This manuscript presents an interesting and timely study that constructs a 15-year (2009–2023) typical meteorological year dataset for Australia (TMY-MER) based on NASA's MERRA-2 reanalysis data. The main contribution is a multi-step calibration pipeline that classifies sky regimes, corrects solar radiation errors, and interpolates error parameters to 353 locations via IDW, followed by DIRINT direct-diffuse separation and building surface radiation testing. While the research objective aligns well with the growing demand for spatially dense climatic data in urban energy simulation, I have several major reservations regarding the methodological robustness. Specifically, questions remain regarding potential data leakage in the IDW validation, the absence of comparative baselines against established bias-correction methods, and an unaddressed temporal mismatch with the reference TMY files. Addressing these concerns is essential to ensure the reliability and reproducibility of the proposed dataset. Consequently, I recommend a major revision, with specific comments outlined below.

(1) The authors use 122 traditional TMY stations to compute error metrics and then apply Inverse Distance Weighting (IDW) to interpolate these errors across 353 sites. However, the independence of the subsequent validation is questionable. It is unclear whether the test locations used for validation (e.g., Figure 7e) were strictly withheld from the 122 reference stations used to generate the IDW weight matrix. If test stations were already part of the calibration set, validating against them amounts to circular reasoning, which inherently yields artificially low errors and fails to demonstrate genuine out-of-sample predictive accuracy. Given that Australian stations are heavily clustered along the southeastern coast while inland areas are extremely sparse, the authors must implement a rigorous spatial cross-validation scheme to demonstrate that the IDW-based framework is truly generalizable across data-sparse regions.

(2) While the proposed classification-based error correction framework is interesting, the manuscript lacks essential comparative benchmarks to substantiate its superiority and necessity. The authors only compare the raw TMY-MER against the recalibrated TMY-MER, without testing against established baseline methods widely used. Without such comparisons, it is impossible to evaluate whether the

multi-step clear-sky classification framework provides significant performance gains over simpler, standard approaches. Furthermore, in Section 3.6, the authors compare DIRINT-decomposed TMY-MER data against baseline TMY data decomposed via the BRL model. This introduces a critical confounding variable, as it conflates MERRA-2 meteorological deviations with intrinsic algorithmic differences between DIRINT and BRL; the same decomposition model must be applied to both datasets to ensure an apples-to-apples comparison.

(3) There is a fundamental temporal inconsistency between the datasets being compared that undermines the error assessment in Section 2.2. The TMY-MER dataset is constructed from modern reanalysis data spanning 2009–2023, whereas the authors fail to specify the exact time horizon of the 122 traditional TMY stations used as the ground trut reference. Traditional Australian TMY files (e.g., ACDB/NatHERS benchmarks) typically rely on historical observations from earlier periods (such as 1970–2005). Because Australia has experienced notable decadal climatic shifts, altered cloud regimes, and extreme weather over the last two decades, differences between these datasets naturally reflect real climate change and decadal variability rather than intrinsic simulation flaws in MERRA-2. Forcing modern reanalysis data to align with an outdated historical baseline risks artificially suppressing recent climate realities, and the authors must explicitly clarify this temporal discrepancy and justify how true model error is separated from decadal trends.

In addition, there are several minor discrepancies, unit errors, and presentation issues that should be addressed to improve the technical quality of the manuscript. These specific points are summarized below:

(1) In Section 3.3 and across the captions of Figures 3–6, peak radiation intensities are reported in kWh/m^2 (e.g., describing peak intensity approaching 1000 kWh/m^2). Instantaneous or hourly solar irradiance should be expressed in power units (W/m^2), as hourly cumulative values of that magnitude are physically impossible. Please

systematically verify and correct these units throughout the text and figure labels.

(2) Several equations and symbols failed to render properly. Specifically, the summation and variables in Equation (2) and Equations (5)–(6) contain broken characters and apparent misprints, and the Nomenclature section includes multiple unrendered symbol placeholders. The entire mathematical notation requires a thorough proofreading.

(3) In Section 2.2 (Step 1), clear-sky days are identified based on a predefined relative error threshold (τ). However, the manuscript never explicitly specifies the numerical value of τ adopted in this study or how it was chosen. Please provide this exact threshold and briefly justify its selection.

(4) In Section 2.1.2, the text refers to extracting seven key parameters from an undefined "AMY dataset" (presumably Actual Meteorological Year). However, Table 1 actually lists 8 meteorological parameters alongside GHI (amounting to 9 parameters). Please define the acronym AMY at its first mention and reconcile the variable count with Table 1.

(5) In Section 2.1.2 (Step 5), the smoothing procedure at monthly boundaries is described as applying cubic spline interpolation to replace the 24-hour interface data with linearly distributed values. This description is contradictory, and smoothing over an entire 24-hour window risks distorting the natural diurnal cycle and peak values of temperature and radiation. Please clarify the exact interpolation procedure applied at the month boundaries.