Developing a Spatially Distributed Typical Meteorological Year Dataset for Solar Radiation Assessment in Australia

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

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

[2]School of Architecture, Tsinghua University, Beijing, 100084, China

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

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

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

[6]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

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

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

&These authors contributed equally: Jingpeng Fu, Pingan Ni

*Corresponding author:

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

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

Abstract: Spatially distributed Typical Meteorological Year (TMY) products are important for solar-radiation evaluation and building-performance simulations, particularly in regions where conventional station-based TMY products have limited spatial coverage and delayed updates. This study develops TMY-MER, a recent-period MERRA2-based typical-year product for GHI-centered solar-radiation evaluation across Australia. Hourly MERRA2 data from 2009 to 2023 were used to construct typical-year weather files for 353 representative city-coordinate or local-government-area locations. A weather-type-conditioned GHI discrepancy assessment framework was introduced to distinguish clear-sky and overcast conditions. Weather-type-conditioned GHI bias parameters were derived by comparing TMY-MER with a reference TMY product at 122 comparison locations and were spatially interpolated using inverse distance weighting (IDW) to generate TMY-MER-IDW. The quantitative assessment focuses on GHI discrepancies rather than full multi-variable validation, while non-GHI variables are retained from MERRA2 as contextual variables for weather-file construction. The results show that TMY-MER exhibits relatively small GHI discrepancies under clear-sky conditions, generally within 5% on an annual basis, whereas overcast conditions show larger deviations, with maximum discrepancies reaching approximately 40% due to uncertainties in cloud and aerosol representation. DIRINT-based decomposition and Ladybug Tools building-cluster simulations were further conducted as application-oriented analyses to examine downstream usability and simulation consistency. The resulting TMY-MER and TMY-MER-IDW products provide a spatially distributed basis for GHI-centered solar-radiation analysis and building-scale solar-radiation applications across Australia.

Keywords: Atmospheric reanalysis datasets; Typical meteorological year; Solar radiation; Weather type segmentation

Nomenclature

Abbreviations

AOD	Aerosol Optical Depth	LBT	Ladybug Tools
CF	Cloud Fraction	LD	Downward longwave radiation
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 MERRA2 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 TMYx	MRE_G	Mean relative error of global annual data
CSD_{MER}^m	Monthly clear sky days in the MERRA2	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 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 serve as essential technological support for achieving carbon neutrality targets (Dincer and Aydin, 2023; 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).

The building sector accounts for nearly 30% of global energy consumption (Publishing, 2013), and accurate, compatible weather data serve 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; Wu et al., 2023a). Therefore, the precise assessment of solar radiation data are fundamental prerequisites 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 can 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 become a critical breakthrough.

1.2 Literature review

TMY datasets and construction methods. 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 Typical Meteorological Months (TMM), TMY datasets effectively reflect long-term climate patterns while minimizing the influence of short-term climatic anomalies (Tambula et al., 2023). These datasets have been widely utilized to support building energy consumption simulations and solar energy resource assessments by providing reliable climatic inputs (Li et al., 2017; Wang 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 and 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 and Urban, 1995; Wilcox and 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.

Development and limitations of conventional TMY methods. The effectiveness of the Sandia method for TMY generation has been validated in solar energy assessments in Turkey and building energy consumption simulations across multiple climatic zones (Li et al., 2023; Pusat et al., 2015). Concurrently, national datasets have been successively released in countries such as Argentina (Bre and Fachinotti, 2016), Australia (Ren et al., n.d.), and Canada (Berardi and Jafarpur, 2020). However, existing

TMY datasets primarily rely on ground-based meteorological observations. Due to the sparse distribution of observation stations (Cao et al., 2020; Hao et al., 2020) and the widespread absence of solar radiation sensors (Bre et al., 2021; Cao et al., 2022), traditional ground-based TMY datasets suffer from limited spatial coverage.

Advancements and applications of reanalysis datasets. With advances in meteorological data assimilation technologies, gridded atmospheric reanalysis datasets have effectively addressed the spatiotemporal limitations of traditional meteorological observations by integrating multisource measurements with numerical model outputs (Abatzoglou, 2013; Tang et al., 2019). Against this backdrop, the MERRA2 reanalysis dataset developed by NASA's Global Modeling 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, MISR satellites, and AERONET ground stations, have substantially enhanced the simulation accuracy of climate system interactions (Randles et al., 2017). MERRA2 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 modeling (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), demonstrating significant multidisciplinary research value.

Limitations and biases in MERRA2 radiation data. As a key data source for long-term radiation studies, MERRA2 has effectively filled the spatiotemporal gaps left by earlier ground-based observations; however, its radiation data exhibit systematic biases. Studies have shown that these biases primarily originate from deficiencies in simulating aerosol–cloud interactions. On one hand, assimilation errors in aerosol optical depth (AOD) (Chakraborty and 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, including global horizontal irradiance (GHI), direct normal irradiance (DNI), and diffuse horizontal irradiance (DHI), 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.

Current validation efforts and unresolved limitations. Validation studies of MERRA2 radiation data have achieved notable regional advances. Chinese researchers identified a systematic overestimation of GHI, attributing it to the underestimation of cloud cover and AOD. However, the generalizability of these findings is limited by analyses focusing on single years (Du et al., 2022) and long-term studies that do not distinguish between different weather types (Cao et al., 2022). In contrast, validation efforts in Australia revealed an overestimation of AOD and an underestimation of radiation (Isaza et al., 2024), although the sparse distribution of observation sites constrains the characterization of spatial variability across diverse climatic zones. Existing research exhibits limitations in both spatial-temporal coverage and analytical dimensions, highlighting the need for integrated analyses across multiple regions, extended time periods, and various weather conditions to comprehensively elucidate the error characteristics of MERRA2 solar radiation data.

1.3 Research objectives and contributions

This study integrates MERRA2 reanalysis data with typical-year weather-file generation methods to develop TMY-MER, a spatially distributed MERRA2-based typical-year product for GHI-centered solar-radiation evaluation in Australia. The study focuses on product generation, product-level GHI discrepancy assessment, and application-oriented simulation consistency, rather than full multi-variable validation. The major contributions of this study are:

(1) Development of a spatially distributed MERRA2-based typical-year product. Based on 15 years of MERRA2 data from 2009 to 2023, TMY-MER provides broad regional coverage for representative locations across Australia. The dataset is not

intended to increase the native spatial resolution of MERRA2, but to support solar-radiation-related applications where conventional station-based TMY products have limited coverage or delayed updates.

(2) Weather-type-conditioned GHI discrepancy assessment and bias-parameter interpolation using inverse distance weighting (IDW). A clear-sky identification method was applied to distinguish GHI discrepancies under clear-sky and overcast conditions. Weather-type-conditioned GHI bias parameters were derived at 122 comparison locations and subsequently interpolated to the 353 representative TMY-MER locations using IDW, thereby generating TMY-MER-IDW. The resulting workflow is interpreted as a product-level GHI discrepancy assessment and an application-oriented adjustment procedure, rather than as independent validation against ground-truth observations.

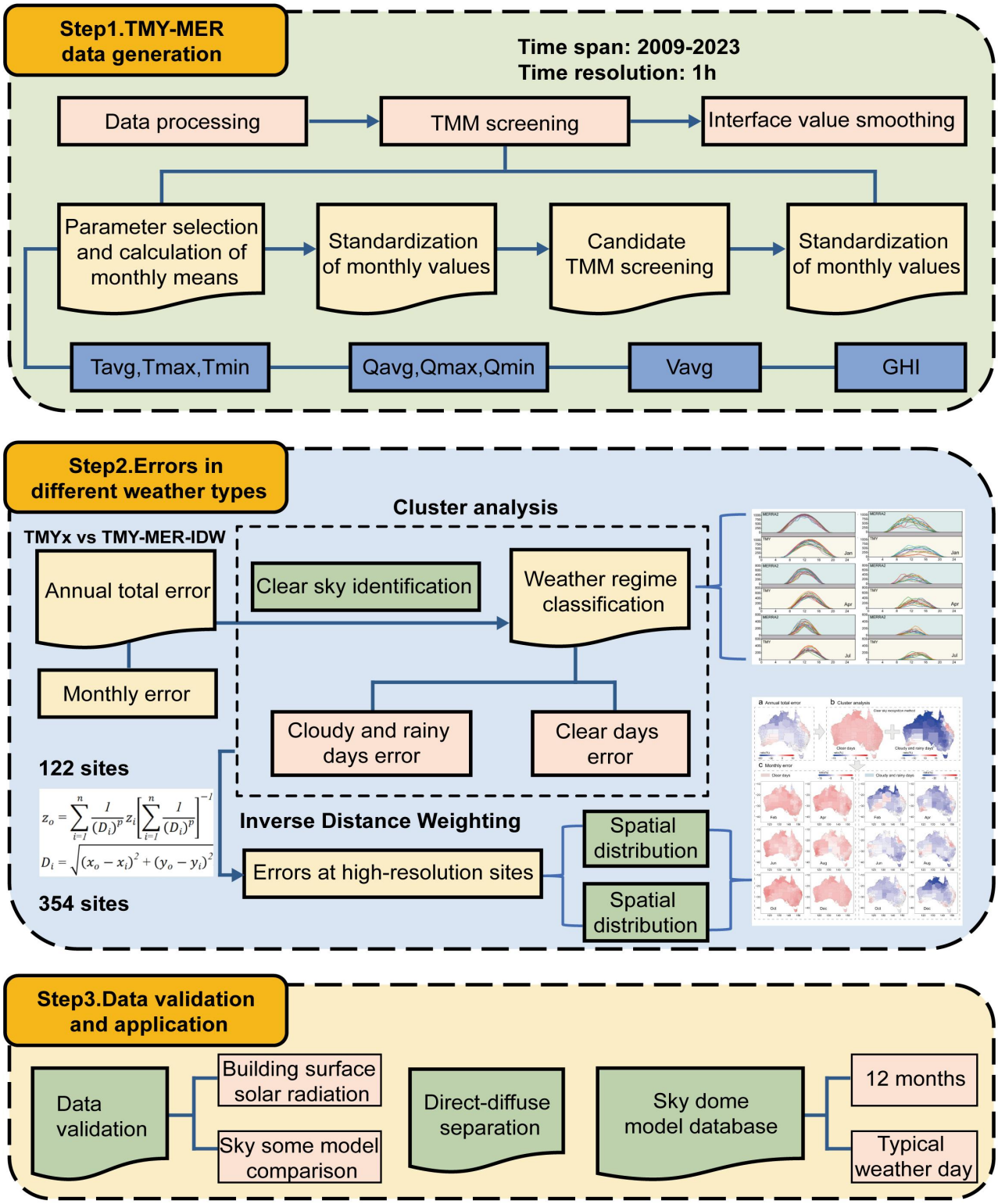


Figure 1: Research framework

2 Methods and materials

Figure 1 illustrates the framework of this study, which aims to construct a spatially distributed MERRA2-based Typical Meteorological Year dataset for GHI-centered solar-radiation evaluation in Australia. The framework consists of three main modules: TMY-MER data generation, weather-type-conditioned GHI discrepancy assessment and IDW-based bias-parameter interpolation, and application-oriented solar-radiation modeling.

Step 1: Based on MERRA2 reanalysis data from 2009 to 2023, hourly meteorological parameters were extracted for 353 representative city-coordinate or local-government-area locations across Australia. Using a TMY generation method that combines parameter normalization and weighted scoring, twelve TMMs were selected from the historical multi-year dataset. Cubic spline interpolation was applied to smooth transitions at monthly boundaries, thereby improving data continuity and generating the initial TMY-MER weather files.

Step 2: To assess systematic GHI discrepancies in MERRA2-derived solar radiation, a weather-type-conditioned assessment method incorporating clear-sky identification was used. Clear-sky and overcast days were distinguished based on the mean relative error index ($\boldsymbol{MRE_I}$) between the daily maximum solar radiation intensity ($\boldsymbol{I_{d,max}}$) and the monthly maximum solar radiation intensity ($\boldsymbol{I_{m,max}}$). Weather-type-conditioned GHI bias parameters were then derived separately for clear-sky and overcast conditions. To extend the derived GHI bias parameters from the 122 comparison locations to the 353 representative locations, IDW interpolation was applied, thereby generating the TMY-MER-IDW dataset.

Step 3: The adjusted TMY-MER-IDW weather files were further examined through application-oriented solar-radiation modeling. The Direct Insolation Separation Model (DIRINT) was applied to decompose GHI into DNI and DHI, and the resulting radiation components were used in Ladybug Tools (LBT) building-cluster simulations. The simulated building-surface solar-radiation distributions and sky-dome luminance patterns were compared with those obtained from the reference TMY product to assess application-level consistency, rather than to provide independent validation of dataset accuracy.

2.1 Generation of TMY-MER

2.1.1 Data sources

MERRA2, developed by NASA's Global Modeling and Assimilation Office, is a global atmospheric reanalysis dataset that extends from 1980 onward, with a native spatial resolution of 0.5° × 0.625° and hourly temporal resolution for the variables used in this study (Gelaro et al., 2017). By incorporating bias-corrected satellite aerosol observations, clear-sky radiation information, and observed precipitation for land surface process simulations, MERRA2 improves the representation of atmospheric composition and land–atmosphere interactions. Its radiation calculation and data assimilation are implemented within the GEOS-5 system, with aerosol-related assimilation and evaluation described in previous MERRA2 studies (Randles et al., 2017; Buchard et al., 2017). In this study, 15 consecutive years of MERRA2 data from 2009 to 2023 were selected to construct a recent-period MERRA2-based typical-year product. This period is shorter than the conventional 30-year climatological period and is therefore interpreted as representing recent climate conditions rather than a long-term climatological normal. The variables extracted from MERRA2 include mean dry-bulb temperature ($\boldsymbol{T_{avg}}$), maximum dry-bulb temperature ($\boldsymbol{T_{max}}$), minimum dry-bulb temperature ($\boldsymbol{T_{min}}$), mean total column water vapor ($\boldsymbol{Q_{avg}}$), maximum total column water vapor ($\boldsymbol{Q_{max}}$), minimum total column water vapor ($\boldsymbol{Q_{min}}$), mean relative humidity ($\boldsymbol{RH_{avg}}$), mean wind speed ($\boldsymbol{V_{avg}}$), and hourly GHI. These variables were used for typical-month selection and weather-file construction, whereas the quantitative discrepancy assessment and subsequent bias-parameter adjustment focused on GHI.

2.1.2 Method for generating the TMY-MER

This study focuses on Australia and uses representative city-coordinate or local-government-area locations as the spatial

units for weather-file generation. Hourly meteorological data for 353 representative locations were extracted from the MERRA2 reanalysis database for the period 2009–2023. These locations should not be interpreted as ground-based weather stations or as native MERRA2 grid cells. A recent approach to TMY construction originally developed for ERA5 reanalysis data was adapted to select twelve representative TMMs from historical multi-year records, forming a complete annual weather sequence for each location (Wu et al., 2023b). The detailed selection procedure is described below.

(1) Parameter selection and monthly value calculation: Nine selection parameters were extracted from the annual meteorological year (AMY) dataset spanning 2009 to 2023, including average, maximum, and minimum dry-bulb temperature; average, maximum, and minimum total column water vapor; average relative humidity; average wind speed; and global horizontal irradiance. Here, AMY refers to the annual meteorological year extracted from the MERRA2 hourly dataset for each candidate year. For each parameter (i), the monthly value ($\boldsymbol{X_{i,m,y}}$) was calculated for each month ($\boldsymbol{m}$) and year ($\boldsymbol{y}$).

(2) Normalization of monthly values: Based on the 15-year AMY dataset, the mean ($\overline{\boldsymbol{X}}_{\boldsymbol{i,m}}$) and standard error ($\boldsymbol{S_{i,m}}$) were computed for each parameter and month. Monthly values were then standardized using Eq. (1) to obtain the normalized values ($\boldsymbol{\eta_{i,m,y}}$). This normalization places variables with different physical units and magnitudes on a comparable dimensionless scale.

(3) Selection of candidate typical meteorological months: A month was considered a candidate typical month if the normalized values ($\boldsymbol{\eta_{i,m,y}}$) for all nine selection parameters were less than 1. Multiple candidate months could be identified under this criterion. This criterion indicates that the candidate month remains within the accepted relative discrepancy range defined in the selection procedure; it does not imply that the original meteorological variables themselves are bounded by 1.

(4) Weighted score calculation and selection of typical meteorological months: Weights were assigned to the nine selection parameters based on their relative importance in solar-radiation-oriented typical-month selection, as shown in Table 1. For each candidate month, a weighted score ($\boldsymbol{D_m}$) 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
Temperature of the dry bulb	$\boldsymbol{T_{avg}}$ (Average temperature of the dry bulb)	2/20
	$\boldsymbol{T_{max}}$(Max. temperature of the dry bulb)	1/20
	$\boldsymbol{T_{min}}$(Min. temperature of the dry bulb)	1/20
Total Quantity of Vapors	$\boldsymbol{Q_{avg}}$ (Average Total Quantity of Vapors)	2/20
	$\boldsymbol{Q_{max}}$(Max. Total Quantity of Vapors)	1/20
	$\boldsymbol{Q_{min}}$(Min. Total Quantity of Vapors)	1/20
Relative humid	$\boldsymbol{RH_{avg}}$(Average of the relative humid)	1/20
Wind speed	$\boldsymbol{V_{avg}}$ (Average wind speed)	1/20
Solar radiation	GHI(Average of the total horizontal radiation)	10/20

Note: The typical-month selection uses nine selection parameters in total, including eight non-GHI meteorological parameters and one GHI parameter. The weights sum to 1.

(5) Smoothing at monthly boundaries: To address discontinuities at the transitions between the 12 selected months, a cubic spline interpolation model (McKinley and 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 files provide spatially distributed coverage for representative locations across Australia. The 353 locations are representative city-coordinate or local-government-area points rather than ground-based weather stations or newly generated high-resolution grid nodes. Locations falling within the same native MERRA2 grid cell should be interpreted as sharing the same underlying MERRA2 grid-cell meteorological representation unless an interpolation or bias-adjustment procedure is explicitly applied.

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

where i denotes the selection parameter index, m denotes the month, and y denotes the candidate year. $X_{i,m,y}$ is the monthly value of the i-th selection parameter in month m of year y. $\overline{X}_{i,m}$ is the multi-year mean value of the i-th parameter for month m calculated from the 15-year AMY series, and $S_{i,m}$ is the corresponding standard error. The normalized value $\eta_{i,m,y}$ represents the relative deviation of a candidate month from the multi-year monthly reference value for a given parameter.

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

where D_m represents the weighted score used for candidate typical-month selection, n is the total number of selection parameters, and K_i is the weight assigned to the i-th selection parameter, as shown in Table 1. In this study, n = 9. The term $\eta_{i,m,y}$ is the normalized value of the i-th parameter for month m in year y, as defined in Eq. (1). The candidate month with the lowest D_m value was selected as the typical meteorological month.

2.2 GHI discrepancy assessment and IDW-based bias-parameter interpolation

Although ERA5 and MERRA2 data are widely used because of their broad spatial coverage, fine temporal granularity, and long time series, existing research has shown that reanalysis-based radiation data may contain systematic discrepancies. This study evaluates GHI discrepancies between TMY-MER and the reference TMY product under different weather types. The comparison is treated as an inter-product GHI discrepancy assessment rather than independent validation against ground-truth observations, because the reference TMY product is itself a statistically constructed dataset. To support weather-type-conditioned adjustment, the relatively stable clear-sky discrepancy is used to derive the baseline GHI bias parameter (MRE_B), which is then used to interpret and adjust the more uncertain overcast-condition discrepancies. The weather-type-conditioned GHI discrepancy assessment and IDW-based bias-parameter interpolation process includes the following five steps:

① Weather-type classification based on clear-sky identification. Widely used TMY products, such as TMYx, CSWD, and IWEC, were considered as reference TMY products, with TMYx used as the reference example in this study. First, the mean relative error index (MRE_I) between the daily maximum solar radiation intensity ($I_{d,max}$) and the monthly maximum solar radiation intensity ($I_{m,max}$) was calculated. Days with MRE_I values below a predefined operational threshold (τ) were classified as clear-sky-like days, while the remaining days were treated as overcast or cloud-affected days. This classification defines relative radiation-pattern classes for product-level GHI discrepancy assessment, rather than independently validated sky-condition labels. The corresponding monthly counts of clear-sky-like days were then recorded for the reference TMY product and TMY-MER.

② Monthly clear-sky GHI bias quantification. The smaller value between CSD_X^m and CSD_{MER}^m for each month was selected as the baseline number of clear-sky days, denoted as CSD_{bas}^m. The relative difference between the monthly average GHI values of the two datasets was evaluated. Radiation differences exceeding the baseline clear-sky GHI bias parameter (MRE_B) were defined as clear-sky residual bias (MRE_O). Based on MRE_O, the selected clear-sky-day GHI data were either reduced or increased to ensure that the monthly maximum clear-sky radiation 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.

③ Annual and monthly overcast-condition GHI bias calculation. Due to the high uncertainty in daily maximum solar radiation intensity ($I_{d,max}$) under overcast conditions, the total monthly solar radiation (Q_m) was adopted as a more robust metric for discrepancy assessment. Local monthly relative bias (MRE_L) and overall annual relative bias (MRE_G) were evaluated beyond the baseline bias parameter (MRE_B). (MRE_L) was used to scale GHI for overcast days to improve consistency in monthly totals, while (MRE_G) was retained to preserve intermonthly differences in solar-radiation totals between TMY-MER and the reference

TMY product.

④ Bias-anomaly detection and adjustment. To avoid the situation where $\boldsymbol{CSD_X^m}$ exceeds $\boldsymbol{CSD_{MER}^m}$ 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 $\boldsymbol{CSD_{bas}^m}$ threshold, anomaly detection was implemented. For each month, the original $\boldsymbol{I_{d,max}}$ were re-evaluated against the $\boldsymbol{I_{m,adj}}$. If any $\boldsymbol{I_{d,max}}$ exceeded $\boldsymbol{I_{m,adj}}$, the adjusted hourly averages from the revised $\boldsymbol{CSD_{bas}^m}$ dataset were substituted. Following this anomaly detection and adjustment process, the $\boldsymbol{MRE_T}$ was reassessed.

⑤ Spatial interpolation of GHI bias parameters. Weather-type-conditioned GHI bias parameters derived from the 122 comparison locations, including the baseline clear-sky-day count $\boldsymbol{CSD_{bas}^m}$ and the GHI bias parameters $\boldsymbol{MRE_B}$, $\boldsymbol{MRE_L}$, $\boldsymbol{MRE_G}$ and $\boldsymbol{MRE_T}$, were spatially interpolated to the 353 representative TMY-MER locations using IDW, as defined in Eqs. (3)–(4). This step generates the TMY-MER-IDW dataset for the 353 representative locations.

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

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

In the IDW model, $\boldsymbol{z_o}$ represents the estimated value, $\boldsymbol{z_i}$ denotes the observed attribute at sample point $\boldsymbol{i}$, $\boldsymbol{D_i}$ is the distance between the target location and the sample point, and $\boldsymbol{p}$ is the power parameter controlling the smoothing effect, typically set to 2 to minimize the mean absolute error.

Because the IDW procedure interpolates parameters derived from the comparison locations, discrepancy reductions at locations involved in deriving these parameters should be interpreted as fitted or interpolated agreement relative to the reference TMY product, rather than independent predictive validation at unseen locations.

2.3 Application-oriented direct–diffuse decomposition and simulation workflow

To address the limitation that the MERRA2-derived weather files provide GHI as the primary solar-radiation variable, this study first applied the DIRINT decomposition model (Erbs et al., 1982) from the pvlib library (https://github.com/pvlib/pvlib-python) to derive DNI and DHI from GHI. By introducing the $\Delta K_t^{'}$ stability index, the DIRINT model adjusts the estimated DNI according to the temporal stability and dynamic characteristics of the GHI time series, as defined in Eqs. (5)–(7). The model also accounts for relevant atmospheric and environmental conditions, such as pressure and temperature, and has been widely used for direct–diffuse radiation decomposition (Engerer, 2015; Lave et al., 2015). Because DIRINT is an empirical decomposition model and may introduce model-specific uncertainty, the derived DNI and DHI components are treated as application-oriented variables rather than independently validated radiation components.

The derived radiation components were then used in the LBT-based simulation workflow. To examine the application potential of the adjusted TMY-MER-IDW weather files, two building clusters characterized by complex shading relationships or distinctive layouts were selected as application cases. Surface solar-radiation simulations were conducted using both the reference TMY product and the corresponding TMY-MER-IDW data within the same LBT/Radiance modeling environment (Ni et al., 2023; Yue et al., 2024). For each case, the same building geometry, simulation settings, and post-processing procedure were applied to both datasets to ensure that the comparison reflected differences in the input weather files rather than differences in model configuration.

The resulting surface irradiance distributions and sky-dome luminance patterns were compared to assess application-level consistency within the same modeling workflow. The comparison focused on whether TMY-MER-IDW could reproduce the main spatial characteristics and seasonal variation patterns of building-surface solar exposure relative to the reference TMY product. However, because both datasets were processed through the same simulation engine and may share common modeling assumptions, this comparison is interpreted as a simulation-consistency assessment and application case study, rather than

independent validation of dataset accuracy.

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

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

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

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

2.4 Research area

This study selects Australia as a typical research region, primarily based on its unique resource conditions and research design requirements. First, the photovoltaic development potential in low-density urban areas. The low-density building layout significantly reduces the shading effect, thereby improving the effective operating hours and unit area efficiency of photovoltaic systems. Research shows that the photovoltaic potential in low-density areas can reach 1 to 6.6 times the total energy consumption, significantly higher than in medium-to-high-density areas (Geng et al., 2024). Second, the need to verify climate and geographic diversity. Australia spans from tropical to temperate climates and features a variety of terrains, including coastal plains, mountains, and deserts. Its climate and geographic characteristics (such as deep 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.

Third, the urgency of energy transition and data support. 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 spatial context, geographic illustrations are employed in the following sections. All maps of Australia used in this study are sourced from OpenStreetMap.

3 Results

3.1 Data sample documentation

A total of 353 TMY-MER weather files were generated for representative locations across Australia. These locations were defined using administrative-center or city-coordinate information associated with Australian local-government areas. For each representative location, the geographic coordinate was mapped to the nearest native MERRA2 grid cell, and the corresponding MERRA2 hourly meteorological variables were extracted from that grid cell. Therefore, the 353 locations should be interpreted as grid-referenced representative sampling points derived from the MERRA2 reanalysis field, rather than ground-based weather stations, independent observational sites, or newly generated high-resolution grid nodes. Because the native spatial resolution of MERRA2 is 0.5° × 0.625°, multiple representative locations may be associated with the same MERRA2 grid cell. In such cases, these locations share the same underlying MERRA2 meteorological representation before any subsequent bias-parameter interpolation or product-level adjustment is applied. Each generated TMY-MER weather file contains 8,760 hourly records, including 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 selected typical meteorological month. Sample records from an example TMY-MER file are shown in Table 2. The weather files were also converted to EnergyPlus Weather (EPW)

format for downstream building-performance and solar-radiation simulations. In this study, the quantitative discrepancy assessment focuses on GHI, while the other meteorological variables are retained from MERRA2 as contextual variables to preserve the weather-file structure and support application-oriented simulation workflows.

Table 2 Sample TMY-MER data records for Bega Valley, Australia

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

3.2 Evaluation of relative errors in total solar radiation

From the perspective of annual total GHI, although the MERRA2-derived TMY-MER dataset exhibits clear geographical heterogeneity in its product-level differences, the overall magnitude remains within 15%. Monthly analysis further reveals distinct seasonal variations, as presented in Fig. 2. During the Australian summer from December to February, the GHI differences show strong spatial heterogeneity. TMY-MER generally exhibits positive biases along the eastern, southern, and southwestern coastal regions, whereas the northern and northwestern coastal regions show negative biases, with the maximum negative deviation reaching approximately 30%. In inland Australia, the magnitude of difference is relatively small and predominantly positive, with most locations showing positive deviations below 10%. During the transitional seasons of spring and autumn, negative deviations remain evident in March and November in the northern and northwestern coastal regions, while other months show varying degrees of positive bias.

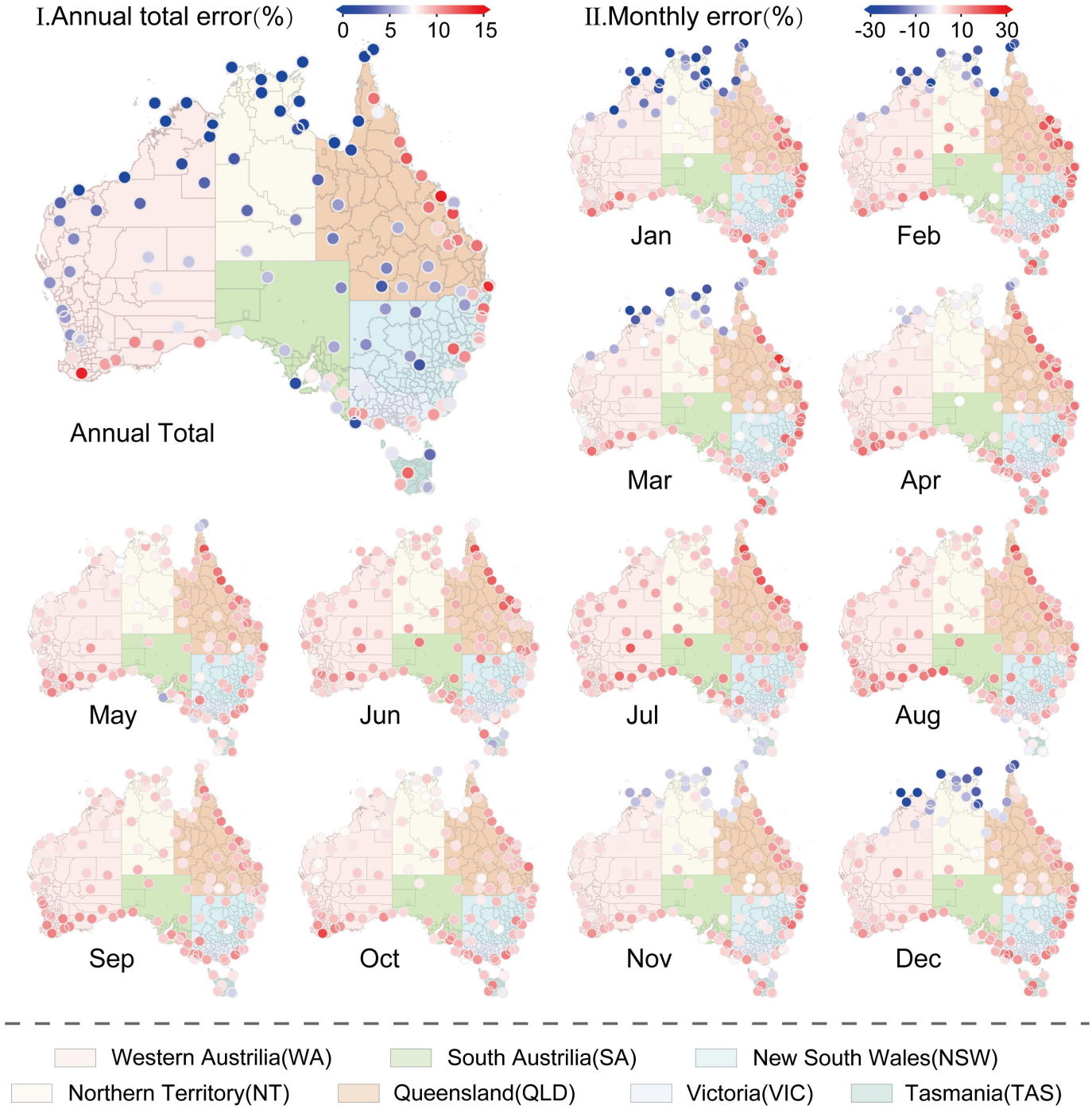


Figure 2: Error distribution of TMY-MER solar radiation data

The spatial distribution pattern indicates that the differences are generally larger in the eastern and southern coastal regions, with autumn showing higher magnitudes than spring. Inland regions exhibit smaller deviations, generally not exceeding 10%. During winter from June to August, TMY-MER shows positive biases across most comparison regions, with the largest deviation observed in the northwestern coastal region, reaching approximately 30%, followed by the southern coastal regions. The central region also shows higher magnitudes than in other seasons, exceeding 20% in some areas. It is noteworthy that the positive bias in the southwestern coastal region during winter is lower than that observed in the same region during summer. Overall, the comparison across spatial and temporal scales suggests that more urbanized areas, such as the southeastern coastal region of Australia and the southwestern part of Western Australia, tend to exhibit more pronounced positive GHI differences relative to the reference TMY product.

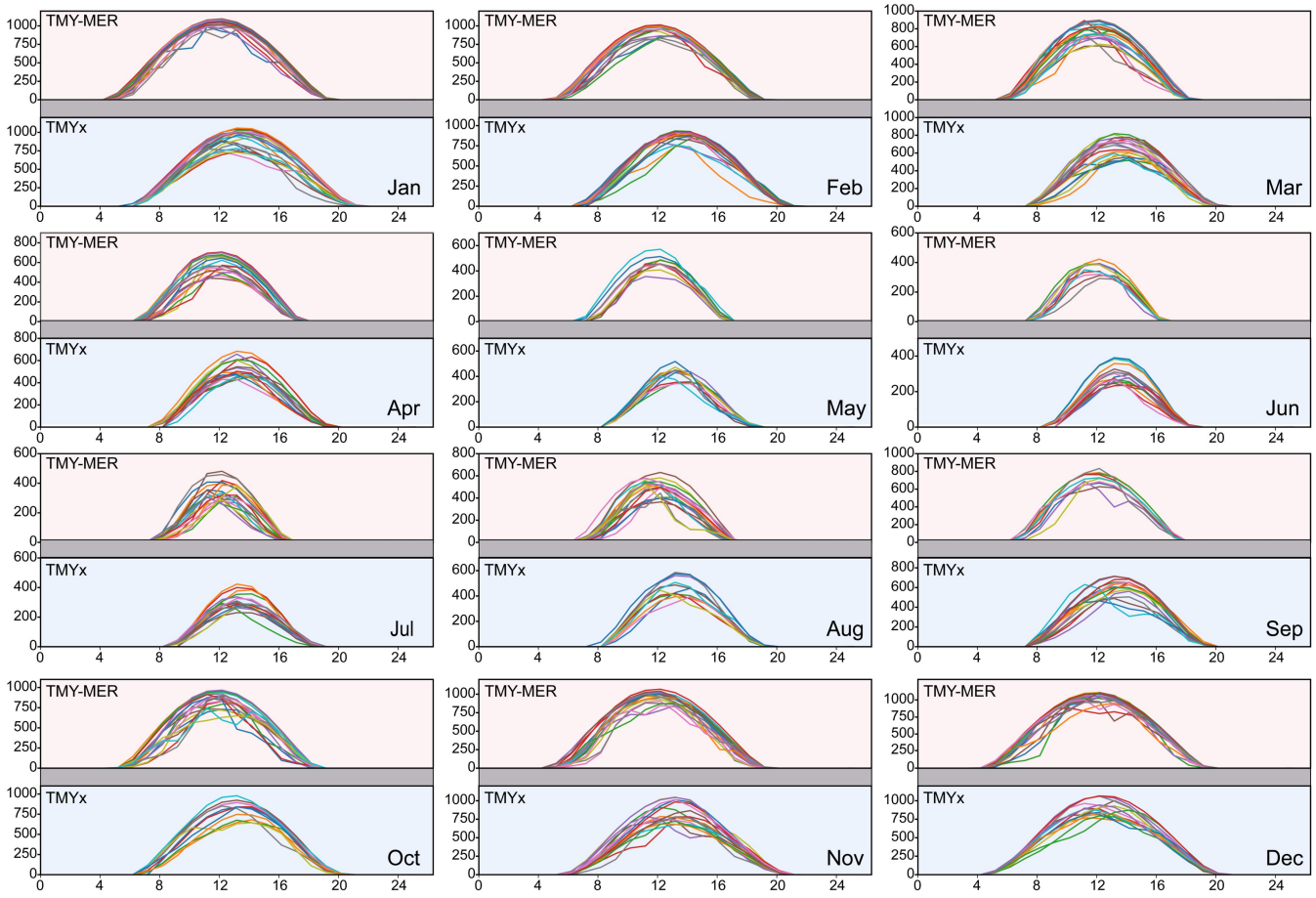


Figure 3: Illustrative hourly GHI profiles of clear-sky-like days at representative comparison location P38 (Unit: W/m²). Each curve represents the hourly GHI distribution of one classified clear-sky-like day.

3.3 Weather type identification

A clear-sky detection algorithm was applied to classify solar radiation parameters from the TMYx and TMY-MER meteorological datasets into clear and overcast weather types. To ensure 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 TMYx 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 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 W/m². In contrast, during winter, for example, in June and July, daily peak values typically fall below 600 W/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 TMYx and TMY-MER under overcast conditions exceeds 500 W/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.

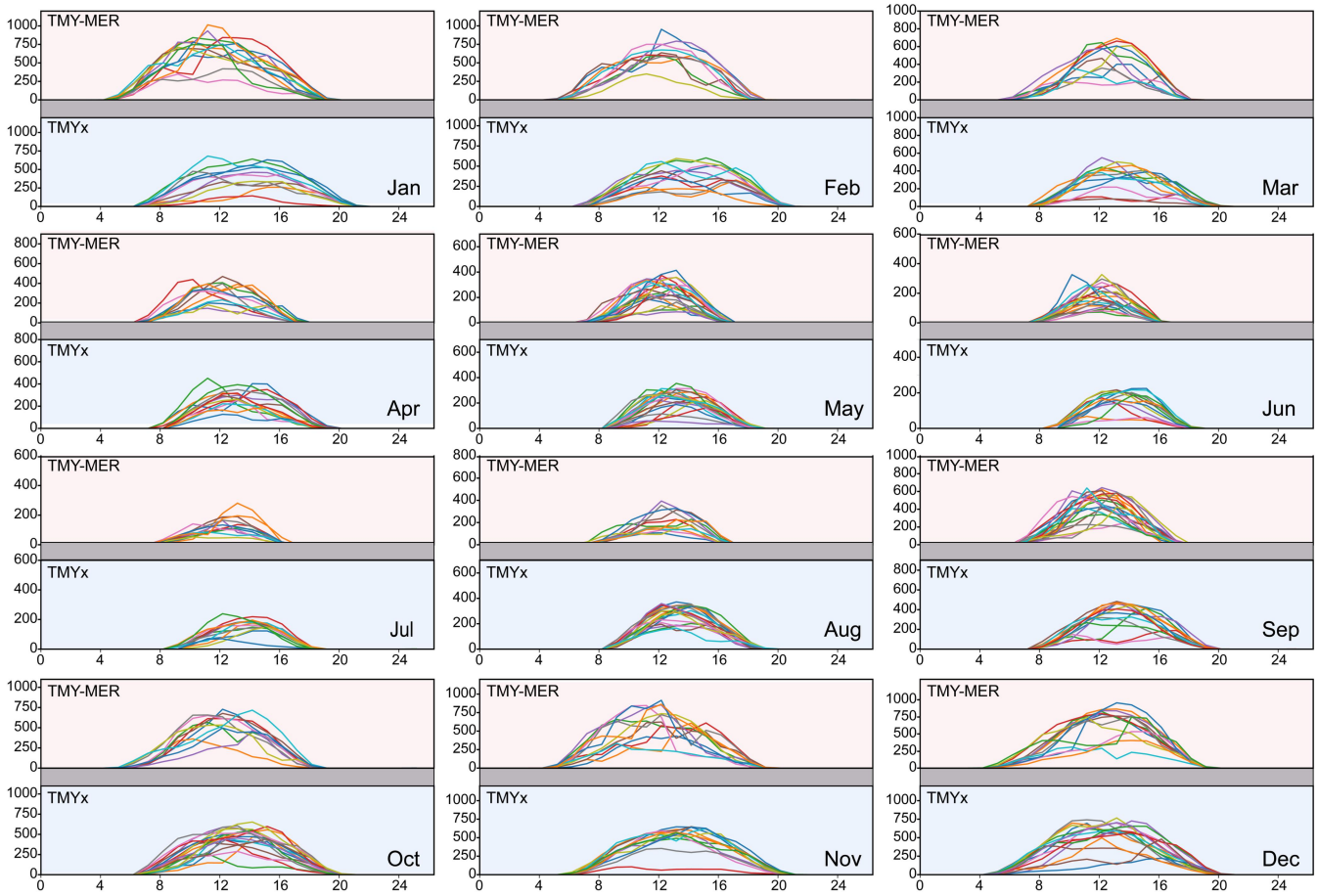


Figure 4: Illustrative hourly GHI profiles of overcast or cloud-affected days at representative comparison location P38 (Unit: W/m²). Each curve represents the hourly GHI distribution of one classified overcast or cloud-affected day.

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 TMYx. 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, 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 122 comparison locations, 24 representative comparison locations were selected to analyze 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 systematic 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.

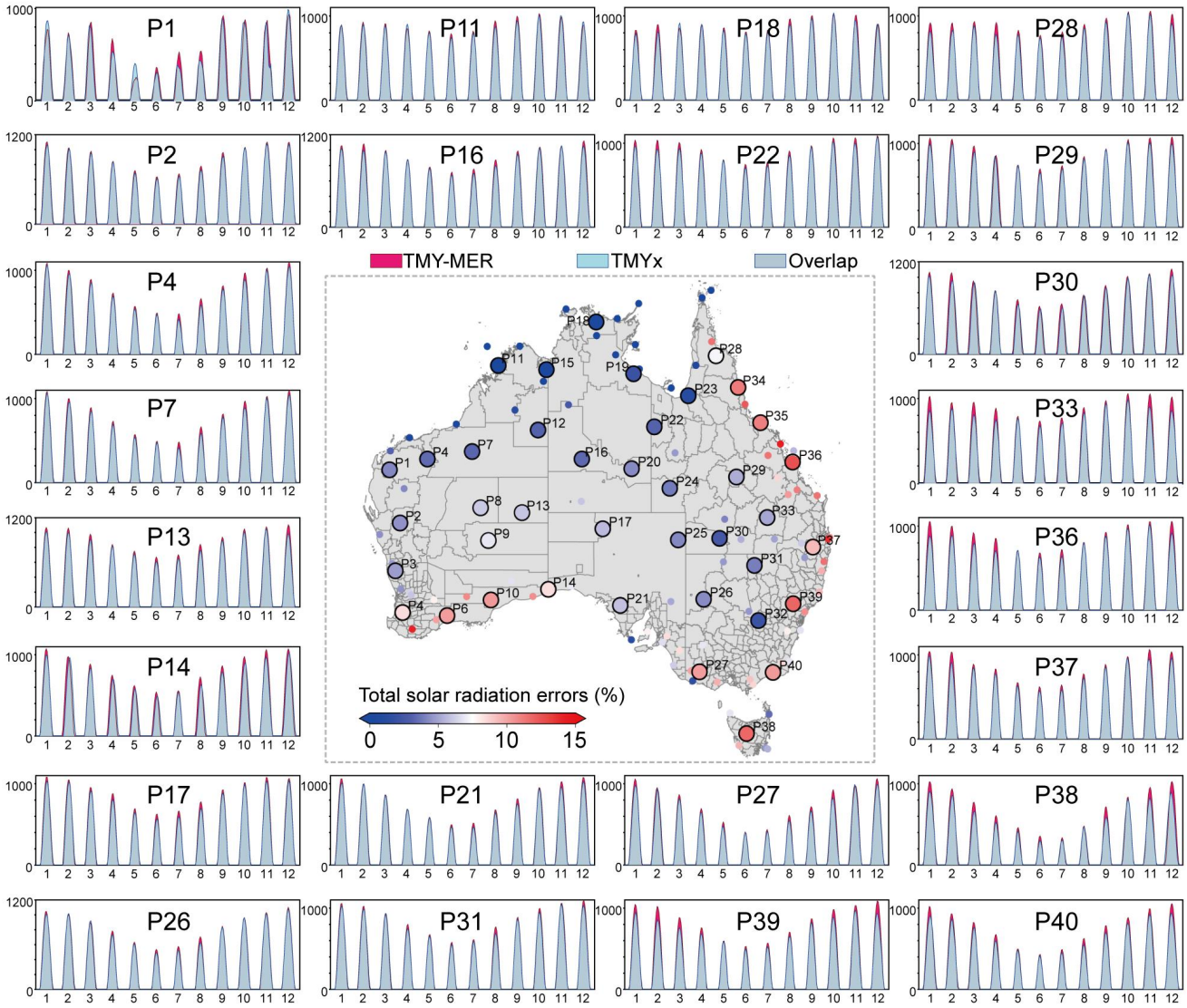


Figure 5: Monthly solar radiation errors across different regions under clear-sky conditions at various sites (Unit: W/m²)

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 TMYx datasets. Substantial overestimations exceeding 40% are observed during specific months at several comparison locations, 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.

Spatial analysis reveals pronounced regional differences in the distribution of solar radiation estimation errors. In the western coastal regions (e.g., P27, P35, P36, P37, P38, P39), TMY-MER consistently overestimates solar radiation, particularly during the wet season from November to March. These overestimations are likely linked to seasonal cloud variability, high urban density, and elevated aerosol emissions in these areas. In the central inland regions (e.g., P13, P16, P17, P20), 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 MERRA2. However, during cloudy winter months, notable overestimations are observed, and seasonal variations in bias are evident. For example, P13 shows substantial overestimation in December and March under cloudy conditions, while P17 exhibits slight underestimation in spring and summer. In the northern region (e.g., P11, P18, P28), TMY-MER tends to overestimate solar radiation during summer,

whereas underestimation is observed in other months at sites including P11, P18, P19, P23, and P28, for instance, in autumn at P11; in both summer and autumn at P18 and P23; in spring and summer at P19; and from October to December at P28. In the southern region (e.g., P4, P6, P10, P14, P21, P27, P40), moderate overestimations occur during overcast winter and adjacent months, although the overall magnitude of bias is relatively limited. These spatial patterns highlight the heterogeneity of TMY-MER performance across climatic zones and emphasize the influence of local atmospheric conditions on radiation modeling accuracy.

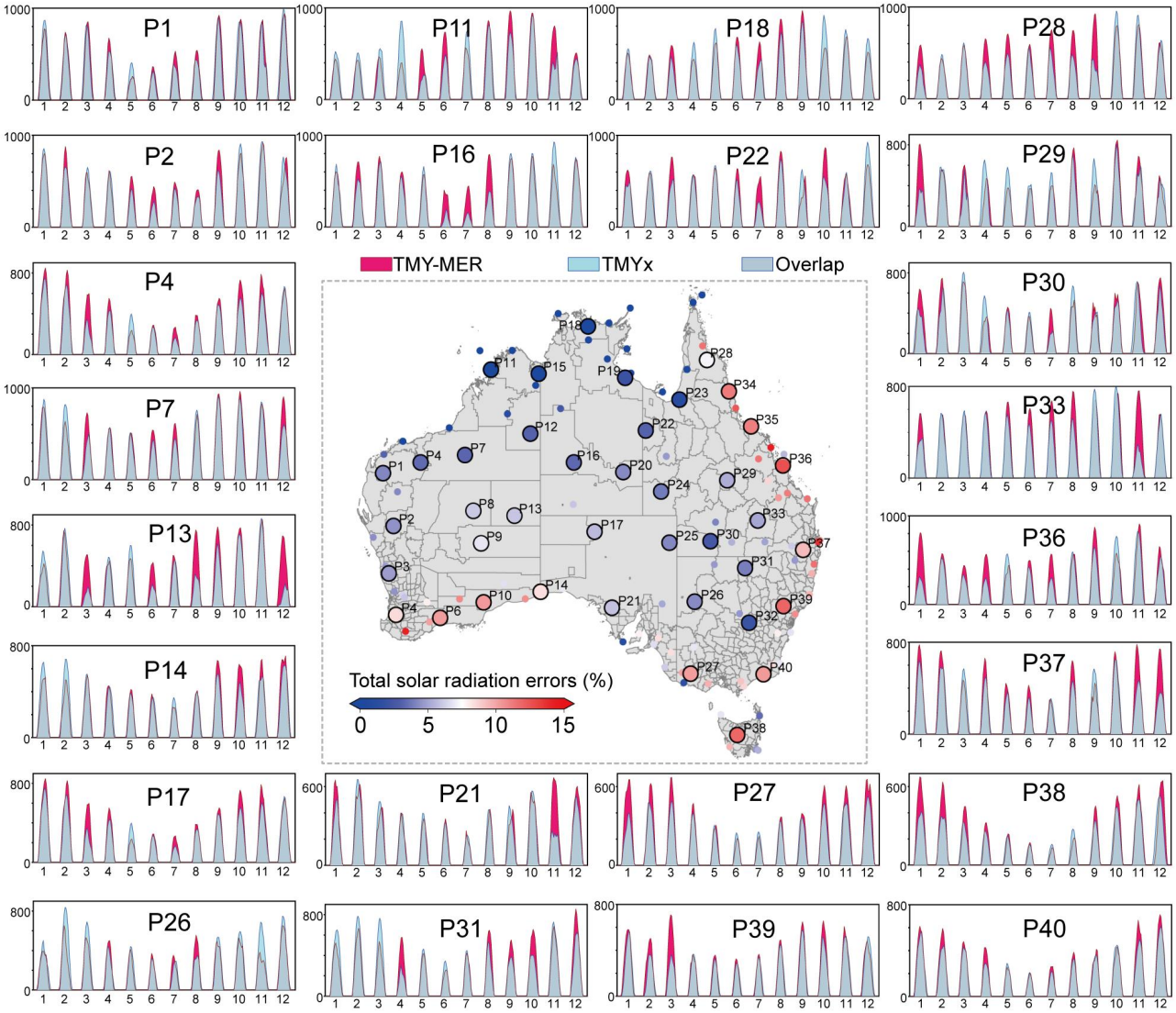


Figure 6: Monthly analysis of solar radiation errors across different regions under overcast conditions at various sites (Unit: W/m²)

3.4.2 Spatiotemporal patterns of IDW-interpolated GHI bias estimates

By comparing GHI data from the 122 locations where the reference TMY product is available with their corresponding TMY-MER counterparts across Australia, weather-type-conditioned GHI bias parameters were derived under clear-sky and overcast conditions. These bias parameters were then spatially interpolated from the 122 comparison locations to the 353 representative TMY-MER locations using IDW. This procedure produced spatially distributed IDW-interpolated GHI bias estimates for all representative locations under both weather scenarios, as illustrated in Fig. 7a and c. Because the IDW step is based on bias parameters derived from comparison with the reference TMY product, the adjusted TMY-MER-IDW results should be interpreted as fitted or interpolated product-level agreement achieved through the bias-adjustment and interpolation procedure,

rather than as independent validation at unseen locations. Under clear-sky conditions, the annual relative GHI bias estimates derived from the TMY-MER–reference TMY comparison ranged from 0% to 6%, with TMY-MER generally showing a positive discrepancy. Smaller discrepancies were mainly observed in the northwestern part of Cape York Peninsula, parts of southwestern Australia, and the eastern coastal belt between approximately 30°S and 40°S, whereas larger deviations appeared in central Cape York, southern South Australia, southern Victoria, Tasmania, and parts of the arid interior. Under overcast conditions, the discrepancy range expanded to approximately −40% to 40%, indicating greater uncertainty associated with cloud representation. Negative discrepancies were more evident in tropical northern Australia, Tasmania, and several inland areas, while positive discrepancies were mainly distributed along the southwestern and eastern coastal zones and parts of southeastern Victoria and South Australia.

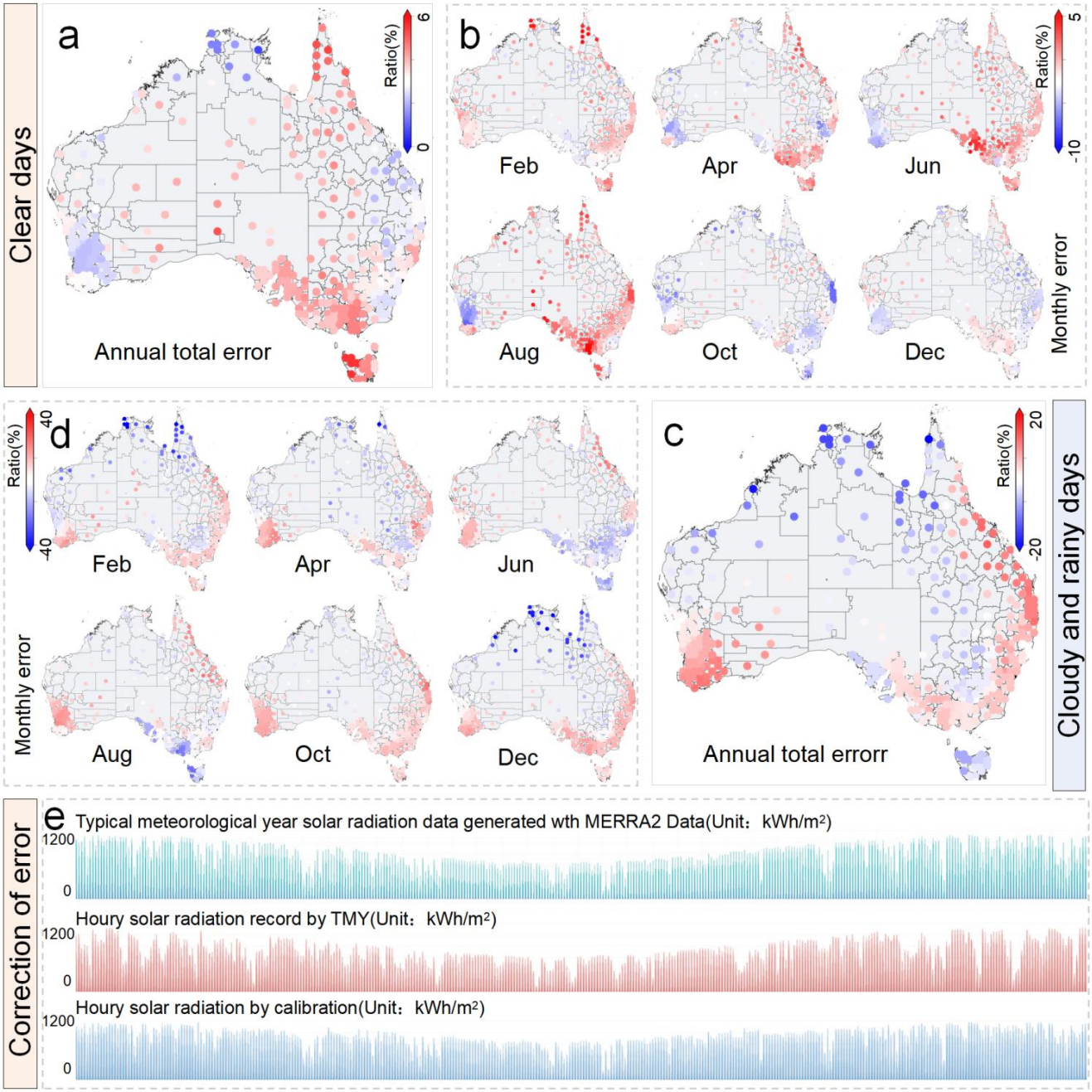


Figure 7. Spatiotemporal distribution of IDW-interpolated relative GHI bias estimates used for TMY-MER-IDW adjustment: (a) annual clear-sky relative GHI bias; (b) monthly clear-sky relative GHI bias; (c) annual overcast relative GHI bias; (d) monthly overcast relative GHI bias; (e) illustrative product-level adjustment example at comparison location P37.

Temporal analysis, as shown in Fig. 7b and d, further reveals clear seasonal variation in the spatial distribution of the IDW-interpolated GHI bias estimates. Under clear-sky conditions from April to August, TMY-MER showed mild positive deviations, generally within 5%, across southeastern and northern coastal regions, while southwestern coastal areas showed negative biases reaching up to 10%. From October to December, the discrepancies were generally smaller and more frequently negative, with larger deviations concentrated in coastal areas and smaller deviations occurring inland. In February, discrepancies approached zero across most regions, except for localized positive deviations over Cape York and the northern part of the Northern Territory. Under overcast conditions, seasonal variability was more pronounced: in summer, areas north of the Tropic of Capricorn experienced substantial negative discrepancies of approximately −30% to −40%, whereas southern regions generally showed positive discrepancies; in winter, southwestern coastal Australia and Tasmania showed negative discrepancies of approximately −20% to −30%, while many other regions showed positive deviations; and in the transitional seasons, widespread positive discrepancies of approximately 20% to 40% were observed across many coastal regions except northern Australia, whereas inland areas generally showed smaller or slightly negative deviations. The comparison at location P37 (Fig. 7e) serves as an illustrative example of the adjustment behavior of the weather-type-conditioned correction and IDW-based bias-parameter interpolation, rather than as independent validation.

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 TMYx. Figure 8b further evaluates the total radiation errors under clear-sky and overcast conditions, based on a sky condition classification approach. Figure 8c and d 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 APY Lands in South Australia, and the Petermann-Simpson region in the Northern Territory, the positive bias during winter is markedly reduced compared to other seasons. Conversely, in certain eastern coastal regions of Western Australia, including Irwin, the model exhibits 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 shows a slight underestimation, as do the SA Western subregion of the Victoria Desert in South Australia and the Corangamite, Ararat Surrounds, and St Arnaud areas in Victoria 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 Western Australia, and the Victoria River region in the Northern Territory. In contrast, central regions display relatively minor deviations. In winter and transitional months (April to October), moderate positive biases occur in eastern coastal Queensland, eastern Tasmania, and southwestern Western Australia, while most other regions are dominated by significant negative biases. Specifically, underestimations in northern coastal and adjacent inland regions during transitional months range from 30% to 40%, exceeding those observed in winter. In southeastern coastal areas, winter underestimation reaches its peak, with values up to 45%, surpassing those in the transitional season as well.

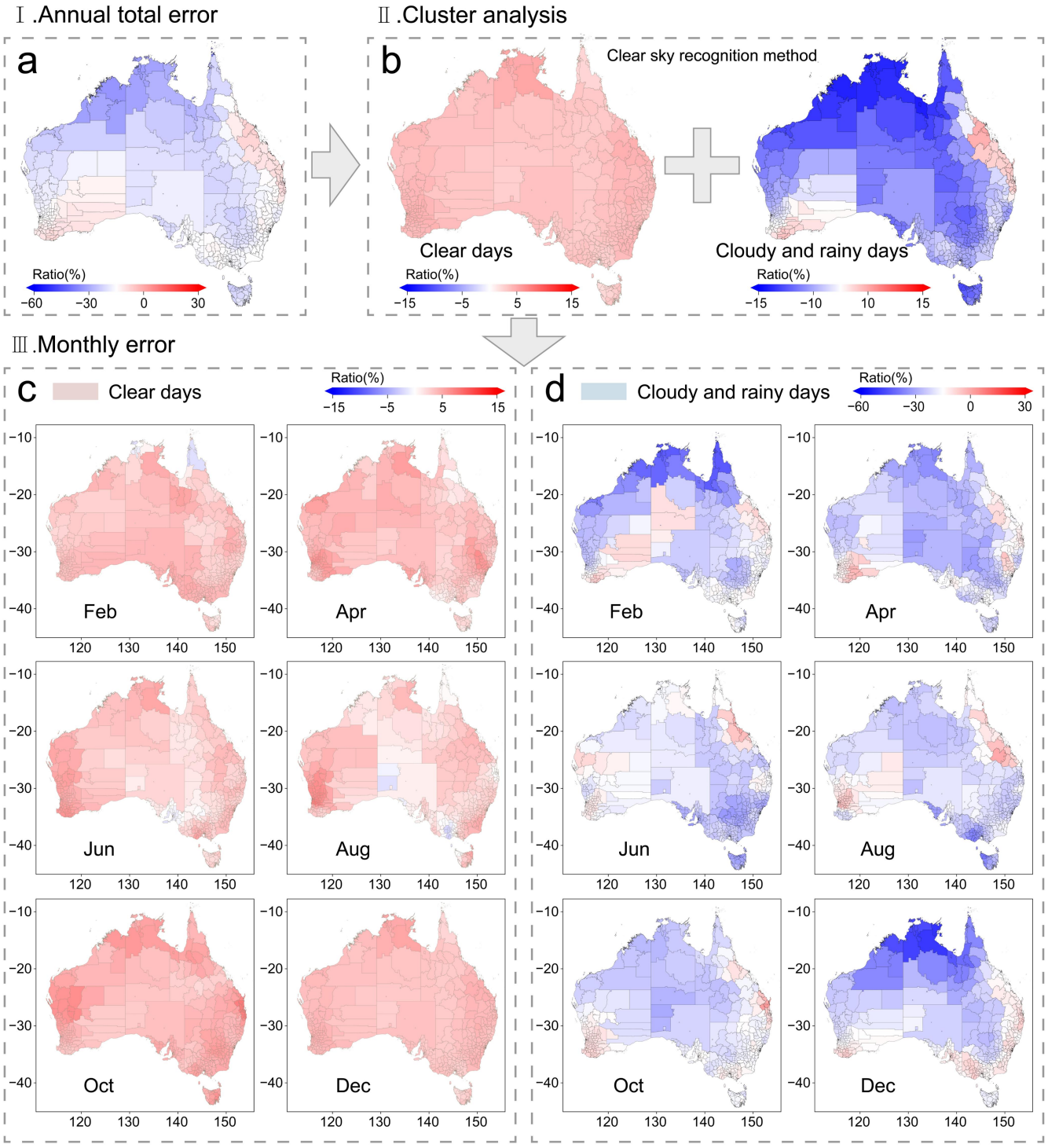


Figure 8: Distribution of average errors across district and county boundaries; (a) Annual solar radiation error; (b) Annual solar radiation error for clear days and cloudy/rainy days following weather type clustering; (c) Monthly solar radiation error on clear days; (d) Monthly solar radiation error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

3.6 Application-oriented direct–diffuse decomposition and sky-radiance modeling

After constructing the TMY-MER and TMY-MER-IDW datasets for GHI-centered solar-radiation evaluation, the DIRINT model was applied to decompose GHI into DNI and DHI for downstream application-oriented modeling. Because DIRINT is an empirical decomposition model and may introduce additional uncertainty, the derived DNI and DHI components are treated as application-oriented radiation variables rather than independently validated radiation components.

Figure 9 illustrates the DIRINT-derived DNI/DHI time series, component-level comparison, and sky-radiance modeling outputs for representative cases. In the component-level comparison, DNI and DHI from TMY-MER-IDW were derived from GHI using the DIRINT model, whereas the corresponding DNI and DHI components from the reference TMY product were derived using the BRL model (Lemos et al., 2017; Ridley et al., 2010). The comparison shows that GHI differences between TMY-MER-IDW and the reference TMY product are relatively small in the selected months, with deviations of approximately −6.0% in January and 1.7% in July. Larger deviations are observed for the DIRINT-derived DNI component, approximately −17.9% and −29.3% in the same two months, while DHI shows smaller positive deviations of approximately 3.6% and 5.4%. These component-level differences reflect the additional uncertainty introduced by empirical direct–diffuse decomposition, especially for DNI.

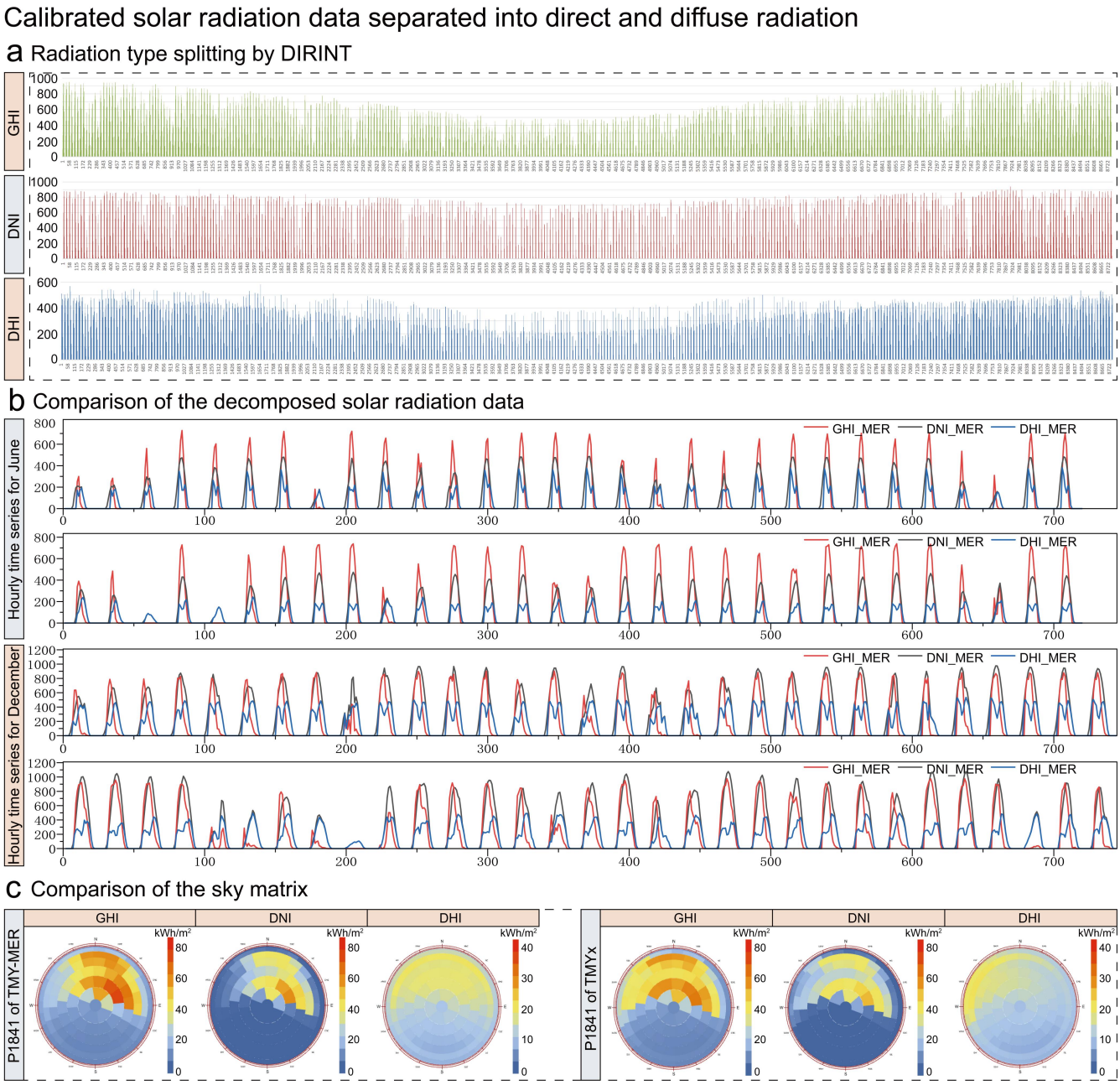


Figure 9: Direct–diffuse radiation decomposition and sky-radiance modeling based on TMY-MER-IDW: (a) DIRINT-derived DNI and DHI time series at site P1814; (b) comparison of GHI, DNI, and DHI components between TMY-MER-IDW and the reference TMY product; (c) sky-radiance models generated from the decomposed radiation components.

The sky-radiance models generated from the decomposed radiation components show broadly similar spatial patterns between TMY-MER-IDW and the reference TMY product, supporting the use of TMY-MER-IDW for application-oriented sky-radiance modeling and building-scale solar-radiation simulations. However, because DNI and DHI are derived through empirical decomposition rather than directly corrected or independently validated, the results in Fig. 9 should be interpreted as an application-level consistency illustration rather than independent validation of the original MERRA2 GHI, the DIRINT-derived radiation components, or the TMY-MER-IDW dataset.

4 Discussion

4.1 Research contributions

This study developed a spatially distributed MERRA2-based typical-year dataset (TMY-MER) for solar-radiation evaluation across Australia, aiming to address the limited spatial coverage and update frequency of conventional station-based TMY products (Gelaro et al., 2017). Unlike previous studies that mainly identified systematic biases in MERRA2-derived solar radiation (Cao et al., 2022; Du et al., 2022), this study further characterizes the spatiotemporal patterns of GHI discrepancies and their potential contributing factors under different weather conditions. By using MERRA2 data from 2009 to 2023, TMY-MER provides a recent-period typical-year product for GHI-centered solar-radiation evaluation, reducing reliance on sparse and infrequently updated ground-based TMY sources (Marion and Urban, 1995; Ren et al., n.d.; Song et al., 2007; Isaza et al., 2024). The dataset is positioned as a spatially distributed product for representative locations, rather than as a product that increases the native spatial resolution of MERRA2, and is intended to support solar-radiation-related analysis and building-scale applications.

A weather-type decomposition method based on clear-sky identification was proposed to distinguish GHI discrepancies under clear-sky and overcast conditions. The analysis indicates that TMY-MER shows relatively small positive deviations under clear-sky conditions, with annual mean discrepancies generally below 5%. In contrast, discrepancies under overcast conditions are substantially larger, reaching up to 40%, which is mainly associated with uncertainties in cloud and aerosol representation in reanalysis data. These findings are consistent with previous studies that identified cloud cover and aerosol-related processes as important sources of radiation bias in reanalysis products (Cao et al., 2022; Du et al., 2022). Spatially, the most pronounced overestimations occur in densely populated southeastern coastal regions, with annual cumulative discrepancies reaching approximately 30%, whereas inland areas generally show smaller deviations, typically below 10%. Seasonally, larger discrepancies are observed during winter overcast periods, while summer clear-sky conditions show smaller deviations, generally below 5%.

Methodologically, the combination of weather-type decomposition and IDW interpolation of bias parameters provides a structured framework for assessing and spatially extending GHI discrepancy information from comparison locations to 353 representative locations across Australia. This framework improves the interpretability of MERRA2-based GHI discrepancies under different sky conditions and supports the generation of the spatially distributed TMY-MER-IDW product for solar-radiation evaluation. However, because the IDW procedure is based on bias parameters derived from comparison with a reference TMY product, the resulting discrepancy reduction should be interpreted as product-level agreement rather than independent validation against ground-truth observations. Within this scope, the proposed workflow offers a transferable approach for applying reanalysis-based typical-year data to regions with diverse climatic conditions and provides data support for solar-energy assessment, building-surface irradiance simulation, and urban thermal-environment analysis.

4.2 Discussion on error sources

This study systematically investigates the mechanisms contributing to errors in MERRA2 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 MERRA2 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, 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 Queensland, wet-season (December to March) overcast conditions are dominated by deep convective systems. MERRA2'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).

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%. Urbanization further exacerbates these inaccuracies via 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 not detected by reanalysis systems, introducing additional overestimations in radiative flux. Moreover, in temperate regions such as the southern highlands of Victoria, 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 not captured by the MERRA2 model.

4.3 Application case study and simulation consistency

To examine the applicability of the adjusted TMY-MER-IDW dataset in building-scale solar-radiation simulations, two representative building clusters with complex shading conditions were selected. Monthly surface solar-irradiance distributions were simulated using both the reference TMY product and the adjusted TMY-MER-IDW dataset at a spatial resolution of 0.5 m × 0.5 m. Accordingly, this analysis is interpreted as an application case study and simulation-consistency assessment, rather than as independent validation of dataset accuracy.

Cluster 1 (Fig. 10a), located in Collie, Western Australia, was used to compare the surface solar-radiation distributions generated from the two datasets under a relatively simple building configuration. The results show that the irradiance patterns simulated using TMY-MER-IDW are generally consistent with those derived from the reference TMY product in terms of both spatial distribution and magnitude. The annual total relative deviation remained below 10%. For the representative months of March, June, September, and December, the deviations were generally below 10%, except in September, when the deviation increased to approximately 16%. Cluster 2 (Fig. 10b), located in Brisbane, Queensland, represents a more complex urban configuration with stronger mutual shading among buildings. Despite the increased geometric complexity, the surface solar-radiation distributions simulated using TMY-MER-IDW remained broadly consistent with those obtained from the reference TMY product. The annual cumulative relative deviation was below 5%, with winter deviations remaining below 5% and deviations in other representative months generally below 3%. These results suggest that the adjusted TMY-MER-IDW dataset can reproduce the main spatial characteristics of surface solar exposure within the same simulation workflow.

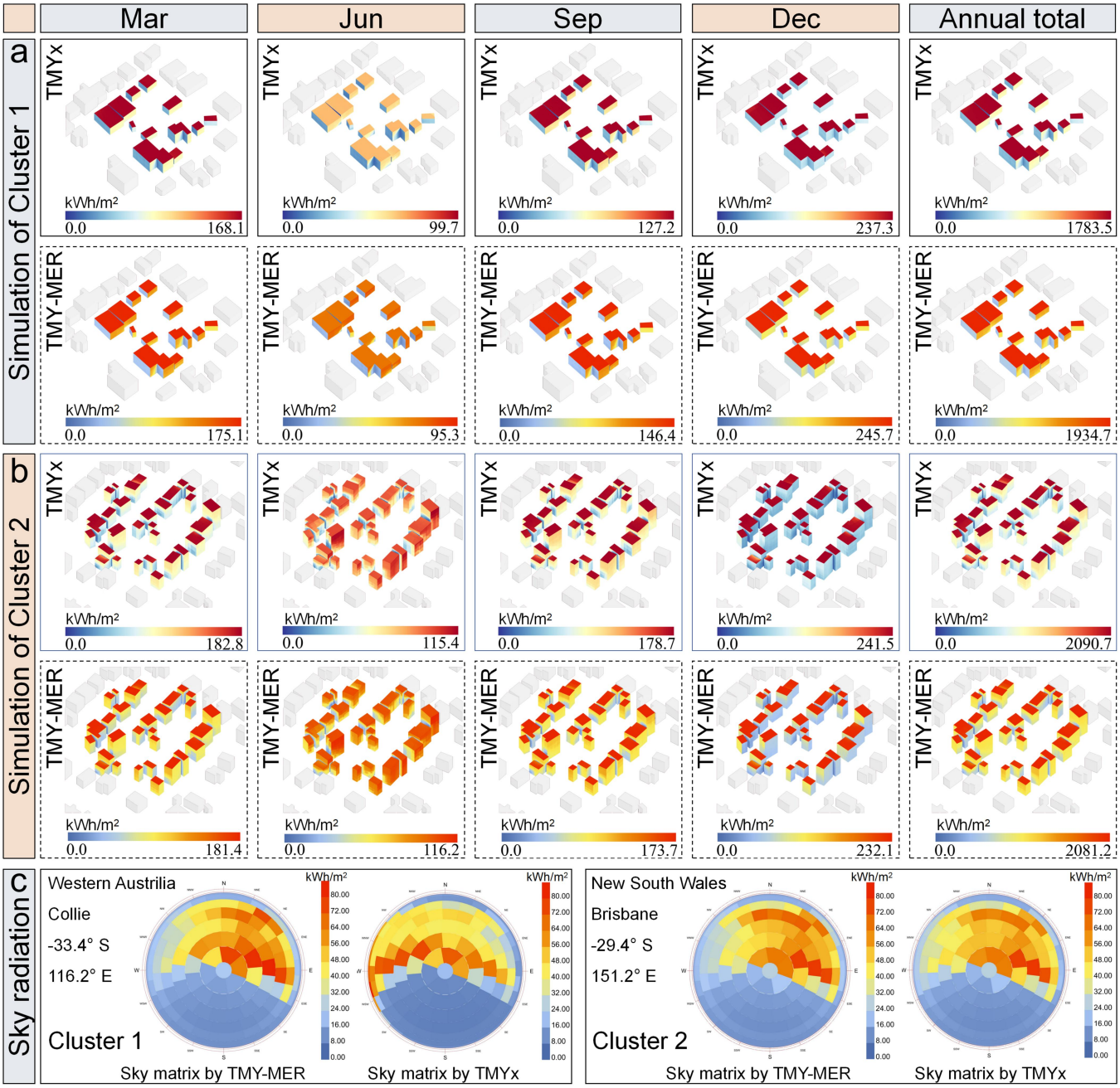


Figure 10: Application of adjusted meteorological data in building cluster simulations. (a) Surface solar radiation distribution for Building Cluster 1 based on TMYx and TMY-MER; (b) Surface solar radiation distribution for Building Cluster 2 based on TMYx and TMY-MER; (c) Comparison of distribution characteristics of sky radiation models.

Figure 10c further compares the sky-radiance models generated from TMY-MER-IDW and the reference TMY product for the two building clusters. For Cluster 1, the differences between the two datasets are mainly reflected in irradiance magnitude during winter, whereas seasonal variations in the transitional and summer periods are more evident in the spatial distribution of sky radiance. For Cluster 2, the sky-radiance models exhibit nearly identical spatial patterns across the selected seasons, with only minor differences in intensity. Overall, the similarity in surface solar-irradiance distributions and sky-radiance patterns indicates application-level consistency between TMY-MER-IDW and the reference TMY product within the same LBT/Radiance simulation workflow. However, because both datasets were processed using the same simulation engine, these results should be interpreted as indicators of simulation consistency and downstream applicability, rather than as independent evidence of dataset accuracy.

Based on the adjusted TMY-MER-IDW data, numerical sky-radiance models were constructed to support monthly and

representative-day solar-radiation simulations. A set of 36 monthly sky-radiance models was generated based on monthly GHI, DNI, and DHI components (Fig. 11a). In addition, another set of 36 representative-day sky-radiance models was constructed for the solar equinoxes and solstices, including the days immediately preceding and following each event (Fig. 11b). These models provide flexible input conditions for building-scale solar-radiation simulations under different temporal scenarios. By selecting monthly or representative-day sky-radiance models according to specific simulation objectives, the proposed workflow can support analyses of total, direct, and diffuse solar exposure on building surfaces, as well as representative seasonal assessments for urban thermal-environment and photovoltaic-related applications. Therefore, this modeling framework demonstrates the downstream usability of the adjusted TMY-MER-IDW data in application-oriented solar-radiation simulations. However, because the sky-radiance models are generated within the same LBT/Radiance simulation workflow, the results should be interpreted as evidence of simulation consistency and practical applicability, rather than as independent validation of dataset accuracy.

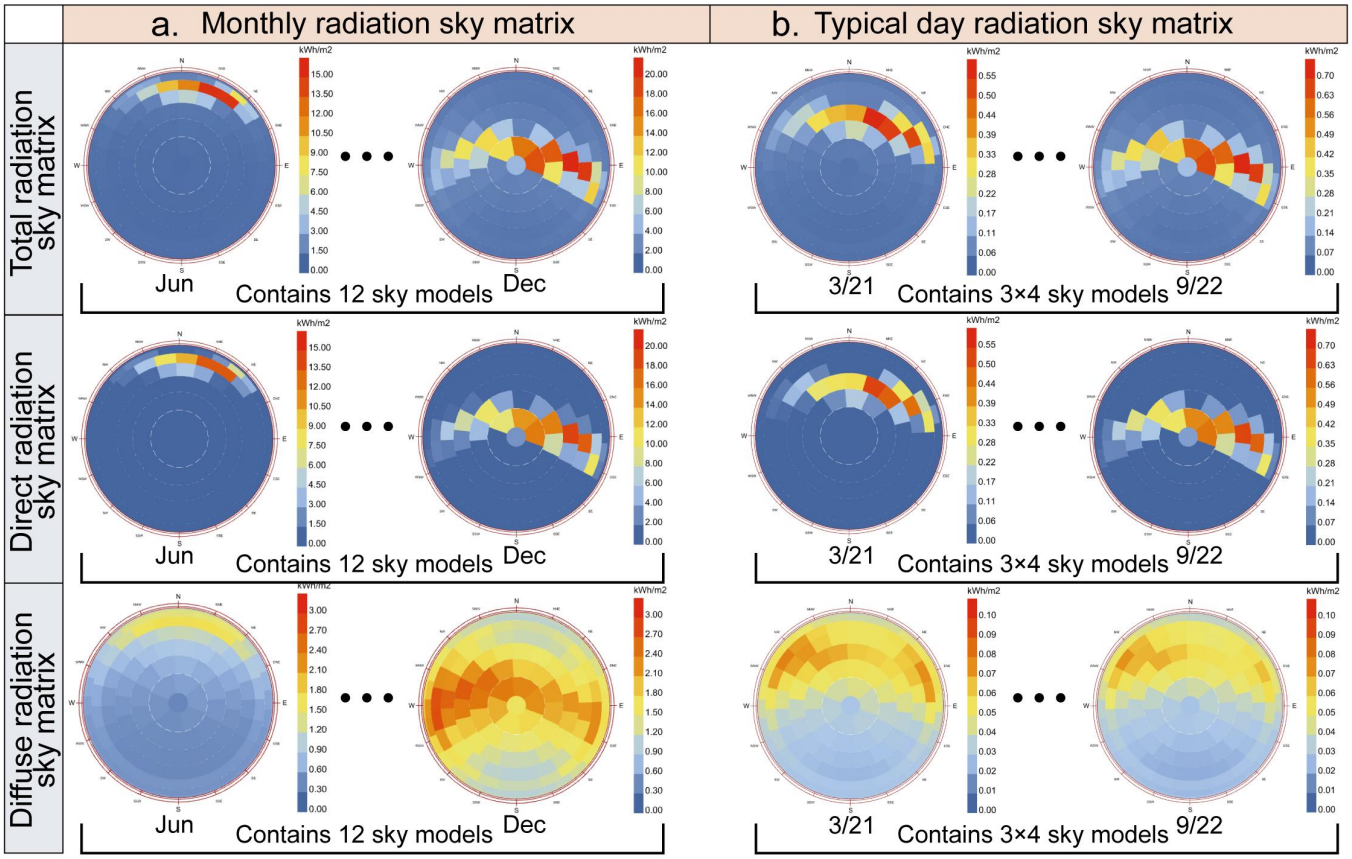


Figure 11: Construction of multi-condition sky-radiance models based on TMY-MER-IDW: (a) monthly radiation sky matrix; (b) representative-day radiation sky matrix.

4.4 Limitations and future work

Several limitations should be noted. First, the dataset is based on a recent 15-year period from 2009 to 2023. Although this period is relevant for contemporary solar-radiation applications, it is shorter than the conventional 30-year climatological period and may influence typical-month selection. Future work should compare the 15-year recent-period dataset with longer MERRA2 baselines, such as a 30-year period or the full MERRA2 record available since 1980.

Second, the quantitative discrepancy assessment and bias adjustment in this study focus on GHI. Other meteorological variables, including dry-bulb temperature, relative humidity, wind speed, and wind direction, are retained from MERRA2 as contextual variables and are not independently corrected or validated. Future work should evaluate non-GHI variables and inter-variable coherence, particularly the coupling among GHI, temperature, humidity, wind, and cloud-related conditions.

Third, the comparison with the reference TMY product should be interpreted as an inter-product comparison rather than

independent ground-truth validation. The reference TMY product is itself a statistically constructed dataset, and therefore the reported discrepancies represent product-level differences rather than true observational errors. Future work should conduct independent validation against high-quality ground-based solar-radiation observations where such data are available.

Fourth, because IDW is used to interpolate weather-type-conditioned GHI bias parameters derived from 122 comparison locations, the IDW-adjusted results do not by themselves demonstrate independent predictive performance at unseen locations. Future work should conduct leave-one-out or spatial-block cross-validation to evaluate the generalizability of the interpolation-based adjustment. In addition, the clear-sky threshold τ used in weather-type classification may influence the derived bias parameters, and sensitivity analysis of τ should be examined in future work.

Finally, the 353 representative locations are administrative-center or city-coordinate-based locations mapped to the nearest native MERRA2 grid cells, rather than ground-based weather stations or newly generated high-resolution grid nodes. Therefore, multiple representative locations may share the same underlying MERRA2 grid-cell meteorological representation.

5 Data availability

The TMY-MER-IDW 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 spatially distributed MERRA2-based typical-year product (TMY-MER) for GHI-centered solar-radiation evaluation across Australia. A weather-type-conditioned GHI discrepancy assessment framework was introduced by integrating clear-sky identification, weather-type classification, and IDW-based GHI bias-parameter interpolation. The comparison with the reference TMY product is interpreted as an inter-product GHI discrepancy assessment rather than independent validation against ground-truth observations. DIRINT decomposition and building-scale solar-radiation simulations were further used as application-oriented analyses to examine downstream usability and simulation consistency.

(1) A recent-period TMY-MER dataset was generated for 353 representative locations across Australia. The dataset provides broad spatial coverage for GHI-centered solar-radiation applications, but it does not increase the native spatial resolution of MERRA2 and should not be interpreted as a high-resolution gridded dataset. Non-GHI variables are retained from MERRA2 as contextual variables to maintain the weather-file structure and support downstream simulation workflows, but they are not independently corrected or validated in the present study.

(2) The weather-type-conditioned assessment showed that TMY-MER generally produced small positive GHI discrepancies under clear-sky conditions, with annual mean discrepancies typically below 5%. Under overcast conditions, the discrepancies became substantially larger, with maximum deviations reaching approximately 40%, likely associated with uncertainties in cloud and aerosol representation in reanalysis data. Spatially, larger positive discrepancies were observed in several coastal regions, particularly in southeastern and southwestern coastal areas, whereas inland areas generally showed smaller discrepancies.

(3) The IDW-based interpolation of weather-type-conditioned GHI bias parameters supported the generation of the adjusted TMY-MER-IDW product. Because the interpolated parameters were derived from comparison with the reference TMY product, the resulting discrepancy reduction should be understood as product-level agreement within the proposed adjustment framework, rather than independent predictive validation at unseen locations.

(4) The DIRINT-based decomposition of GHI into DNI and DHI provided derived radiation components for application-oriented modeling. The subsequent sky-radiance modeling and building-scale solar-radiation simulations showed that TMY-MER-IDW could reproduce the main spatial characteristics and seasonal patterns of surface solar exposure within the same LBT/Radiance workflow. These results support the use of TMY-MER-IDW for application-oriented solar-radiation modeling,

while the derived DNI and DHI components remain subject to uncertainty introduced by empirical direct–diffuse decomposition.

(5) Future work should further evaluate and improve the dataset by incorporating independent ground-truth validation, leave-one-out or spatial-block cross-validation of IDW-based bias-parameter interpolation, evaluation of non-GHI meteorological variables and inter-variable coherence, sensitivity analysis of the clear-sky threshold τ, and comparison between the 15-year recent-period baseline and longer MERRA2 climatological baselines.

Author contributions.

JF and PN contributed equally to this work. JF and PN conceived the research idea, developed the methodology, collected and processed the MERRA2 reanalysis data, and drafted the manuscript with revisions. ZY and BH supervised the research, secured funding, and managed project resources. DP supported the building-cluster simulation application case and provided critical feedback on the application assessment. 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.: Development of gridded surface meteorological data for ecological applications and modelling, Int. J. Climatol., 33, 121–131, https://doi.org/10.1002/joc.3413, 2013.

Aliana, A., Chang, M., Østergaard, P. A., Victoria, M., and Andersen, A. N.: 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, 2022.

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

Bre, F. and Fachinotti, V. D.: 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, 2016.

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

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

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., and Yu, H.: The MERRA-2 aerosol reanalysis, 1980 onward. Part II: Evaluation and case studies, J. Clim., 30, 6851–6872, https://doi.org/10.1175/JCLI-D-16-0613.1, 2017.

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

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

Cao, Q., Liu, Y., Sun, X., and Yang, L.: 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, 2022.

Chakraborty, T. and Lee, X.: Large differences in diffuse solar radiation among current-generation reanalysis and satellite-derived products, J. Clim., 1–52, https://doi.org/10.1175/JCLI-D-20-0979.1, 2021.

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., and Saleska, S. R.: Amazon forest biogeography predicts resilience and vulnerability to drought, Nature, 631, 111–117, https://doi.org/10.1038/s41586-024-07568-w, 2024.

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

Dincer, I. and Aydin, M. I.: New paradigms in sustainable energy systems with hydrogen, Energy Convers. Manage., 283, 116950, https://doi.org/10.1016/j.enconman.2023.116950, 2023.

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

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

Engerer, N. A.: 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, 2015.

Erbs, D. G., Klein, S. A., and Duffie, J. A.: Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation, Sol. Energy, 28, 293–302, https://doi.org/10.1016/0038-092X(82)90302-4, 1982.

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

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

Fu, J., Ni, P., Yan, Z., He, B.-J., Prasad, D., and Qin, G, Lei, F., Yue, Y., & Yan, J.: A Spatially Distributed MERRA2-based Typical-Year Dataset for Solar-Radiation Evaluation in Australia (Version v1.1), Zenodo, https://doi.org/10.5281/zenodo.15479502, 2025.

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

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., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

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

Hall, I. J., Prairie, R. R., Anderson, H. E., and Boes, E. C.: Generation of a typical meteorological year, Sandia Labs., Albuquerque, NM (USA), 1978.

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

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

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

00594-5, 2024.

Isaza, A., Evans, J. P., Kay, M., Prasad, A., and Brember, S.: 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, 2024.

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

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

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

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

Lemos, L. F. L., Starke, A. R., Boland, J., Cardemil, J. M., Machado, R. D., and Colle, S.: 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, 2017.

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

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

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

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

Lund, H.: The design reference year user's manual, thermal insulation laboratory, Tech. Univ. Den.: Lyngby, 1995.

Marion, W. and Urban, K.: User's manual for TMY2s (typical meteorological years) - derived from the 1961-1990 national solar radiation data base, https://doi.org/10.2172/87130, 1995.

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., and Yan, X.: 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, 2023.

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

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

Publishing, O.: Transition to sustainable buildings: strategies and opportunities to 2050, 1st ed., Organization for Economic Cooperation & Development, Paris, 1 pp., 2013.

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

Pusat, S., Ekmekçi, İ., and Akkoyunlu, M. T.: 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, 2015.

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

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

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., Sourdeval, O., Stier, P., Tesche, M., van Diedenhoven, B., and Wendisch, M.: Constraining the twomey effect from satellite observations: Issues and perspectives, Atmos. Chem. Phys., 20, 15079–15099, https://doi.org/10.5194/acp-20-15079-2020, 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., and Flynn, C. J.: The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation, J. Clim., 30, 6823–6850, https://doi.org/10.1175/JCLI-D-16-0609.1, 2017.

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

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., and Lauret, P.: Modelling of diffuse solar fraction with multiple predictors, Renewable Energy, 35, 478–483, https://doi.org/10.1016/j.renene.2009.07.018, 2010.

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

Song, F., Zhu, Q., Wu, R., Jiang, Y., Xiong, A., Wang, B., Zhu, Y., and Li, Q.: Meteorological data set for building thermal environment analysis of China, Build. Simul., 2007.

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

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

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

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

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

Wilcox, S. and Marion, W.: Users manual for TMY3 data sets, Tech. Rep., 2008.

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

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

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

Yoo, A.: Australian energy market operator (AEMO), 2023.

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

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