My review comments are structured as follows: **Overall Assessment**, **Major Strengths**, and **Recommendations for Improvement**.

I. Overall Assessment

This paper presents a systematic evaluation of the water balance consistency of 47 state-of-the-art hydrological datasets (precipitation, evapotranspiration, runoff, and soil moisture) using 8,294 independent combinations. The methodology is rigorous, the data coverage is extensive, and the study holds significant scientific and practical value. It reveals a widespread lack of water balance consistency in current global hydrological datasets and provides an in-depth analysis of the spatial patterns, influencing factors, and temporal trends. The manuscript is well-structured, the methods are transparent, and the results are credible. I recommend **acceptance after minor revisions**.

II. Major Strengths

1.High Novelty: This is the first study to systematically assess the consistency of multi-source, multi-variable hydrological datasets from a water balance perspective, filling a critical gap in the current literature.

2.Methodological Rigor:

a.The use of independent dataset combinations effectively avoids spurious consistency arising from the use of the same model or forcing data.

b.The use of adjusted R^2 as the consistency metric mitigates errors introduced by unit inconsistencies between variables.

c.The application of SHAP for factor attribution enhances the interpretability of the results.

3.Comprehensive Data Coverage: The inclusion of gauge-based, satellite-based, and reanalysis products ensures broad spatiotemporal coverage and strong representativeness.

4.Insightful and Actionable Results:

a.Clearly identifies the strengths and weaknesses of different data sources across various regions and climatic conditions.

b.Highlights the significant impact of soil moisture data depth on consistency.

c.Reveals an improvement in dataset consistency in mid-to-high latitude regions of the Northern Hemisphere in recent decades.

III. Recommendations for Improvement

1. Clarifications in the Methods Section

Handling Soil Moisture Depth Differences: While the manuscript states that ΔSM represents "change," the response of soil moisture at different depths to P-ET-R varies. It would be beneficial to clarify if any normalization or sensitivity analysis was performed for ΔSM across different depths.

Temporal Scale Analysis: The significant differences in consistency between daily and annual scales warrant further discussion of the underlying physical mechanisms (e.g., high noise at daily scales, strong smoothing effects at annual scales).

2. Deepening the Results and Discussion

Root Causes of Low Consistency: Beyond the mentioned observational errors and model structures, could factors like surface-groundwater exchange or human activities (e.g., irrigation, reservoir regulation) also contribute? Expanding the discussion on this point would be valuable.

Mechanisms Behind Spatial Consistency Patterns: For instance, is the low consistency in high-latitude regions linked to insufficient representation of processes like snowpack and permafrost? Further interpretation in the context of existing literature is recommended.

3. Figures and Presentation

Figure 1: The meaning of the asterisk * and dashed lines in the boxplots should be explicitly stated in the figure caption.

Figure 2: The grey areas, indicating "multiple datasets show similar performance or low consistency," would benefit from having the specific thresholds for "similar" and "low" defined in the caption or figure.

Supplementary Material: Briefly mentioning the names of the best/worst performing datasets from Figures S13–S28 in the main text would help readers quickly grasp key findings.

4. Language and Formatting

Some sentences are quite long; breaking them up would improve readability.

Terminology should be checked for consistency (e.g., unified use of "gauge-based" vs. "station-based").

IV. Recommendation

Recommendation: Minor Revision

This manuscript makes a pioneering contribution to the evaluation of hydrological datasets. It is scientifically sound, its conclusions are robust, and it provides crucial insights for hydrological model development, data fusion, and climate change research. I recommend acceptance after the authors address the points above.
