General comments

This manuscript develops TMY-MER, a typical meteorological year (TMY) dataset for 353 locations across Australia based on MERRA-2 reanalysis data (2009–2023). The authors apply a clear-sky / overcast weather classification, derive monthly bias parameters by comparing against an existing TMY dataset at 122 locations, and use inverse distance weighting (IDW) to extrapolate these bias parameters to 353 sites. The topic is relevant and the integration of weather-type-conditioned bias correction with reanalysis-based TMY generation is potentially useful. However, in my view the manuscript has fundamental issues in dataset definition, methodological description, scope of correction, and validation independence.

1. **Only GHI is corrected.** All error assessment and correction throughout Sections 2.2, 3.2–3.5, and 4.2 concern GHI only; the other meteorological variables in the output file (DBT, RH, WS, WD in Table 2) are passed through from raw MERRA-2. The weighting scheme in Table 1 also assigns GHI a 50% weight in typical-month selection, making the underlying selection itself solar-radiation-oriented rather than balanced across variables.
2. **“High resolution”.** MERRA-2's native resolution is 0.5° × 0.625° (line 156); the 353 "sites" are an oversampling of this coarse grid. In addition, the physical definition of a "site" is internally inconsistent (line 170 calls them weather stations “approximately 353 weather stations across these counties”, line 290 calls them city points from the World Cities Database), and it is likely that many sites share a single MERRA-2 grid cell.
3. **Independent validation.** The entire "validation" framework compares TMY-MER against an existing TMY product, which is itself a derived statistical construct rather than ground truth. The IDW correction is moreover fitted on the same 122 stations that subsequently appear within the 353 "validation" sites, producing data leakage. The Ladybug Tools building simulation, presented in the abstract as evidence of accuracy, is in fact a consistency check between two datasets fed into the same engine.
4. **The methodological description is insufficient.** The construction of TMY-MER, the status of IDW as a bias-correction step rather than a display step, the meaning of "AMY," the unspecified MREI threshold τ, and inconsistencies in Table 1 are difficult to follow from the current text.

In its current form, the manuscript reads more as "a climatological bias-correction analysis of MERRA-2 GHI against an existing Australian TMY product, with IDW spatial extrapolation" than as an independently validated, multi-variable, high-resolution TMY dataset. To meet ESSD's standards, the authors must either (a) substantially broaden the scope of correction and validation to support the current TMY framing, or (b) narrow the framing. And the manuscript requires substantial new analysis, such as independent ground-truth validation, cross-validation of the IDW step, evaluation or correction of non-GHI variables, and quantitative comparison against existing TMY products, and rewriting for the methodology description. I therefore recommend a major revision

Detailed comments

1. Issues with the Definition of the Dataset

1.1 The physical definition of the 353 "sites" is unclear and internally inconsistent

Line 167–168 states the dataset is extracted "from the MERRA-2 reanalysis database for approximately 353 weather stations across these counties," while lines 289–291 state "The grid was selected based on the open-source World Cities Database, which covers major cities in Australia." These two descriptions are incompatible: city centroids from a cities database are not weather stations. The reader cannot determine whether a "site" is (a) a weather ground station, (b) a city centroid, (c) a MERRA-2 native grid cell, or (d) a city coordinate point.

This ambiguity is particularly serious because MERRA-2's native spatial resolution is 0.5° × 0.625° (≈ 55 × 70 km, line 156), which is much coarser than typical inter-city distances. Multiple "sites" therefore very likely share the same underlying MERRA-2 grid cell, and the effective number of independent data points is substantially smaller than 353. The authors must (i) define unambiguously what a site is, with a coordinate file; (ii) report how many of the 353 sites map to distinct MERRA-2 grid cells; and (iii) clarify how sites that share a single grid cell differ from one another, if at all.

1.2 "High-resolution" in the title overstates the actual spatial resolution

The title, abstract, and Section 4.1 repeatedly describe TMY-MER as "high-resolution." However, the underlying MERRA-2 data have a native spatial resolution of 0.5° × 0.625° (line 156). Distributing the data to 353 discrete points does not constitute high spatial resolution. Also, it is unclear why TMY-MER is published as 353 discrete points rather than as a continuous Australia-wide grid. The authors should justify this choice, or consider releasing a gridded version. I recommend replacing "high-resolution" with "spatially distributed" or "regionally extensive" throughout the manuscript.

1.3 The 15-year base period is short

The 2009–2023 window (line 160) is shorter than the conventional 30-year climatological period typically used for TMY construction, and it overlaps a strong warming trend, which may bias the resulting "typical" months warm. A brief justification of the chosen window length, and a sensitivity discussion of how a longer window (which MERRA-2 makes available back to 1980) would change results, would strengthen the paper.

2. Issues with Data Quality and Validation

2.1 The "TMY" label is inconsistent with the actual scope of correction

A Typical Meteorological Year is, by definition, a physically coherent multi-variable record in which solar radiation, temperature, humidity, and wind are mutually consistent. This inter-variable coherence is what enables EnergyPlus, Ladybug Tools, and similar engines to produce meaningful building simulations. The manuscript's title and abstract claim a TMY dataset, but all error assessment and correction (Sections 2.2, 3.2–3.5, 4.2) and all reported error metrics concern GHI only. Temperature, humidity, wind speed, and wind direction (Table 2) are passed through from raw MERRA-2 without any

error evaluation or correction. This will create a physical-consistency problem: applying a bias correction to GHI alone, while leaving cloud-correlated variables (RH, DBT) uncorrected, breaks the multi-variable coherence that justifies the TMY concept. After correction, high-GHI hours may no longer correspond to low-RH, low-cloud conditions in the raw MERRA-2 record. Building simulations using this dataset may therefore produce inconsistent results, especially in cooling-dominated coastal climates where DBT–GHI coupling matters.

2.2 Quantitative comparison against existing TMY products is missing

The introduction (lines 76–84) mentions multiple existing TMY products and methods (Sandia, Danish, Festa-Ratto, CSWD; TMY3, IWEC; Argentine, Australian, Canadian datasets) and the ERA5-based global TMY of Wu et al. (2023), but no quantitative comparison against any of these is provided. Of particular concern is that a national TMY dataset for Australia already exists (Ren et al. 2021). The authors must demonstrate quantitatively where TMY-MER improves upon or differs from existing products. A summary table of methodological differences would also help the general reader.

2.3 There is no independent validation; the entire "validation" framework seems circular

The manuscript does not contain a clearly separated and independent validation section for the dataset. Section 2.3 is titled “Data validation and scalability”, but it mainly describes building-cluster simulations, DIRINT decomposition, and sky-model generation. The main quantitative evaluation in Section 3 compares TMY-MER with an existing TMY product, which should be interpreted as inter-product comparison rather than independent validation. The IDW-adjusted results further demonstrate fitting to the reference TMY dataset rather than out-of-sample performance, because no leave-one-out or spatial-block cross-validation is provided. Therefore, the current manuscript lacks independent validation against ground-based solar-radiation observations or other external reference datasets.

Throughout the paper, "validation" consists of comparing TMY-MER against an existing TMY dataset (Figures 2, 5, 6, 7, 8). However, the TMY dataset is not ground truth, it is also a statistical construct derived from sparse station interpolation and typical-month selection. Comparing two derived products yields product–product agreement, not accuracy. The errors reported throughout Section 3 should therefore be understood as discrepancies between two model-based products, not as estimates of true bias. Australia provides several genuinely independent solar-radiation data sources. And the author should consider perform independent validation against those datasets as ground-truth.

This problem is compounded by data leakage in the IDW step. The 122 comparison stations are used to derive the bias parameters MREB, MREO, MREL, and MREG, and these parameters are then IDW-extrapolated to all 353 stations (Eqs. 3–4). The 122 stations are a subset of the 353; their post-correction "errors" are by construction near zero. Figure 7e at station P37 illustrates this circularity directly — the "calibrated" curve matches TMY because the correction was fitted to match TMY. This is not validation; it is fitting quality. The authors should use leave-one-out (or spatial-block) cross-validation of the IDW correction. The author should withhold subsets of the 122 calibration stations and demonstrate that the IDW-extrapolated correction performs well at unseen locations.

2.4 The Ladybug Tools building cluster simulation is an application, not a validation

Section 4.3 currently presents the LBT building-cluster simulation as evidence of dataset accuracy ("Validation through building cluster simulations shows that the optimized dataset achieves over 90% consistency with traditional TMY data," abstract line 36–37). This framing is misleading. Two datasets fed into the same Radiance/LBT engine producing similar surface irradiance patterns demonstrates consistency, not accuracy — if both datasets carry the same systematic bias under cloudy conditions, the simulation will reproduce that bias in both runs and still show "90% consistency."

The authors should either (i) move Section 4.3 to Supplementary Information as Reviewer 1 suggested, or (ii) explicitly relabel it as an "Application case study" and remove "validation" claims from the abstract, conclusions, and Section 4.3 itself.

3. Issues with Methodology Description

Reviewer 1 has documented these methodological-description issues in detail; I concur with most of the methodological-description issues raised by Reviewer 1 and do not repeat them here.