Review for “SUPER v2: A 3-Hourly Global Precipitation Dataset Optimized for Sparse Data Challenges” By Zhang et al.

The manuscript presents SUPER v2, an updated version of the 2022 SUPER global precipitation product, which merges multiple satellite- and reanalysis-based precipitation datasets into a unified global 3-hourly dataset using a linear uncertainty-weighted least-squares framework built on Quality Control (QC) and Correlated Triple Collocation (CTC) error estimation. The principal modification relative to the original version is the reduction of candidate input datasets to an “optimal” subset of four products, motivated by the finding that strong interdependence and error cross-correlation among precipitation datasets can degrade merging performance under linear estimation and damp extreme variability. The method retains monthly bias correction, rain/no-rain adjustment, and additive linear error assumptions, while the 3-hourly product is derived by redistributing daily totals according to IMERG temporal structure (with ERA5 fallback when necessary). The merged product is evaluated against gauge observations in China and parts of Africa using correlation, RMSE, bias, detection skill, and extreme precipitation indices, with reported improvements over individual parent datasets and enhanced stability compared to SUPER v1. The authors argue that the revised framework reduces redundancy, improves robustness, and remains independent of direct gauge input during merging, thereby offering a globally applicable precipitation dataset intended for data-sparse regions.

While the dataset is carefully constructed and evaluated, I am not convinced that SUPER v2 represents a sufficiently substantive methodological or dataset-level advance beyond the 2022 version to warrant publication in ESSD. The core framework remains a linear least-squares merging approach with QC/TC-based weighting, and the primary modification—reducing candidate inputs to mitigate interdependence—reflects expected behavior under multicollinearity rather than a conceptual breakthrough. The selection of candidate datasets based on gauge performance in China introduces regional tuning into what is presented as a globally applicable framework, and the manuscript does not sufficiently demonstrate the robustness of this selection across independent climatic regimes. Moreover, recent nonlinear precipitation merging approaches are not meaningfully discussed or benchmarked, limiting the assessment of added value relative to current state-of-the-art methods. For these reasons, I recommend rejection for publication in ESSD. Major comments are listed below.

1. **Linear framework and expected behavior under multicollinearity**
   The merging strategy remains fundamentally linear, combining monthly-corrected daily products using least-squares weights derived from QC/TC uncertainty estimates. Under such a linear structure, multicollinearity among correlated precipitation products naturally leads to instability and variance damping. Therefore, the finding that reducing redundant inputs improves performance is largely a predictable property of linear estimators rather than strong evidence of methodological advancement. Given that precipitation processes are highly nonlinear and exhibit heteroskedastic and intensity-dependent errors, the reliance on linear assumptions may represent a limitation rather than a novelty.

2. **Lack of comparison with modern nonlinear merging approaches**
   Although the manuscript acknowledges the existence of neural network approaches, the evaluation primarily benchmarks against individual products and MSWEP. It does not assess

whether the proposed linear framework is competitive with more recent nonlinear machine-learning-based fusion methods that can capture regime-dependent biases, storm-type differences, or intensity-dependent error structures. Including at least one representative nonlinear baseline would substantially strengthen the claim of added value, particularly for precipitation where error distributions are non-Gaussian and strongly nonlinear.

3. **Limited demonstration of global generalizability**
   The merged product is evaluated against gauge observations in China and parts of Africa, which provides useful regional validation. However, these regions represent specific hydroclimatic regimes and cannot be assumed to reflect global precipitation diversity. Performance in these domains does not necessarily extend to tropical convection-dominated regions, high-latitude snow-dominated environments, or complex mountainous terrain such as the Andes or maritime Southeast Asia. Broader cross-regional validation or sensitivity analysis would be required to substantiate claims of global applicability.

4. **Regionally tuned candidate selection**
   Certain candidate datasets are selected based on their performance against gauge observations in China. This introduces a regionally tuned element into what is presented as a globally applicable framework. Dataset performance under East Asian monsoon conditions may not translate to other climatic regimes, potentially limiting global generalizability. The robustness of the selected subset across independent regions is not sufficiently demonstrated.

5. **Limited added value of the 3-hourly product**
   The downscaling step redistributes daily SUPER v2 totals into 3-hourly intervals using IMERG's temporal distribution (with ERA5 fallback when IMERG indicates zero rainfall). While operationally reasonable, this approach implies that sub-daily variability is largely inherited from IMERG/ERA5 rather than independently reconstructed through multi-source merging. Consequently, the incremental added value of the 3-hourly product relative to bias-adjusted IMERG may be limited and should be more clearly justified.

6. **Overstatement regarding sparse gauge independence**
   The manuscript emphasizes that the merged product is not affected by sparse gauge problems because no independent gauge observations are used in the merging framework. However, the absence of direct gauge input does not imply independence from gauge-related structural biases, as several parent datasets incorporate gauge calibration or evaluation in their development. The manuscript should clarify this distinction and moderate claims regarding immunity to sparse gauge conditions.