Public justification (visible to the public if the article is accepted and published):

Dear authors,

Thank you for your responses to the review comments.

Rev.#1 pointed out specifically 'The most severe shortcoming is the complete absence of validation against independent observational data.' Rev.#2 and Rev.#3 also have the same concerns but formulated in different ways. You talked around this point but did not address it by providing independent validation of your results. You need to demonstrate at the least that for some of the areas where droughts were predicted that you can show indeed they occurred by other data sources.

There is a recent global high resolution drought analysis, you may compare your results with their findings. see: https://arxiv.org/abs/2509.20195

There are also several recent independent datasets (soil moisture, evapotranspiration, precipitation) from remote sensing and machine learning methods that may be used.

The same is also for the point raised by Rev.#1, on the predictive skill at lead times (up to 180 days). This also needs to be validated by independent data. Rev.#2 also raised the same issue.

On the 'critical parameters for VMD—the number of modes (K) and the penalty factor (α)' by Rev.#1, and Rev.#3, a clear explanation is needed.

I hope you can address those points substantially in your revision.

Best wishes

Bob Su

Editor HESS

1. Rev.#1 pointed out specifically 'The most severe shortcoming is the complete absence of validation against independent observational data.' Rev.#2 and Rev.#3 also have the same concerns but formulated in different ways. You talked around this point but did not address it by providing independent validation of your results. You need to demonstrate at the least that for some of the areas where droughts were predicted that you can show indeed they occurred by other data sources. There is a recent global high resolution drought analysis, you may compare your results with their findings. see: https://arxiv.org/abs/2509.20195 There are also

several recent independent datasets (soil moisture, evapotranspiration, precipitation) from remote sensing and machine learning methods that may be used.

Response 1: We sincerely thank the editor and reviewers for their insightful and constructive comments. The primary reason for selecting the ERA5 reanalysis dataset lies in its comprehensive advantages in temporal continuity, spatial consistency, and the integration of multiple meteorological variables. ERA5 has been widely used in global and regional hydro-meteorological studies and has demonstrated strong stability and reliability in key variables such as evapotranspiration, precipitation, and soil moisture. Therefore, it is commonly adopted as a fundamental data source for the development of regional drought indices and related methodological research.

In global-scale studies, Xu et al. (2024) used ERA5 data for global-scale drought prediction, demonstrating that ERA5 data can effectively predict drought events worldwide. Filipović et al. (2021) constructed a regional soil moisture prediction system using ERA5 data and drought predictions through the LSTM model, proving ERA5's independent application capability in local drought forecasting. Gupta et al. (2024) proposed a deep learning model based on ERA5 data for global drought prediction, highlighting the potential and value of ERA5 data in global drought forecasting.

In a regional study, Gao et al. (2023) conducted a spatiotemporal analysis of cloud water resources in the Huaihe River Basin using ERA5 data, showing that ERA5 can effectively capture variations in cloud water resources, providing reliable data support for water resource management and drought prediction in the region. Zhang et al. (2021) used ERA5-Land data to construct a Soil Moisture Deficit Index (SWDI) for evaluating agricultural drought in the Huaihe River Basin, and through comparison with other data sources, validated the applicability of ERA5-Land data to drought analysis for the region. Additionally, Li et al. (2025) modelled hydrological drought in the Huaihe River Basin using ERA5 data, and the study demonstrated that ERA5 data can effectively predict drought events in the region, proving its applicability.

We fully acknowledge that relying solely on the ERA5 reanalysis dataset for model development and validation may introduce the risk that the model learns the internal structure of the data rather than the actual drought processes. Therefore, during the revision process, we additionally incorporated an independent dataset, namely the global gridded evapotranspiration dataset (1982–2024) from the National Tibetan Plateau Data Center (https://data.tpdc.ac.cn), to perform cross-comparative analyses of ERA5 variables and the drought indices derived from them at the data level, thereby providing important supporting evidence for the original findings.

First, at the level of fundamental meteorological variables, the DEDI index is derived from actual evapotranspiration (AET) and potential evapotranspiration (PET), and the consistency between these variables directly affects the reliability of the index. Therefore, a systematic comparison of AET and PET from the ERA5 and Global Gridded datasets was conducted. The spatial distribution of Pearson correlation coefficients between the two datasets was calculated over the Huaihe River Basin. The results indicate a high level of consistency at the regional scale, with correlation coefficients ranging from 0.77 to 0.85 for AET and from 0.80 to 0.90 for PET, and with broadly consistent spatial patterns. These findings suggest that, although ERA5 does not explicitly represent anthropogenic processes such as irrigation, its key hydro-meteorological variables can still effectively capture evapotranspiration variability at the regional scale, showing strong agreement with the independent dataset.

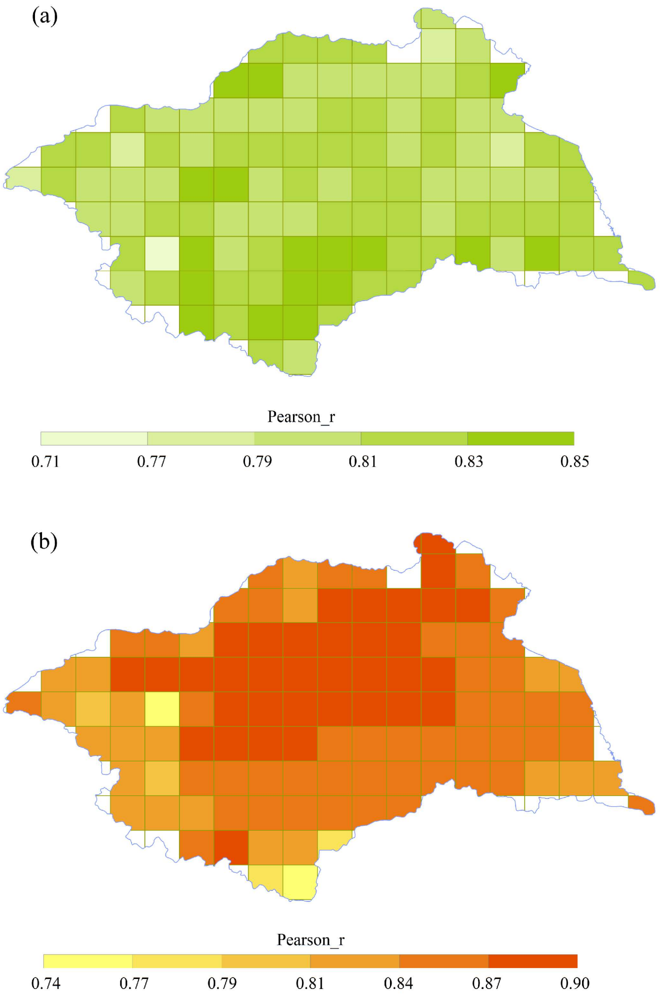

Figure 1. Spatial consistency analysis of evapotranspiration data between ERA5 and the Global Gridded dataset: (a) actual evapotranspiration (AET); (b) potential evapotranspiration (PET).

On this basis, we further conducted a comparative analysis at the drought index level. The DEDI index was constructed separately using the ERA5 and Global Gridded datasets, and statistical tests were performed to evaluate their consistency. As shown in Figure 2, the two datasets exhibit good agreement across different drought categories. For mild, moderate, and severe drought conditions, the data points are generally distributed along the 1:1 reference line, indicating a high degree of consistency within these ranges. Although slight deviations are observed under extreme drought conditions, the overall trend remains consistent with the reference line and does not affect the overall correspondence between the two datasets. Overall, the strong correlation between them suggests that the DEDI index constructed from ERA5 can reliably capture drought evolution, rather than merely reflecting the internal characteristics of the reanalysis data.

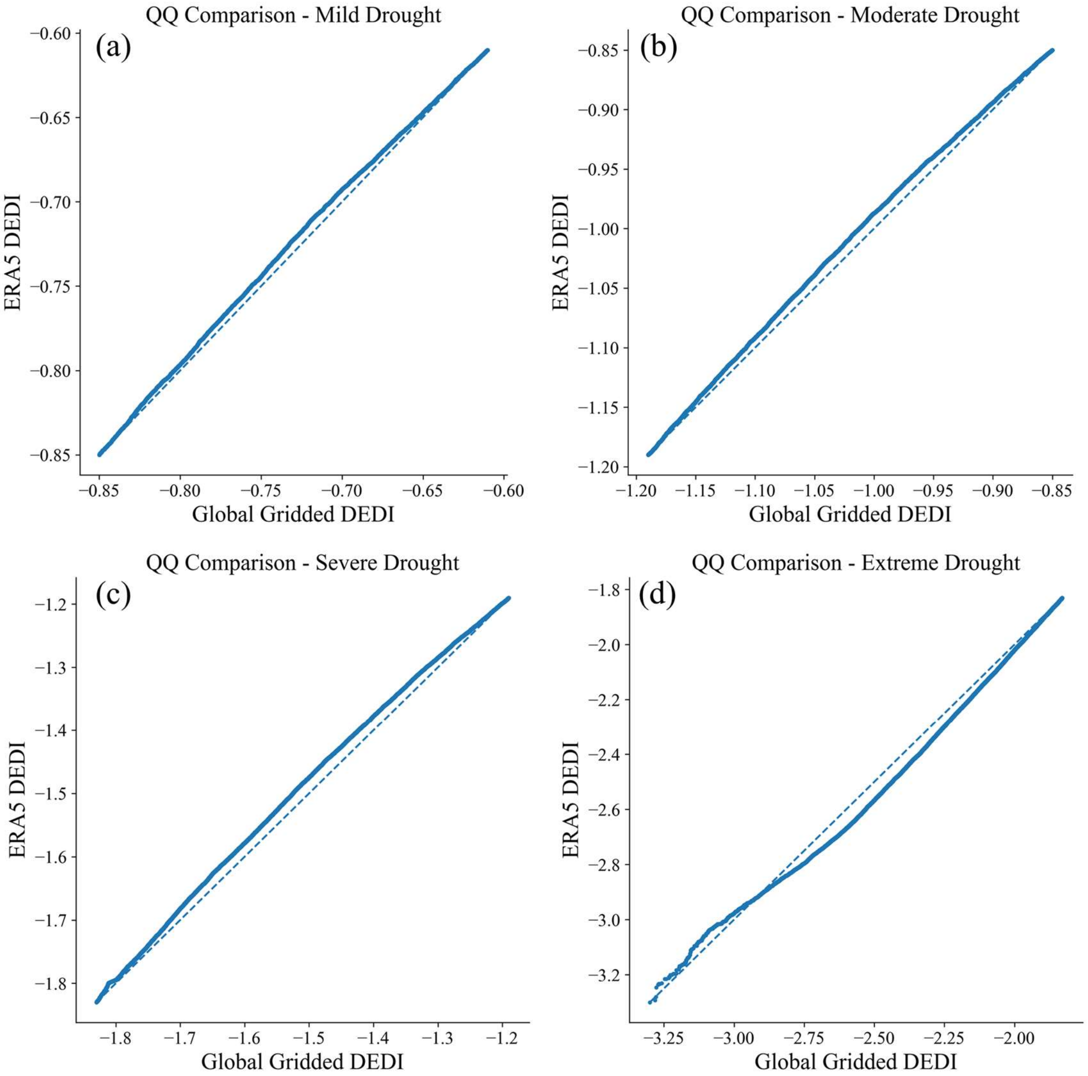

Figure 2. QQ plot of the DEDI index derived from ERA5 and the Global Gridded dataset.

From the boxplot results in Figure 3, the overall range of the DEDI distributions from the two datasets is largely consistent, both spanning approximately from −2.0 to 2.0, indicating good comparability in terms of drought and wetness intensity. The medians of the two datasets are also close, suggesting a similar central tendency. In addition, the interquartile range (IQR) and overall spread are comparable, indicating similar variability. Although the median of the DEDI derived from the Global Gridded dataset is slightly positive, while that from ERA5 is closer to zero, the overall distribution structure remains consistent, with no evident systematic bias.

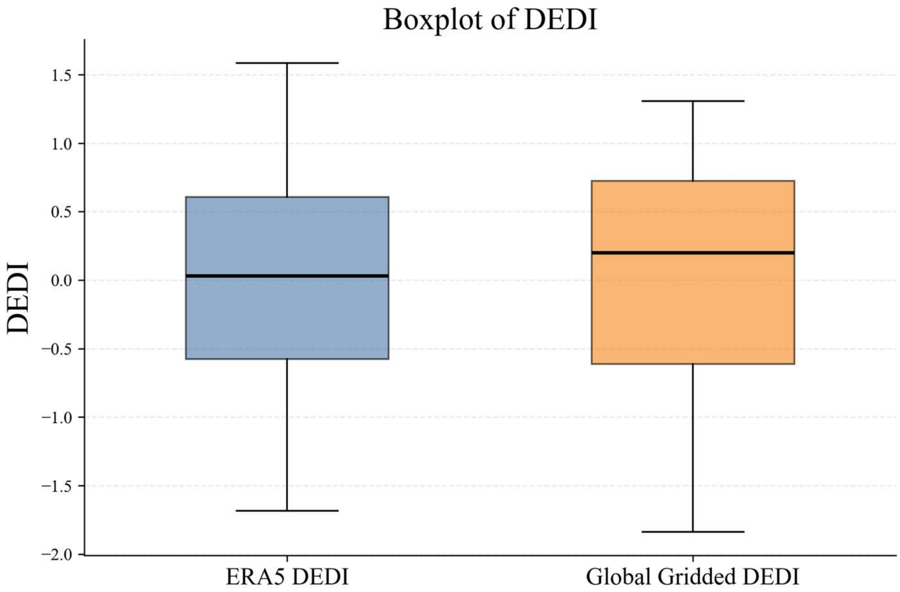

Figure 3. Boxplot of the DEDI distributions from ERA5 and the Global Gridded dataset.

Table 1 presents a comparison of drought occurrence frequencies at different drought levels between the ERA5 and Global Gridded datasets. The results show that the frequency distributions across different drought categories are generally similar between the two datasets, with mild drought frequencies of 13.48% and 11.72%, moderate drought frequencies of 7.41% and 7.98%, severe drought frequencies of 3.69% and 4.88%, and extreme drought frequencies of 2.98% and 3.51%, respectively. Overall, the Global Gridded dataset exhibits slightly higher occurrence frequencies across all drought categories than ERA5; however, the differences are small and the trends are consistent. These findings indicate strong consistency in the statistical characteristics of drought intensity classification between the two datasets, further demonstrating that the DEDI index derived from ERA5 can reliably represent regional drought conditions rather than merely reflecting the internal characteristics of the reanalysis data.

Table 1. Comparison of drought occurrence frequencies at different drought levels between datasets

Drought Category	Mild Drought	Moderate Drought	Severe Drought	Extreme Drought
ERA5 Dataset	13.48%	7.41%	3.69%	2.98%
Global Gridded Dataset	11.72%	7.98%	4.88%	3.51%

Based on the combined results from Figures 1–3 and Table 1, although some differences exist between the two datasets under extreme conditions, they exhibit strong consistency in overall distribution patterns and statistical structures. This further demonstrates that the DEDI index constructed from ERA5 can reliably capture drought variability, rather than merely reflecting the internal characteristics of the reanalysis data.

It should be noted that the purpose of the above supplementary analysis is to validate the ERA5 data and the drought index from the perspective of consistency with an independent dataset, thereby enhancing the reliability of the study. Therefore, this part has been included as supplementary material. In addition, it is clarified in the discussion that future work will incorporate in-situ observations and multi-source remote sensing products to more systematically evaluate model performance under real drought events. Through these improvements, the data reliability of this study has been further strengthened.

Cai, H., Wang, K. (2025). China Homogenized (1961-2024) and Global Gridded (1982-2024) Evapotranspiration Dataset. National Tibetan Plateau / Third Pole Environment Data Center. https://doi.org/10.11888/Terre.tpdc.303108.

Xu, L., Zhang, X., Wu, T., Yu, H., Du, W., & Chen, N. (2024). Global prediction of flash drought using machine learning. Geophysical Research Letters, 51(21). https://doi.org/10.1029/2024GL111134.

Filipović, N., Brdar, S., Mimić, G., Marko, O., & Crnojević, V. (2021). Regional soil moisture prediction system based on Long Short Term Memory network. Biosystems Engineering. https://doi.org/10.1016/j.biosystemseng.2021.11.019

Gupta, B. B., Gaurav, A., Attar, R. W., Arya, V., Bansal, S., Alhomoud, A., & Chui, K. T. (2024). Advance drought prediction through rainfall forecasting with hybrid deep learning model. SCIENTIFIC REPORTS, 14(1), 30459. https://doi.org/10.1038/s41598-024-80099-6

Zhang, R., Lu, L., Ye, Z., Huang, F., Li, J., Wei, L., Mao, T., Xiong, Z., & Wei, S. (2021). Assessment of Agricultural Drought Using Soil Water Deficit Index Based on ERA5-Land Soil Moisture Data in Four Southern Provinces of China. AGRICULTURE-BASEL, 11(5), 411. https://doi.org/10.3390/agriculture11050411

Li, M., Yao, Y., Feng, Z., and Ou, M.: Hydrological drought prediction and its influencing features analysis based on a machine learning model, Nat. Hazards Earth Syst. Sci., 25, 4299 – 4316, https://doi.org/10.5194/nhess-25-4299-2025, 2025.

Gao, J., Feng, J., Cao, Y., & Zheng, X. (2023). Evaluation of Cloud Water Resources in the Huaihe River Basin Based on ERA5 Data. ATMOSPHERE, 14(8), 1253. https://doi.org/10.3390/atmos14081253

2.The same is also for the point raised by Rev.#1, on the predictive skill at lead times (up to 180 days). This also needs to be validated by independent data. Rev.#2 also raised the same issue.

Response 2: Thank you for your valuable comments. We understand the concerns of the reviewers and the editor regarding long lead-time prediction capability. To address this issue, we have supplemented the revised manuscript with additional explanations from two perspectives: mechanistic interpretation and independent data validation.

First, it should be clarified that the 180-day prediction discussed in this study is not a deterministic weather forecast in the traditional sense, but rather a prediction of drought state evolution. Compared with weather systems, drought formation and development exhibit stronger temporal accumulation effects and persistence, and are influenced by slow-varying processes such as soil moisture deficits, delayed groundwater recharge, persistent evapotranspiration anomalies, and regional water balance conditions. Therefore, at medium- to long-term scales, the model does not learn specific precipitation or weather events, but rather the persistence and evolution patterns of drought states over time. Previous studies have shown that deep learning models primarily capture the persistence and stage-wise evolution characteristics of drought, rather than extrapolating specific weather processes (Esit et al., 2021; Rahmati et al., 2024). Based on this understanding, we agree that the predictive capability of this study is more appropriately described as drought state prediction rather than traditional weather forecasting.

At the same time, we acknowledge that explanations based solely on conceptual and mechanistic perspectives are insufficient to support long lead-time results. Therefore, in this revision, we further supplemented the study with drought event validation using independent datasets. Specifically, a typical spring drought event in the Huaihe River Basin in 2000 was selected as a case study, and an independent dataset—the China Homogenized (1961–2024) and Global Gridded (1982–2024) Evapotranspiration Dataset—was introduced.The DEDI index was constructed separately based on ERA5 and the independent dataset, and drought identification results at different lead times (7, 15, 30, 60, 120, and 180 days) were compared grid-by-grid and time step by time step over the same spatial grids. The hit rate was used as a consistency metric: if both datasets simultaneously identified drought or non-drought at the same grid and time, it was counted as a hit; otherwise, it was counted as a miss.

Table 2 shows that the average hit rates for different lead times are 0.80, 0.89, 0.81, 0.83, 0.77, and 0.79, respectively, with an overall range of 0.77–0.89, indicating consistently high agreement. Specifically, for short- to medium-term lead times (7–60 days), the hit rates are all above 0.80, with the highest value (0.89) observed at the 15-day lead time, suggesting the strongest consistency between the two datasets at this timescale. For longer lead times (120 and 180 days), although the hit rates show a slight decrease, they remain relatively high at 0.77 and 0.79, without significant degradation, indicating that the model maintains reasonable stability even at a six-month scale. Overall, the hit rates exhibit only minor fluctuations with increasing lead time and show no evident declining trend, demonstrating the robustness of the model across different temporal scales.These results indicate that even under long lead-time conditions, the two independent datasets maintain good overall consistency in drought state identification, further supporting the reliability of the DEDI index constructed from ERA5 in capturing drought evolution processes at medium- to long-term scales.

Table 2. Accuracy of drought identification at different lead times.

7	15	30	60	120	180
0.80	0.89	0.81	0.83	0.77	0.79

Therefore, our understanding is that what the model captures at the 30–180 day scale is not a deterministic forecasting capability for specific weather systems, but rather the persistence and evolution signals of drought states. These signals arise partly from the slow-varying nature and hydrological memory of drought processes, and partly from the inherent statistical persistence in the time series. To avoid overinterpretation, we have further moderated the relevant statements in the revised manuscript and explicitly clarified that the model reflects the predictability of drought states at medium- to long-term scales, rather than directly predicting specific meteorological events. Meanwhile, the newly introduced independent dataset provides additional support for this medium- to long-term predictive capability from an external data perspective.

Esit, M., S. Kumar, A. Pandey, D. M. Lawrence, I. Rangwala, and S. Yeager, 2021: Seasonal to multi-year soil moisture drought forecasting. npj Climate and Atmospheric Science, 4, 16, https://doi.org/10.1038/s41612-021-00172-z.

Rahmati, M., Amelung, W., Brogi, C., Dari, J., Flammini, A., Bogena, H., et al. (2024). Soil moisture memory: State-of-the-art and the way forward. *Reviews of Geophysics*, 62, e2023RG000828. https://doi.org/10.1029/2023RG000828

3.On the 'critical parameters for VMD—the number of modes (K) and the penalty factor (α)' by Rev.#1, and Rev.#3, a clear explanation is needed.

Response 3: Thank you for your valuable comments. Regarding the key parameters of VMD—namely the number of modes (K) and the penalty factor (α)—we have added detailed explanations in Section 3.2 (VMD), as follows:

In this study, two key parameters of VMD need to be predefined: the penalty factor α (bandwidth constraint parameter) and the number of modes K (i.e., the number of IMF components). The penalty factor α is empirically determined based on the length of the time series, with its value set within 1.5–2.0 times the sample length, aiming to balance frequency band separation capability and decomposition stability. When α is small, the bandwidth of each IMF becomes large, which may lead to spectral overlap among different modes and reduce the physical interpretability of the decomposition results; whereas an excessively large α overly constrains the bandwidth, making the decomposition more sensitive to noise. In this study, α is finally set to 1.75 times the sample length.

In addition, the number of modes K is determined based on the frequency distribution characteristics of the decomposed signals and preliminary experiments. When K is less than 7, some IMF components exhibit significant spectral aliasing, making it difficult to effectively separate signals at different time scales. When K exceeds 7, the center frequencies of adjacent IMFs become too close and the energy distribution becomes more dispersed, leading to redundancy or noise-dominated modes, which reduces the stability and physical interpretability of the decomposition. Therefore, K is uniformly set to 7 to ensure that the main frequency information of the original DEDI series is effectively separated while avoiding redundancy caused by over-decomposition. Preliminary experiments at multiple representative grid points demonstrate that this parameter combination achieves stable decomposition performance and good predictive capability. To maintain methodological consistency and avoid potential spatial overfitting, the same VMD parameter settings are applied uniformly to all grid points across the study area.