Developing a High-Resolution Typical Meteorological Year Dataset for Solar Radiation Evaluation in Australia

Jingpeng Fu[1,2,3,8], Pingan Ni[1,2,3,8], Deo Prasad[3], Guojin Qin[4], Fuming Lei[1], Yingjun Yue[1], Jiaqing Yan[1], Zengfeng Yan[1,2*], Bao-Jie He[5,6,7*]

[1]School of Architecture, Xi'an University of Architecture and Technology, Xi'an, 710055, China

[2]State Key Laboratory of Green Building, Xi'an, 710055, China

[3]School of Built Environment, University of New South Wales, Kensington 2052, NSW, Australia

[4]School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

[5]Centre for Climate-Resilient and Low-Carbon Cities, School of Architecture and Urban Planning, Ministry of Education, Key Laboratory of New Technology for Construction of Cities in Mountain Area, Chongqing University, Chongqing 400045, China

[6]School of Architecture, Design and Planning, The University of Queensland, Brisbane 4072, Australia

[7]CMA Key Open Laboratory of Transforming Climate Resources to Economy, Chongqing, 401147, China

[8]These authors contributed equally

*Corresponding author:

Zengfeng Yan, Email - yanzengfeng@xauat.edu.cn

Bao-Jie He, Email - baojie.unsw@gmail.com

Abstract: High spatiotemporal resolution typical meteorological year (TMY) data are essential for building energy modelling and urban climate studies. However, conventional TMY datasets, limited by sparse ground-based station coverage and infrequent updates, fail to meet the demands of detailed urban-scale simulations. To overcome these limitations, this study uses Australia as a case study and develops a new high-resolution dataset, the TMY derived from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2), hereafter referred to as TMY-MER. A novel weather classification approach was introduced, utilizing a mean relative error index derived from the ratio of daily to monthly maximum solar radiation to identify clear-sky conditions. Uncertainty errors were spatially interpolated using the inverse distance weighting (IDW) method. The results reveal several limitations in the previously generated TMY datasets. TMY-MER demonstrates stable accuracy under clear-sky conditions, with annual average errors below 5%, while under cloudy conditions, influenced by cloud simulation bias, errors can reach up to 50%. Spatially, annual solar irradiance is overestimated by 30% in southeastern coastal urban clusters, while errors in inland regions remain below 10%. Temporally, the peak error during cloudy winter periods reaches 30%, whereas summer clear-sky errors are under 5%. Further analysis using the direct-diffuse separation model indicates a systematic overestimation of diffuse horizontal irradiance (DHI) within 6%, and an approximately 20% negative bias in direct normal irradiance (DNI). Validation through building cluster simulations shows that the optimized dataset achieves over 90% consistency with traditional TMY data, with monthly mean errors below 5%. The multidimensional error assessment framework significantly enhances the reliability of reanalysis data for use in complex climate zones, supporting dynamic energy system planning and urban thermal environment modelling.

Keywords: atmospheric reanalysis datasets; typical meteorological year; solar radiation; direct-diffuse separation

Nomenclature

Abbreviations

AOD	Aerosol Optical Depth	LBT	Ladybug Tools
BRL	Boland-Ridley-Laurent Model	LD	Downward longwave radiation
CF	Cloud Fraction	MERRA2	Modern-Era Retrospective Analysis for Research and Applications, Version 2
DBT	Dry Bulb Temperature	RH	Relative Humidity
DHI	Diffuse horizontal irradiance	TMY	Typical Meteorological Year
DIRINT	Direct Insolation Separation Model	TMY-MER	TMY based on MERRA-2 data
DNI	Direct normal irradiance	TPVs	Tibetan Plateau Vortices
GHI	Global horizontal irradiance	TY	Actual year of a TMY
IDW	Inverse Distance Weighting	WS	Wind Speed

Symbols

CSD_X^m	Monthly clear sky days in the TMY	MRE_G	Mean relative error of global annual data
CSD_{MER}^m	Monthly clear sky days in the MERRA-2	MRE_I	Mean relative error between $I_{d,max}$ and $I_{m,max}$
CSD_{bas}^m	Baseline number of clear sky days	MRE_L	Mean relative error of local monthly data
D_i	Distance	MRE_O	Mean relative error of Observation
D_m	A criterion	p	Power of distance
G_{on}	The corrected solar constant	Q_m	Total monthly solar radiation
i	Number of parameters	$S_{i,m}$	Standard error of all years
$I_{d,max}$	Daily maximum solar radiation intensity	$X_{i,m,y}$	Average value of parameter i of month *m* in year *y*
$I_{m,adj}$	Adjusted monthly maximum solar radiation intensity	$\overline{X}_{i,m}$	Average value of all years
$I_{m,max}$	Monthly maximum solar radiation intensity	z_o	Estimates calculated by spatial interpolation
K_i	Weights of the *i*-th meteorological parameter	z_i	Attribute values for (i=1, 2, 3...n) samples in the spatial difference calculation
K_t	The clearness index	θ_z	Zenith angle
$K_t^{'}(t)$	The modified clearness index at the current time step	$\eta_{i,m,y}$	Standardized results of the *i*-th parameter of the *m*-th month in year *y*
MRE_A	Mean Relative Error of Adjusted Global Data	τ	Clear-sky MRE_I threshold
MRE_B	Mean relative error of climate change	$\Delta K_t^{'}$	The stability correction index

1 Introduction

1.1 Research background

Solar radiation is the fundamental energy source driving the operation of the Earth's system, and its spatiotemporal distribution profoundly influences the climate system, ecological cycles, and the sustainable development of human society. As the core carrier of renewable energy, solar energy development and utilization not only represent critical pathways for addressing energy crises and climate change but also serves as essential technological support for achieving carbon neutrality targets (Dincer & Aydin, 2023; Olauson et al., 2016; Pfeifroth et al., 2024). Based on solar radiation meteorological datasets, the assessment of solar energy resources can be enabled, and the analysis of energy yields for photovoltaic (PV) and concentrated solar power facilities can be conducted (Bright, 2019; Korany et al., 2016). Moreover, solar radiation is a major factor shaping the urban thermal environment. By forecasting radiative energy, the strategic layout of buildings and green spaces can be optimized to enhance energy efficiency and mitigate the urban heat island effect (Jamei et al., 2020; Khan et al., 2024).

The building sector accounts for nearly 30% of global energy consumption (Falchetta et al., 2024), and accurate, compatible weather data serves as the foundation for building performance simulations (Dong et al., 2021). Simultaneously, reliable solar radiation data are essential for a wide range of scientific activities, including heritage conservation, meteorology, atmospheric sciences, and agricultural applications (Paulescu et al., 2022; Ray et al., 2024; Sarr et al., 2024; R. Wu et al., 2023). Therefore, achieving a more precise assessment of solar radiation data is a fundamental prerequisite for optimizing energy planning and improving the built environment. However, the spatiotemporal distribution of solar radiation is influenced by multiple factors such as geographic location, seasonal variability, atmospheric conditions, and terrain characteristics, leading to inherent complexity and uncertainty in its distribution (Ni et al., 2023). Although mainstream typical meteorological year (TMY) datasets reflect long-term climatic characteristics, they suffer from limitations such as insufficient spatial coverage and delayed updates (Cao et al., 2022). Addressing these evaluation challenges, the construction of meteorological datasets with strong spatiotemporal representativeness has been a critical breakthrough.

1.2 Literature review

TMY datasets are constructed by synthesizing long-term meteorological observations to generate a representative artificial year that captures the statistical characteristics of climate variability (Cui et al., 2017). By selecting twelve TMY datasets effectively reflect long-term climate patterns while minimizing the influence of short-term climatic anomalies (Machard et al., 2024; Tambula et al., 2023). These datasets have been widely used to support building energy consumption simulations and solar energy resource assessments by providing reliable climatic inputs (W. Li et al., 2017; Wang, Jia, et al., 2025; Yu et al., 2024). The generation of TMY datasets typically incorporates multiple key meteorological parameters, including dry-bulb temperature, dew-point temperature, and solar radiation (Sun et al., 2020). Various methodologies have been proposed for TMY construction, including the Sandia method (Hall et al., 1978), the Danish method (Lund, 1995), the Festa-Ratto method (Festa & Ratto, 1993), and the China Standard Weather Data (CSWD) method (Song et al., 2007). Prominent global TMY datasets include the U.S. TMY series (Marion & Urban, 1995; Wilcox & Marion, 2008), the International Weather for Energy Calculations (IWEC) series and the CSWD developed by the China Meteorological Administration (Song et al., 2007). Nevertheless, existing TMY datasets are often outdated and suffer from insufficient temporal resolution, limiting their ability to meet the growing demand for high-accuracy, real-time meteorological information required by modern building performance simulations and dynamic solar energy assessments.

The effectiveness of the Sandia method for TMY generation has been validated in solar energy assessments in Turkish building energy consumption simulations across multiple climatic zones (H. Li et al., 2023; Pusat et al., 2015). Concurrently, national datasets have been successively released in countries such as Argentina (Bre & Fachinotti, 2016), Australia (Ren et

al., 2021), and Canada (Berardi & Jafarpur, 2020). Despite their widespread adoption, conventional TMY methods continue to face several important limitations. In particular, most existing TMY datasets rely heavily on ground-based meteorological observations. However, the sparse distribution of observation stations (Cao et al., 2020; Hao et al., 2020), combined with the widespread lack of solar radiation sensors (Bre et al., 2021; Cao et al., 2022), significantly constrains their spatial coverage and reduces their ability to represent regional climatic variability.

To overcome these spatial and temporal limitations, reanalysis datasets have become an increasingly important resource in recent years. Leveraging advances in data assimilation techniques, these gridded datasets integrate multiple sources of measurements with numerical model outputs, thereby improving spatial resolution and temporal continuity (Abatzoglou, 2013; Tang et al., 2019). Against these, the MERRA-2 reanalysis dataset developed by NASA's Global Modelling and Assimilation Office (NASA GES DISC, 2025) has emerged as a benchmark in the field due to its technological breakthroughs. Its hourly temporal resolution and improved assimilation algorithms, which incorporate aerosol data from AVHRR, MODIS and MISR satellites, and AERONET ground stations, substantially enhance the simulation accuracy of climate system interactions (Randles et al., 2017). MERRA-2 has been extensively applied across a wide range of disciplines, including climate dynamics research (e.g., diagnoses of tropical cyclone dissipation mechanisms and interdecadal evolution of Tibetan Plateau vortices) (Lin et al., 2024; Xie et al., 2024); land surface process modelling (e.g., permafrost model optimization and ocean surface net radiation calculations) (Liang et al., 2022; Qin et al., 2020), and the characterization of extreme weather events (Chen et al., 2024; Horan et al., 2024; Pulliainen et al., 2020). These applications demonstrate the significant multidisciplinary research value of MERRA-2. This paper primarily analyzes the solar radiation data from the dataset.

Nonetheless, while MERRA-2 has filled many observational gaps, its radiation data are not without flaws. Studies have shown that these biases primarily originate from deficiencies in simulating aerosol–cloud interactions. On the one hand, assimilation errors in aerosol optical depth (AOD) (Chakraborty & Lee, 2021) and the underestimation of low cloud cover (Camberlin et al., 2023) contribute to the overestimation of solar radiation. On the other hand, inadequate parameterization of cloud microphysical processes exacerbates radiation estimation errors over complex terrains (e.g., the Tibetan Plateau) and monsoon regions (Qi et al., 2024). Although researchers have identified deviations in radiation parameters through point-based validations (Aliana et al., 2022; Feng et al., 2023), most existing studies are constrained by short temporal records or small sample sizes (Du et al., 2022; Isaza et al., 2024), limiting their ability to capture the spatiotemporal heterogeneity of errors.

Efforts to validate MERRA-2 radiation data have achieved some regional progress, but several limitations remain unresolved. In China, researchers have consistently reported an overestimation of global horizontal irradiance (GHI), largely due to underrepresentation of cloud cover and AOD in the model (Cao et al., 2022; Du et al., 2022). Yet, these findings are often based on single-year analyses or long-term assessments that do not differentiate between weather types. In contrast, studies in Australia have highlighted an overestimation of AOD and an underestimation of solar radiation (Isaza et al., 2024), though sparse observational networks limit broader spatial conclusions. These discrepancies underscore the need for more comprehensive validation studies that span diverse regions, extended time periods, and varying meteorological conditions in order to improve more accuracy and applicability of MERRA-2 radiation data.

1.3 Research innovations

This study innovatively integrates reanalysis data with TMY generation methods, using MERRA-2 as the primary data source to develop TMY-MER, a nationwide TMY dataset for Australia. Through a spatiotemporal error decomposition model, the multidimensional distribution patterns of solar radiation errors in TMY-MER are systematically characterized. The major innovations and contributions of this study are:

(1) Development of a TMY dataset with high spatial coverage. Based on the most recent 15 years of MERRA-2 data, TMY-MER overcomes the limitations of traditional TMY datasets that rely on sparse ground-based observations, providing timely and high-resolution support for dynamic energy planning and urban thermal environment modelling.

(2) Multidimensional solar radiation error analysis. A novel weather-type classification method based on clear-sky identification is proposed to separately analyse solar radiation errors under clear and overcast conditions, revealing the spatiotemporal distribution patterns of errors. Coupled with inverse distance weighting (IDW) interpolation, the spatial generalization of error parameters is achieved, enabling the optimization of solar radiation data across all TMY-MER sites.

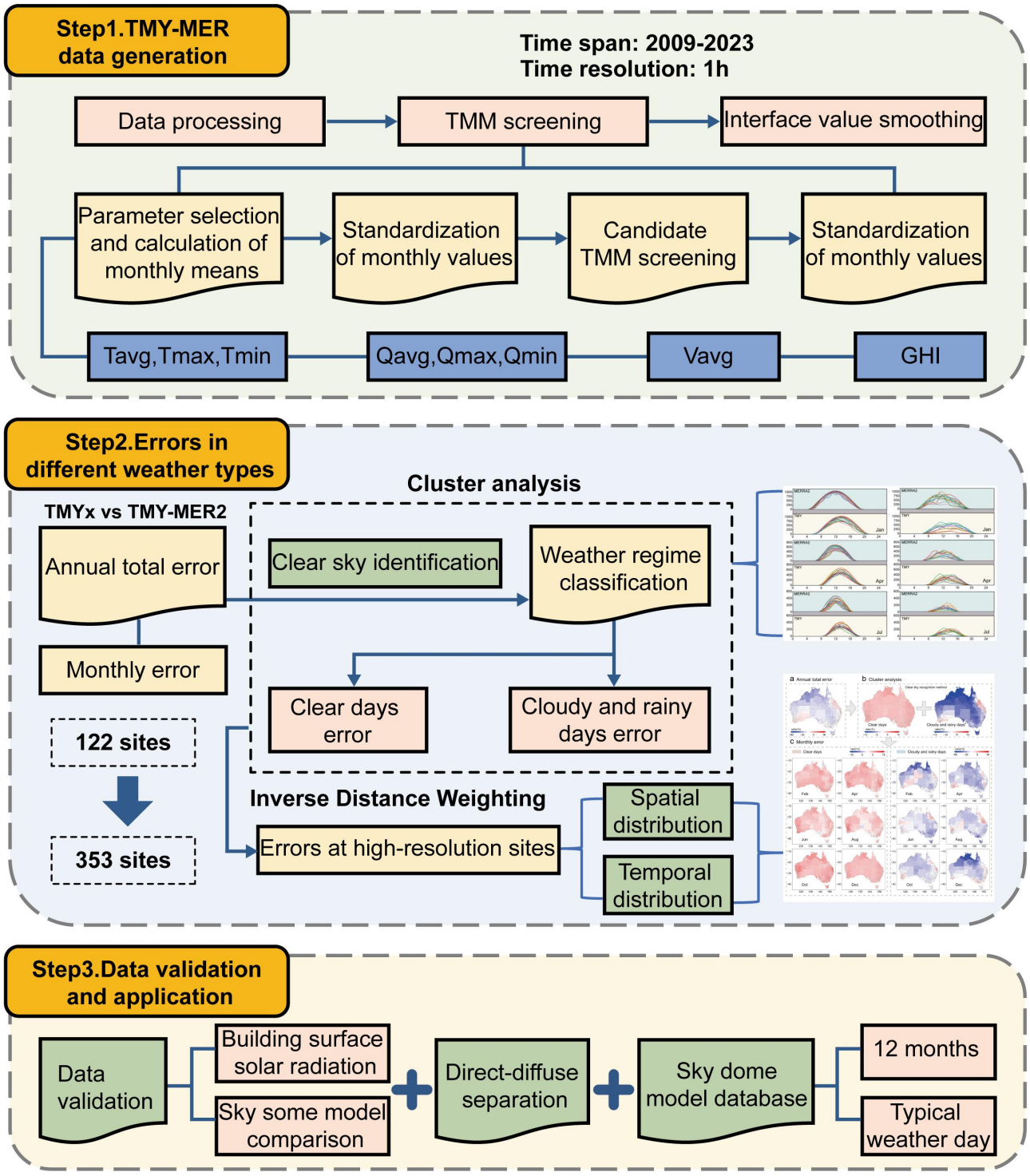


Figure 1. Research framework

2 Methods and materials

Figure 1 illustrates the core framework of this study, which aims to construct a high-accuracy dataset, named TMY-MER, encompassing three major modules: data generation, error assessment, and application validation.

Step 1: Based on MERRA-2 reanalysis data for hourly meteorological parameters for 2009 to 2023, from 353 sites across Australia were extracted. Using a TMY generation method that combines parameter normalization and weighted calculation, twelve TMMs were selected from the historical multiyear dataset. A cubic spline interpolation model was applied to smooth transitions at monthly boundaries, thereby enhancing data continuity and constructing a TMY-MER dataset covering all county-level regions in Australia.

Step 2: To address the systemic errors in MERRA-2 solar radiation data, a classification-based assessment method incorporating clear-sky identification was proposed. Clear and cloudy days were distinguished based on the relative error (*MREI*) between the daily maximum solar radiation intensity ($I_{d,max}$) and the monthly maximum solar radiation intensity ($I_{m,max}$). Monthly adjustments were applied to the radiation intensity on clear days, while the radiation data for cloudy days were adjusted based on the monthly total radiation. To ensure spatial coverage accuracy, IDW interpolation was employed to generalize the error parameters, enabling error correction across 353 TMY-MER sites in Australia.

Step 3: The accuracy of the adjusted data was validated by simulating building surface solar radiation on the Ladybug Tools (LBT) platform, comparing the results between the original TMY dataset and the adjusted TMY-MER dataset. The DIRINT model was applied to separate the GHI into direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI). Additionally, a multi-condition sky model database was further developed to enable efficient evaluation of solar energy distribution under various complex environmental scenarios.

2.1 Generation of TMY-MER

2.1.1 Data sources

MERRA2 is a new-generation reanalysis dataset, with spatiotemporal coverage extending from 1980 to the present. It offers a spatial resolution of 0.5°×0.625° and a temporal resolution as fine as one hour (Gelaro et al., 2017). By assimilating bias-corrected satellite aerosol observations and clear-sky solar radiation data and using observed precipitation to drive land surface process simulations, MERRA-2 significantly enhances the ability to accurately reproduce atmospheric parameters. Its core algorithms integrate the Chou-Suarez shortwave radiation scheme and the latest satellite data assimilation techniques from the GEOS-5 system (Randles et al., 2017). The total aerosol optical depth (AOD) of the dataset has been independently validated and is shown to be reliable (Buchard et al., 2017). In this study, 15 consecutive years of data from 2009 to 2023 were selected as the foundational dataset for constructing the TMY database. The key meteorological variables extracted from MERRA-2 for developing a high-resolution, multi-dimensional solar radiation dataset include: average dry-bulb temperature (T_{avg}), maximum dry-bulb temperature (T_{max}), minimum dry-bulb temperature (T_{min}), mean total column water vapor (Q_{avg}), maximum total column water vapor (Q_{max}), minimum total column water vapor (Q_{min}), average relative humidity (RH_{avg}), average wind speed (V_{avg}), and hourly GHI.

2.1.2 Method for generating the TMY-MER

This study focuses on the Australian region, using county-level boundaries as the basis for spatial segmentation. Hourly meteorological data from the past 15 years (2009–2023) were extracted from the MERRA-2 reanalysis database for approximately 353 weather stations across these counties, including solar radiation measurements. Based on this dataset, the data were processed using well-established methodologies for generating TMY datasets. A recent approach to TMY construction, originally developed for ERA5 reanalysis data, was adapted in this study to select 12 meteorological months

from historical multi-year records, forming a complete annual weather sequence for each station (Y. Wu et al., 2023). The detailed selection procedure for these 12 typical months is outlined below.

(1) Parameter selection and monthly mean calculation: Seven key parameters were extracted from the AMY dataset spanning 2009 to 2023. For each parameter (i), the monthly value ($X_{i,m,y}$) was calculated for each month (m) and year (y).

(2) Normalization of monthly values: Based on the 15-year dataset, the mean ($\overline{X}_{i,m}$) and standard error ($S_{i,m}$) were computed for each parameter and month. Monthly values were then standardized using Equation (Eq.) (1) to obtain the normalized values (($\eta_{i,m,y}$).

(3) Selection of typical meteorological months as candidates: A month was considered a typical candidate month if the normalized values ($\eta_{i,m,y}$) for all seven parameters were less than 1. Multiple candidate months could be identified under this criterion.

(4) Weighted score calculation and selection of typical meteorological months: Weights were assigned to the seven parameters based on their relative importance (Table 1). For each candidate month, a weighted score (D_m) was calculated using Eq. (2). The month with the lowest score was subsequently selected as the typical meteorological month.

Table 1. Factors and their weights for selection

Meteorological Parameter	Selection parameter	Weight
Dry-bulb temperature	$\boldsymbol{T_{avg}}$ (Average dry-bulb temperature)	2/20
	$\boldsymbol{T_{max}}$(Maximum dry-bulb temperature)	1/20
	$\boldsymbol{T_{min}}$(Minimum dry-bulb temperature)	1/20
Total quantity of vapors	$\boldsymbol{Q_{avg}}$ (Average total quantity of vapors)	2/20
	$\boldsymbol{Q_{max}}$(Maximum total quantity of vapors)	1/20
	$\boldsymbol{Q_{min}}$(Minimum total quantity of vapors)	1/20
Relative humidity	$\boldsymbol{RH_{avg}}$(Average relative humidity)	1/20
Wind speed	$\boldsymbol{V_{avg}}$ (Average wind speed)	1/20
Solar radiation	GHI(Global horizontal irradiance)	10/20

(5) Smoothing at monthly boundaries: To address discontinuities at the transitions between the 12 selected months, a cubic spline interpolation model (McKinley & Levine, n.d.) was applied to the 24-hour interface data. The original 24 interface values were replaced by a linearly distributed set of smoothed values.

The resulting TMY-MER station distribution demonstrates strong spatial representativeness. Each county-level region in Australia is covered by at least one meteorological station. Moreover, areas with higher urban density feature finer subdivisions and denser TMY-MER station distributions, ensuring that the extracted datasets more accurately capture the full meteorological characteristics of each region.

$$\eta_{i,m,y} = \frac{X_{i,m,y} - \overline{X}_{i,m}}{S_{i,m}} \quad (1)$$

where i is the number of parameters, m is the number of months, and y is the number of years, $X_{i,m,y}$ is the average value of parameter i in month m of year y, $\overline{X}_{i,m}$ is the average value of all years, and $S_{i,m}$ is the standard error of all years.

$$D_m = \sum_{i=0}^{n} K_i \left|\eta_{i,m,y}\right| \quad (2)$$

where D_m is a criterion, K_i is the weight of the *i-th* meteorological parameter, and $\eta_{i,m,y}$ are the standardized results of the *i-th* parameter of the *m-th* month in year y, as shown in Eq. (1).

2.2 Multi-dimensional error assessment

Although ERA-5 and MERRA-2 data are widely used in many studies due to their advantages, such as high spatial resolution, finer temporal granularity, and long time series, existing research has shown that reanalysis data may contain errors that significantly exceed those caused by climate change. This study further evaluates the multidimensional errors in GHI between TMY-MER and traditional TMY datasets, defining potential climate change and observational errors based on different weather types. To address the multidimensional errors in MERRA-2 data, the clear-sky error, which is relatively small and well-defined, is set as the baseline error (MRE_B). After identifying the baseline errors caused by climate change, they are further used to assess the highly uncertain overcast day errors. The error assessment process mainly includes the following five steps:

Step 1: Weather type classification was based on clear-sky identification. Widely available, low-resolution meteorological data were selected as baseline data, with TMY serving as an example in this study. First, the relative error (MRE_I) between daily maximum solar radiation intensity ($I_{d,max}$) and monthly maximum solar radiation intensity ($I_{m,max}$) was calculated. Days with an error below a predefined threshold (τ) were classified as clear-sky days, and the number of clear-sky days for both datasets was denoted as CSD_X^m and CSD_{MER}^m. These values were ranked according to MRE_I size, and the corresponding MRE_I and daily sequences were recorded.

Step 2: Monthly clear day error quantification. The smaller value between CSD_X^m and CSD_{MER}^m for each month was selected as the reference number of clear days, denoted as CSD_{bas}^m. The relative error between the monthly average solar radiation values of the two datasets was evaluated. Radiation discrepancies exceeding the climate change baseline error (MRE_B) were defined as clear-sky observation errors (MRE_O). Based on MRE_O, the selected clear day data were either reduced or increased to ensure that the monthly maximum radiation on clear days in TMY-MER remained within the bounds defined by MRE_B. The recalibrated monthly maximum solar radiation intensity ($I_{m,adj}$) was then recalculated for subsequent anomaly detection.

Step 3: Annual and monthly overcast day error calculation. Due to the high uncertainty in solar radiation intensity ($I_{d,max}$) under overcast conditions, the total monthly solar radiation (Q_m) was adopted as a more robust metric for error assessment. Local monthly relative errors (MRE_L) and overall annual relative errors (MRE_G) were evaluated beyond the baseline error (MRE_B). MRE_L was used to scale solar radiation for overcast days, ensuring better consistency in monthly totals, while MRE_G was retained to preserve the intermonth solar radiation differences between TMY-MER and TMY datasets, reflecting the effects of climate change.

Step 4: Error anomaly detection and adjustment. To avoid the situation where CSD_X^m exceeds CSD_{MER}^m but requires a substantial upward adjustment of TMY-MER for cloudy days, which could result in excessively high sunny day solar radiation intensities outside the CSD_{bas}^m threshold, anomaly detection was implemented. For each month, the original $I_{d,max}$ were re-evaluated against the $I_{m,adj}$. If any $I_{d,max}$ exceeded $I_{m,adj}$, the adjusted hourly averages from the revised CSD_{bas}^m dataset were substituted. Following the anomaly detection and adjustment process, the MRE_T was reassessed.

Step 5: Spatial interpolation of error parameters. Error parameters derived from low-resolution station comparisons were interpolated to enhance the spatial accuracy of the TMY-MER dataset. IDW interpolation (Ding et al., 2025; Wang, Yao, et al., 2025; Yanto et al., 2017), as defined in Eqs. (3) and (4), was applied to calculate the weighted average of error parameters from N neighbouring reference stations. The spatial distributions of CSD_{bas}^m, MRE_B, MRE_L, MRE_G and MRE_T were then computed for all stations.

$$z_o = \sum_{i=1}^{n} \frac{1}{(D_i)^p} z_i \left[\sum_{i=1}^{n} \frac{1}{(D_i)^p} \right]^{-1} \tag{3}$$

$$D_i = \sqrt{(x_o - x_i)^2 + (y_o - y_i)^2} \tag{4}$$

where z_o represents the estimated value, z_i denotes the observed attribute at sample point i, D_i is the distance between the target location and the sample point, and p is the power parameter controlling the smoothing effect, typically set to *2* to minimize the mean absolute error.

2.3 Data validation and scalability

To verify the accuracy and reliability of the adjusted solar radiation data derived from the TMY-MER dataset, two building clusters characterized by complex shading relationships or unique layouts were selected as research subjects. These clusters encompass diverse building types, height variations, and spatial configurations, effectively simulating the varied impacts on solar radiation observed in real urban environments. Using the LBT platform (Ni et al., 2023; Yue et al., 2024), surface solar radiation simulations were conducted for the selected clusters using both TMY meteorological data and the corresponding TMY-MER data. During the simulations, the distribution characteristics of solar radiation on building surfaces were meticulously recorded and compared. In parallel, sky dome models were generated for each meteorological dataset, and a comparative analysis of their luminance distributions was conducted. This analysis systematically examined luminance variations across different azimuths and elevation angles, as well as trends in overall luminance values. To enhance computational efficiency and meet the demands of complex simulations, a multi-dimensional sky model database was developed, introducing an innovative data storage structure to keep vectorized parameters and solar radiation matrices. This module can parse solar radiation data from weather files, invoke the Radiance engine, and process complex simulation scenarios involving multiple temporal dimensions and radiation types in batches, thereby meeting the demands of advanced simulations.

To address the limitation of MERRA-2, which only includes GHI data, this study utilized the DIRINT model (Ineichen et al., 1992) from the pvlib library (https://github.com/pvlib/pvlib-python) as a method for splitting total radiation into direct, normal, and horizontally diffuse components. By introducing the $\Delta K_t^{'}$ stability index, the model can adaptively adjust the estimated DNI based on the dynamic characteristics of the GHI time series, thereby enhancing its adaptability under varying weather conditions, as shown in Eqs. (5)-(7). The DIRINT model is suitable for a range of environmental conditions, including different pressure and temperature conditions, making it widely applicable globally (Engerer, 2015; Lave et al., 2015).

$$K_t = \frac{GHI}{G_{on} \cdot cos(\theta_z)} \tag{5}$$

$$DNI = \frac{GHI - DHI}{cos(\theta_z)} \tag{6}$$

$$\Delta K_t^{'} = K_t^{'}(t) - K_t^{'}(t-1) \tag{7}$$

where K_t is the clearness index, GHI is the global horizontal irradiance, G_{on} is the corrected solar constant, and θz is the solar zenith angle. DNI is the direct normal irradiance, DHI is the diffuse horizontal irradiance. $\Delta K_t^{'}$ is the stability correction index, $K_t^{'}(t)$ is the modified clearness index at the current time step, and $K_t^{'}(t-1)$ is the value at the previous time step.

2.4 Research area

Australia was selected as a representative study region due to its unique resource conditions and the requirements of the research design. One of the main reasons is the high photovoltaic development potential in low-density urban areas. The dispersed building layout in these areas significantly reduces shading effects, thereby increasing the effective operating hours and unit-area efficiency of photovoltaic systems. Studies have shown that the photovoltaic potential in low-density areas can reach 1 to 6.6 times total energy consumption, which is significantly higher than that in medium- to high-density urban areas (Geng et al., 2024).

Another important consideration is the need to verify climate and geographic diversity. Australia spans from tropical to cool temperate climates and features a variety of terrains, including coastal plains, mountains, and deserts (Hughes, 2011). Its climate and geographic characteristics, such as sizeable convective clouds in the tropical monsoon region, stratocumulus clouds in the temperate oceanic climate, and clear-sky conditions in desert regions) provide an effective means to validate the applicability of reanalysis data in different environments, enhancing the universality of the error assessment methods.

A further justification for the selection of Australia is the urgency of energy transition and the availability of supporting data. As one of the countries with the richest solar resources in the world (with an average annual solar radiation of 2000–2500 kWh/m²), the Australian government plans to achieve a renewable energy share of 82% by 2030 (Yoo, 2023). High-precision solar radiation data is crucial for photovoltaic system planning and grid scheduling, and the scarcity of traditional ground-based observation data further underscores the value of constructing the TMY-MER dataset in this study. To support the subsequent analysis with a spatial context, geographic illustrations are employed in the following sections. All maps of Australia used in this study are sourced from OpenStreetMap (Herfort et al., 2023).

3 Results

3.1 Data sample documentation

Table 2. Sample of TMY-MER data records for Bega Valley , AUS

MONTH	DAY	HOUR	DBT (°C)	RH (%)	WS (m/s)	WD (°)	GHI (W/m²)	AMY (y)
1	1	0	16.205	35.388	249.906	2.983	0	2015
1	1	1	16.549	36.583	251.552	2.926	0	2015
1	1	2	16.666	37.435	252.302	2.764	0	2015
1	1	3	16.730	37.865	254.672	2.548	2.330	2015
1	1	4	16.758	37.944	261.009	2.487	104.781	2015
1	1	5	17.242	37.699	268.873	2.250	311.625	2015
1	1	6	18.626	36.774	273.367	2.053	529.250	2015
1	1	7	20.608	34.667	277.232	1.832	726.750	2015
1	1	8	22.397	32.984	279.595	1.114	888.750	2015
1	1	9	23.585	31.771	267.137	0.804	1000.000	2015
1	1	10	24.584	30.769	237.193	0.795	1054.500	2015
1	1	11	24.931	31.477	206.241	0.945	1048.500	2015
1	1	12	24.814	32.454	183.234	1.292	976.250	2015
1	1	13	24.354	33.058	166.891	1.829	841.250	2015
1	1	14	23.572	33.659	152.768	2.642	647.500	2015
1	1	15	22.462	34.374	142.915	3.087	419.500	2015
1	1	16	21.184	34.677	135.434	3.116	200.438	2015
1	1	17	19.943	34.475	131.307	2.857	57.766	2015
1	1	18	18.906	34.060	125.832	2.496	1.343	2015
1	1	19	18.327	34.016	116.899	2.061	0	2015
1	1	20	18.033	33.948	113.845	1.672	0	2015
1	1	21	17.793	33.928	108.520	1.245	0	2015
1	1	22	17.511	33.873	99.784	0.979	0	2015
1	1	23	17.321	33.826	90.687	0.775	0	2015

A total of 353 TMY-MER data sites were generated for the Australian region. The grid was selected based on the open-source World Cities Database, which covers major cities in Australia. The geographic distribution of the TMY-MER dataset spans 6 states and 2 major territories, covering a total of 353 local government areas in Australia. For each grid, the TMY includes data for 8,760 hours. The complete meteorological variables include month, day, hour, dry bulb temperature (°C), relative humidity (%), wind speed (m/s), wind direction (°), global horizontal irradiance (W/m²), and the actual year corresponding to the typical meteorological month (y). The complete data records for TMY-MER are listed in Table 2. Additionally, the weather data was converted to the EPW format for building performance simulation.

3.2 Evaluation of relative errors in total solar radiation

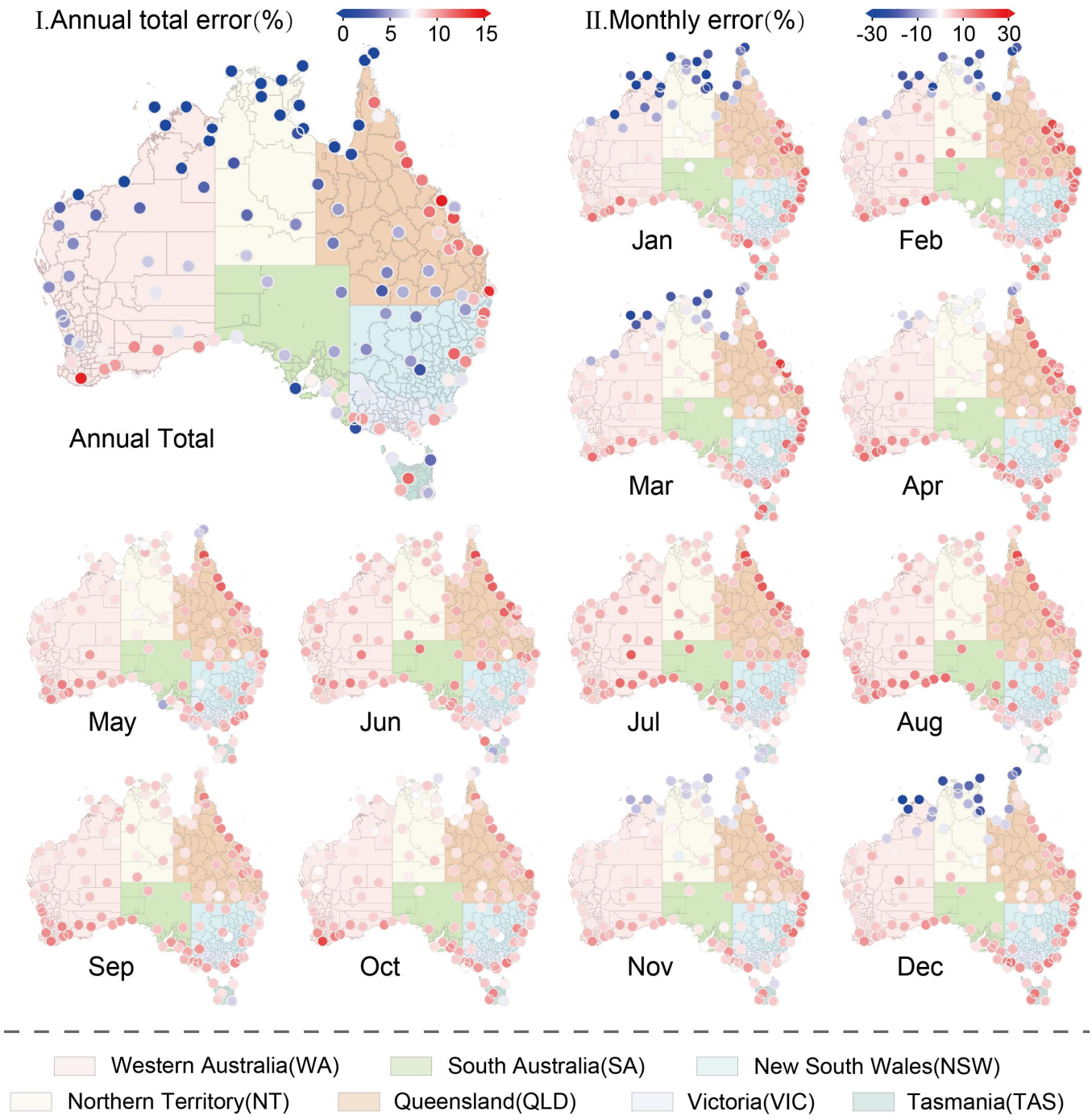


Figure 2. Error distribution of TMY-MER solar radiation data in Australia.

This study extracted the TMY solar radiation data from 122 TMY sites in Australia. The corresponding nearest neighbor TMY-MER data sites were selected for each TMY site. The TMY-MER solar radiation data generated were compared with the corresponding TMY data. The annual total errors and the error distribution for each month are shown in Fig. 2. Regarding the total solar radiation over the entire year, although the MERRA-2 dataset exhibits significant geographical variability in its error distribution, the error magnitude remains within 15%. Monthly time series analysis reveals clear seasonal variations in error characteristics. During the Australian summer (December to February), radiation errors exhibit notable spatial heterogeneity. The TMY-MER dataset consistently shows a positive bias along the eastern, southern, and southwestern coasts, while the northern and northwestern coastal regions display negative biases, with the maximum reaching 30%. In the inland areas of Australia, the radiation error magnitude is relatively low and predominantly

positive, with the vast majority of sites showing positive bias not exceeding 10%. In the transitional seasons of spring and autumn, March and November (months approaching summer) in the northern and northwestern coastal areas still show negative bias, while other months exhibit varying degrees of positive bias.

The spatial distribution pattern is as follows: the error magnitude is largest in the eastern and southern coastal areas, with autumn showing higher error magnitudes than spring; the inland areas exhibit smaller error magnitudes, not exceeding 10%. During winter (June to August), the TMY-MER dataset shows positive bias in all monitored regions, with the largest error magnitude observed in the northwest coastal region (with an extreme value of 30%), followed by the southern coastal regions. The central region has a significantly higher error magnitude compared to other seasons, reaching more than 20%. It is noteworthy that the positive bias in the southwestern coastal region during winter is notably lower than the error levels in this region during the summer. A comparison of the spatial and temporal scales reveals that more urbanized areas, such as the southeastern coastal region of Australia and the southwestern part of Western Australia (WA), exhibit more pronounced positive bias characteristics.

3.3 Weather type identification

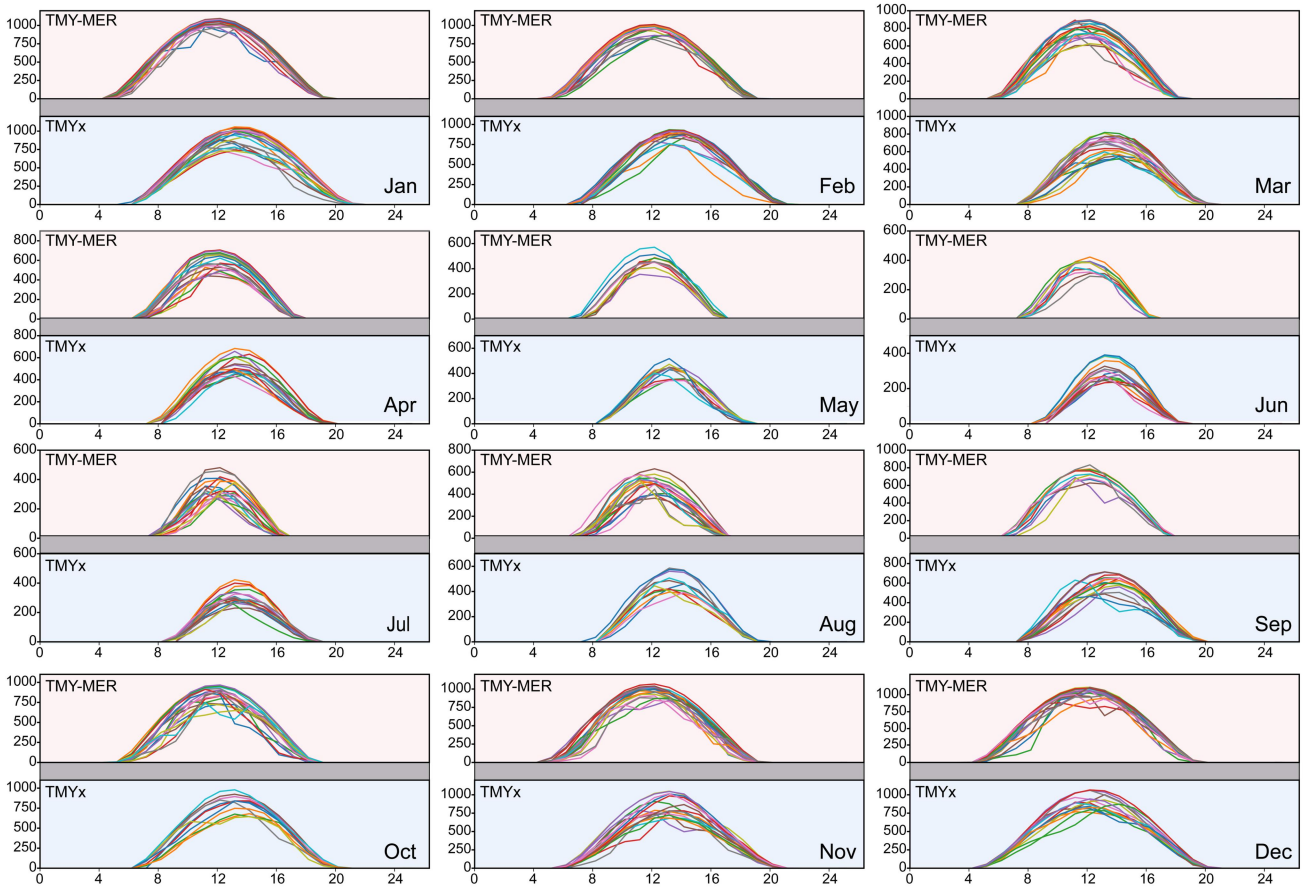


Figure 3. Hourly solar radiation on clear days in Central Highlands, Tasmania (kWh/m²)
Note: Each curve represents the hourly distribution of solar radiation on a particular clear day.

A clear-sky detection algorithm was applied to classify solar radiation parameters from the TMY and TMY-MER meteorological datasets into clear and overcast weather types. To ensure more accurate identification of monthly clear days at each site, a specialized classification model was developed. The study analyzed 122 geographically diverse locations across Australia, using TMY as the primary dataset and TMY-MER as the reference. For each site, monthly statistics were compiled for the number of clear and overcast days, along with corresponding solar radiation data. Hourly solar radiation distribution curves were constructed for each day of the month, allowing for comparative analysis of radiation patterns using

clustering methods. Site P38, located in the Central Highlands of Tasmania (TAS), was selected as a representative example (Fig. 3 and 4).

From a temporal perspective, the monthly radiation curves exhibit clear seasonal variation. In summer, particularly in December and January, the daily radiation distribution range expands significantly, with peak intensities approaching 1000 kWh/m². In contrast, during winter in June and July, daily peak values typically fall below 600 kWh/m², indicating strong alignment between solar radiation variability and seasonal transitions. Weather-type analysis further reveals that, within the same month, clear days consistently exhibit higher solar radiation intensities than overcast days. The hourly radiation curves on clear days are more concentrated and stable, whereas those on overcast days are highly dispersed and variable. For instance, in December, the maximum daily radiation difference between TMY and TMY-MER under overcast conditions exceeds 500 kWh/m², indicating substantial intraday fluctuations. Moreover, although these curves generally follow a diurnal cycle, overcast conditions introduce pronounced nonlinearity due to dynamic cloud interactions affecting solar transmittance.

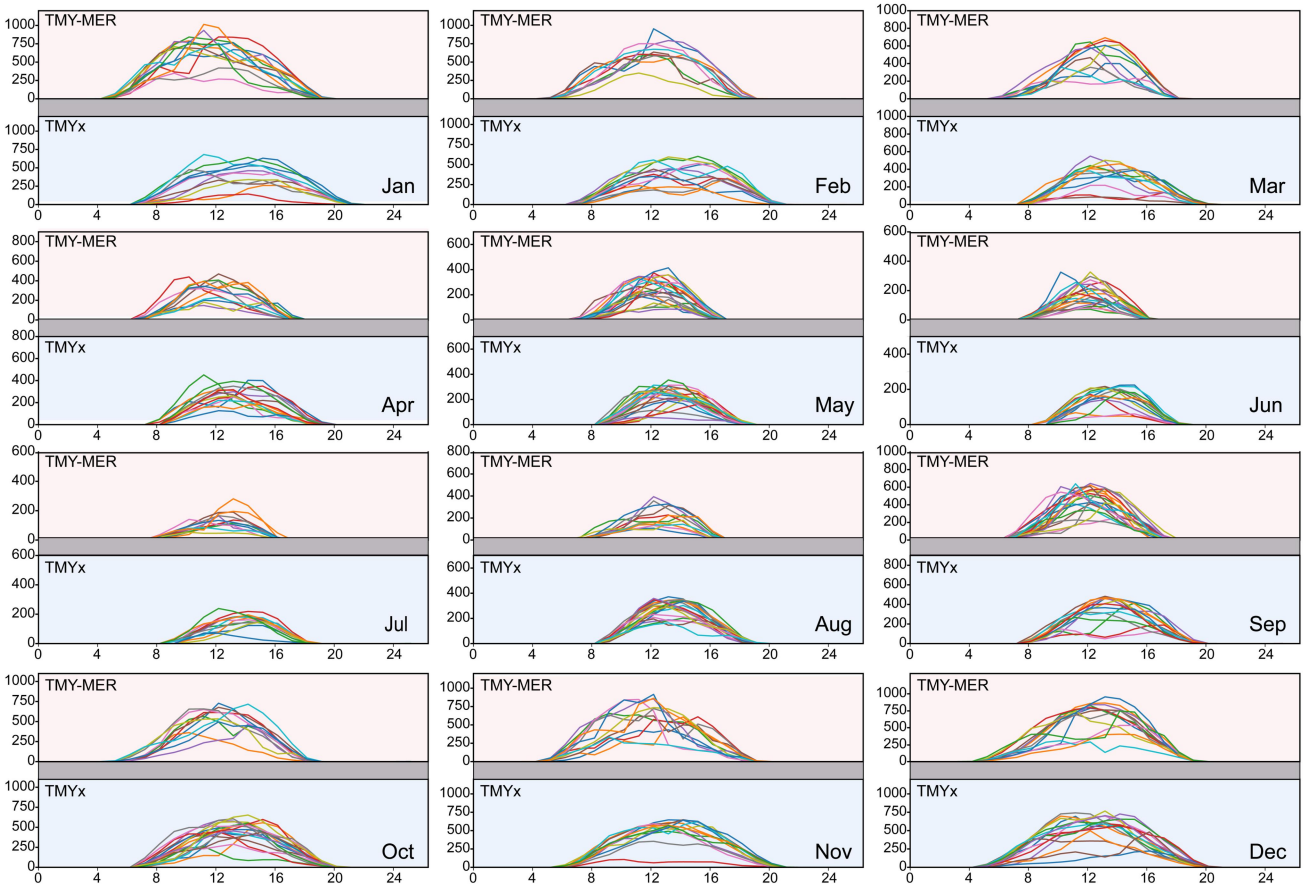


Figure 4. Hourly solar radiation on overcast days in Central Highlands, Tasmania (kWh/m²)
Note: Each curve represents the hourly distribution of solar radiation for a particular overcast day.

Annual comparative results show that TMY-MER performs reliably under clear-sky conditions, with radiation estimates exhibiting a positive bias of less than 10% relative to TMY. However, under overcast conditions, particularly during the summer months (December to February), the bias increases significantly, reaching up to 30-40%. Despite overall consistency between the datasets in the proportion of weather types, notable classification discrepancies emerge during certain periods. For example, at site P38 in August and September (late winter and early spring), significant differences are observed between the datasets. Based on these findings, subsequent site-level error analyses will adopt the monthly average solar radiation under each weather type as the baseline parameter.

3.4 Error analysis by weather type based on clustering

3.4.1 Monthly errors under different weather conditions

Based on the established evaluation methodology, monthly solar radiation errors were calculated at each site using average values under clear and overcast conditions as reference baselines. From the complete dataset, 24 representative meteorological stations were selected to analyse the distribution characteristics of monthly average solar radiation in the TMY-MER dataset under both weather scenarios as illustrated in Figs. 5 and 6. Under clear-sky conditions, TMY-MER consistently exhibits a slight positive bias across most sites, with error characteristics closely resembling those of site P38. Sites P36 and P37 show notably higher levels of overestimation, with annual cumulative errors exceeding 5%. These systemic deviations are mainly attributed to the stable optical properties of the atmosphere during clear weather, where minimal cloud cover and low aerosol concentrations lead to seasonal radiation patterns that govern the bias. Consequently, monthly errors at individual sites remain temporally consistent and spatially uniform, indicating robust model performance under clear-sky conditions.

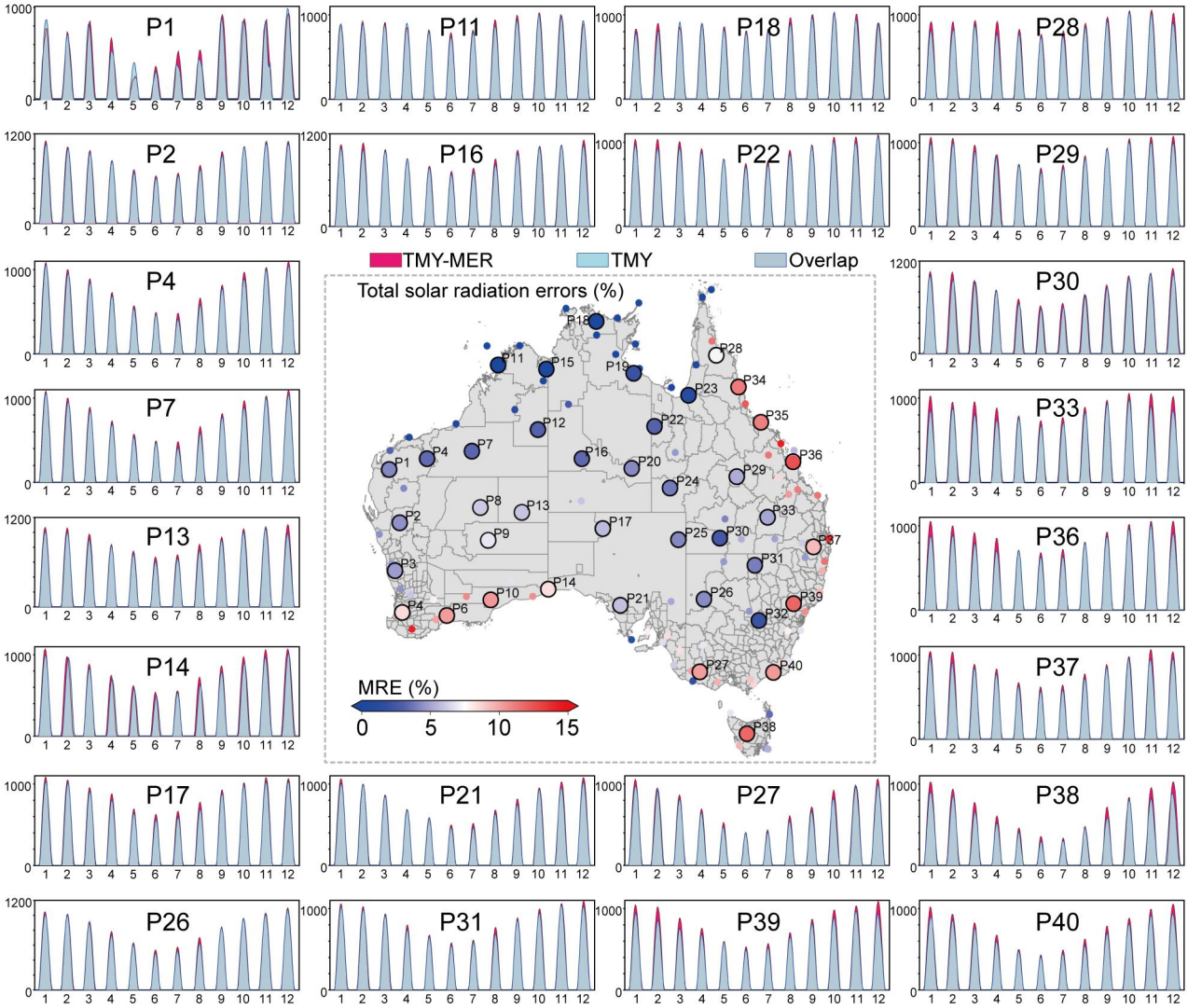


Figure 5. Monthly solar radiation errors across different regions under clear-sky conditions at various sites. (kWh/m²) Source

In contrast, solar radiation errors under overcast conditions are more complex and variable. Greater cloud cover uncertainty, increased aerosol loading, and diverse cloud configurations exert significant influence on radiative transfer, resulting in larger fluctuations in the discrepancies between TMY-MER and TMY datasets. Substantial overestimations exceeding 40% are observed during specific months at several stations, including P13 (March, August, December), P21

(November), P33 (November), and P36 (January). These findings highlight the sensitivity of solar radiation estimates to atmospheric variability under cloudy conditions and suggest that TMY-MER's performance is more susceptible to local meteorological disturbances in such scenarios.

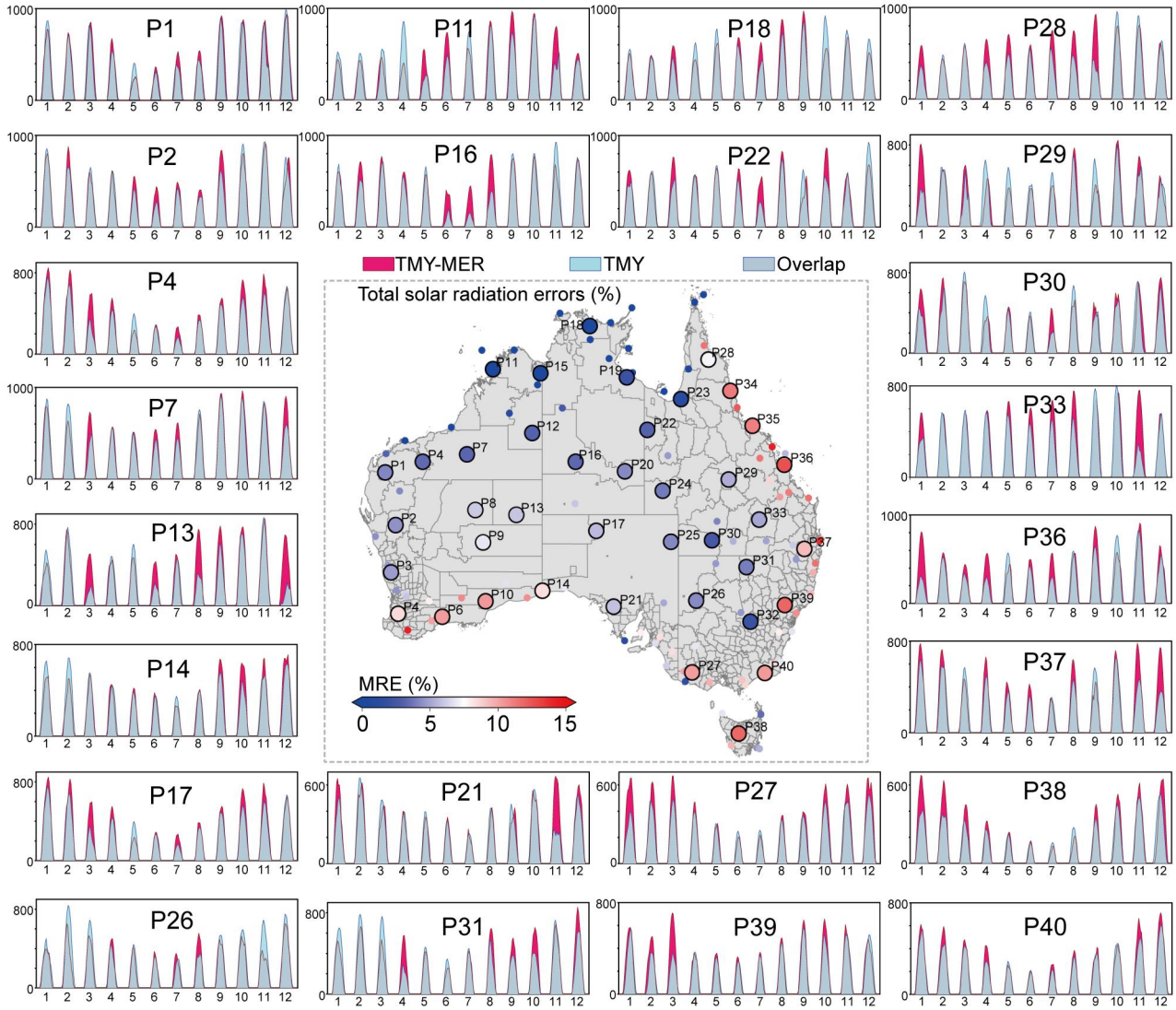


Figure 6. Monthly analysis of solar radiation errors across different regions under overcast conditions at various sites (kWh/m²).

Spatial analysis reveals pronounced regional differences in the distribution of solar radiation estimation errors. In the western coastal regions—such as stations P27, P36, and P37—TMY-MER consistently overestimates solar radiation, particularly during the wet season from November to March. These overestimations are likely linked to seasonal cloud variability and aerosol emissions in these areas. In the central inland regions—including but not limited to P13, P16, and P17—discrepancies between TMY-MER and reference datasets remain relatively small under clear-sky conditions, due to limited anthropogenic activity and low urban density, which reduce the influence of cloud cover and aerosols on reanalysis products such as MERRA-2. However, during cloudy winter months, notable overestimations are observed, and seasonal variations in bias are evident. For example, site P13 shows substantial overestimation in December and March under cloudy conditions, while P17 exhibits slight underestimation during spring and summer. At northern sites—namely P11, P18, and P28—the TMY-MER tends to overestimate solar radiation during summer, while in other months, most sites exhibit underestimation. For instance, P11 shows underestimation in autumn; P18 and P23 exhibit underestimation in both summer and autumn; P19 underestimates during spring and summer; and P28 underestimates from October to December. In southern regions, locations such as P4, P14, and P27 show varying degrees of overestimation of solar radiation during summer and

adjacent months under overcast and rainy conditions, whereas the errors in solar radiation during winter are relatively minor. These spatial patterns highlight the heterogeneity of TMY-MER performance across climatic zones and emphasize the influence of local atmospheric conditions on radiation modelling accuracy.

3.4.2 Spatiotemporal distribution of high-resolution solar radiation errors

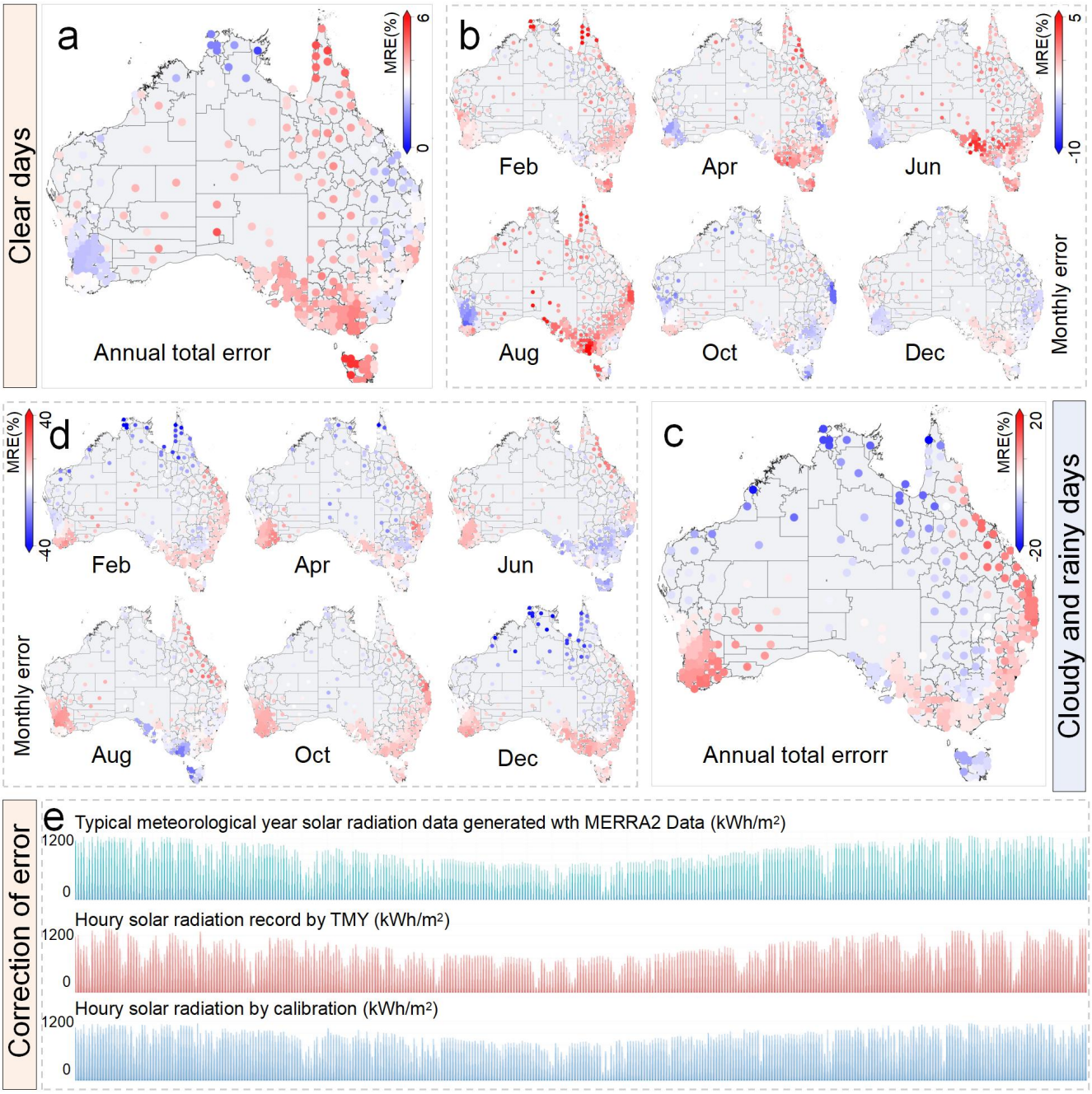


Figure 7. Spatiotemporal distribution of high-resolution solar radiation errors after IDW interpolation; (a) Annual solar radiation error on clear days; (b) Monthly solar radiation error on clear days; (c) Annual solar radiation error on cloudy and rainy days; (d) Monthly solar radiation error on cloudy and rainy days;(e) Annual solar radiation error error assessment at station P37.

By comparing solar radiation data from 122 TMY stations and their corresponding TMY-MER counterparts across Australia, solar radiation biases under clear-sky and overcast conditions were derived using a cluster-based weather-type classification. Using the IDW method, a weighted mean error was computed for each TMY-MER site based on its nearest bias-adjusted reference stations. This enabled the derivation of the spatial distribution of solar radiation errors across all 353 TMY-MER stations under both weather scenarios, as illustrated in Fig. 7a and 7c. Weather-type based annual radiation

analysis revealed that, under clear-sky conditions, the relative bias between the TMY-MER and TMY datasets ranged from 0% to 6%, with TMY-MER consistently exhibiting a positive bias. Spatially, smaller errors were found in the northwestern Cape York Peninsula, the southwestern region, and the eastern coastal belt between 30°S and 40°S. In contrast, larger deviations appeared in central Cape York, southern South Australia (SA), southern Victoria (VIC), TAS, and arid interior regions. Under overcast conditions, the bias range expanded to between -40% and 40%, with negative biases dominating the tropical north, TAS, and most inland areas, while positive biases were observed along the southwestern and eastern coasts, as well as in southeastern VIC and SA.

Temporal analysis, as illustrated in Fig. 7b and 7d, further reveals clear seasonal patterns in the spatial distribution of bias. Under clear-sky conditions from April to August, TMY-MER exhibited mild positive deviations, generally within 5%, across southeastern and northern coastal regions, while southwestern coastal areas showed negative biases reaching up to 10%. Between October and December, radiation biases were mostly negative and of smaller magnitude, with larger deviations concentrated in coastal zones and smaller ones inland. In February, biases approached zero across most regions, except for a localized positive deviation of approximately 5% over Cape York and the northern part of the NT. Under overcast conditions, seasonal bias variability became more pronounced. During summer, areas north of the Tropic of Capricorn experienced substantial negative biases ranging from –30% to –40%, whereas southern regions generally exhibited positive biases, with coastal zones showing more prominent errors compared to inland areas. In winter, the southwestern coastal region of Australia, including TAS, showed negative biases between -20% and -30%, while most other regions displayed positive biases, again with greater deviations along the coast. During transitional seasons, widespread positive biases ranging from 20% to 40% were observed throughout most coastal regions, excluding northern Australia, whereas inland areas generally exhibited smaller or slightly negative deviations. Validation at a representative station, site P37 (Fig. 7e), confirmed that the adjusted TMY-MER dataset effectively corrected the systematic biases observed in the original dataset across all 8,760 hourly time steps throughout the year.

3.5 Regional mean biases

Figure 8a presents the mean annual solar radiation bias across Australian administrative regions, comparing two meteorological datasets: TMY-MER and TMY. Figure 8b further evaluates the total radiation errors under clear-sky and overcast conditions, based on a sky condition classification approach. Figure 8c and 8d display the monthly distributions of regional mean solar radiation bias under these two distinct weather types, respectively. Under clear-sky conditions, the TMY-MER dataset generally overestimates solar radiation, with a distinct spatial pattern characterized by larger biases along the coastal regions and smaller biases in inland areas. Temporally, while monthly error trends remain relatively consistent, pronounced seasonal variations are evident across different regions. For instance, in inland zones such as the ‘Outback’ and Anangu Pitjantjatjara Yankunytjatjara Lands in SA, and the Petermann-Simpson region in the NT, the positive bias during winter is markedly reduced compared to other seasons. Conversely, in certain eastern coastal regions of WA, including Irwin, the model exhibits a more significant overestimation during winter; however, the maximum underestimation remains below 15%. Notably, a limited number of regions exhibit negative biases during specific months. For example, in February, the Cape York area in Queensland (QLD) shows a slight underestimation, as does the western subregion of the Victoria Desert in SA and the Corangamite, Ararat Surrounds, and St Arnaud areas in VIC during August. However, in all these cases, the maximum underestimation remains below 15%.

The monthly bias patterns under overcast conditions reveal more complex spatial and seasonal dynamics. During the summer months (e.g., December and February), substantial underestimations, up to 60%, are observed in the northern coastal areas, such as Kununurra and Halls Creek in WA and the Victoria River region in the NT. In contrast, central regions display relatively minor deviations. In winter and transitional months (April to October), moderate positive biases occur in eastern coastal QLD, eastern TAS, and southwestern WA, while most other regions are dominated by significant negative biases.

Specifically, underestimations in northern coastal and adjacent inland regions during the transitional months range from 30% to 40%, exceeding those observed in winter alone. In southeastern coastal areas, winter underestimation reaches its peak, with values up to 45%, surpassing those in the transitional season as well.

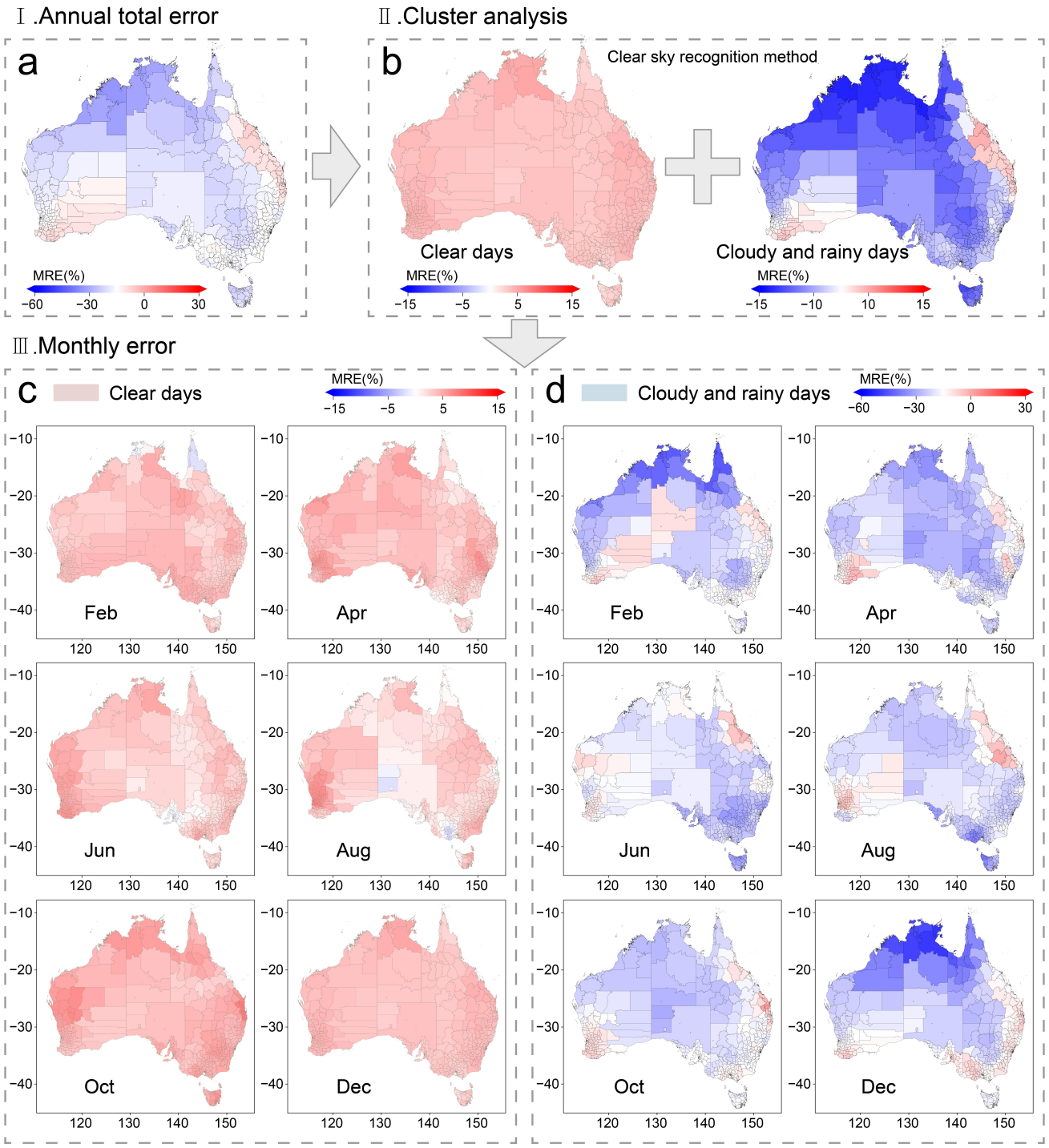


Figure 8. Distribution of average errors across district and county boundaries; (a) Annual solar radiation error; (b) Annual solar radiation error for clear days and cloudy/rainy days following weather type clustering; (c) Monthly solar radiation error on clear days; (d) Monthly solar radiation error on cloudy and rainy days.

3.6 Direct-diffuse decomposition and error analysis

After constructing and precisely adjusting the TMY-MER dataset, the DIRINT model was applied to decompose GHI into DNI and DHI. Figure 9a illustrates the decomposition at site P1814, which is located in the Bega Valley in New South Wales (NSW), and produced a complete 8,760-hour time series of DNI and DHI derived from the TMY-MER solar radiation data. Both DNI and DHI exhibit pronounced seasonal variability, with summer peaks reaching approximately 900 kWh·m^{-2}

and 500 kWh·m^{-2}, respectively, while winter maxima decrease to around 700 kWh·m^{-2} and 400 kWh·m^{-2}. Figure 9b presents the hourly decomposition results for January and July using the DIRINT model, alongside corresponding DNI and DHI profiles generated from the TMY dataset using the BRL model (Lemos et al., 2017; Ridley et al., 2010). Despite differences in weather-type distributions caused by climate variability and extreme events, the GHI, DNI, and DHI curves show substantial agreement across both datasets. The total GHI errors for January and July were –6.0% and 1.7%, respectively; for DNI, –17.9% and –29.3%; and for DHI, 3.6% and 5.4%. While GHI and DHI showed relatively minor deviations, DNI exhibited a consistent negative bias in TMY-MER, especially during winter. Figure 9c compares the annual sky radiance models based on TMY-MER and TMY. The GHI sky models demonstrate strong spatial agreement across azimuthal directions, with TMY-MER slightly overestimating irradiance values, typically within 5%. The DNI distributions are also spatially aligned; however, TMY-MER systematically overestimates values, especially near the zenith, indicating greater summer overestimation. Additionally, the TMY-MER model shows higher DHI values in the northeastern sky sector compared to TMY, with the overestimation limited to under 20%.

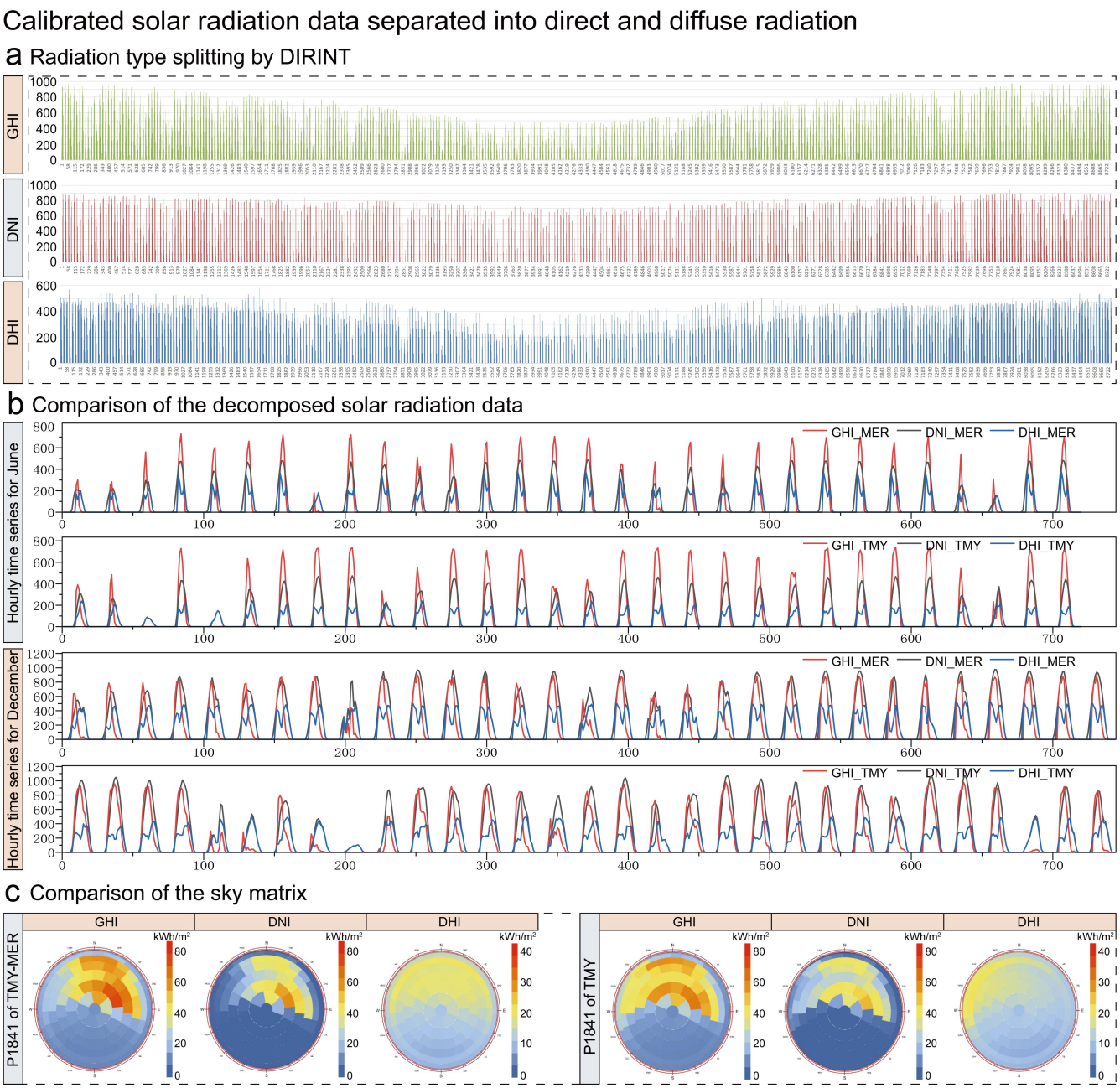


Figure 9. Separation of direct and diffuse solar radiation components and error analysis; (a) Direct and diffuse separation of solar radiation data at P1814 station using TMY-MER; (b) Comparison of data following direct and diffuse separation for TMY-MER and TMY; (c) Sky radiation models following direct and diffuse separation for TMY-MER and TMY.

4 Discussion

4.1 Research contributions

This study has developed a high-resolution typical meteorological year dataset for the entirety of Australia, based on MERRA-2 reanalysis data, to address the spatial and temporal limitations of conventional TMY datasets in capturing localized meteorological variability (Gelaro et al., 2017). Unlike previous studies that merely identified systematic biases in MERRA-2-derived solar radiation (Cao et al., 2022; Du et al., 2022), this work systematically characterizes the spatiotemporal distribution and potential contributing factors of these errors, and implements targeted error assessment through an innovative methodology. By integrating data from 2009 to 2023, TMY-MER overcomes the reliance on outdated ground-based observations used in earlier datasets (Marion & Urban, 1995; Song et al., 2007). With enhanced spatial and temporal resolution, TMY-MER accurately captures the effects of terrain and microclimates on solar irradiance, offering reliable data for dynamic energy planning and urban thermal environment modelling.

A novel weather-type decomposition technique based on clear-sky identification was proposed to classify and correct radiation errors under different sky conditions. The analysis reveals that TMY-MER slightly overestimates solar radiation under clear-sky conditions, with annual mean errors below 5%. In contrast, under overcast conditions, errors become substantially larger, up to 40%, due to inaccuracies in cloud and aerosol modelling. These results support the hypothesis that cloud cover and aerosols are primary contributors to radiation bias, as suggested in previous studies (Cao et al., 2022; Du et al., 2022). Spatial analysis reveals that the most pronounced overestimations occur in densely populated southeastern coastal regions of Australia, with annual cumulative errors as high as 30%, while inland areas generally show smaller deviations, typically below 10%. From a seasonal perspective, the largest overestimations are observed during winter overcast periods, with errors reaching a maximum of around 30%, whereas errors under summer clear-sky conditions remain minimal and typically fall below 5%.

The integration of weather-type decomposition with IDW for parameter generalization enables TMY-MER to constrain total annual errors to within 15%. This error assessment framework ensures comprehensive spatial coverage across all 353 local government areas, addressing the challenge posed by the sparse distribution of conventional meteorological stations in Australia. Furthermore, the proposed methodology is both scalable and transferable, making it suitable for regions with complex climatic conditions and enhancing the global applicability of reanalysis datasets in both scientific research and practical planning contexts.

4.2 Discussion on error sources

This study systematically investigates the mechanisms contributing to errors in MERRA-2 solar radiation data through a spatiotemporal analysis. These biases primarily stem from the interplay between intrinsic limitations of the reanalysis model and the diverse environmental conditions across different regions of Australia. In industrialized eastern cities such as Sydney and Melbourne, elevated aerosol concentrations resulting from intensive anthropogenic activities suppress cloud droplet growth via the Twomey effect (Quaas et al., 2020), thereby prolonging cloud lifetime and enhancing cloud albedo (Jia et al., 2021). The absence of coupled aerosol-cloud dynamic feedback in MERRA-2 leads to systematic underestimation of cloud cover under overcast conditions, which in turn causes significant overestimation of solar radiation, reaching up to 40% in certain cases. This issue is particularly pronounced in urban areas during summer. For example, in Mount Gambier, SA, the lack of cumulus parameterization during the summer season results in radiation overestimations ranging from 20% to 40%, consistent with previous findings highlighting the dominant role of cloud and aerosol interactions (Du et al., 2022). In tropical monsoon regions such as northern QLD, wet-season (December to March) overcast conditions are dominated by deep convective systems. MERRA-2's simplified treatment of vertical cloud structure (Stamatis et al., 2022) likely

contributes to persistent underestimation of solar radiation during this period. In contrast, during the dry season, the model tends to underestimate low-level cloud cover, leading to excessive surface irradiance estimates (Camberlin et al., 2023).

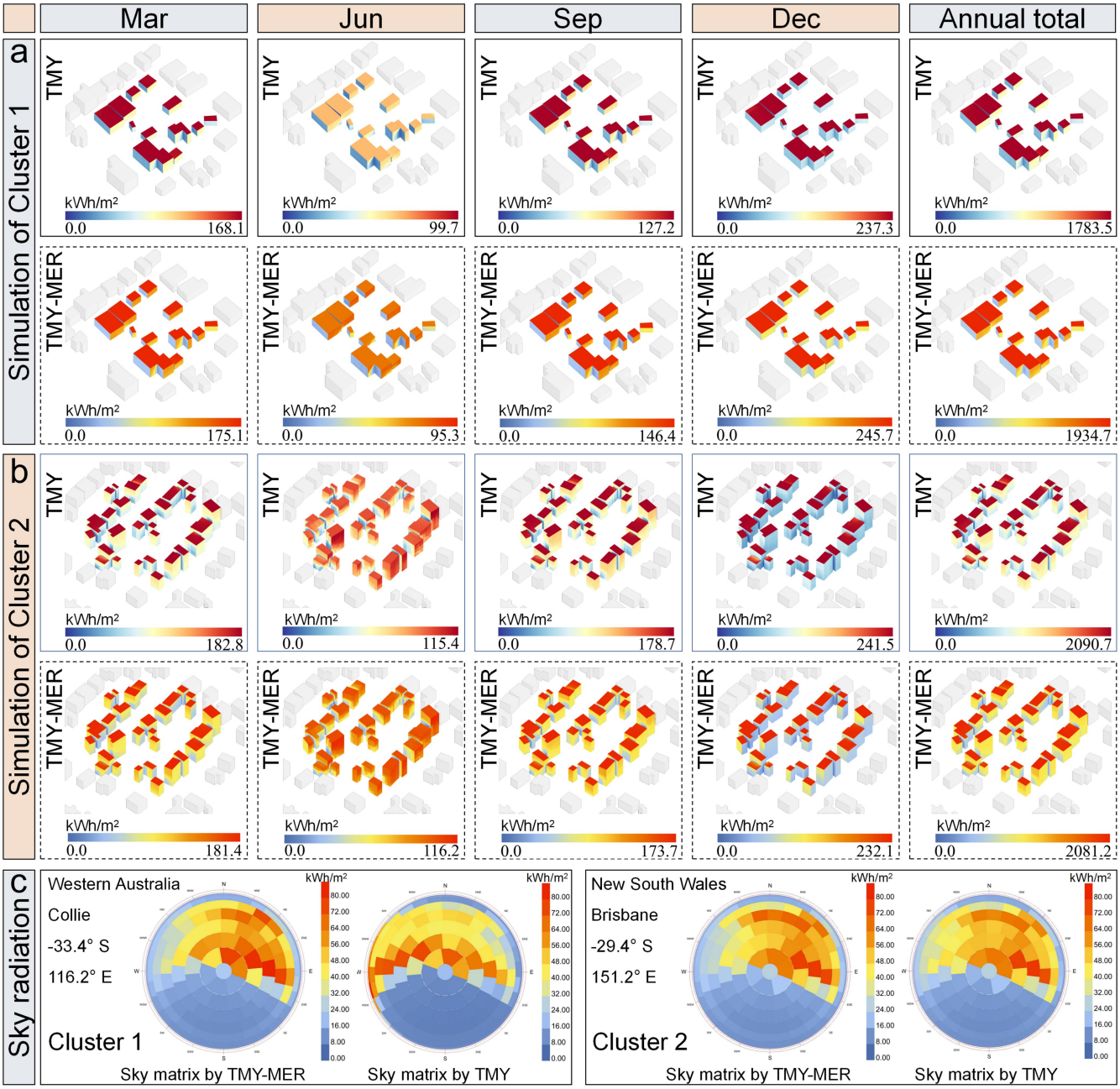


Figure 10. Application of adjusted meteorological data in building cluster simulations: Comparison of solar radiation results from TMY and TMY-MER; (b) Surface solar radiation distribution for Building Cluster 2 based on TMY and TMY-MER; (c) Comparison of sky radiation model distribution characteristics.

Under clear-sky conditions, southeastern coastal regions generally show a modest positive bias of around 5%. Meanwhile, overcast conditions in southwestern and eastern coastal areas are associated with more pronounced overestimations, with deviations reaching up to 20%. This discrepancy is likely due to the increased difficulty in accurately simulating radiative transfer processes under cloudy conditions. Urbanization further exacerbates these inaccuracies through the intensification of the urban heat island effect (Gao et al., 2025), which enhances local thermal gradients and convective activity. The resulting vertical moisture transport can trigger localized cumulus cloud formation when ascending air masses reach the condensation level. In densely populated areas, such clouds are often small in scale and are not adequately represented by reanalysis systems due to simplifications in their physical parameterizations. As a result, these clouds frequently go undetected, introducing additional overestimations in radiative flux. Moreover, in temperate regions such as the southern highlands of VIC, solar radiation is frequently overestimated by 10% to 20%, possibly due to localized cloud

formation driven by diurnal temperature variations (Han et al., 2025), which is similarly not well captured by the MERRA-2 data.

4.3 Data validation and application potential

To evaluate the applicability of the adjusted TMY-MER dataset, this study selected two representative building clusters with complex shading characteristics. Using both the unadjusted TMY and the adjusted TMY-MER datasets, monthly solar irradiance distributions were simulated at a spatial resolution of 0.5 m × 0.5 m. Cluster 1 (Fig. 10a) is located in Collie, WA. Irradiance on building surfaces was modelled using both datasets, and the results show strong agreement in both spatial distribution and intensity. The annual total deviation remained below 10%. For the representative months of each season March, June, September, and December, errors were generally under 10%, except in September, where the deviation increased to 16%. Cluster 2 (Fig. 10b), located in Brisbane, QLD, features a more complex urban geometry. Despite this, the two datasets produced closely matching irradiance patterns. The annual cumulative deviation did not exceed 5%, with winter deviations remaining below 5% and those in other months generally below 3%.

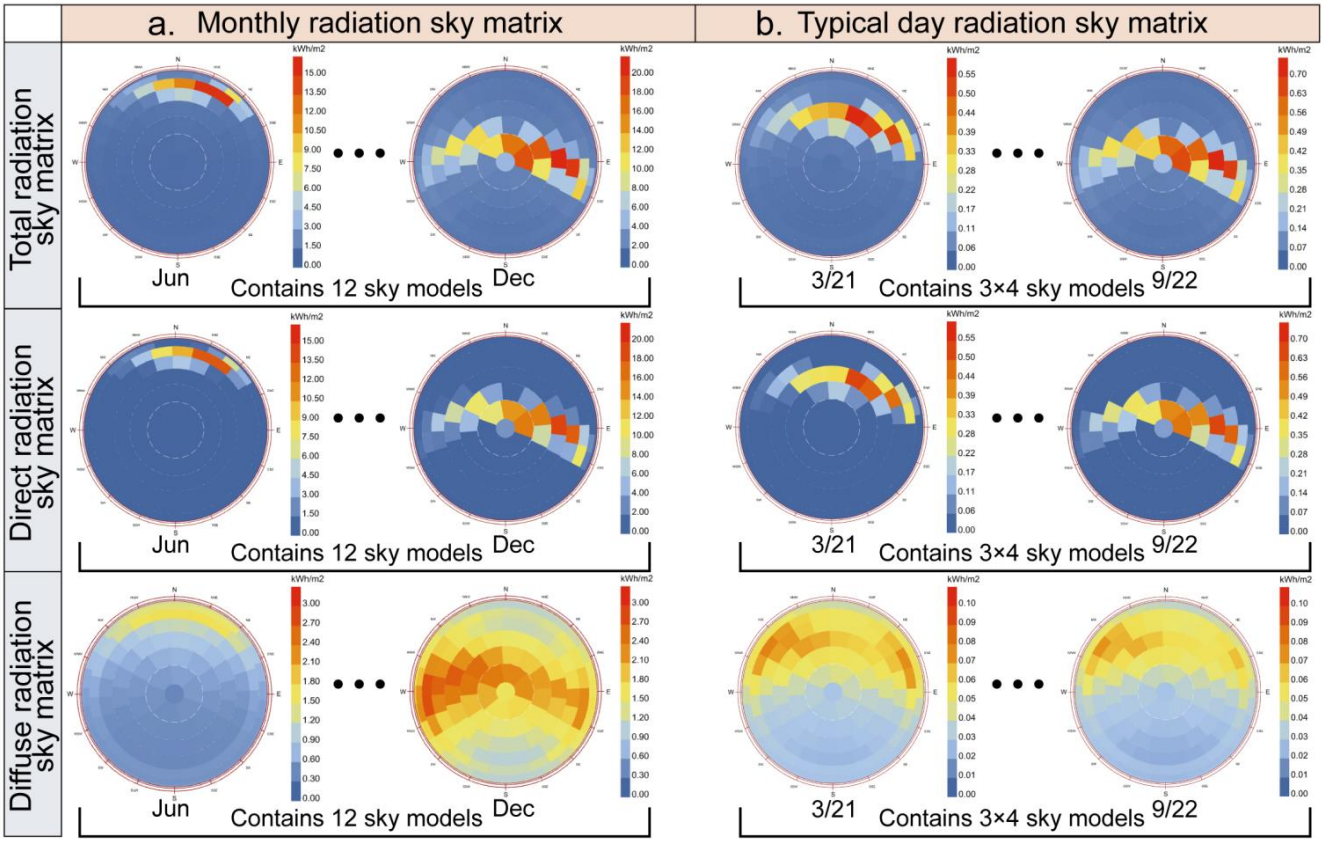


Figure 11. Database construction of multi-condition sky models; (a) Monthly radiation sky matrix; (b) Typical day radiation sky matrix.

Figure 10c illustrates the sky radiation models based on TMY-MER and TMY datasets for Cluster 1 and Cluster 2. Notable differences in sky radiance distribution can be observed between the two datasets. For Cluster 1, discrepancies during winter are primarily reflected in the irradiance magnitude, whereas variations in the transitional and summer seasons are more apparent in spatial distribution. In contrast, the sky radiance models for Cluster 2 show nearly identical spatial layouts across all seasons, with only minor differences in intensity. After error assessment, TMY-MER tends to slightly overestimate surface solar irradiance compared to TMY. Despite this, the TMY-MER-simulated spatial distribution of surface irradiance closely matches that of TMY, with differences remaining within acceptable uncertainty margins. These deviations are likely due to climate variability. The observed consistency in luminance patterns and magnitudes in the sky

models derived from both datasets provides compelling validation of the accuracy and reliability of the adjusted TMY-MER dataset.

TMY-MER can be employed as a high-resolution, globally available TMY dataset. It serves as a viable alternative to conventional TMY data for solar radiation modelling applications. With improved spatial resolution and broader geographic coverage, it provides a more accurate and efficient data foundation for scientific research and engineering applications. This, in turn, promotes technological innovation and advancements in solar energy utilization and related fields, underscoring its significant theoretical and practical value. Figure 11a illustrates the development of a comprehensive set of 36 numerical sky radiance models based on monthly GHI, DNI, and DHI data, addressing the growing demand for assessments across various temporal scales, such as monthly evaluations and characteristic days. Figure 11b shows another set of 36 models constructed to represent conditions around the solar equinoxes and solstices, including the days immediately preceding and following each event. This multi-dimensional modelling framework effectively addresses limitations in previous studies, which often lacked the capacity to simulate complex radiative scenarios on a case-by-case basis. By selecting and combining appropriate sky radiance models based on specific computational requirements, researchers can efficiently evaluate solar energy distributions under diverse conditions within a single simulation run. For instance, this includes monthly analyses of total, direct, and diffuse solar radiation on building surfaces in energy studies, or assessments of total solar exposure during representative winter and summer days in urban thermal environment studies.

5 Data availability

The TMY-MER database can be accessed in the Zenodo dataset from the following DOI: https://doi.org/10.5281/zenodo.15479502 (Fu et al., 2025).

6 Conclusions

This study developed a high-resolution TMY dataset, termed TMY-MER, based on MERRA-2 reanalysis data and covering the entire Australian continent. To address systematic errors inherent in reanalysis datasets and those arising from climate variability, a calibration approach was implemented through clear-sky identification and weather type classification, further refined using spatial interpolation techniques to enhance data accuracy. Validation results demonstrate that the calibrated dataset exhibits high reliability in complex urban environment simulations, offering substantial support for dynamic energy planning and urban climate research.

(1) Under clear-sky conditions, the TMY-MER dataset demonstrated a slight positive bias in solar radiation estimates, with annual mean errors typically below 5%. In contrast, overcast conditions introduced substantially larger deviations, primarily due to inaccuracies in cloud and aerosol representation, with maximum errors reaching 40%. Spatially, the most significant overestimations were observed in densely urbanized southeastern coastal regions, where typical annual biases ranged from 10% to 15%, and localized peaks approached 30%. Inland areas showed smaller errors, mostly between 5% and 10%. From a temporal perspective, the largest deviations occurred during winter overcast periods, while summer clear-sky conditions exhibited consistently high accuracy.

(2) Decomposition of GHI into DNI and DHI using the DIRINT model revealed a systematic underestimation in DNI, with monthly biases of −17.9% in January and −29.3% in July, and a consistent overestimation in DHI ranging from 3.6% to 5.4%. Hourly irradiance patterns and sky dome distributions remained highly consistent across datasets, confirming the strong temporal fidelity of TMY-MER in capturing diurnal solar variation.

(3) Based on the adjusted TMY-MER data, a suite of 36 numerical sky radiance models was constructed to represent monthly and typical-day conditions, incorporating GHI, DNI, and DHI components. This model library enables efficient and

high-resolution solar simulations under diverse environmental scenarios, offering robust analytical support for building energy modelling , urban thermal environment evaluation, and photovoltaic system optimization.

Author contributions.

JF and PN contributed equally to this work. JF and PN conceived the research idea, developed the methodology, collected and processed the MERRA-2 reanalysis data, and drafted the manuscript with revisions. ZY and BH supervised the research, secured funding, and managed project resources. DP validated data through building simulations and provided critical feedback. GQ and FL implemented spatial interpolation and software support. YY and JY performed weather-type classification and data quality checks. All authors participated in writing, reviewing, and editing the manuscript.

Competing interests.

The authors declare that they have no conflict of interest.

References

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, *33*(1), 121–131. https://doi.org/10.1002/joc.3413

Aliana, A., Chang, M., Østergaard, P. A., Victoria, M., & Andersen, A. N. (2022). Performance assessment of using various solar radiation data in modelling large-scale solar thermal systems integrated in district heating networks. *Renewable Energy*, *190*, 699–712. https://doi.org/10.1016/j.renene.2022.03.163

Berardi, U., & Jafarpur, P. (2020). Assessing the impact of climate change on building heating and cooling energy demand in canada. *Renewable and Sustainable Energy Reviews*, *121*, 109681. https://doi.org/10.1016/j.rser.2019.109681

Bre, F., E Silva Machado, R. M., Lawrie, L. K., Crawley, D. B., & Lamberts, R. (2021). Assessment of solar radiation data quality in typical meteorological years and its influence on the building performance simulation. *Energy and Buildings*, *250*, 111251. https://doi.org/10.1016/j.enbuild.2021.111251

Bre, F., & Fachinotti, V. D. (2016). Generation of typical meteorological years for the argentine littoral region. *Energy and Buildings*, *129*, 432–444. https://doi.org/10.1016/j.enbuild.2016.08.006

Bright, J. M. (2019). Solcast: Validation of a satellite-derived solar irradiance dataset. *Solar Energy*, *189*, 435–449. https://doi.org/10.1016/j.solener.2019.07.086

Buchard, V., Randles, C. A., Da Silva, A. M., Darmenov, A., Colarco, P. R., Govindaraju, R., Ferrare, R., Hair, J., Beyersdorf, A. J., Ziemba, L. D., & Yu, H. (2017). The MERRA-2 aerosol reanalysis, 1980 onward. Part II: Evaluation and case studies. *Journal of Climate*, *30*(17), 6851–6872. https://doi.org/10.1175/JCLI-D-16-0613.1

Camberlin, P., Togbedji, C. F., Pergaud, J., Berger, A., Aellig, R., Fink, A. H., Knippertz, P., Moron, V., & Philippon, N. (2023). The representation of dry-season low-level clouds over western equatorial africa in reanalyses and historical CMIP6 simulations. *Climate Dynamics*, *61*(5), 2815–2837. https://doi.org/10.1007/s00382-023-06714-w

Cao, Q., Liu, Y., Lyu, K., Yu, Y., Li, D. H. W., & Yang, L. (2020). Solar radiation zoning and daily global radiation models for regions with only surface meteorological measurements in China. *Energy Conversion and Management*, *225*, 113447. https://doi.org/10.1016/j.enconman.2020.113447

Cao, Q., Liu, Y., Sun, X., & Yang, L. (2022). Country-level evaluation of solar radiation data sets using ground measurements in China. *Energy*, *241*, 122938. https://doi.org/10.1016/j.energy.2021.122938

Chakraborty, T., & Lee, X. (2021). Large differences in diffuse solar radiation among current-generation reanalysis and satellite-derived products. *Journal of Climate*, 1–52. https://doi.org/10.1175/JCLI-D-20-0979.1

Chen, S., Stark, S. C., Nobre, A. D., Cuartas, L. A., De Jesus Amore, D., Restrepo-Coupe, N., Smith, M. N., Chitra-Tarak, R., Ko, H., Nelson, B. W., & Saleska, S. R. (2024). Amazon forest biogeography predicts resilience and vulnerability to drought. *Nature*, *631*(8019), 111–117. https://doi.org/10.1038/s41586-024-07568-w

Cui, Y., Yan, D., Hong, T., Xiao, C., Luo, X., & Zhang, Q. (2017). Comparison of typical year and multiyear building simulations using a 55-year actual weather data set from China. *Applied Energy*, *195*, 890–904. https://doi.org/10.1016/j.apenergy.2017.03.113

Dincer, I., & Aydin, M. I. (2023). New paradigms in sustainable energy systems with hydrogen. *Energy Conversion and Management*, *283*, 116950. https://doi.org/10.1016/j.enconman.2023.116950

Ding, W., Feng, X., Li, N., & Fang, Y. (2025). A method for combing multi-beam bathymetric data to correct the deep sea sub-bottom profile. *Scientific Reports*, *15*(1), 13947. https://doi.org/10.1038/s41598-025-99323-y

Dong, B., Widjaja, R., Wu, W., & Zhou, Z. (2021). Review of onsite temperature and solar forecasting models to enable better building design and operations. *Building Simulation*, *14*(4), 885–907. https://doi.org/10.1007/s12273-020-0759-2

Du, Y., Shi, H., Zhang, J., Xia, X., Yao, Z., Fu, D., Hu, B., & Huang, C. (2022). Evaluation of MERRA-2 hourly surface solar radiation across China. *Solar Energy*, *234*, 103–110. https://doi.org/10.1016/j.solener.2022.01.066

Engerer, N. A. (2015). Minute resolution estimates of the diffuse fraction of global irradiance for southeastern australia. *Solar Energy*, *116*, 215–237. https://doi.org/10.1016/j.solener.2015.04.012

Falchetta, G., Cian, E. D., Pavanello, F., & Wing, I. S. (2024). Inequalities in global residential cooling energy use to 2050. *Nature Communications*, *15*(1), 7874.

Feng, C., Zhang, X., Xu, J., Yang, S., Guan, S., Jia, K., & Yao, Y. (2023). Comprehensive assessment of global atmospheric downward longwave radiation in the state-of-the-art reanalysis using satellite and flux tower observations. *Climate Dynamics*, *60*(5), 1495–1521. https://doi.org/10.1007/s00382-022-06366-2

Festa, R., & Ratto, C. F. (1993). Proposal of a numerical procedure to select reference years. *Solar Energy*, *50*(1), 9–17. https://doi.org/10.1016/0038-092X(93)90003-7

Fu, J., Ni, P., Yan, Z., He, B.-J., Prasad, D., and Qin, G, Lei, F., Yue, Y., & Yan, J.: A high-resolution Typical Meteorological Year dataset for solar radiation evaluation in Australia (v1.0), Zenodo, https://doi.org/10.5281/zenodo.15479502, 2025

Gao, W., Liu, J., Li, S., Xu, K., Wang, M., & Xia, Z. (2025). Effect of urban morphology on local-scale urban heat island intensity under varying urbanisation: A case study of wuhan. *Sustainable Cities and Society*, 106328. https://doi.org/10.1016/j.scs.2025.106328

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Da Silva, A. M., Gu, W., … Zhao, B. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Geng, X., Xie, D., & Gou, Z. (2024). Optimizing urban block morphologies for net-zero energy cities: Exploring photovoltaic potential and urban design prototype. *Building Simulation*, *17*. https://doi.org/10.1007/s12273-024-1104-y

Hall, I. J., Prairie, R. R., Anderson, H. E., & Boes, E. C. (1978). *Generation of a typical meteorological year* (No. SAND-78-1096C; CONF-780639-1). Sandia Labs., Albuquerque, NM (USA). https://www.osti.gov/biblio/7013202

Han, X., Su, J., Deng, X., Peng, N., & Lan, W. (2025). Evaluating diurnal cloud cycle and its radiative effects from CMIP6. *Atmospheric Research*, *318*, 107985. https://doi.org/10.1016/j.atmosres.2025.107985

Hao, D., Asrar, G. R., Zeng, Y., Zhu, Q., Wen, J., Xiao, Q., & Chen, M. (2020). DSCOVR/EPIC-derived global hourly and daily downward shortwave and photosynthetically active radiation data at 0.1° × 0.1° resolution. *Earth System Science Data*, *12*(3), 2209–2221. https://doi.org/10.5194/essd-12-2209-2020

Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., & Zipf, A. (2023). A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. *Nature Communications*, *14*(1), 3985. https://doi.org/10.1038/s41467-023-39698-6

Horan, M. F., Kucharski, F., Johnson, N., & Ashfaq, M. (2024). Winter precipitation predictability in central southwest asia and its representation in seasonal forecast systems. *Npj Climate and Atmospheric Science*, *7*(1), 1–13. https://doi.org/10.1038/s41612-024-00594-5

Hughes, L. (2011). Climate change and australia: Key vulnerable regions. *Regional Environmental Change*, *11*(S1), 189–195. https://doi.org/10.1007/s10113-010-0158-9

Ineichen, P., Perez, R. R., Seal, R. D., Maxwell, E. L., & Zalenka, A. (1992). Dynamic global-to-direct irradiance conversion models. *Ashrae Trans*, *98*(1), 354–369.

Isaza, A., Evans, J. P., Kay, M., Prasad, A., & Brember, S. (2024). Impacts of 2019–20 australian bushfires on solar photovoltaic generation using high-resolution simulations. *Solar Energy*, *284*, 113025. https://doi.org/10.1016/j.solener.2024.113025

Jamei, E., Ossen, D. R., Seyedmahmoudian, M., Sandanayake, M., Stojcevski, A., & Horan, B. (2020). Urban design parameters for heat mitigation in tropics. *Renewable and Sustainable Energy Reviews*, *134*, 110362. https://doi.org/10.1016/j.rser.2020.110362

Jia, H., Ma, X., Yu, F., & Quaas, J. (2021). Significant underestimation of radiative forcing by aerosol–cloud interactions derived from satellite-based methods. *Nature Communications*, *12*(1), 3649. https://doi.org/10.1038/s41467-021-23888-1

Khan, A., College, L., Khorat, S., Khatun, R., Anand, P., & Santamouris, M. (2024). Rooftop photovoltaic solar panels warm up and cool down cities. *Nature Cities*, *1*(11), Article 11. https://doi.org/10.1038/s44284-024-00137-2

Korany, M., Boraiy, M., Eissa, Y., Aoun, Y., Abdel Wahab, M. M., Alfaro, S. C., Blanc, P., El-Metwally, M., Ghedira, H., Hungershoefer, K., & Wald, L. (2016). A database of multi-year (2004–2010) quality-assured surface solar hourly irradiation measurements for the egyptian territory. *Earth System Science Data*, *8*(1), 105–113. https://doi.org/10.5194/essd-8-105-2016

Lave, M., Hayes, W., Pohl, A., & Hansen, C. W. (2015). Evaluation of global horizontal irradiance to plane-of-array irradiance models at locations across the United States. *IEEE Journal of Photovoltaics*, *5*(2), 597–606. https://doi.org/10.1109/JPHOTOV.2015.2392938

Lemos, L. F. L., Starke, A. R., Boland, J., Cardemil, J. M., Machado, R. D., & Colle, S. (2017). Assessment of solar radiation components in brazil using the BRL model. *Renewable Energy*, *108*, 569–580. https://doi.org/10.1016/j.renene.2017.02.077

Li, H., Huo, Y., Fu, Y., Yang, Y., & Yang, L. (2023). Improvement of methods of obtaining urban TMY and application for building energy consumption simulation. *Energy and Buildings*, *295*, 113300. https://doi.org/10.1016/j.enbuild.2023.113300

Li, W., Zhou, Y., Cetin, K., Eom, J., Wang, Y., Chen, G., & Zhang, X. (2017). Modelling urban building energy use: A review of modelling approaches and procedures. *Energy*, *141*, 2445–2457. https://doi.org/10.1016/j.energy.2017.11.071

Liang, H., Jiang, B., Liang, S., Peng, J., Li, S., Han, J., Yin, X., Cheng, J., Jia, K., Liu, Q., Yao, Y., Zhao, X., & Zhang, X. (2022). A global long-term ocean surface daily/0.05° net radiation product from 1983–2020. *Scientific Data*, *9*(1), 337. https://doi.org/10.1038/s41597-022-01419-x

Lin, Z., Guo, W., Ge, J., Yao, X., & Su, D. (2024). Interdecadal change of tibetan plateau vortices during the past 4 decades and its possible mechanism. *Climate Dynamics*, *62*(5), 2971–2989. https://doi.org/10.1007/s00382-023-07045-6

Lund, H. (1995). The design reference year user's manual, thermal insulation laboratory. *Technical University of Denmark: Lyngby*.

Machard, A., Salvati, A., P. Tootkaboni, M., Gaur, A., Zou, J., Wang, L. L., Baba, F., Ge, H., Bre, F., Bozonnet, E., Corrado, V., Luo, X., Levinson, R., Lee, S. H., Hong, T., Salles Olinger, M., Machado, R. M. E. S., Da Guarda, E. L. A., Veiga, R.

K., … Holzer, P. (2024). Typical and extreme weather datasets for studying the resilience of buildings to climate change and heatwaves. *Scientific Data*, *11*(1), Article 1. https://doi.org/10.1038/s41597-024-03319-8

Marion, W., & Urban, K. (1995). *User's manual for TMY2s (typical meteorological years)—Derived from the 1961-1990 national solar radiation data base* (No. NREL/TP-463-7668, 87130, ON: DE95004064; p. NREL/TP-463-7668, 87130, ON: DE95004064). https://doi.org/10.2172/87130

McKinley, S., & Levine, M. (1998). *Cubic spline interpolation*. College of the Redwoods.

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC): GES DISC – Goddard Earth Sciences Data and Information Services Center, [online] Available at: https://disc.gsfc.nasa.gov/, last access: 21 May 2025.

Ni, P., Yan, Z., Yue, Y., Xian, L., Lei, F., & Yan, X. (2023). Simulation of solar radiation on metropolitan building surfaces: A novel and flexible research framework. *Sustainable Cities and Society*, *93*, 104469. https://doi.org/10.1016/j.scs.2023.104469

Olauson, J., Ayob, M. N., Bergkvist, M., Carpman, N., Castellucci, V., Goude, A., Lingfors, D., Waters, R., & Widén, J. (2016). Net load variability in nordic countries with a highly or fully renewable power system. *Nature Energy*, *1*(12), Article 12. https://doi.org/10.1038/nenergy.2016.175

Paulescu, M., Badescu, V., Budea, S., & Dumitrescu, A. (2022). Empirical sunshine-based models vs online estimators for solar resources. *Renewable and Sustainable Energy Reviews*, *168*, 112868. https://doi.org/10.1016/j.rser.2022.112868

Pfeifroth, U., Drücke, J., Kothe, S., Trentmann, J., Schröder, M., & Hollmann, R. (2024). SARAH-3 – satellite-based climate data records of surface solar radiation. *Earth System Science Data*, *16*(11), 5243–5265. https://doi.org/10.5194/essd-16-5243-2024

Pulliainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., Ikonen, J., Takala, M., Cohen, J., Smolander, T., & Norberg, J. (2020). Patterns and trends of northern hemisphere snow mass from 1980 to 2018. *Nature*, *581*(7808), 294–298. https://doi.org/10.1038/s41586-020-2258-0

Pusat, S., Ekmekçi, İ., & Akkoyunlu, M. T. (2015). Generation of typical meteorological year for different climates of turkey. *Renewable Energy*, *75*, 144–151. https://doi.org/10.1016/j.renene.2014.09.039

Qi, Q., Wu, J., Gueymard, C. A., Qin, W., Wang, L., Zhou, Z., Niu, J., & Zhang, M. (2024). Mapping of 10-km daily diffuse solar radiation across China from reanalysis data and a machine-learning method. *Scientific Data*, *11*(1), 756. https://doi.org/10.1038/s41597-024-03609-1

Qin, Y., Zhang, P., Liu, W., Guo, Z., & Xue, S. (2020). The application of elevation corrected MERRA2 reanalysis ground surface temperature in a permafrost model on the qinghai-tibet plateau. *Cold Regions Science and Technology*, *175*, 103067. https://doi.org/10.1016/j.coldregions.2020.103067

Quaas, J., Arola, A., Cairns, B., Christensen, M., Deneke, H., Ekman, A. M. L., Feingold, G., Fridlind, A., Gryspeerdt, E., Hasekamp, O., Li, Z., Lipponen, A., Ma, P.-L., Mülmenstädt, J., Nenes, A., Penner, J. E., Rosenfeld, D., Schrödner, R., Sinclair, K., … Wendisch, M. (2020). Constraining the twomey effect from satellite observations: Issues and perspectives. *Atmospheric Chemistry and Physics*, *20*(23), 15079–15099. https://doi.org/10.5194/acp-20-15079-2020

Randles, C. A., Da Silva, A. M., Buchard, V., Colarco, P. R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., Shinozuka, Y., & Flynn, C. J. (2017). The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. *Journal of Climate*, *30*(17), 6823–6850. https://doi.org/10.1175/JCLI-D-16-0609.1

Ray, B., Lasantha, D., Beeravalli, V., Anwar, A., Nurun Nabi, M., Sheng, H., Rashid, F., & Muyeen, S. M. (2024). A comprehensive framework for effective long-short term solar yield forecasting. *Energy Conversion and Management: X*, *22*, 100535. https://doi.org/10.1016/j.ecmx.2024.100535

Ren, Z., Tang, Z., & James, M. (2021). *Typical meteorological year weather files for building energy modelling*. Australia's National Science Agency.

Ridley, B., Boland, J., & Lauret, P. (2010). Modelling of diffuse solar fraction with multiple predictors. *Renewable Energy*, *35*(2), 478–483. https://doi.org/10.1016/j.renene.2009.07.018

Sarr, A., Soro, Y. M., Tossa, A. K., & Diop, L. (2024). A new approach for modelling photovoltaic panel configuration maximizing crop yield and photovoltaic array outputs in agrivoltaics systems. *Energy Conversion and Management*, *309*, 118436. https://doi.org/10.1016/j.enconman.2024.118436

Song, F., Zhu, Q., Wu, R., Jiang, Y., Xiong, A., Wang, B., Zhu, Y., & Li, Q. (2007). Meteorological data set for building thermal environment analysis of China. *Building Simulation*.

Stamatis, M., Hatzianastassiou, N., Korras-Carraca, M. B., Matsoukas, C., Wild, M., & Vardavas, I. (2022). Interdecadal changes of the MERRA-2 incoming surface solar radiation (SSR) and evaluation against GEBA & BSRN stations. *Applied Sciences*, *12*(19), 10176. https://doi.org/10.3390/app121910176

Sun, J., Li, Z., Xiao, F., & Xiao, J. (2020). Generation of typical meteorological year for integrated climate based daylight modelling and building energy simulation. *Renewable Energy*, *160*, 721–729. https://doi.org/10.1016/j.renene.2020.07.024

Tambula, S., Musademba, D., & Chihobo, C. H. (2023). Impact of spatial and annual solar variability on the economic viability of non-tracking solar heating systems. *Energy Reports*, *10*, 1379–1386. https://doi.org/10.1016/j.egyr.2023.08.016

Tang, W., Yang, K., Qin, J., Li, X., & Niu, X. (2019). A 16-year dataset (2000–2015) of high-resolution (3 h, 10 km) global surface solar radiation. *Earth System Science Data*, *11*(4), 1905–1915. https://doi.org/10.5194/essd-11-1905-2019

Wang, M., Jia, Z., Tao, L., Wang, W., & Xiang, C. (2025). Optimizing the tilt angle of kinetic photovoltaic shading devices considering energy consumption and power generation—Hong kong case. *Energy and Buildings*, *326*, 115072. https://doi.org/10.1016/j.enbuild.2024.115072

Wang, M., Yao, J., Chang, H., Liu, R., Xu, N., Liu, Z., Gong, H., Zheng, H., Wang, J., Guo, X., Cao, Y., Zhao, Y., & Lu, H. (2025). Underground well water level observation grid dataset from 2005 to 2022. *Scientific Data*, *12*(1), 728. https://doi.org/10.1038/s41597-025-04799-y

Wilcox, S., & Marion, W. (2008). Users manual for TMY3 data sets. *Technical Report*.

Wu, R., Ma, Y., Xie, H., Wang, X., Hokoi, S., Yue, Y., Hu, J., Mu, B., & Li, Y. (2023). Influence of solar radiation on the surface loss of the painted statues in grotto no. 3 of maijishan grottoes, China. *Building and Environment*, *240*, 110416. https://doi.org/10.1016/j.buildenv.2023.110416

Wu, Y., An, J., Gui, C., Xiao, C., & Yan, D. (2023). A global typical meteorological year (TMY) database on ERA5 dataset. *Building Simulation*, *16*(6), 1013–1026. https://doi.org/10.1007/s12273-023-1015-3

Xie, X., Song, X., Oltmanns, M., Li, Y., Qian, Q., & Wei, Z. (2024). Observing the air-sea turbulent heat flux on the trajectory of tropical storm danas. *Journal of Oceanology and Limnology*, *42*(5), 1425–1437. https://doi.org/10.1007/s00343-024-3192-z

Yanto, Livneh, B., & Rajagopalan, B. (2017). Development of a gridded meteorological dataset over java island, indonesia 1985–2014. *Scientific Data*, *4*(1), 170072. https://doi.org/10.1038/sdata.2017.72

Yoo, A. (2023). *Australian energy market operator (AEMO)*. https://policycommons.net/artifacts/4341069/australian-energy-market-operator-aemo/5149727/

Yu, Y., Chou, J., Yao, X., & Ma, N. (2024). Generation and application of typical meteorological year data for PV system potential assessment: A case study in China. *Journal of Building Engineering*, *86*, 108831. https://doi.org/10.1016/j.jobe.2024.108831

Yue, Y., Yan, Z., Ni, P., Lei, F., & Qin, G. (2024). Promoting solar energy utilization: Prediction, analysis and evaluation of solar radiation on building surfaces at city scale. *Energy and Buildings*, *319*, 114561. https://doi.org/10.1016/j.enbuild.2024.114561