0$$

$$I = \sum_{m=1}^{12} I_m$$

where *Tm* is the monthly mean temperature in °C. The correction exponent *a* is given by

$$a = 6.75 \times 10^{-7} I^3 - 7.71 \times 10^{-5} I^2 + 1.792 \times 10^{-2} I + 0.49239$$

The unadjusted potential evapotranspiration is

$$PET_u = 16\left(\frac{10 T_m}{I}\right)^a \quad (\mathrm{mm/month})$$

and is corrected by the day length and the number of days per month:

$$PET = PET_u \times \frac{N_m}{12} \times \frac{d_m}{30}$$

where *Nm* is the mean day length (hours) and *dm* is the number of days in month *m*. The day length is obtained as

$$N_m = \frac{24}{\pi}\arccos\left(-\tan\phi \cdot \tan\delta\right)$$

$$\delta = 0.409\sin\left(\frac{2\pi J}{365} - 1.39\right)$$

Where ϕ is the latitude, δ is the solar declination, and *J* is the Julian day of the middle of the month. When $Tm \leq 0$, PET is set to zero.

Using the estimated PET, the monthly water balance is computed as

$$D(t) = P(t) - PET(t)$$

The accumulated water balance over a time scale of *k* months is

$$D_k(t) = \sum_{i=0}^{k-1} D(t-i)$$

Because the accumulated water balance can take negative values, a three-parameter log-logistic distribution is fitted to $D_k(t)$ over the 1991–2020 reference period using the probability weighted moments method (Vicente-Serrano et al., 2010). The cumulative probability is

$$F(x) = \left[1 + \left(\frac{\alpha}{x-\gamma}\right)^{\beta}\right]^{-1}, \quad x > \gamma$$

where α, β, and γ are the scale, shape, and location parameters, respectively. Finally, the SPEI is obtained by transforming the cumulative probability *F* through the inverse cumulative distribution function of the standard normal distribution:

$$SPEI = \Phi^{-1}(F(D_k))$$

2.3 Performance Metrics

To compare the gridded reanalysis products with station observations, bilinear interpolation was applied to derive the grid value at each station location (Zhan et al., 2016; Shah and Mishra, 2014). Area-based weights were assigned to the four grid cells surrounding each station, and grid cells with missing values were excluded from the weighted average. When all four surrounding grid cells were missing, the value of the nearest valid grid cell was used. This procedure was applied separately to the SPI/SPEI-ERA5-Land and SPI/SPEI-ERA5-Drought gridded datasets, using only time steps common to the observed and reanalysis series.

Using these matched series, the agreement between the observation- and reanalysis-based SPI/SPEI was quantified. Aggregate metrics combining all stations and time steps, as well as station-, time-, and climate-zone-specific metrics, were computed. For the observed $\mathrm{o_i}$ and reanalysis $\mathrm{e_i}$ values

$(\mathrm{i} = 1, \ldots, \mathrm{N})$, the Nash–Sutcliffe efficiency (NSE), Pearson correlation coefficient (R), root mean square error (RMSE), and bias are defined as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{N} \left(y_{obs,i} - y_{pred,i}\right)^2}{\sum_{i=1}^{N} \left(y_{obs,i} - \bar{y}_{obs}\right)^2}$$

$$R = \frac{\sum \left(o_i - \bar{o}\right)\left(e_i - \bar{e}\right)}{\sqrt{\sum \left(o_i - \bar{o}\right)^2 \cdot \sum \left(e_i - \bar{e}\right)^2}}$$

$$RMSE = \sqrt{\frac{1}{N} \sum \left(o_i - e_i\right)^2}$$

$$Bias = \frac{1}{N} \sum \left(e_i - o_i\right)$$

NSE, originally proposed by Nash and Sutcliffe (1970) and widely adopted for evaluating hydrological and drought-index simulations (e.g., Chen et al., 2019; Aschale et al., 2026), measures the fraction of the observed variance explained by the reanalysis. An NSE of unity indicates perfect agreement, whereas an NSE of zero implies that the reanalysis performs no better than the observed mean; negative values indicate performance worse than the climatological mean. R describes the strength of the linear association between the observed and reanalysis time series and is insensitive to systematic differences in magnitude (Wilks, 2011). RMSE measures the overall magnitude of the errors and is more sensitive to large, infrequent errors (Wilks, 2011; Chai, 2014). Bias quantifies the mean systematic offset between the two series, indicating whether the reanalysis tends to over- or under-estimate the index on average (Wilks, 2011). Bias is defined as the mean of (reanalysis − observation); thus, a positive bias indicates systematic over-estimation and a negative bias under-estimation.

Because drought monitoring ultimately depends on the correct identification of drought conditions, the ability of the reanalysis-based indices to detect dry (drought) conditions was additionally evaluated using contingency-table-based skill scores, following the terminology of Wilks (2011). Four statistics were considered: the probability of detection (POD), the false alarm ratio (FAR), the critical success index (CSI, also known as the threat score), and the Heidke skill score (HSS). POD is the fraction of observed drought months that are successfully detected by the reanalysis; FAR is the fraction of detected drought months that are in fact false alarms; CSI combines POD and FAR into a single measure and represents the proportion of correctly classified months among all months that were either observed or detected as drought, thereby treating misses and false alarms symmetrically; HSS expresses the fraction of correct monthly classifications relative to the correct classifications expected by chance, so it is a more demanding measure than the plain hit rate and is suitable for comparing skill across datasets with different base rates.

Drought months were identified using the conventional SPI/SPEI drought-class thresholds of McKee et al. (1993), as adopted for SPEI by Vicente-Serrano et al. (2010), with a drought month defined as SPI or SPEI ≤ − 1.00 (moderate drought or more severe). Unlike many previous studies that impose a minimum duration (e.g., two or more consecutive months), no minimum-duration criterion was applied in the main analysis; each month with SPI or SPEI ≤ − 1.00 in the observed series was classified as a drought month. This choice was made to avoid underestimating drought occurrence by excluding short episodes, given that the purpose of this study is a sensitivity evaluation of the two sets of reanalysis-based drought indices (SPI/SPEI-ERA5-Land and SPI/SPEI-ERA5-Drought). As a sensitivity check, we additionally repeated the categorical evaluation by applying a 2-month minimum-duration criterion. Categorical metrics (POD, FAR, CSI, HSS) were computed on a monthly basis, whereas drought event characteristics were derived from continuous drought runs. Here, N is the number of drought events (continuous runs of drought months), average drought duration $(ADD) = \frac{1}{N}\sum_{j=1}^{N} D_j$ is the mean event duration (with D_j the duration of event j), and the average drought severity $(ADS) = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{D_j}\sum_{t\in E_j}|I_t|\right)$ is the mean event severity, defined as the average absolute index value within an event. The contingency table was then constructed for each dataset, index, time scale, and climate zone.

Finally, to provide an actionable diagnostic, we cross-checked the extreme outliers (|z| > 3) against the ERA5-Drought quality flags. For each calendar month, the provided Shapiro-Wilk significance mask (Shapiro & Wilk, 1965; Keune et al., 2025) (0 = non-significant, 1 = significant) was applied to the corresponding time step. We then calculated the ratio of extreme outliers (separately for z > 3 and z < -3) occurring in grid cells with a quality flag of 0, to assess the structural link between extreme anomalies and low-quality cells over the 1981–2020 period.

3. Results

3.1 Overall performance

Table 1 summarizes the overall continuous performance of SPI/SPEI-ERA5-Land and SPI/SPEI-ERA5-Drought in reproducing observed SPI and SPEI at 1-, 3-, 6-, and 12-month scales, pooled across all 154 stations and 480 months over the northern Korean Peninsula during 1981–2020.

Overall, SPI/SPEI-ERA5-Land consistently outperformed SPI/SPEI-ERA5-Drought across all eight index–scale combinations, exhibiting higher NSE and R values, as well as lower RMSE and Bias. Notably, ERA5-Land-based indices maintained a relatively stable performance across scales (NSE ranging from 0.51 to 0.64), whereas ERA5-Drought-based indices showed a marked degradation at longer accumulation periods. For instance, the NSE of SPEI-ERA5-Drought dropped from 0.44 at the 1-month scale to 0.37 at the 6-month scale, and the SPI-ERA5-Drought index even turned negative (− 0.19) at the 6-month scale, contrasting sharply with the robust performance of SPI-ERA5-Land (NSE = 0.54). However, this degradation was not monotonic for SPEI: the NSE of SPEI-ERA5-Drought recovered to 0.50 at the 12-month scale, higher than at the 3-month (0.38) and 6-month (0.37) scales.

The performance gap between the two sets of indices was particularly pronounced for SPI compared to SPEI. While both sets of indices showed a small positive mean bias, the ERA5-Drought indices exhibited a larger systematic overestimation (Bias up to + 0.17 for SPI-6 and SPI-12) relative to the ERA5-Land indices (Bias ≤ + 0.09). Furthermore, the RMSE for the ERA5-Drought indices reached its maximum (1.09) at SPI-6, highlighting their limitations in capturing mid-term drought variability compared to the finer-resolution ERA5-Land-based indices.

Table 1
Overall performance of ERA5-Land- and ERA5-Drought based Indices in reproducing observed SPI and SPEI over the northern Korean Peninsula during 1981–2020 (N = 73,920).

Time scale	Dataset	SPI				SPEI			
		NSE	R	RMSE	Bias	NSE	R	RMSE	Bias
1	ERA5-Land	0.63	0.82	0.61	0.02	0.64	0.82	0.60	0.02
	ERA5-Drought	0.29	0.72	0.85	0.09	0.44	0.75	0.75	0.01
3	ERA5-Land	0.58	0.79	0.66	0.05	0.60	0.8	0.64	0.04
	ERA5-Drought	0.23	0.69	0.89	0.09	0.38	0.73	0.79	0.04
6	ERA5-Land	0.54	0.78	0.68	0.07	0.57	0.78	0.66	0.04
	ERA5-Drought	-0.19	0.60	1.09	0.17	0.37	0.73	0.79	0.07
12	ERA5-Land	0.51	0.77	0.68	0.09	0.54	0.77	0.66	0.05
	ERA5-Drought	0.13	0.66	0.91	0.17	0.50	0.77	0.69	0.08

3.2 Categorical performance for drought detection

Table 2 reports the categorical performance of the ERA5-Land-based and ERA5-Drought-based indices in detecting drought events defined as SPI/SPEI ≤ − 1.0 at 1-, 3-, 6-, and 12-month scales, pooled over all 154 stations and 480 months.

The POD was higher for the ERA5-Land-based indices than for the ERA5-Drought-based indices at seven of eight index–scale combinations. POD for the ERA5-Land indices ranged from 0.48 (SPI-12) to 0.65 (SPEI-1), whereas the ERA5-Drought indices ranged from 0.46 (SPI-12) to 0.61 (SPEI-1). At SPEI-12, POD was 0.50 for both sets of indices. The largest POD gap was at SPI-1 (0.64 vs. 0.59, gap = 0.05), and the smallest was at SPEI-12.

The FAR showed a mixed pattern. The ERA5-Land indices had lower FAR at three combinations (SPEI-1, SPEI-3, and SPEI-6), while the ERA5-Drought indices had lower FAR at three combinations (SPI-1: 0.35 vs. 0.37; SPI-6: 0.42 vs. 0.43; SPI-12: 0.45 vs. 0.46); SPI-3 and SPEI-12 were tied (0.40 vs. 0.40 and 0.43 vs. 0.43, respectively). FAR increased with accumulation scale for both sets of indices: from 0.37 (ERA5-Land, SPI-1) to 0.46 (SPI-12), and from 0.33 (ERA5-Land, SPEI-1) to 0.43 (SPEI-12).

The CSI and HSS both favored the ERA5-Land-based indices. CSI for the ERA5-Land indices ranged from 0.34 (SPI-12) to 0.49 (SPEI-1), and CSI for the ERA5-Drought indices ranged from 0.34 (SPI-12) to 0.46 (SPEI-1). HSS for the ERA5-Land indices ranged from 0.44 (SPI-12) to 0.59 (SPEI-1), and HSS for the ERA5-Drought indices ranged from 0.43 (SPI-12) to 0.55 (SPEI-1). At SPEI-12, CSI was 0.36 and HSS was 0.46 for both sets of indices.

All four categorical metrics deteriorated with increasing accumulation scale for both sets of indices. The highest POD, CSI, and HSS values occurred at 1-month scales, and the lowest at 12-month scales.

Table 2
Categorical performance of ERA5-Land and ERA5-Drought in detecting drought events (SPI/SPEI ≤ − 1.0) at 1-, 3-, 6-, and 12-month scales.

Time scale	Dataset	SPI				SPEI			
		POD	FAR	CSI	HSS	POD	FAR	CSI	HSS
1	ERA5-Land	0.64	0.37	0.47	0.57	0.65	0.33	0.49	0.59
	ERA5-Drought	0.59	0.35	0.45	0.56	0.61	0.36	0.46	0.55
3	ERA5-Land	0.60	0.40	0.43	0.53	0.61	0.37	0.45	0.55
	ERA5-Drought	0.58	0.40	0.42	0.52	0.58	0.39	0.42	0.52
6	ERA5-Land	0.56	0.43	0.40	0.49	0.57	0.40	0.41	0.51
	ERA5-Drought	0.55	0.42	0.39	0.49	0.55	0.42	0.39	0.49
12	ERA5-Land	0.48	0.46	0.34	0.44	0.50	0.43	0.36	0.46
	ERA5-Drought	0.46	0.45	0.34	0.43	0.50	0.43	0.36	0.46

As a sensitivity check, we repeated the categorical evaluation using a 2-month minimum-duration criterion (Table S2). Applying this stricter criterion generally reduced the absolute skill scores for both products, reflecting the increased difficulty of detecting persistent droughts. However, the relative ranking remained largely unchanged: the ERA5-Land-based indices consistently maintained higher POD, CSI, and HSS than the ERA5-Drought-based indices across most index–scale combinations (e.g., SPEI-3 HSS: 0.54 vs. 0.51; SPEI-1 HSS: 0.44 vs. 0.39). Notably, at SPI-3, both products performed nearly identically (HSS = 0.50), indicating that the advantage of ERA5-Land is more pronounced for short-term droughts and largely disappears at the 6–12 month scales.

3.3 Climate-zone-specific performance

The 154 stations were stratified into four climate zones (NIZ, MIZ, WCZ, ECZ) to examine whether the reanalysis performance varied systematically with climate regime. Figure 2 summarizes the continuous evaluation (NSE and R) as a bubble heatmap, and Fig. 3 presents the drought-event characteristics (N, ADD, ADS) as boxplots.

Across all four climate zones and all eight index–scale combinations, SPI/SPEI-ERA5-Land consistently yielded higher NSE values than SPI/SPEI-ERA5-Drought. The ECZ exhibited the strongest reproducibility, with ERA5-Land NSE values of 0.57–0.65 (R = 0.79–0.83), whereas the NIZ showed the weakest, with ERA5-Land NSE values of 0.46–0.60 (R = 0.73–0.80). The MIZ and WCZ were intermediate, with ERA5-Land NSE ranging from 0.48 to 0.66 and from 0.50 to 0.65, respectively.

The Land–Drought gap was most pronounced at the intermediate time scales. Over the MIZ, SPI-ERA5-Land achieved an NSE of 0.56 at SPI-6, whereas SPI-ERA5-Drought yielded − 0.90 at the same scale. Comparable negative NSE values for ERA5-Drought occurred over the NIZ at SPI-3 (− 0.90; the same value as MIZ–SPI-6) and over the WCZ at SPI-12 (− 0.36). At the 1-month scale, NSE was positive for both sets of indices in most zones, with the sole exception of SPEI-ERA5-Drought in the MIZ (− 0.11). At the 12-month scale, the ERA5-Drought indices remained positive in the MIZ (SPI-12: 0.44) but turned negative in the WCZ (SPI-12: −0.36). Pearson correlation coefficients were comparatively insensitive to climate zone. For ERA5-Land, R ranged from 0.73 to 0.83; for ERA5-Drought, from 0.48 to 0.82. The MIZ, WCZ, and ECZ showed similar R values, whereas the NIZ was lowest.

Drought events were identified by run theory with a threshold of SPI/SPEI ≤ − 1.0 (Yevjevich, 1967; Sharma and Panu, 2012; McKee et al., 1993). An event was defined as a continuous run of months with SPI/SPEI ≤ − 1.0 and was terminated when the index exceeded − 1.0 in the following month; thus, months with an index greater than − 1.0 separate consecutive events. For each station and climate zone, N, ADD, and ADS were extracted. Figure 3 presents the boxplots for the 1- and 6-month scales; the corresponding boxplots for the 3- and 12-month scales are provided in Fig. S1. The per-station values averaged over all 154 stations are summarized for each index–scale combination in Table S3, and the mean event characteristics for all zone–scale combinations are listed in Table S4.

When averaged over all 154 stations, the ERA5-Land-based indices reproduced the observed event count and average duration reasonably well at most scales, but tended to overestimate the average severity. For example, at the 6-month scale, ERA5-Land yielded N = 28.7 vs. 30.4, ADD = 2.44 vs. 2.35 months, and ADS = 1.22 vs. 1.14 for SPI, and N = 28.3 vs. 31.4, ADD = 2.49 vs. 2.40 months, and ADS = 1.09 vs. 1.08 for SPEI. The ERA5-Drought-based indices showed a similar tendency but with a stronger overestimation of ADS, particularly at the 12-month scale (SPI-12: ADS = 2.94 vs. 2.28; SPEI-12: ADS = 2.82 vs. 2.32).

The zone-stratified boxplots showed that the ADS overestimation was not confined to a single zone but appeared across the NIZ, MIZ and WCZ. For example, at SPEI-6 over the NIZ, observed ADS was 1.15, whereas SPEI-ERA5-Drought yielded 1.65 (Δ = +0.50) and SPEI-ERA5-Land 1.07 (Δ=−0.08). At the 1–6 month scales, the ERA5-Drought indices overestimated ADS in all zones, generally more strongly than the ERA5-Land indices. At the 12-month scale, however, both products substantially overestimated ADS in the MIZ and WCZ (e.g., SPI-12: Δ ≈ +1.1 for both datasets). Over the ECZ, both sets of indices reproduced ADS within approximately ± 0.3 at the 1- and 6-month scales.

Overall, the distributions of all three event characteristics from the ERA5-Land indices were closer to the observed distributions than those from the ERA5-Drought indices, most notably for ADS at SPI-1, where

ERA5-Drought overestimated ADS in the WCZ and ECZ.

In addition to continuous metrics and drought event characteristics, the categorical skill in detecting drought months (POD, FAR, CSI, and HSS) varied systematically across climate zones (Table 3). The ERA5-Land-based indices consistently exhibited higher POD, CSI, and HSS values than the ERA5-Drought-based indices across almost all zones and index-scales. However, the FAR showed a mixed pattern: the ERA5-Land indices generally exhibited a larger FAR than the ERA5-Drought indices for SPI, whereas they exhibited a smaller FAR for SPEI. Both sets of indices generally showed the highest detection skill in the MIZ and WCZ at the shorter accumulation scales (especially the 1-month scale), whereas the ECZ did not show a clear advantage at the longer scales. At SPI-12 and SPEI-12, the HSS for the ECZ was essentially tied with that for the NIZ (0.47/0.47 and 0.49/0.49, respectively). This contrasts with the continuous metrics, where the ERA5-Land-based indices in the ECZ region consistently showed the highest NSE. In other words, the advantage of the ECZ region is more evident in continuous reproduction than in categorical drought-month detection at long scales.

For instance, at SPEI-1, SPEI-ERA5-Land achieved the highest HSS of 0.63 in the WCZ and 0.62 in the MIZ, while effectively minimizing false alarms (FAR = 0.29 and 0.30, respectively). In contrast, the NIZ generally showed the lowest categorical performance. For the ERA5-Land indices in the NIZ, HSS decreased with increasing accumulation scale—from 0.56 at SPI-1 to 0.47 at SPI-12—and was lower than in the other zones at three of the four SPI scales, indicating that the complex topography of the northern inland region significantly degrades the classification of drought months, particularly at longer accumulation periods. Furthermore, for both sets of indices, categorical metrics systematically deteriorated as the accumulation scale increased from 1 to 12 months. The highest CSI and HSS values occurred at the 1-month scale, whereas the 12-month scale showed the lowest skill and the highest FAR overall (e.g., FAR = 0.46 for SPI-ERA5-Land at SPI-12).

Table 3
Climate-zone-specific categorical performance of ERA5-Land- and ERA5-Drought-based Indices in detecting drought months (SPI/SPEI ≤ -1.0). Values (L/D) are presented as ERA5-Land / ERA5-Drought-based Indices.

Index	Time scale	Metric	NIZ (L/D)	MIZ (L/D)	WCZ (L/D)	ECZ (L/D)
SPI	1	POD	0.62 / 0.60	0.66 / 0.62	0.67 / 0.61	0.65 / 0.61
		FAR	0.38 / 0.37	0.34 / 0.32	0.33 / 0.30	0.37 / 0.35
		CSI	0.45 / 0.44	0.49 / 0.48	0.50 / 0.48	0.47 / 0.46
		HSS	0.56 / 0.55	0.60 / 0.59	0.61 / 0.59	0.58 / 0.57
	3	POD	0.59 / 0.58	0.62 / 0.60	0.59 / 0.57	0.63 / 0.62
		FAR	0.42 / 0.41	0.38 / 0.37	0.39 / 0.39	0.39 / 0.38
		CSI	0.42 / 0.41	0.45 / 0.45	0.43 / 0.42	0.45 / 0.46
		HSS	0.51 / 0.51	0.55 / 0.55	0.53 / 0.52	0.56 / 0.56
	6	POD	0.57 / 0.55	0.60 / 0.57	0.56 / 0.55	0.59 / 0.58
		FAR	0.45 / 0.45	0.39 / 0.38	0.39 / 0.38	0.41 / 0.41
		CSI	0.39 / 0.39	0.44 / 0.42	0.42 / 0.41	0.42 / 0.42
		HSS	0.48 / 0.48	0.53 / 0.52	0.52 / 0.51	0.52 / 0.52
	12	POD	0.54 / 0.53	0.50 / 0.47	0.48 / 0.46	0.52 / 0.51
		FAR	0.44 / 0.44	0.41 / 0.41	0.41 / 0.40	0.44 / 0.44
		CSI	0.38 / 0.38	0.37 / 0.35	0.36 / 0.36	0.37 / 0.37
		HSS	0.47 / 0.47	0.46 / 0.45	0.46 / 0.46	0.47 / 0.47
SPEI	1	POD	0.62 / 0.59	0.68 / 0.62	0.68 / 0.63	0.64 / 0.61
		FAR	0.34 / 0.37	0.30 / 0.34	0.29 / 0.30	0.35 / 0.36
		CSI	0.47 / 0.44	0.53 / 0.47	0.54 / 0.50	0.48 / 0.45
		HSS	0.57 / 0.53	0.62 / 0.56	0.63 / 0.60	0.58 / 0.55
	3	POD	0.60 / 0.60	0.64 / 0.60	0.60 / 0.57	0.63 / 0.60
		FAR	0.37 / 0.42	0.34 / 0.37	0.37 / 0.39	0.35 / 0.36
		CSI	0.45 / 0.42	0.48 / 0.45	0.45 / 0.42	0.47 / 0.45
		HSS	0.55 / 0.51	0.58 / 0.54	0.54 / 0.51	0.57 / 0.55
	6	POD	0.55 / 0.54	0.59 / 0.58	0.57 / 0.55	0.60 / 0.57
		FAR	0.41 / 0.44	0.38 / 0.38	0.38 / 0.38	0.37 / 0.38

Index	Time scale	Metric	NIZ (L/D)	MIZ (L/D)	WCZ (L/D)	ECZ (L/D)
		CSI	0.40 / 0.38	0.44 / 0.43	0.43 / 0.41	0.45 / 0.42
		HSS	0.49 / 0.47	0.53 / 0.52	0.52 / 0.51	0.55 / 0.52
	12	POD	0.54 / 0.55	0.52 / 0.52	0.51 / 0.51	0.53 / 0.53
		FAR	0.40 / 0.41	0.40 / 0.42	0.41 / 0.41	0.41 / 0.41
		CSI	0.40 / 0.40	0.39 / 0.38	0.38 / 0.37	0.39 / 0.39
		HSS	0.49 / 0.49	0.48 / 0.47	0.48 / 0.47	0.49 / 0.49

3.4 Spatial distribution of NSE

Figure 4 maps the NSE distribution between observed and reanalysis-based SPI and SPEI across the northern Korean Peninsula for each index–scale combination.

The ERA5-Land-based indices (Figs. 4a–d for SPI, Figs. 4i–l for SPEI) showed predominantly positive NSE, with NSE ranging from approximately 0.0 to 0.82 (mean 0.51–0.64). Slightly higher values appeared over the WCZ and ECZ, while slightly lower values appeared over the NIZ. A localized near-zero minimum occurred at the NIZ station near 127.13°E, 40.97°N (NSE = − 0.005 at SPI-12, Rangrim station), appearing as a white patch within the NIZ. SPI-1 and SPEI-1 showed higher NSE (mean 0.63–0.64) than SPI-12 and SPEI-12 (0.51–0.54).

The ERA5-Drought-based indices (Figs. 4e–h for SPI, Figs. 4m–p for SPEI) showed a more heterogeneous and substantially weaker pattern, with NSE ranging from − 7.88 to 0.82 (mean − 0.25 to 0.49). Negative NSE was widespread rather than localized: interpolated grid minima reached − 0.99 to − 1.00 over the SPI-3, SPI-6, and SPI-12 panels and − 0.97 to − 1.00 over SPEI-1, SPEI-3, and SPEI-6. Only SPEI-12 remained entirely positive (grid minimum 0.088). On average, the ECZ consistently showed higher NSE than the inland zones. The spatial contrast between the ERA5-Land indices and the ERA5-Drought indices was largest at SPI-6 and SPI-12 and smallest at SPEI-12. Note that the zone-mean NSE values in Section 3.3 are station-based averages, whereas the maps in Fig. 4 show interpolated grid-cell values; the localized negative cells in the maps can therefore be smoothed out in the station-based zone averages, so the two representations are complementary rather than contradictory.

Comparing SPI and SPEI, the negative NSE of the ERA5-Drought indices was more widely distributed for the SPI series, whereas it was relatively alleviated for the SPEI series, with all grid cells remaining positive at SPEI-12. In contrast, the ERA5-Land indices showed only small differences between SPI and SPEI, with both indices exhibiting a similar level of positive NSE distribution.

3.5 Representative time series and the 2001 case study

Figure 5 shows the observed time series and those of SPI/SPEI-ERA5-Land and SPI/SPEI-ERA5-Drought at the stations with the maximum and minimum station-wise NSE-Land for each of SPI-1, SPI-6, SPEI-1,

and SPEI-6 (left and right columns, respectively). The best-performing station was Kyonghung (ECZ) for SPI-1 and SPEI-1, and Sinhung (ECZ) for SPI-6 and SPEI-6. At Kyonghung, SPI-ERA5-Land achieved an NSE of 0.73 at SPI-1, and SPEI-ERA5-Land achieved 0.75 at SPEI-1, with R of 0.87 at both scales; SPI-ERA5-Drought was substantially lower for SPI-1 (NSE = − 0.30 vs. 0.73), although the gap narrowed at SPEI-1 (SPEI-ERA5-Drought: 0.69 vs. SPEI-ERA5-Land: 0.75). At Sinhung, SPI-ERA5-Land and SPI-ERA5-Drought performed almost identically for SPI-6 (NSE = 0.71 vs. 0.72; R = 0.86 vs. 0.86) and SPEI-6 (NSE = 0.73 vs. 0.71; R = 0.87 vs. 0.86), indicating that the two products converged at this station at the 6-month scale. Notably, at stations where ERA5-Land performed well, ERA5-Drought generally performed well too; in particular, at Sinhung for SPI-6, SPI-ERA5-Drought (NSE = 0.72) slightly exceeded SPI-ERA5-Land (0.71), demonstrating that it can match or even outperform ERA5-Land.

The worst-performing station was Rangrim (NIZ) for SPI-1, SPI-6, and SPEI-6, and Taehung-2 (ECZ) for SPEI-1. At Rangrim, NSE for the ERA5-Land indices was low across scales and that for the ERA5-Drought indices was even lower at SPI-1 (NSE = − 1.03 vs. 0.18), although the two products were similar at SPI-6 (0.09 vs. 0.09) and SPEI-6 (0.17 vs. 0.15). At Taehung-2 (SPEI-1), the ERA5-Land and ERA5-Drought indices were also comparable (NSE = 0.16 vs. 0.15; R = 0.58 vs. 0.58). Notably, Taehung-2 is assigned to the ECZ but lies at a high elevation (896 m; 128.87°E, 41.05°N; Table S1); its poor performance is therefore more likely related to elevation and local topographic effects than to the ECZ as a whole. Overall, the worst stations were concentrated in the NIZ and ECZ, with the ERA5-Land and ERA5-Drought indices diverging most strongly at SPI-1 over Rangrim. Because these four stations are illustrative rather than a representative sample of their zones, zone-level conclusions should rely on the full station set.

As shown in Fig. 5, the ERA5-Drought indices (red) systematically deviated from the observed series (black) at extreme values—underestimating drought peaks and overestimating wet extremes—whereas the ERA5-Land indices (blue) tracked the observations more closely. This tendency was most pronounced at the worst-performing stations (e.g., Rangrim), where large negative NSE values coincided with the strongest extreme-value mismatches.

Because the best- and worst-performing stations in Fig. 5 were selected by NSE-Land, this comparison is structurally biased toward the best and worst performance of ERA5-Drought. To examine this selection bias, we additionally analyzed one randomly selected station per climate zone (Fig. S2). At these randomly chosen stations, the two products largely converged, with a mean NSE of 0.585 for ERA5-Land and 0.491 for ERA5-Drought, indicating that the extreme cases in Fig. 5 do not represent the typical station-level behavior.

The 2001 spring drought was the most spatially extensive drought on record over the DPRK, affecting 68.6% of the national territory (Kim et al., 2025). To evaluate this event across climate zones, Fig. 6 compares station observations, ERA5-Land, and ERA5-Drought at five stations representing the WCZ (Sariwon and Haeju), MIZ (Koksan), ECZ (Chonnae), and NIZ (Ryongrim) throughout 2001 using the 1- and 6-month SPI and SPEI (3- and 12-month results in Fig. S3).

At the 1-month scale, both products broadly reproduced the April 2001 dry peak but underestimated its magnitude (e.g., at Sariwon, observed − 3.25 vs. ERA5-Land − 2.73 and ERA5-Drought − 2.17; at Haeju, − 3.45 vs. −2.78 and − 2.32). A clearer phase difference appears in August–September, when the observations remained below − 1.0 (Sariwon: −1.29 and − 1.03) but ERA5-Drought returned to near-normal values (− 0.31 and − 0.13), indicating a premature termination of the drought in the reanalysis. ERA5-Land tracked the observed trajectory more steadily at most sites. However, the error direction was not uniform across all stations. At Chonnae, ERA5-Drought overestimated drought for SPI-1 (May: −3.94 vs. observed − 1.67; November: −3.10 vs. −1.13), whereas SPEI-1 remained relatively stable.

The 6-month panels reveal the central limitation: while dry conditions persisted well beyond spring in the observations, both products rebounded toward near-zero or positive values after June, missing the long tail of the drought (e.g., at Sariwon, observed SPI-6 remained below − 1.0 from March to June, whereas both products exceeded − 1.0 after May). Conversely, some stations showed errors in the opposite direction. At Ryongrim for SPEI-6, ERA5-Drought produced − 7.72 in April (observed − 1.44), and at Koksan for SPEI-6, it produced − 5.14 in June (observed − 0.52), indicating a strong overestimation of drought at these sites. This combination of failed persistence and localized overestimation explains the degradation of skill at longer scales.

The SPI–SPEI comparison shows that both indices produced qualitatively similar patterns—an early rebound and a missed long tail—although errors near the drought onset were somewhat smaller for SPEI-1. This similarity indicates that the key limitation lies in the representation of accumulated drought persistence, not in the choice of index.

The 3- and 12-month results (Fig. S3) extend this picture. At the 3-month scale, ERA5-Land reproduced the May peak comparatively well, whereas ERA5-Drought underestimated it, and the two products occasionally diverged in error direction. At the 12-month scale, the discrepancy was maximized: observations maintained negative anomalies throughout the year, but neither product captured the signal, and ERA5-Drought even produced extreme positive anomalies in spring (e.g., Sariwon SPI-12: +7.12 in March 2001), opposite in phase to observations.

Overall, the 2001 case study confirms the pattern identified in Section 3.3: both reanalyses capture short-term dry peaks but increasingly fail to reproduce drought persistence as the timescale lengthens. ERA5-Land generally outperformed ERA5-Drought in the stability of the year-round trajectory and the reproduction of extreme values, although this superiority was not universal. For example, ERA5-Drought showed higher NSE than ERA5-Land for SPI-3 and SPEI-3 at Haeju and Koksan, for SPEI-1 at Chonnae and Koksan, and for SPEI-12 at Koksan.

3.6 Frequency and distribution of extreme index values

Figure 7 summarizes the frequency of |z| > 3 values in the observed, ERA5-Land, and ERA5-Drought indices across all eight index-scale combinations. Overall, ERA5-Drought produced extreme values at 4–15 times the rate of the observations, whereas ERA5-Land remained close to the observed frequency

across all eight combinations (ERA5-Land: 0.01–0.42% vs. observed: 0.04–0.45%). The largest discrepancy occurred for SPEI-3 (0.69% vs. 0.04%) and SPEI-6 (0.99% vs. 0.07%). The SPI-ERA5-Drought frequencies ranged from 1.17% (SPI-3) to 2.03% (SPI-6), compared with 0.28–0.45% in the observations.

The zone-wise breakdown (Figs. 7c–f) shows that the excess extremes in ERA5-Drought indices were not uniformly distributed but concentrated in the MIZ and WCZ. At the 6-month scale (panel e), ERA5-Drought produced $|z| > 3$ values for SPI at rates of 3.31% in the MIZ and 2.75% in the WCZ, compared with 0.25% and 0.44% in the observations. At the 12-month scale (Fig. 7f), the WCZ rate for SPI reached 3.65% (observed: 0.44%). A comparable concentration appeared for SPEI at the 6-month scale, with ERA5-Drought reaching 2.13% in the MIZ and 1.27% in the NIZ (observed: 0.08% and 0.02%). The sign distribution (Table S5) further reveals a systematic asymmetry: ERA5-Drought predominantly produced positive extremes for SPI (e.g., 1350 vs. 148 at SPI-6) but negative extremes for SPEI (e.g., 100 vs. 408 at SPEI-3). The observed extremes are also not balanced in sign—for example, observed SPI-6 has 225 positive versus 37 negative extremes, whereas SPI-1 has 88 positive versus 242 negative extremes—reflecting the scale-dependent skewness inherent in the gamma-fitted SPI. However, the asymmetry in ERA5-Drought is much larger and more scale-dependent than in the observations, indicating that the ERA5-Drought extremes are amplified by input biases rather than simply reflecting the intrinsic skewness of the index.

4. Discussion

In Table 1, both reanalyses showed a positive bias across all scales, with ERA5-Drought exhibiting a larger offset than ERA5-Land. As a pooled average, the sign and magnitude may vary at individual stations. Rather than being an artifact of a few extreme values, the bias is systematic: it persists across stations and scales and reflects a systematic offset in the precipitation forcing that is only partly absorbed by the standardization of the index. This highlights the inherent uncertainty of reanalysis-based indices and the need for bias correction in operational monitoring. Note that ERA5 precipitation is often underestimated over complex terrain (Aschale et al., 2026; Oukaddour et al., 2025); the net wet bias found here therefore does not imply that ERA5 precipitation is uniformly overestimated, but rather that the bias is region- and scale-dependent.

The ERA5-Land-based indices outperformed the ERA5-Drought-based indices across all eight index-scale combinations and at every timescale (Tables 1 and 2). While its finer resolution (0.1° vs. 0.25°) and elevation correction (Muñoz-Sabater et al., 2021) are direct explanations, resolution alone does not guarantee better precipitation, as shown in Morocco (Oukaddour et al., 2025) and West Africa (Bodjrenou et al., 2025). Two factors are more decisive. First, ERA5-Land provides precipitation directly from high-resolution coupled integrations, whereas ERA5-Drought relies on coarser inputs (Keune et al., 2025). Second, the ERA5-Drought indices contain biases and extreme values that lower their NSE regardless of resolution. In the study region's sparse network and complex terrain, these factors make the ERA5-Land indices more reliable, consistent with the tendency of coarse global products to blur local patterns (Le et al., 2026).

The most prominent pattern is the sharp degradation of SPI-ERA5-Drought performance as the timescale lengthens. Longer indices accumulate monthly errors, and extreme outliers in the ERA5-Drought indices further inflate the variance, driving NSE downward; clipping these extremes restores NSE to positive (Table S6). Importantly, the failing scale is not a universal SPI property but depends on product- and climate-specific error structures; for example, ERA5-Land showed negative NSE at short scales in Sicily (Aschale et al., 2026).

Unlike SPI, the difference between the two sets of indices for SPEI was much smaller despite the difference in PET estimation methods (e.g., SPEI-12 NSE 0.54 vs. 0.50). In this East Asian monsoon region, the standardized water balance is mainly governed by precipitation, and the PET term contributes relatively little to drought anomalies (Vicente-Serrano et al., 2010). Although ERA5-Drought uses Penman-Monteith PET (Keune et al., 2025) and ERA5-Land uses Thornthwaite PET, the standardized SPEI signals from the two products remain close, consistent with the relatively small influence of PET method choice on SPEI in humid regions (Stagge et al., 2014; Beguería et al., 2014). This is not universal, however: in an arid basin in Iran, SPEI underestimated drought at the 3-month scale (Gorjizade and Hafshejani, 2026), and the temperature-only Thornthwaite method can misrepresent PET in some climates (Stagge et al., 2014; Beguería et al., 2014). These limitations should be considered when interpreting the SPEI results, particularly in this humid monsoon region.

The performance of the indices varied systematically by climate zone, being highest in the maritime ECZ and lowest in the topography-dominated NIZ, followed by the MIZ and WCZ (Fig. 2). Negative NSE values for the ERA5-Drought indices were concentrated in the inland and western coastal regions where representing topographic forcing and local convective precipitation on a coarse grid is most difficult (e.g., MIZ at SPI-6, NIZ at SPI-3, and WCZ at SPI-12), whereas the ECZ maintained positive values across all scales under more uniform maritime precipitation. This spatial pattern is consistent with findings that gridded products perform better in plains than in mountains (Oukaddour et al., 2025) and that greater topographic influence leads to larger station-level discrepancies in reanalysis (Bodjrenou et al., 2025; Aschale et al., 2026). After extreme-value clipping, the negative NSE values in the inland and western coastal zones turned positive (Table S7), indicating that the variance-based NSE metric in these regions was strongly affected by a small number of extreme outliers. However, this improvement was confined to NSE; the categorical detection skill (POD, FAR, CSI, HSS) at the − 1.0 threshold and the bias remained largely unchanged (Table S6). This is particularly relevant because these are precisely the regions where observation density is lowest and reanalysis is most needed.

The frequency analysis provides direct evidence for the propagation mechanism proposed in the Introduction. The excess extremes in ERA5-Drought are not randomly distributed but concentrated in the MIZ and WCZ, where complex topography and coastal effects make precipitation representation most challenging. The sign asymmetry—positive extremes dominating SPI but negative extremes dominating SPEI—suggests that the coarse-resolution precipitation input and the divergent PET scheme distort the upper and lower tails of the fitted distribution in different ways. The observed extremes are also not balanced in sign: for example, observed SPI-6 has 225 positive versus 37 negative extremes, whereas

SPI-1 has 88 positive versus 242 negative extremes. Quantitatively, the theoretical |z| > 3 rate under a standard normal distribution is 0.27%, whereas the observed rates are 0.45% for SPI-1 (about 1.6 times the theoretical value) and 0.35% for SPI-6; this modest excess is expected because SPI tails derived from gamma fits are not exactly normal. Nevertheless, the asymmetry in ERA5-Drought is much larger and more scale-dependent than in the observations, indicating that the ERA5-Drought extremes are not simply a reflection of the intrinsic skewness of the index.

Note that Keune et al. (2025) do not recommend clipping at |z| > 3; instead, they recommend masking regions based on the provided quality flags (e.g., Shapiro-Wilk test). In this study, clipping was performed solely as a diagnostic sensitivity test to examine how extreme outliers affect the variance-based NSE metric, and it is not proposed as an operational remedy. Clipping the simulated extremes to ± 3 (and also tested at ± 3.5 and ± 4.0; Tables S6, 7) improved the NSE of the ERA5-Drought indices across all scales—for example, the NSE of SPI-ERA5-Drought at SPI-6 turned from negative (− 0.19) to positive (+ 0.41)—while the ERA5-Land indices were virtually unaffected (NSE change < 0.01). However, this improvement was confined to the variance-based NSE; the categorical detection skill at the operational threshold (− 1.0) remained essentially unchanged, and the bias persisted.

Even after clipping, ERA5-Land maintained a higher NSE at all scales. The categorical detection skill (POD, FAR, etc.) remained unchanged because the threshold (− 1.0) lies outside the clipped range, whereas the bias was partially reduced (e.g., SPI-6 Bias_D decreased from 0.17 to 0.11; Table S6). The remaining positive bias indicates that the degradation of ERA5-Drought is driven by systematic precipitation errors, not just extreme outliers. Thus, while the quality mask effectively filters unreliable cells, it does not address the underlying bias. For these regions, bias-correcting the ERA5-Drought precipitation or using higher-resolution data (e.g., ERA5-Land) is a more fundamental solution than clipping or masking alone.

To provide a more actionable diagnostic, we cross-checked the extreme outliers (|z| > 3) against the ECMWF quality flags over the 1981–2020 period (Table S8). A clear contrast emerges between SPI and SPEI. For SPI, while the baseline quality-fail area accounts for 34.45% of valid points, 43.57% of extreme drought outliers (z < − 3) occur in these failed cells. In contrast, for SPEI, the quality-fail area is lower (24.50%), and the overlap ratio for extreme droughts drops to 23.99%. This indicates that SPEI, by incorporating the temperature-driven PET term, dampens the precipitation errors that cause extreme negative spikes in complex terrains. Nevertheless, for SPI, the strong concentration of negative outliers in quality-fail cells validates that extreme drought anomalies are structurally linked to reanalysis limitations.

Therefore, rather than relying on post-hoc clipping, applying the ECMWF quality mask is a much more robust solution for operational drought monitoring. For these flagged regions where ERA5-Drought struggles with complex terrain, future operational systems should either use bias-corrected data or rely on higher-resolution reanalysis (such as ERA5-Land) to ensure reliable drought monitoring.

Several limitations apply. (i) The 154 stations are unevenly distributed and sparse in mountainous inland regions, potentially under-representing these areas. (ii) No significance tests or confidence intervals were computed for the performance metrics, so the reported differences are descriptive rather than statistically confirmed, and the uncertainty of the distribution fitting (Guttman, 1999) was not quantified. (iii) The 2001 case study, though extended to all four climate zones, remains limited to a few sites and cannot fully capture the drought's spatial heterogeneity. (iv) ERA5-Drought quality flags (Shapiro-Wilk, p0) were cross-checked in the Discussion (Table S8) but were not applied to the main NSE/POD calculations. (v) The nominal N (73,920) overstates the effective degrees of freedom due to autocorrelation and overlapping accumulation windows. (vi) The distribution fitting used a 1991–2020 reference period for consistency with ERA5-Drought, so results may depend on this choice (Guttman, 1999).

Despite these limitations, the results have clear practical implications. First, product choice matters. Because both observations and ERA5-Land dataset use the identical Thornthwaite method for PET estimation—whereas ERA5-Drought relies on its own scheme—methodological consistency may have structurally favored ERA5-Land. Consistent with findings that dataset choice drives drought uncertainty more than index selection (Huang et al., 2026), the choice between these two products had a decisive impact on NSE here. Second, in the DPRK's limited observation network, reanalysis provides essential continuity for hydrological studies (Bodjrenou et al., 2025), making high-resolution ERA5-Land with strong detection skill suitable for operational monitoring (Aschale et al., 2026). ERA5-Drought remains a convenient ready-made dataset, but should be used cautiously for 6–12 month SPI until its extreme-value and bias characteristics are improved.

Future research should incorporate additional meteorological observations that account for topographic effects to fill the validation gaps in the mountainous inland regions, and recompute SPEI directly from raw ERA5 inputs using alternative PET estimation methods (e.g., Penman-Monteith) to assess how different PET schemes affect the bias characteristics currently present in the fixed ERA5-Drought product. Furthermore, given the sparse observation network and missing data periods in the DPRK, reconstructing (or gap-filling) the observed drought time series using high-resolution reanalysis data (ERA5-Land) is proposed as an important future research task.

5. Conclusion

This study provides a comprehensive, long-term evaluation of ERA5-Land and ERA5-Drought for drought monitoring in the DPRK. Based on 154 stations from 1981 to 2020, we conclude that the SPI/SPEI-ERA5-Land indices are consistently more reliable than the SPI/SPEI-ERA5-Drought indices for reproducing observed SPI and SPEI across all timescales.

The primary limitations of the ERA5-Drought indices are their systematic biases and a few extreme outliers, which are particularly detrimental at intermediate and long accumulation periods (6–12 months) and in topographically complex inland zones. While clipping these extremes improved its performance, the persistence of bias indicates that raw ERA5-Drought data are insufficient for accurate

drought severity assessment without bias correction. Furthermore, the methodological consistency of the Thornthwaite PET method in both the observations and ERA5-Land structurally favors the latter for SPEI calculation.

For practical applications in the DPRK, high-resolution ERA5-Land is recommended as a robust alternative for operational drought early warning systems. Given the sparse observation network, future efforts should integrate ERA5-Land to reconstruct historical drought series and improve the bias characteristics of ready-made products like ERA5-Drought.

Declarations

Acknowledgements We thank the State Hydro-Meteorological Administration of the DPRK (SHMA) for providing the in-situ monthly precipitation and mean temperature observations. We also acknowledge the Copernicus Climate Change Service (C3S) for providing the ERA5-Land reanalysis dataset and the ERA5-Drought drought indices through the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/).

Author Contributions Yong-Sik Ham: Writing – original draft, methodology, visualization, investigation, Data curation; Wei He: Writing – review & editing, validation, Supervision, Resources; Kyong-Bok Sonu: visualization, Data curation.

Funding

This research did not receive any specific funding.

Competing Interests

The authors declare no conflicts of interest.

Data availability

The ERA5-Land and ERA5-Drought datasets are publicly available from the Copernicus Climate Data Store (CDS). The station observations used in this study were provided by the State Hydro-Meteorological Administration (SHMA) of the DPRK and may be available from the corresponding author upon reasonable request.

References

1. Abu Arra A, Birpınar ME, Şişman E (2025) Evaluating ERA5-LAND and IMERG-NASA products for drought analysis: Implications for sustainable water resource management. Sustainability 17:7529. https://doi.org/10.3390/su17167529
2. Aschale TM, Palazzolo N, Buonacera G, Cancelliere A, Peres DJ (2026) An evaluation of ERA5-Land reanalysis data for drought monitoring: Comparison with drought indices from observational

datasets in Sicily. Hydrol Res 57:100016

3. Beguería S, Vicente-Serrano SM, Reig F, Latorre B (2014) Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. Int J Climatol 34:3001–3023. https://doi.org/10.1002/joc.3887
4. Bodjrenou R, Sintondji LO, Comandan F (2025) Revealing the spatiotemporal precipitation patterns of ECMWF fifth-generation reanalyses since the mid-20th century in West-Africa. Environ Adv 20:100636. https://doi.org/10.1016/j.envadv.2025.100636
5. Ceppi A, Achite M, Toubal AK, Caloiero T (2025) Mapping drought characteristics in northern Algerian Basins using the ERA5-Land dataset. Sci Rep 15:10720. https://doi.org/10.1038/s41598-025-95418-8
6. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. Geosci Model Dev 7:1247–1250. https://doi.org/10.5194/gmd-7-1247-2014
7. Chen S, Gan TY, Tan X, Shao D, Zhu J (2019) Assessment of CFSR, ERA-Interim, JRA-55, MERRA-2, NCEP-2 reanalysis data for drought analysis over China. Clim Dyn 53:737–757. https://doi.org/10.1007/s00382-018-04611-1
8. Choukri M, Naimi M, Chikhaoui M (2026) Comparative analysis of CHIRPS and ERA5-Land for precipitation and drought assessment in Morocco (1987–2016). Theor Appl Climatol 157:245. https://doi.org/10.1007/s00704-026-06180-4
9. Fernandes FAC, dos Santos LOF, Nunes NC, Machado NG, Biudes MS (2024) Quantifying droughts in Mato Grosso with SPI and SPEI. Theor Appl Climatol 155:9751–9766. https://doi.org/10.1007/s00704-024-05212-1
10. Gorjizade A, Hafshejani LD (2026) Multiscale drought dynamics in a semi-arid region of Iran: A reanalysis-based assessment using run theory. Results Eng 31:111575. https://doi.org/10.1016/j.rineng.2026.111575
11. Guttman NB (1999) Accepting the Standardized Precipitation Index: A calculation algorithm. J Am Water Resour Assoc 35:311–322
12. Harishnaika N, Mohammed Aslam MA, Arpitha M, Ahmed SA, Kondagadupula M, Pasha MA (2026) Reanalysis of ERA5-land precipitation, PET, SPI and potential water deficit: A detailed examination of Indian monsoon changes by relating hydro-climatic stress in Karnataka. Phys Chem Earth Parts A/B/C 144:104615
13. Huang K, Zhang H, Cui G, Wang Y, Yin M, Du J (2026) Drought propagation in China: Uncertainties originate more from dataset choice than drought index selection. Atmos Res 330:108555. https://doi.org/10.1016/j.atmosres.2025.108555
14. Keune J, Di Giuseppe F, Barnard C, Damasio da Costa E, Wetterhall F (2025) ERA5–Drought: Global drought indices based on ECMWF reanalysis. Sci Data 12:616. https://doi.org/10.1038/s41597-025-04896-y

15. Kim H-U, Jong S-I, Om K-C, Paek U-S, Sonu K-B, Tak J-Y (2025) Spatiotemporal variability of drought and related atmospheric circulation patterns over the Democratic People's Republic of Korea. J Earth Syst Sci 134:63. https://doi.org/10.1007/s12040-024-02475-0
16. Le T, Sun QC, Tran TD, Li H, Choy S, Kuleshov Y (2026) Characterizing high-resolution uneven drought conditions in Vietnam for more reliable hazard assessment. Earth Syst Environ. https://doi.org/10.1007/s41748-026-01050-5
17. Lloyd-Hughes B, Saunders MA (2002) A drought climatology for Europe. Int J Climatol 22:1571–1592. https://doi.org/10.1002/joc.846
18. McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. In: Preprints, 8th Conference on Applied Climatology. American Meteorological Society, Anaheim, CA, pp 179–184
19. Mo KC, Long LN, Xia Y, Yang SK, Schemm JE, Ek M (2011) Drought indices based on the Climate Forecast System Reanalysis and ensemble NLDAS. J Hydrometeorol 12:181–205. https://doi.org/10.1175/2010JHM1310.1
20. Muñoz-Sabater J, Dutra E, Agustí-Panareda A, Albergel C, Arduini G, Balsamo G, Boussetta S, Choulga M, Harrigan S, Hersbach H, Martens B, Miralles DG, Piles M, Rodríguez-Fernández NJ, Zsoter E, Buontempo C, Thépaut J-N (2021) ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth Syst Sci Data 13:4349–4383. https://doi.org/10.5194/essd-13-4349-2021
21. Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I — A discussion of principles. J Hydrol 10:282–290
22. Om K-C, Ren G, Kim K-H, Pak Y-I, Jong S-I, Kil H-N (2022) Observed trends in extreme temperature events over northern part of the Korean Peninsula during 1960–2019 and a comparative overview. Atmos Res 270:106061. https://doi.org/10.1016/j.atmosres.2022.106061
23. Om K-C, Ren G, Jong S-I, Ham Y-S, Hong C-W, Kang-Chol O, Tysa SK (2024) Spatial and temporal patterns in observed extreme precipitation change over northern part of the Korean Peninsula. J Geophys Res Atmos 129:e2023JD039305. https://doi.org/10.1029/2023JD039305
24. Oukaddour K, Fakir Y, Le Page M (2025) Assessment of five global gridded precipitation estimates over a southern Mediterranean basin (Tensift, Morocco). Geomatics Nat Hazards Risk 16:2468850. https://doi.org/10.1080/19475705.2025.2468850
25. Shah R, Mishra V (2014) Evaluation of the reanalysis products for the monsoon season droughts in India. J Hydrometeorol 15:1575–1591. https://doi.org/10.1175/JHM-D-13-0103.1
26. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52:591–611. https://doi.org/10.2307/2333709
27. Sharma TC, Panu US (2012) Prediction of hydrological drought durations based on Markov chains: Case of the Canadian prairies. Hydrol Sci J 57:705–722. https://doi.org/10.1080/02626667.2012.672741

28. Stagge JH, Tallaksen LM, Xu C-Y, Van Lanen HAJ (2014) Standardized precipitation-evapotranspiration index (SPEI): Sensitivity to potential evapotranspiration model and parameters. In: Hydrology in a Changing World. IAHS Publication 363, Copernicus GmbH, pp 367–373
29. Thornthwaite CW (1948) An approach toward a rational classification of climate. Geogr Rev 38:55–94
30. Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A multiscalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index. J Clim 23:1696–1718. https://doi.org/10.1175/2009JCLI2909.1
31. Wałęga A, Ceppi A, Chiaravalloti F, De Marco A, Kochanek K, Młyński D, Caloiero T (2026) Validation of ERA5-Land precipitation data for meteorological drought assessment in the Upper Vistula Basin, Central Europe. Meteorol Appl 33:e70206. https://doi.org/10.1002/met.70206
32. Wilhite DA, Glantz MH (1985) Understanding: the drought phenomenon: the role of definitions. Water Int 10:111–120
33. Wilhite DA (2000) Drought as a natural hazard: Concepts and definitions. In: Wilhite DA (ed) Drought: A Global Assessment. Routledge, pp 3–18
34. Wilks DS (2011) Statistical Methods in the Atmospheric Sciences, 3rd edn. Academic Press, Oxford
35. WMO(2012) Standardized Precipitation Index User Guide. World Meteorological Organization, Geneva
36. Yevjevich V (1967) An Objective Approach to Definitions and Investigations of Continental Hydrologic Droughts (Hydrology Papers No. 23). Colorado State University, Hydrology and Water Resources Program, Fort Collins, CO
37. Zhan W, Guan K, Sheffield J, Wood EF (2016) Depiction of drought over sub-Saharan Africa using reanalyses precipitation data sets. J Geophys Res Atmos 121. https://doi.org/10.1002/2016JD024858