Supplement of Hydrol. Earth Syst. Sci., 30, 3925–3944, 2026
https://doi.org/10.5194/hess-30-3925-2026-supplement

Hydrology and Earth System Sciences
Open Access

Supplement of

Daily drought prediction in the Huaihe River Basin using VMD-informer-LSTM

Min Li et al.

Correspondence to: Min Li (limintju@126.com)

Independent cross-dataset validation using the TPDC evapotranspiration dataset

In addition to ERA5, the Global Gridded Evapotranspiration Dataset (1982–2024) from the National Tibetan Plateau Data Center (TPDC) was introduced as an independent gridded evapotranspiration product for cross-dataset validation. This dataset was used to evaluate the consistency of ERA5-derived evapotranspiration variables and the DEDI index with an independent data product.

To further assess the reliability of ERA5-derived DEDI, independent cross-dataset validation was conducted using the TPDC Global Gridded Evapotranspiration Dataset. AET, PET, and DEDI derived from ERA5 were compared with those derived from the TPDC dataset. Pearson correlation coefficients, QQ plots, boxplots, and drought-frequency statistics were used to evaluate the consistency between the two datasets.

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. Figure S1 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.

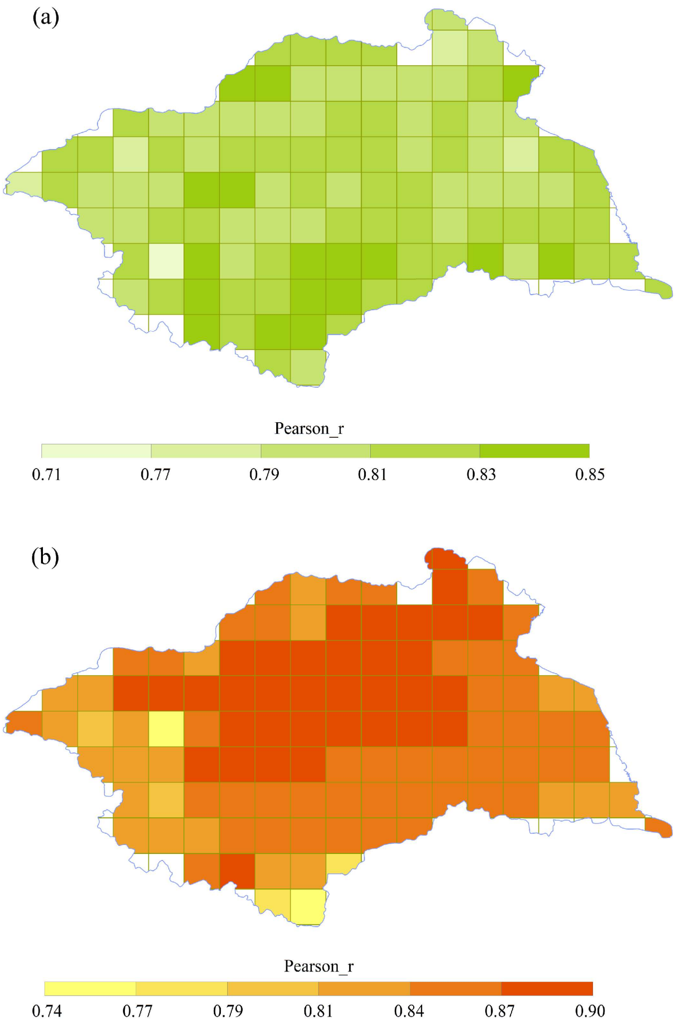
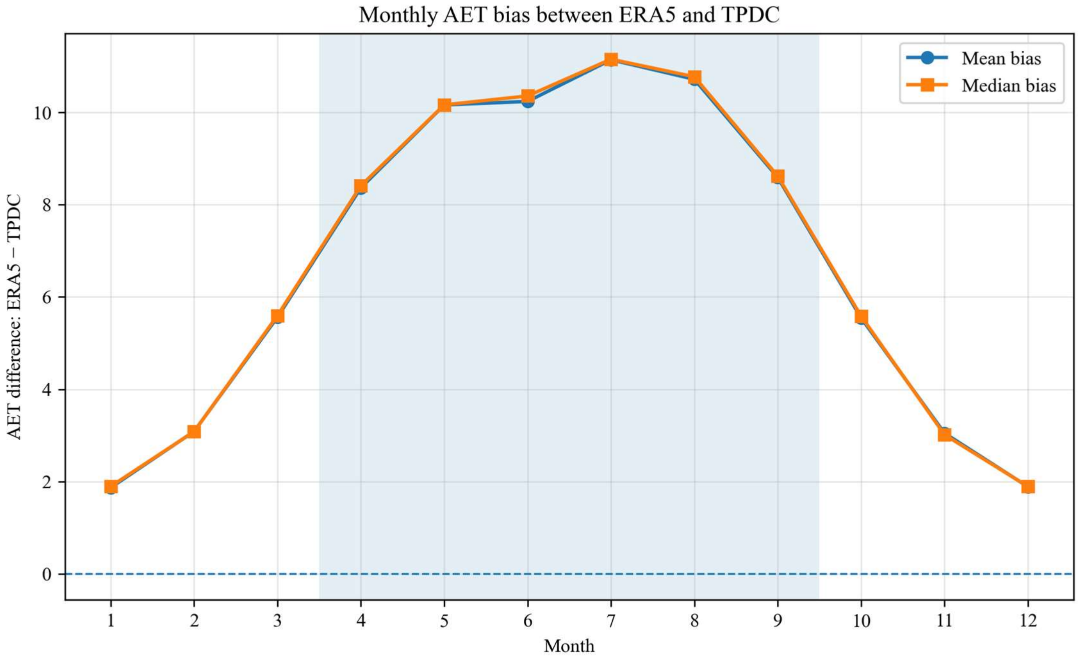


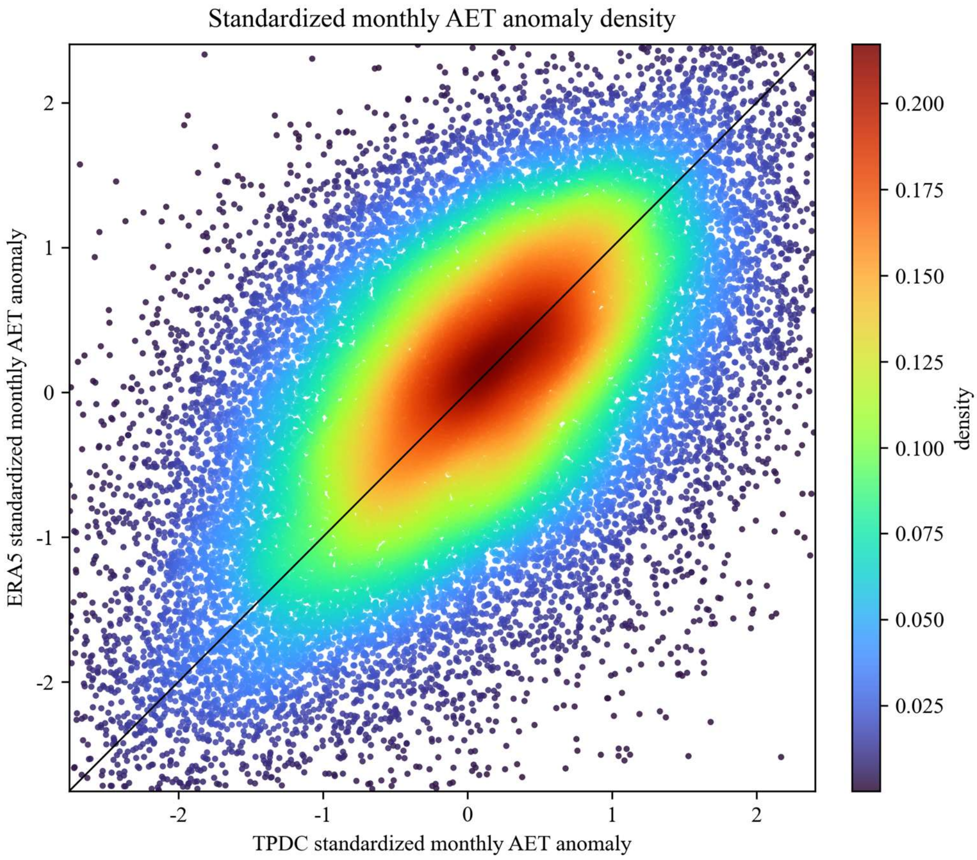
Figure S1. Spatial consistency analysis of evapotranspiration data between ERA5 and the Global Gridded dataset: (a) actual evapotranspiration (AET); (b) potential evapotranspiration (PET).

Cross-comparison of ERA5-derived AET and the independent gridded evapotranspiration dataset

To further evaluate the reliability of ERA5-derived AET from the perspective of seasonal bias and temporal anomaly variation, a cross-comparison was conducted between ERA5-derived AET and the Global Gridded Evapotranspiration Dataset provided by the National Tibetan Plateau Data Center. Since the spatial correlation of AET has already been presented in Figure S1(a), this section focuses on monthly bias characteristics and standardized monthly anomaly variations..



(a) Monthly AET bias between ERA5 and the independent dataset, calculated as ERA5 minus TPDC. The shaded area indicates the crop growing season from April to September.



(b) Density scatter plot of standardized monthly AET anomalies derived from ERA5 and the independent dataset. The black line indicates the 1:1 reference line.

Figure S2. Cross-comparison of ERA5-derived AET and the independent gridded evapotranspiration dataset.

Figure S2(a) further presents the monthly AET bias between ERA5-derived AET and the independent gridded evapotranspiration dataset, calculated as ERA5 minus the independent dataset. The mean and median monthly bias curves are nearly overlapping, indicating that the monthly bias structure is stable and is not dominated by a few extreme values. Meanwhile, the bias shows a clear seasonal pattern during the crop growing season from April to September. This indicates that the relationship between ERA5-derived AET and the independent dataset has a stable seasonal structure. It also suggests that irrigation and agricultural water-management processes should be considered when interpreting AET-related drought results during the growing season.

Figure S2(b) shows the density scatter plot of standardized monthly AET anomalies. Because different evapotranspiration products may differ in algorithms, input data, and variable definitions, direct comparison of absolute AET values may be affected by magnitude differences. Therefore, standardized monthly anomalies were calculated for both datasets to examine their ability to characterize evapotranspiration anomaly variations. The standardized monthly AET anomalies derived from ERA5 and the independent gridded evapotranspiration dataset are mainly distributed along the 1:1 reference line and exhibit a clear positive correspondence. This indicates that ERA5-derived AET can reasonably capture the temporal variability of regional evapotranspiration anomalies, thereby providing additional data-level support for its use in DEDI construction and drought prediction.

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 S3, 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.

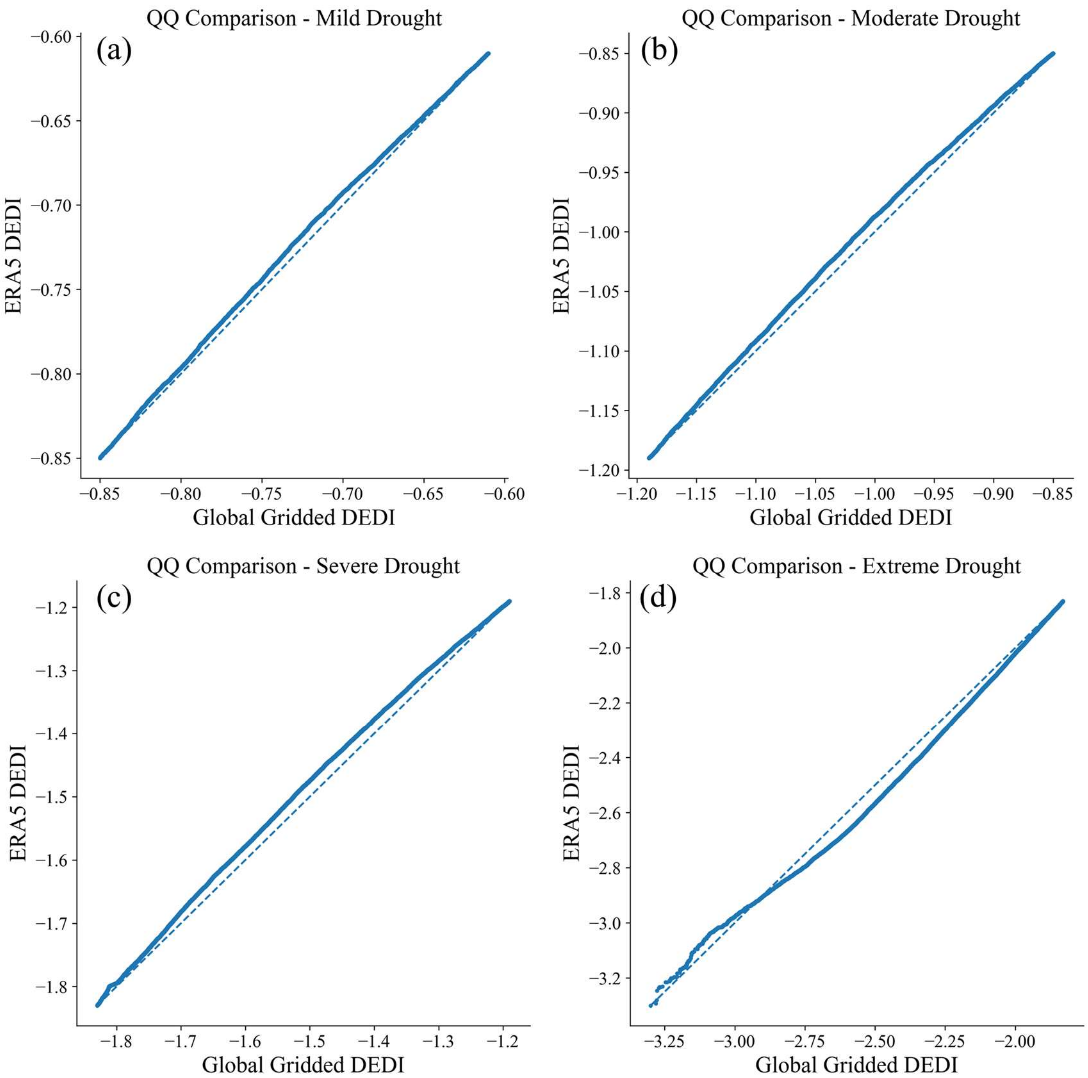


Figure S3. QQ plot of the DEDI index derived from ERA5 and the Global Gridded dataset.

From the boxplot results in Figure S4, 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.

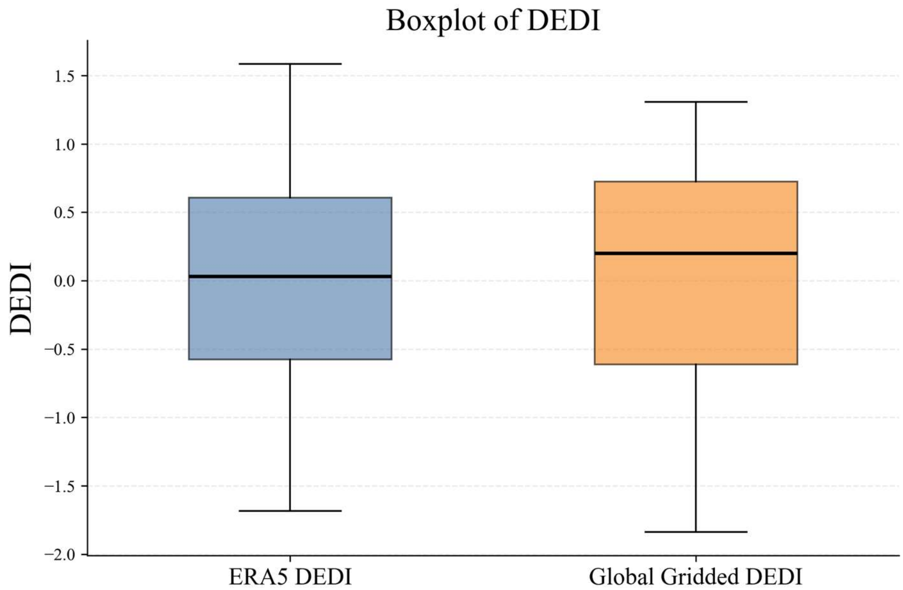


Figure S4. Boxplot of the DEDI distributions from ERA5 and the Global Gridded dataset.

Table S1 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 S1. 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 S1–S4 and Table S1, although some differences exist between the two datasets under extreme conditions, the ERA5-derived evapotranspiration variables and the DEDI index show good correspondence with the independent gridded evapotranspiration product in terms of spatial correlation, standardized anomaly variation, distribution patterns, and drought-category frequencies. These results suggest that the DEDI index constructed from ERA5 can reasonably capture regional drought variability, rather than merely reflecting the internal characteristics of the reanalysis data. The purpose of this analysis is to evaluate the reliability of ERA5-derived evapotranspiration variables and the DEDI index through cross-comparison with an independent gridded evapotranspiration product. Although this cross-dataset validation cannot fully replace direct in-situ observations, it provides additional evidence supporting the reliability of the ERA5-derived DEDI used in this study.

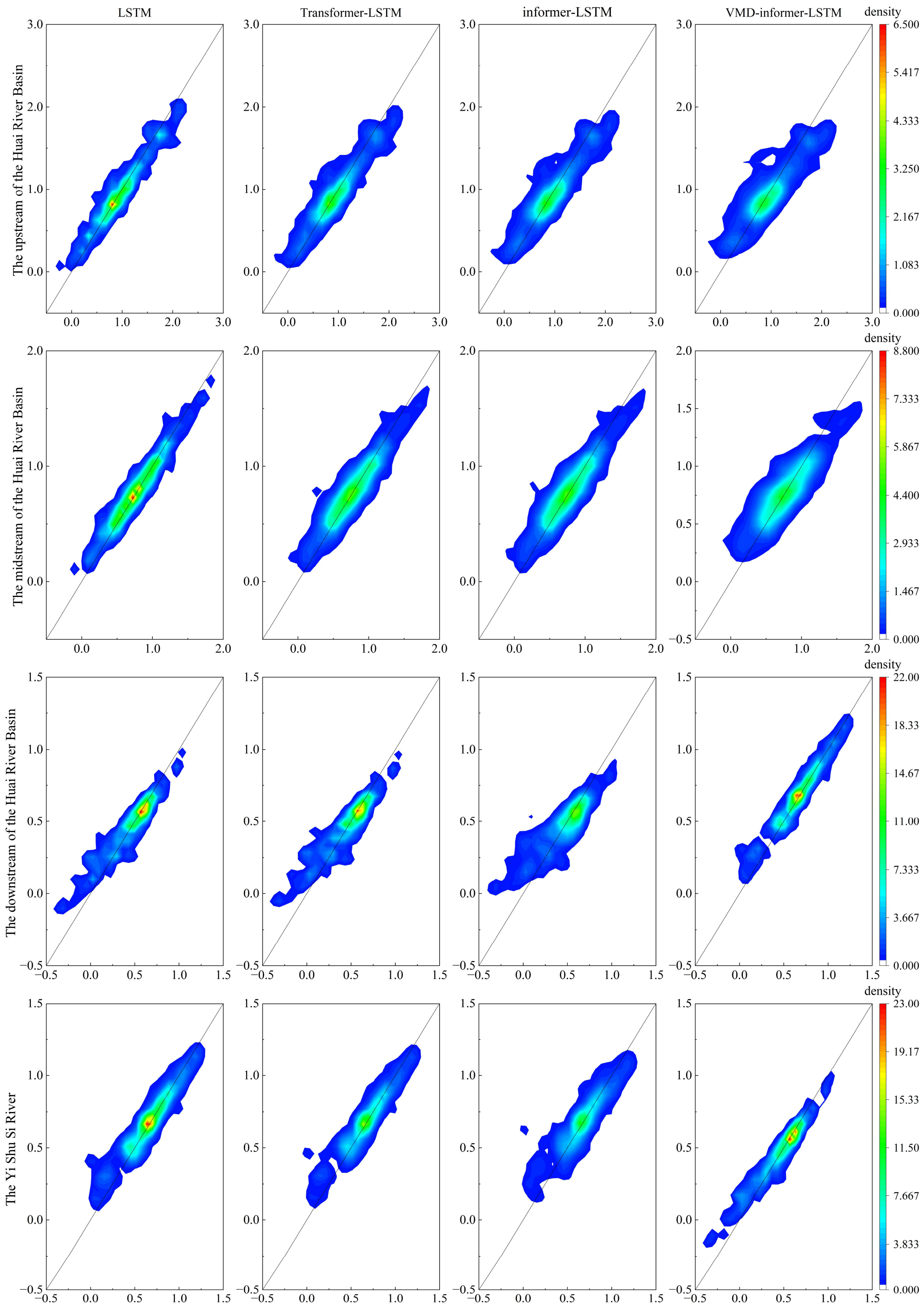


Figure S5. Density scatter plots of model-predicted DEDI against ERA5-derived reference DEDI in different regions of the Huaihe River Basin (x-axis: ERA5-derived reference DEDI; y-axis: model-predicted DEDI). The diagonal line represents the 1:1 reference line, and the color bar indicates point density.