AIMS Geosciences

AIMS Geosciences, 12(2): 499–523.
DOI: 10.3934/geosci.2026019
Received: 28 November 2025
Revised: 27 February 2026
Accepted: 19 March 2026
Published: 12 May 2026

https://www.aimspress.com/journal/geosciences

Research article

Generation of typical meteorological years at 30 sites in Saudi Arabia for solar energy applications

Ashraf Farahat[1,2,3], Harry D Kambezidis[4,5,*], Abdulhaleem Labban[6] and Kosmas A Kavadias[5]

[1] Department of Physics, College of Engineering and Physics, King Fahd University of Petroleum and Minerals, Dhahran SA-31261, Saudi Arabia
[2] Centre of Aviation and Space Exploration, King Fahd University of Petroleum and Minerals, Dhahran SA-31261, Saudi Arabia
[3] Centre of Research Excellence in Renewable Energy, King Fahd University of Petroleum and Minerals, Dhahran SA-31261, Saudi Arabia
[4] Institute of Environmental Research and Sustainable Development, National Observatory of Athens, GR-11810 Athens, Greece
[5] Laboratory of Soft Energies and Environmental Protection, Department of Mechanical Engineering, University of West Attica, GR-12241 Athens, Greece
[6] Department of Meteorology, Faculty of Environmental Sciences, King Abdulaziz University, SA-21589 Jeddah, Saudi Arabia

* **Correspondence:** Email: harry@noa.gr; Tel: +30-210-9649909.

Abstract: Typical Meteorological Years (TMYs) are synthetic annual datasets constructed from long-term observations to represent the characteristic climatic conditions of a location. Rather than reproducing extreme events, a TMY captures the typical variability of key meteorological parameters and is widely used in energy system modeling, building performance simulations, and climate-responsive design. Although TMYs have been developed for many regions worldwide, no comprehensive effort has been made for Saudi Arabia, despite its rapidly growing interest in solar energy. In this study, we present the first derivation of TMYs for thirty sites across the country. Hourly observations of global, direct, and diffuse horizontal irradiance, air temperature, relative humidity, and mean and maximum wind speed were quality-controlled and processed. The modified Sandia National Laboratories method, based on the Finkelstein–Schafer (FS) statistical procedure, was applied to select representative months and assemble each TMY. Graphical diagnostics and statistical indicators were

used to evaluate the agreement between the generated TMYs and long-term climatological averages. The results demosntrated that the constructed TMYs successfully reproduce the typical meteorological behavior at all sites, providing reliable input data for solar-energy assessments and other climate-sensitive applications in Saudi Arabia.

Keywords: typical meteorological years; solar energy applications; Finkelstein-Schafer procedure; Saudi Arabia

1. Introduction

A Typical Meteorological Year (TMY) is a set of meteorological and irradiance parameters mostly consisting of hourly values in a year for a given geographical location (accessed at https://e3p.jrc.ec.europa.eu/articles/typical-meteorological-year-tmy). Furthermore, a TMY consists of a set of months selected from individual years integrated into a complete year. Therefore, a TMY reflects all the climatic information of the location for a period as long as the mean life of the system. The notion of the TMY was first introduced by Hall et al. [1], highlighting its usefulness in climatic applications as it contains a complete year with representative months of the climate for a specific location. The use of a TMY can be valuable in climatological and bio-meteorological studies, solar energy applications, energy design of buildings, and daylight applications. Numerous attempts around the world have been made to create TMYs. These are described below.

In the USA, before the appearance of the TMY notion, the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE, accessed at https://www.ashrae.org) conducted three research projects (report RP-100, [2]; report RP-239, [3]; report RP-364, [4]) during the period 1970–1983, which ended up with creating weather datasets more likely to represent typical weather patterns, a so-called Typical Reference Year (TRY), than a single representative year or an ensemble of months. Those datasets received the abbreviation WYEC (Weather Year for Energy Calculations); they used the TRY format and included solar radiation data (measured, if available, or otherwise calculated via a solar code). In the 1990s, ASHRAE updated the WYEC dataset. The new WYEC datasets were in the TMY format and included illuminance data, data quality, and source flags. The calculated solar radiation data used the Perez model [5]. Seventy-seven WYECs were generated for the USA, known as WYEC2 (version 2). Fifty-one of those WYEC2 sets contained solar radiation data calculated via the Perez model (known as WYEC2W), and twenty-six included measured SOLMET solar data (known as WYEC2T). In 1994, the National Solar Radiation Data Base (NSRDB, accessed at https://rredc.nrel.gov/solar/old_data/nsrdb) was established; this database included two-hundred and thirty-nine TMYs generated from an equal number of meteorological stations using weather data in the period 1961–1990. In 1997, the ASHRAE undertook a research study (report RP-1015 in 2002) for the development of International Weather Years for Energy Calculations (IWECs). These datasets were completed in 2001 and consist of hourly TMYs for two hundred and twenty-seven locations across seventy countries, all outside the USA and Canada. For the development of these IWYECs, up to eighteen years of weather data from the period 1982–1999 were processed using Hall's method. Nowadays, the IWYEC files cover three thousand and twelve locations worldwide (IWYEC2), all outside the USA and Canada, through a research

project of ASHRAE (report RP-1477). These files were derived from data from weather stations around the world and are archived in the Integrated Surface Hourly database maintained by the National Climatic Data Center (NCDC, accessed at https://www.ncdc.noaa.gov).

In the USA, Klein et al. [6] used a similar method to Benseman and Cook's [7] method to construct an average-year dataset of solar radiation and ambient temperature for Madison, Wisconsin. The National Renewable Energy Laboratory (NREL, accessed at https://www.nrel.gov) conducted the first generation of two-hundred and twenty-nine TMYs from an equal number of meteorological stations in the country using data between 1948 and 1980. The second edition of the TMYs (TMY2) was based on data from two-hundred and thirty-nine meteorological stations in the period 1961–1990. The latest edition (TMY3) was based on data from one-thousand and twelve locations in the USA, Guam, Puerto Rico, and US Virgin Islands, covering the period 1976–2005 if data was available, and 1991–2005 otherwise [8]. Moreover, Realpe et al. [9] developed benchmarking for five TMY datasets for CPV-system applications, using data of eighteen years from the meteorological station of Desert Rock, Nevada, and applied the concept of "driver" that has been defined in the frame of the European project ENDORSE [10].

In the regions of Asia, Africa, Australia, and South America, several studies have provided TMYs for various applications. In this respect, Cebecauer and Suri [11] introduced satellite data (the SolarGIS approach, accessed at https://solargis.com) into a database for the generation of TMY when meteorological and/or solar data were not available at a site. Ohunakin et al. [12] developed a TMY for northeastern Nigeria; the same authors created a TMY for Sokoto, northwestern Nigeria, using NASA's satellite imagery information [13]. Additionally, Oko and Ogoloma [14] generated a TMY for Port Harcourt, Nigeria, using data in the period 1983-2002. Chan et al. [15] reviewed different approaches for deriving a TMY for Hong Kong using data of twenty-five years (1979–2003). Later, Chan [16] developed a modified TMY weather file for Hong Kong, taking into account the urban heat-island effect. The same author developed an updated TMY for Hong Kong using a genetic algorithm for different energy applications [17]. Ebrahimpour and Maerefat [18] concluded that the Sandia National Laboratories (SNL) method coupled with the Meteonorm software gives better agreement with the long-term average of measured data during a year. Zang et al. [19] developed TMYs for thirty-five locations in China using the SNL method with own weighting factors and data in the period 1994-2010. Yang et al. [20] and Jiang [21] developed modified versions of the SNL method and applied them for generating TMYs for China, while Sun et al. [22] selected the most suitable TMY using different TMY datasets in China for energy simulations for buildings with daylight utilization. Kim et al. [23] developed TRYs for eighteen locations in south Korea using ISO 15927-4 for building-energy performance. Murphy [24] suggested a modified TMY for PV applications in India, while Farah et al. [25] presented a robust meteorological year weather dataset for the city of Adelaide, Australia. Janjai and Deeyai [26] performed a comparison of three methods (SNL, Danish, and Festa-Ratto) for the generation of a TMY at four sites with tropical climatic conditions in Thailand over the period 1995-2004; they concluded that the best performing method is the Sandia one, thus agreeing with [27]. In addition, Sepúlveda et al. [28] made a comparison of five TMY methodologies (SNL, NREL, TMY3, ASHRAE, and WYEC2 [20,21]) for Chillan, Chile, which was the first attempt to create a TMY in Chile.

Many studies have been carried out with the aim of generating TMYs at various locations in the Mediterranean and the Middle East. Pissimanis et al. [29] were the first to derive a TMY for Athens, Greece. Argiriou et al. [30] provided a newer edition of the TMY-Athens. Petrakis et al. [31] and

Kalogirou [32] generated TMY and TMY2, respectively, for Nicosia, Cyprus, in the period 1986–1992 (the TMY2 simply contained additional hourly values of direct solar radiation). Moreover, Kambezidis et al. [33] developed five versions of TMYs for thirty-three sites (including Athens) in Greece from data in the period 1984–2014 coming from stations belonging to the Hellenic National Meteorological Service; the different TMYs can be used for applications in the fields of meteorology-climatology, agro-meteorology-hydrology, bio-meteorology, photovoltaics, and energy design of buildings. Skeiker [34] introduced a TMY for Damascus, Syria in the period 1981–1990, while, later, Skeiker and Ghani [35] developed an advanced software tool for the generation of a TMY. Pusat et al. [36] created TMYs for eight locations in Turkey, using data in the period 1989–2006 to serve building heating-energy calculations. For the same application, Yilmaz and Ekmekci [37] presented TMYs for forty-one locations in Turkey using meteorological data of those sites in the period 1989–2012. Additionally, Bulut [38] used the Hall's method to create TMYs for various locations in Turkey. In Egypt, Shaltout and Tadros [39] created Typical Solar Years for ten locations in the country. Faiman et al. [40] constructed TMYs for eight locations in Israel covering the period 1989–1999, while Fernàndez et al. [41] generated and evaluated TMY datasets for in-greenhouse and ambient conditions at two Mediterranean coastal sites in Spain.

Another version of the TMY is the Average Meteorological Year (AMY), which consists of all months in a reference period; these months contain average values (hourly, daily, or monthly) of the parameters considered in each month over the reference period. TMYs and AMYs exclude extreme weather events and contain the same number of data points, but their significance differs: A TMY is composed of real observed months that best represent the climatic behavior of the reference period, whereas an AMY contains smoothed values that may deviate from actual conditions. A Short Reference Year (SRY) was also proposed in earlier work [42,43], consisting of selected representative days rather than full months, although its use has largely disappeared due to limited accuracy.

Beyond these classical definitions, the literature shows that TMYs have evolved into a broader family of application-specific reference years. Several researchers have modified the weighting factors in the Finkelstein–Schafer procedure to improve representativeness for particular sectors, including solar-energy modelling and building-performance simulations [1,8,30]. Other researchers have developed specialized variants such as thermal–humid meteorological years for comfort and HVAC (heating, ventilation, and air conditioning) analysis [17], radiation-focused meteorological years for solar-resource assessment [44], and Design Meteorological Years (DMYs) widely used in engineering and building-energy codes, including ASHRAE climate design guidelines [45]. These developments highlight that the construction of a TMY is not universal but depends on the intended application and the relative importance assigned to different meteorological variables. In this context, we focus on generating TMYs optimized for solar-energy assessments in Saudi Arabia.

Despite the usefulness of TMYs, these typical years have some disadvantages [46]. The TMYs consist of Typical Meteorological Months (TMMs), which do not provide information about extreme weather events and, therefore, do not necessarily represent actual conditions at any given time at the location of interest. This occurs because all the data incorporated in the TMY expresses the probability that a parameter value lies in the 50% confidence interval of all the parameter values included in the dataset. Even so, a TMY is a valuable tool because it contains a concise dataset that reflects the climatology of an area.

No attempt has been made to create TMYs for Saudi Arabia. Therefore, we are the first to fill this gap. Moreover, the TMYs are generated for thirty sites across the country, thus covering all possible

climates of the region. Additionally, the TMYs are adapted to the needs of solar energy applications. This means that the parameters selected in the datasets of the thirty sites influence the efficiency of solar collectors such as PVs.

The paper is articulated as follows: In Section 1, we provide an introduction to the topic of the work by referring to relevant studies found in the international literature and state our aims. In Section 2, we give a detailed description of the data used and the way it was processed. In Section 3, we account for the derivation of the thirty TMYs and provide an evaluation of their accuracy. In Section 4, we show the conclusions of the study and provide a discussion for future research in this field. Acknowledgments and References follow.

2. Materials and methods

2.1. Data collection and processing

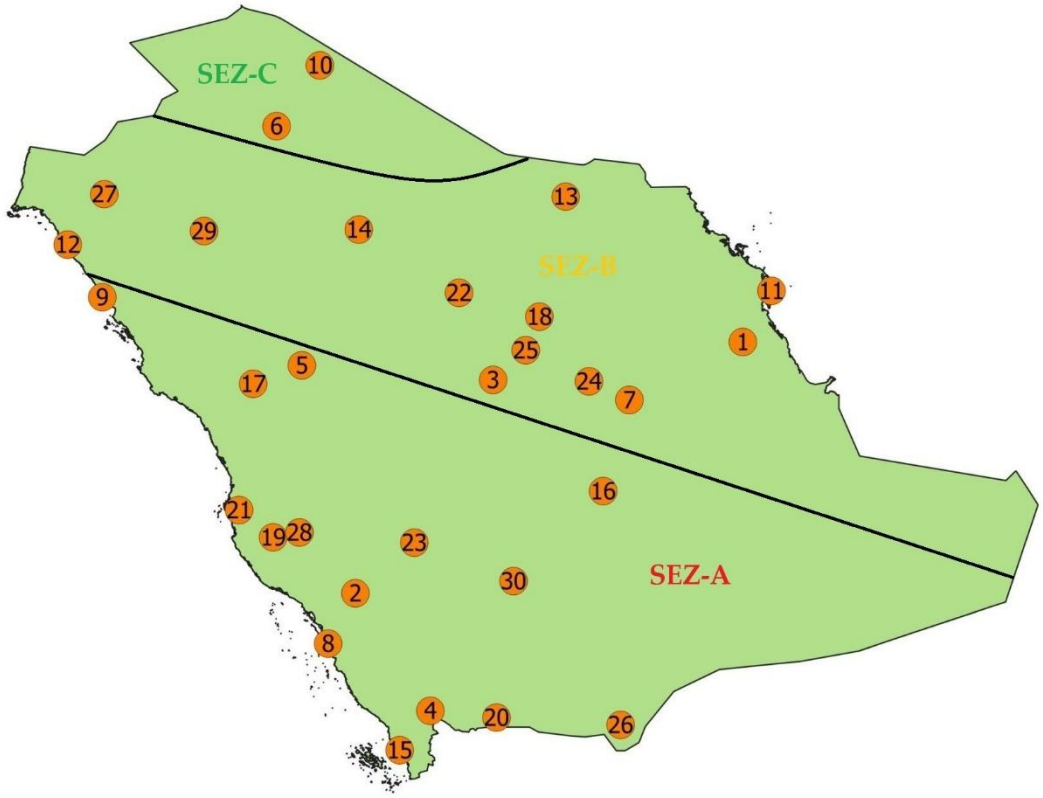

Figure 1. Location of the thirty sites in Saudi Arabia considered in this study. The numbers in the orange circles denote the towns/cities referred to Table 1, column 1. SEZ = solar energy zone.

Table 1 provides information about the thirty sites in Saudi Arabia selected for this study. The information contains the names of the sites (towns/cities), their geographical coordinates, their altitude, and the period of operation of the measuring stations. Figure 1 shows the map of Saudi Arabia with the orange circles indicating the location of the sites; the numbers within the circles correspond to those in Table 1, column 1. The map is divided into three solar energy zones (SEZ), SEZ-A, SEZ-B,

and SEZ-C, in Saudi Arabia. These SEZs denote the optimal tilt for an inclined solar system; for fixed-tilt systems facing south, the optimal tilt for maximum energy retrieval is 20º, 25º, and 30º for SEZ-A, SEZ-B, and SEZ-C, respectively, according to [47]; for 1-axis systems, the solar collector's tilt is 40º, 45º, and 50º, respectively, according to [48].

Table 1. List of the thirty sites within Saudi Arabia considered in this study; λ = geographical longitude (in degrees East); φ = geographical latitude (in degrees North); and z = altitude (in m amsl; amsl = above mean sea level). The geographical coordinates are rounded to the second decimal digit; the altitudes are integers.

##	Town/City	λ	φ	z	Period of measurements
11	Al Ahsa	49.60	25.35	170	29/5/2013–31/8/2021
22	Al Baha	41.64	20.18	1680	1/12/2014–31/3/2021
33	Al Dawadmi	44.47	24.56	955	17/7/2013–31/8/2021
44	Al Farshah	43.18	17.77	1094	18/11/2013–31/8/2021
55	Al Hanakiyah	40.54	24.86	873	28/8/2013–31/8/2021
66	Al Jouf	40.02	29.78	680	1/12/2014–31/8/2021
77	Al Kharj	47.27	24.15	465	23/7/2013–31/8/2021
88	Al Qunfudhah	39.54	24.48	20	5/9/2013–31/7/2021
99	Al Wajh	43.77	26.35	21	26/7/2013–31/8/2021
110	Arar	41.08	19.15	583	1/12/2014–31/8/2021
111	Dammam	36.44	26.26	28	27/5/2013–31/8/2021
112	Duba	40.91	31.03	45	26/7/2013–31/8/2021
113	Hafar Al Batin	50.19	26.39	383	7/10/2013–31/8/2021
114	Hail	35.72	27.34	928	1/1/2015–31/8/2021
115	Jizan	45.96	28.33	65	1/11/2014–31/8/2021
116	Layla	41.71	27.65	567	11/6/2013–31/8/2021
117	Al Madinah	42.55	16.96	636	1/3/2015–31/8/2021
818	Al Majma'ah	46.73	22.28	718	19/7/2013–31/8/2021
119	Makkah	45.42	25.86	277	1/3/2015–31/8/2021
220	Najran	44.54	17.63	1187	13/12/2013–31/8/2021
221	Osfan	39.25	21.89	119	16/9/2013–31/8/2021
222	Buraydah	43.77	26.35	688	3/6/2013–31/8/2021
223	Rania	42.85	21.22	933	1/12/2014–31/8/2021
224	Riyadh	46.44	24.53	895	9/1/2013–31/3/2021
225	Shaqra	45.14	25.17	804	16/7/2013–31/8/2021
226	Sharurah	47.09	17.48	760	4/9/2013–31/8/2021
227	Tabuk	36.48	28.38	781	25/9/2013–31/8/2021
228	Taif	40.49	21.43	1518	5/6/2013–31/8/2021
229	Timaa	38.53	27.62	844	24/7/2013–31/8/2021
330	Wadi Addawasir	44.89	20.43	671	1/8/2013–31/8/2021

The solar resource monitoring stations of Table 1 and Figure 1 are available from the new Renewable Resource Atlas (Renewable Resource Atlas V2) for Saudi Arabia developed by the King Abdullah City for Atomic and Renewable Energy as part of the Renewable Resource Monitoring and Mapping (RRMM; accessed at https://rratlas.kacare.gov.sa/RRMMPublicPortal/?q=en/Home) program in the frame of the Saudi Vision 2030 plan (Saudi Vision 2030). RRMM aimed to monitor and map solar, wind, geothermal, and waste-to-energy resources in the Kingdom. The program included the operation, calibration, and maintenance of a solar resource monitoring network that consisted of forty-six stations scattered across the country. Their operation started in 2013 and ended in 2021 (see column 6, Table 1). Table 2 provides information of the parameters measured by the RRMM network. We created a database for each site; the data [49] consisted of the following hourly mean parameters from the RRMM network: H_g (global horizontal solar irradiance, in Wm^{-2}), H_d (diffuse horizontal solar irradiance, in Wm^{-2}), H_{bh} (direct horizontal solar irradiance, in Wm^{-2}) = H_g–H_d, T (air temperature, in ºC), RH (relative humidity, in %), WS_{ave} (mean wind speed, in ms^{-1}), and WS_{max} (maximum wind speed, in ms^{-1}). These variables were chosen as they influence the performance of solar energy systems (mostly PVs) [50].

Table 2. List of the parameters recorded by most of the RRMM-program stations in Saudi Arabia. The measurement resolution in the fourth column refers to the on-site calculated parameter averages by the data-logging systems. Later, these 1-min values were converted into hourly mean values, which are used in this study. RRMM = Renewable Resource Monitoring and Mapping.

Measurements resolution (min)	Nominal uncertainty	Equipment	Parameter
1	±5%	Pyranometer	Global horizontal irradiance H_g (Wm^{-2})
	±5%	Pyranometer with rotating shadow band	Diffuse horizontal irradiance H_d (Wm^{-2})
	±5%	Pyrheliometer	Direct normal irradiance H_b (Wm^{-2})
	±0.6ºC	Sensor in radiation shield	Air temperature T (ºC)
	±3% to ±7%	Sensor in radiation shield	Relative humidity RH (%)
	±1.1%	Cup anemometer at 3 m height	Wind speed WS (ms^{-1})
	±4 degrees	Wind vane at 3 m height	Wind direction WD (degrees)

After collecting the data deployed in Table 2 for the stations described in Table 1, all hourly values of the parameters at the thirty sites underwent a quality-control test. This process consisted of the following tests: H_g, or H_d, or $H_{bh} > 0$; H_d or $H_{bh} \leq H_g$; $H_g < H_o$ (H_o = solar constant of 1361.1 Wm^{-2} [51]); $T > 0$ ºC; $0\% < RH < 100\%$, WS_{ave} or $WS_{max} > 0$ ms^{-1}, and $WS_{ave} < WS_{max}$. After the quality-control procedure, linear interpolation was applied only when the gaps did not exceed three consecutive hours, in accordance with NREL guidelines for TMY3. In such cases, blank cells in the databases, equipment

misfunctions (cells with values −99), or parameter values not obeying the above inequalities were excluded from further analysis. By doing so, the completeness of the clean datasets in percent of the initial data records is shown in Table A1 (see Appendix A).

Despite the low completeness rate in the clean datasets for some sites (see columns H, J, and L in Table A1), these datasets were considered complete for the statistical analysis applied. This means that the statistical methodology chose clean months with full coverage of hourly values in the seven parameters to constitute the TMY.

2.2. Methodology for generating TMYs

Methodologies exist in the international literature to derive TMYs. The most established approaches [27] include: (i) the Danish method [52]; (ii) the Festa–Ratto methodology [53]; and (iii) the modified Sandia National Laboratories method (MSNL) [33]. The latter is the most widely used and consistently best-performing approach in the international literature, particularly for solar-energy applications [26,27], and was, therefore, adopted in this study. It relies on the Finkelstein–Schafer (FS) statistical procedure [27], which selects twelve TMMs from different years to construct a representative synthetic year.

In recent years, the development of big-data platforms and advanced computational techniques has led to new data-driven approaches for generating TMYs. These include clustering-based methods (e.g., k-means and hierarchical clustering) for identifying representative months or days from large datasets [17,54,55]; machine-learning-assisted selection procedures that optimize variable weighting for specific applications [56]; and hybrid approaches that combine observational data with high-resolution reanalysis products to improve spatial coverage and temporal consistency [57,58]. Several researchers have also proposed application-specific TMY variants, such as thermal–humid meteorological years for building comfort analysis, radiation-focused meteorological years for solar-resource assessment, and design meteorological years used in engineering and HVAC applications. These developments demonstrate that TMY construction has expanded beyond classical statistical methods into the broader domain of data-driven climate analytics.

Despite these advances, the MSNL–FS methodology remains the international benchmark for solar-energy modeling due to its transparency, reproducibility, and extensive validation across climates. Using this method ensures comparability with global TMY datasets (e.g., TMY2 and TMY3) and aligns this work with the most-widely accepted standards in the field. The FS procedure used here consisted of the following steps:

Step 1. The cumulative distribution functions (CDFs) were calculated from the hourly values (if available) of each of the parameters p in the database (p = H_{bh}, H_d, H_g, T, RH, WS_{ave}, WS_{max} in this study) and for each month m and year y over the period of the study (2013–2021); these CDFs are named short-term ones and symbolized as $CDF_{y,m}$. In order to estimate the CDFs for each parameter, the data was grouped into i bins, and then the CDFs were estimated by counting the cases in each bin. The number of the i bins was taken equal to the number of days in each month.

Step 2. Then, the long-term CDFs (symbolised CDF_m) were calculated for each parameter, p, and each site; the CDF_m refers to the long-term composite of all years for the month, m, during the study period (9 years in this work).

Step 3. The mean difference, d, between the long-term CDF, CDF_m, and the short-term one $CDF_{y,m}$ was calculated for each parameter p over the i bins used for the estimation of the CDFs. The FS statistic

for the specific month, m, of the year, y, and the meteorological parameter, p, considered at any site (thirty sites in this study) was calculated from the hourly values of the parameter:

$$FS_{pi}(y,m) = \frac{1}{N} \cdot \sum_{j=1}^{N} \left| CDF_m(p_{ij}) - CDF_{y,m}(p_{ij}) \right| \quad (1)$$

where N is the number of bins (i days) of month m in year y.

Step 4. A variety of TMY methodologies in the literature assign weighting factors, WF, according to the relative influence of each meteorological parameter on the performance of solar-energy systems. Earlier studies (e.g., [19,33,37,59]) consistently place the highest weight on solar radiation components (particularly beam and global irradiance) because they dominate the energy yield of photovoltaic and concentrating solar technologies. Secondary parameters such as temperature, wind speed, and relative humidity are typically assigned moderate or lower weights, reflecting their indirect but non-negligible influence on system efficiency, thermal losses, and operational conditions. In this work, the weighting factors in Table 3 were not adopted verbatim from any single previous study. Instead, they were derived by us following the general weighting philosophy established in earlier TMY-generation research, while adapting the values to the needs of solar-energy applications in Saudi Arabia. The adopted scheme preserved the hierarchy observed in the literature (giving primary emphasis to solar radiation (H_{bh}, H_g, H_d), followed by temperature, wind speed, and relative humidity, while ensuring that the weights summed to unity and reflected the relative contribution of each parameter to the representativeness of the final TMY. To demonstrate consistency with earlier work, the weighting distribution used here aligned closely with the ranges reported in previous TMY studies, where solar radiation parameters typically received 40–60% of the total weight, temperature and wind speed together received 25–40%, and humidity-related variables received 10–20%. Our weighting set (Table 3) fell squarely within these established ranges, while being tailored to the climatic characteristics and solar-resource priorities of the region. WFs were applied to the FS statistical values, one for each of the considered parameters p that corresponds to each month m in the period of the study. In this way, a weighted-sum average (or composite-index) value, $WS_{y,m}$, was calculated; this value was assigned to the respective month m of the year y:

$$WS_{y,m} = \frac{1}{M} \cdot \sum_{i=1}^{M} WF_{p_i} \cdot FS_{p_i,y,m} \quad (2)$$

$$\sum_{i=1}^{M} WF_{p_i} = 1 \quad (3)$$

where M is the number of the parameters p in the database (M = 7 in the present analysis). It should be pointed out here that the WF values were user-defined, based on the importance of each parameter in the database and the contribution of the parameter to the final result (i.e., the TMY). For solar energy applications, researchers have introduced differing WFs, but they give more weight on the solar radiation parameters, as anticipated. Taking this into account, we adopted the weighting factors shown in Table 3. Appendix B provides a comparison between the chosen WFs for the seven parameters and those from other relevant studies [1,29,30,60].

Table 3. Weighting factors, WF_p, for the parameters p considered in this work for the generation of the TMYs across the thirty locations of Saudi Arabia. TMY = Typical Meteorological Year.

p	WF_p
H_{bh}	0.25
H_d	0.10
H_g	0.15
T	0.15
RH	0.10
WS_{ave}	0.15
WS_{max}	0.10
Sum	**1.00**

Step 5. For each month m, the FS statistic was calculated for every year y in the period of the study by including all the appropriate parameters p with their appropriate weighting factors WF_{pi}; a $WS_{y,m}$, was, therefore, derived for the site from Eq. (2). The smaller the $WS_{y,m}$, the better the approximation to a TMM.

Step 6. Applying the above-mentioned procedures to all months and sites during the period of the study, a composite year was formed consisting of the selected months with the smallest $WS_{y,m}$ values. For each month m, the three lowest values of the FS statistic were selected.

Step 7. The Root-Mean-Square Error (RMSE) from the hourly values of parameter p for each month m of every year y in the period of the study, with respect to the mean long-term hourly distribution of the parameter, and the FS statistic, was computed. The RMSE was calculated for the three candidate months; then, for each month, the year corresponding to the lowest RMSE value was selected. In this way, a complete (typical meteorological) year was formed. The statistic RMSE is given by the expression:

$$\text{RMSE}= \sqrt{\sum_{i=1}^{M}(x_{p_i} - \overline{x_{p_i}})^2} \quad (4)$$

where x_{pi} is the value of parameter p_i and $\overline{x_{p_i}}$ is its average value over a range of M bins.

Figure 2 shows the steps of the procedure in the form of a flow chart.

The implementation of the methodology shown in Figure 2 relied on automated processing via Python and Fortran codes, with emphasis on data quality and completeness visualization. Visualization was especially implemented using Pandas, Matplotlib, and Seaborn, following practices recommended in [57] for solar radiation analysis.

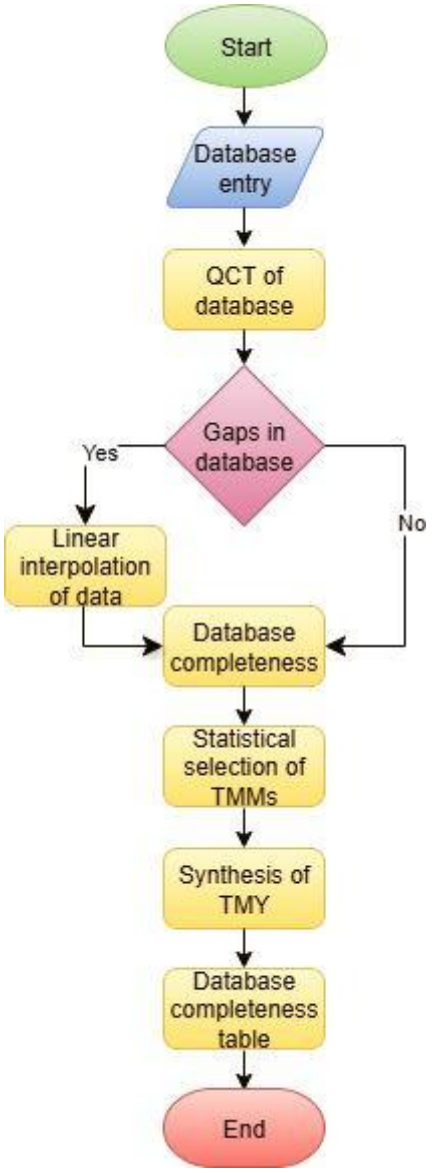

Figure 2. Flow chart of the FS statistical procedure for the generation of TMYs at thirty locations in Saudi Arabia. FS = Finkelstein-Schafer; TMY = Typical Meteorological Year.

3. Results

This section is divided into subsections, which provide the TMYs generated for the thirty sites. We compared them with the AMYs at the same sites and concluded the usefulness of the TMYs for solar energy applications in Saudi Arabia.

3.1. The generated TMYs

Table 4 gives the years selected per month in the derived TMYs for the thirty sites in Saudi Arabia. For all sites, the TMYs consist of TMMs each corresponding to an individual year; for example, the TMM of January for site #1 (Al Ahsa) consists of all valid hourly values of the seven parameters in the year 2016.

Table 4. TMMs from years in the period 2013–2021 selected by the FS statistical procedure to constitute the TMYs of the thirty sites in Saudi Arabia. TMM = Typical Meteorological Month; TMY = Typical Meteorological Year.

Site #	TMM											
	1	2	3	4	5	6	7	8	9	10	11	12
1	2016	2016	2016	2017	2016	2015	2015	2014	2016	2013	2016	2015
2	2020	2016	2018	2018	2016	2018	2016	2018	2015	2015	2015	2014
3	2014	2014	2014	2020	2020	2018	2016	2014	2016	2013	2013	2013
4	2014	2014	2014	2019	2015	2016	2016	2016	2019	2014	2014	2018
5	2014	2020	2015	2020	2020	2018	2016	2014	2014	2013	2019	2014
6	2015	2020	2016	2020	2020	2018	2016	2019	2017	2017	2015	2018
7	2016	2020	2020	2014	2014	2015	2016	2015	2014	2013	2013	2016
8	2015	2016	2014	2018	2018	2018	2018	2016	2016	2016	2017	2013
9	2017	2019	2014	2018	2016	2018	2019	2016	2016	2016	2013	2013
10	2016	2019	2016	2018	2016	2015	2016	2015	2019	2015	2019	2015
11	2016	2014	2014	2017	2016	2018	2016	2016	2013	2015	2017	2013
12	2015	2021	2014	2019	2016	2018	2019	2021	2015	2016	2017	2018
13	2017	2014	2015	2014	2014	2014	2014	2014	2016	2018	2019	2013
14	2017	2020	2021	2020	2020	2018	2019	2016	2017	2017	2019	2015
15	2016	2021	2015	2018	2021	2018	2016	2021	2017	2020	2017	2015
16	2016	2021	2015	2019	2016	2018	2019	2014	2013	2013	2013	2019
17	2021	2016	2016	2018	2018	2018	2015	2016	2017	2016	2017	2018
18	2016	2020	2014	2020	2020	2016	2019	2014	2016	2019	2014	2013
19	2021	2021	2015	2018	2016	2018	2016	2016	2016	2015	2015	2016
20	2021	2021	2014	2019	2015	2015	2016	2016	2014	2014	2015	2019
21	2021	2016	2016	2019	2015	2018	2019	2018	2019	2019	2015	2019
22	2021	2016	2016	2014	2014	2018	2016	2014	2016	2018	2013	2013
23	2021	2020	2015	2019	2020	2016	2016	2016	2015	2019	2017	2015
24	2017	2020	2015	2013	2017	2014	2014	2013	2013	2019	2014	2019
25	2016	2014	2014	2017	2014	2016	2013	2013	2016	2017	2014	2013
26	2020	2014	2014	2019	2020	2017	2019	2019	2014	2014	2014	2019
27	2016	2014	2016	2019	2015	2018	2019	2016	2015	2016	2019	2015
28	2014	2020	2014	2018	2015	2018	2016	2013	2013	2013	2017	2018
29	2014	2021	2021	2019	2014	2018	2019	2021	2019	2014	2019	2018
30	2021	2014	2014	2014	2017	2015	2016	2016	2014	2014	2013	2013

3.2. Evaluation of the generated TMYs

In this section, we aim to provide evidence of the correctness of the TMYs derived. One way to do this was by comparing the percentage difference of the TMYs from the corresponding AMYs on annual basis (e.g., [12,33]). This normalized difference is given by the following expression:

$$\Delta x_p = 100 \cdot \frac{\overline{x_{p,TMY,st}} - \overline{x_{p,2013-2021,st}}}{\overline{x_{p,2013-2021,st}}} \quad (5)$$

where x_p is the value of a parameter p at site st, $\overline{x_{p,TMY,st}}$ is the annual mean TMY value at st, and $\overline{x_{p,2013-2021,st}}$ is its annual mean AMY value for the same site. Figure 3 gives the distribution of the errors Δp across all thirty sites. It is seen that the errors do not exceed ±10% maximum, except for WS_{ave} at the sites of Al Baha (#2) and Shaqra (#25), which show great deviations from the annual ΔWS_{ave} values of the other sites. This occurs because these two stations have many 2–3 hourly missing WS values, which give, after the linear interpolation, quite different TMY and AMY values. Nevertheless, the overall results compare quite well with similar results for Greece [33] with errors in the range –6% to +8% for H_{bh}, ±6% for H_g, –4% to +2.5% for T, ±4% for RH, and ±15% for WS_{ave}.

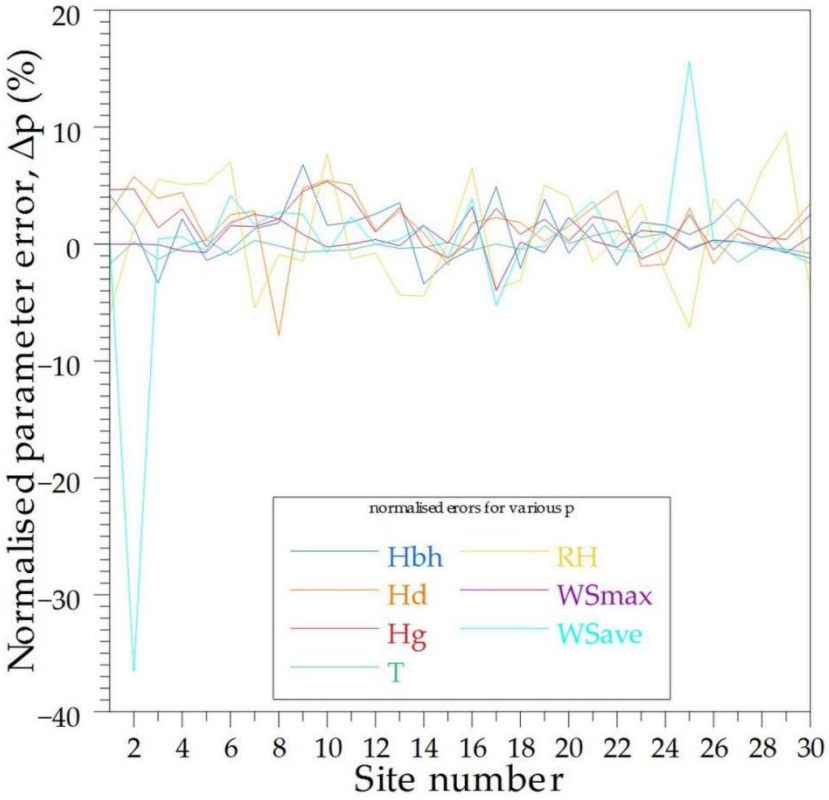

Figure 3. Variation of the Δp errors (in %) for the seven parameters considered in this across the thirty sites in Saudi Arabia. The errors are based on the annual averages of the parameters calculated in the period 2013–2021 from Eq. (5). The numbers of the sites correspond to those in Table 1, column 1.

To re-affirm the above result statistically, the Spearman correlation coefficient r was calculated for all parameter pairs TMY-AMY at all sites. Table 5 shows r's computed from the annual average values of TMY-AMY pairs across all sites in Saudi Arabia and within the period 2013–2021. It is seen that almost all are very close to 1, which means a consistency in the co-variation of both time series in the pair. The deviation in WS_{ave} for sites #2 and 25 shown in Figure 3 is confirmed by the low r values in Table 5. The missing $r_{WSmax,TMY-AMY}$ values for sites #1, 2, 9, and 11 are due to the same WS_{max} values in TMY and AMY series.

Table 5. The Spearman correlation coefficient $r_{p,pair}$ for all parameters p in the TMY-AMY pairs (i.e., $r_{p,TMY-AMY}$) at the thirty sites in Saudi Arabia and over the period 2013–2021. TMY = Typical Meteorological Year; AMY = Average Meteorological Year.

Site #	$r_{Hbh,TMY-AMY}$	$r_{Hd,TMY-AMY}$	$r_{Hg,TMY-AMY}$	$r_{T,TMY-AMY}$	$r_{RH,TMY-AMY}$	$r_{WSmax,TMY-AMY}$	$r_{WSave,TMY-AMY}$
1	0.95	0.80	0.88	1.00	0.82		0.74
2	0.95	0.80	0.91	0.99	0.95		0.21
3	0.97	0.92	0.94	1.00	0.96	0.93	0.88
4	0.99	0.95	0.91	0.99	0.93	0.94	0.85
5	0.92	0.97	0.95	1.00	0.99	0.99	0.96
6	0.83	0.97	0.97	1.00	0.98	0.91	0.93
7	0.97	0.93	0.97	1.00	0.96	0.97	0.93
8	0.96	0.58	0.94	1.00	0.73	0.96	0.96
9	0.86	0.90	0.97	0.99	0.98		0.50
10	0.88	0.98	0.98	1.00	0.97	0.97	0.88
11	0.82	0.89	0.95	1.00	0.96		0.88
12	0.91	0.95	0.97	0.99	0.95	0.90	0.88
13	0.98	0.96	0.98	1.00	0.99	0.96	0.94
14	0.90	0.97	0.97	1.00	0.98	0.94	0.93
15	0.94	0.88	0.93	1.00	0.75	0.97	0.88
16	0.95	0.92	0.94	1.00	0.98	0.96	0.93
17	0.88	0.89	0.95	1.00	0.97	0.93	0.88
18	0.94	0.97	0.97	0.99	0.97	0.95	0.89
19	0.97	0.85	0.96	0.97	0.97	0.97	0.93
20	0.98	0.88	0.78	1.00	0.96	0.99	0.97
21	0.86	0.83	0.95	0.99	0.93	0.69	0.86
22	0.98	0.96	0.99	1.00	0.97	0.94	0.92
23	0.96	0.90	0.94	1.00	0.97	0.87	0.96
24	0.97	0.92	0.93	1.00	0.97	0.97	0.96
25	0.96	0.91	0.95	1.00	0.98	0.96	0.63
26	0.97	0.92	0.81	0.99	0.96	0.95	0.91
27	0.87	0.92	0.94	0.99	0.98	0.98	0.97
28	0.92	0.84	0.92	1.00	0.95	0.99	0.98
29	0.93	0.98	0.99	0.99	0.98	0.98	0.98
30	0.98	0.77	0.82	1.00	0.94	0.89	0.90

3.3. Distribution of errors and correlation coefficients in the solar energy zones

Figure 1 shows the distribution of the thirty sites in the three solar energy zones as defined in [47]. In this section, we aim to show the distribution of the normalized errors across the three SEZs of Saudi Arabia. This was done because the generated TMYs were oriented toward solar energy applications. Figure 4 reproduces Figure 3 by taking into account the location of each site in one of the three SEZs. It is seen that the errors do not depend on the location of the site (and therefore on the SEZ) and are aggregated within a band of maximum width ±10% around the 0% value. This indicates an unbiased

result between the MTYs and AMYs, no matter which SEZ is preferred for the solar energy application. As in Figure 3, sites #2 and #25 show discrepancies with respect to the other sites, as their errors lie outside the ±10% band for WS_{ave}.

Figure 5 is a graphical presentation of the r values shown in Table 5 after filtering them according to the SEZ they belong to. It is interesting to see that lower r's ($r < 0.8$) exist for the WS_{ave}, H_d, and RH parameters (as expected from Table 5) and occur in SEZ-A and SEZ-B.

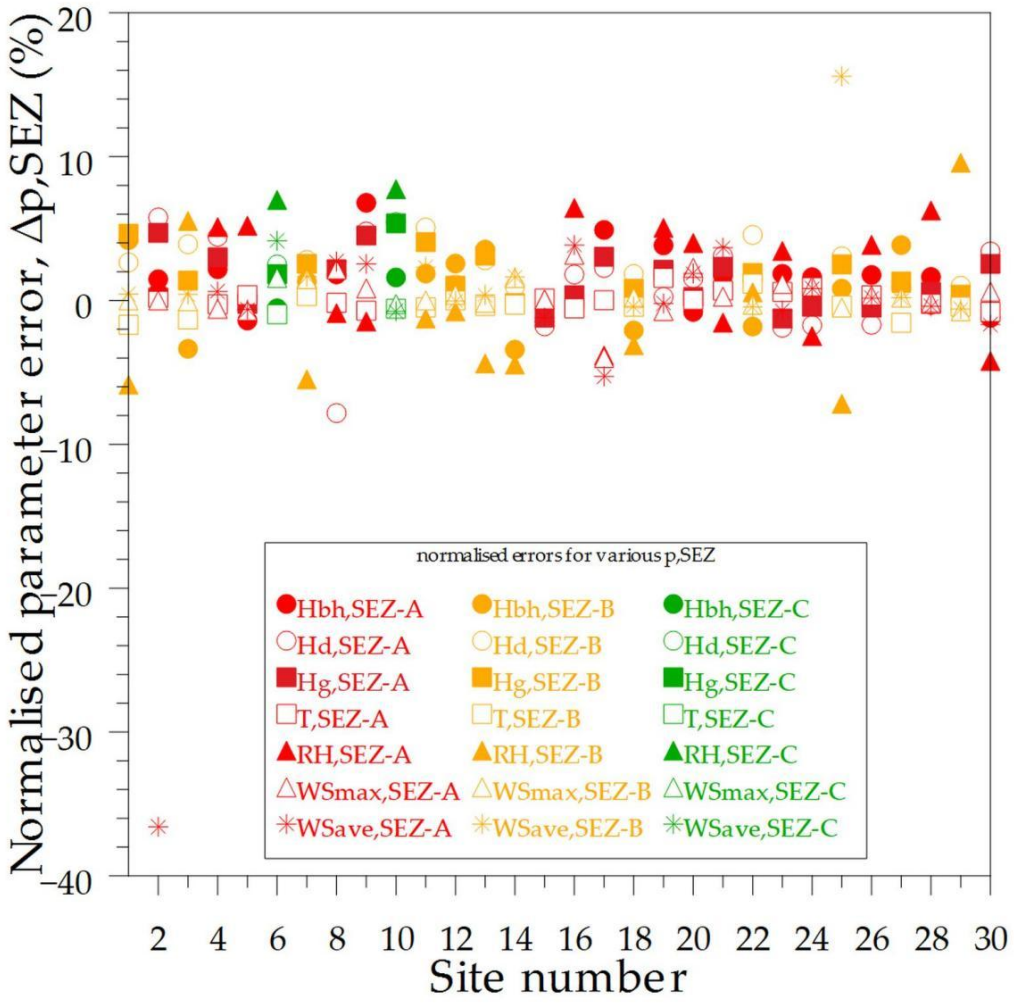

Figure 4. Variation of the Δp errors (in %) for the seven parameters considered in this study across the thirty sites in Saudi Arabia. The errors are based on the annual averages of the parameters calculated in the period 2013–2021 from Eq. (5) and are separated in the solar energy zones (SEZ) they belong to. The numbers of the sites correspond to those in Table 1, column 1.

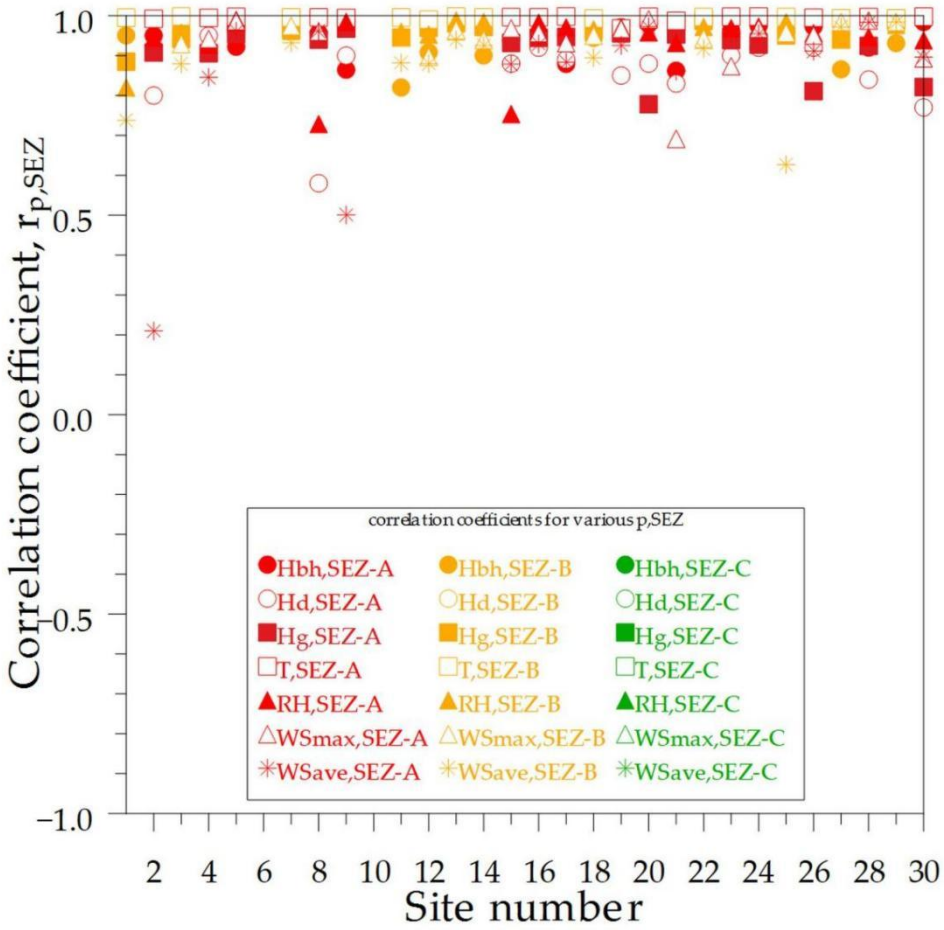

Figure 5. Variation of the Spearman correlation coefficients $r_{p,SEZ}$ for the seven parameters considered in this study across the thirty sites in Saudi Arabia. The correlation coefficients are based on the annual averages of the parameters calculated in the period 2013–2021 and are separated in the solar energy zones (SEZ) they belong to. The numbers of the sites correspond to those in Table 1, column 1.

3.4. Statistical representativeness of the TMYs

An additional step in evaluating the TMYs against the 9-year period averages involved the use of statistical metrics, including the mean, μ, standard deviation, σ, and the Shapiro–Wilk test, SW-test. These indices were selected to assess deviations of the monthly mean values of normalized parameter errors from the expected normal distribution.

In this analysis, the sites within each of the three solar energy zones (SEZs) were treated as ensembles, each comprising the monthly mean values of the respective parameters. Tables 6, 7, and 8 present the statistical indices for SEZ-A, SEZ-B, and SEZ-C, respectively.

It is important to note that a normal distribution curve-defined by its probability density function or PDF assumes μ = 0 and σ = 1. In these Tables, each parameter is denoted by q to avoid confusion with the statistical metric p-value. For the PDF of q to closely resemble a normal distribution, the SW-test value should approach 1, and the associated p-value should exceed 0.05, assuming a 95%

confidence interval. This condition forms the basis of the null hypothesis for normality in the distribution of normalized errors. In essence, the SW-test evaluates how well the parameters' data conforms to a normal distribution.

For SEZ-A (Table 6), the SW-tests and p-values indicate quasi- or pseudo-normality for most Δq parameters, with the exception of ΔH_d, which satisfies both criteria (SW-test ≈ 1 and p-value > 0.05). In SEZ-B (Table 7), all parameters exhibit similar behavior, except for ΔRH. In SEZ-C (Table 8), all Δq parameters demonstrate normality, except for ΔT, which shows pseudo-normal characteristics.

In conclusion, the selected TMYs are representative of the climatic conditions across the evaluated sites. However, a future study incorporating data spanning more than 10 years is recommended to reinforce and validate these findings.

Table 6. Statistical metrics for the normalized parameter errors Δq for sixteen sites in the solar energy zone-A (SEZ-A). The metrics refer to the monthly mean values of parameters q in the period 2013–2021. The means μ and standard deviations σ of Δq's are also provided as additional information. All values are rounded to the second digit. A 95% confidence interval is assumed. SW = Shapiro-Wilk.

Δq	μ	σ	SW-test	p-value
ΔH_{bh}	1.41	7.73	0.98	0.03
ΔH_d	0.66	9.14	1.00	0.69
ΔH_g	1.29	5.38	0.99	0.12
ΔT	0.24	2.92	0.73	<0.001
ΔRH	1.55	10.80	0.99	0.04
ΔWS_{max}	0.35	3.95	0.94	<0.001
ΔWS_{ave}	–1.45	15.11	0.56	<0.001

Table 7. Statistical metrics for the normalized parameter errors Δq for twelve sites in the solar energy zone-B (SEZ-B). The metrics refer to the monthly mean values of the parameters q in the period 2013–2021. The means μ and standard deviations σ of Δq's are also provided as additional information. All values are rounded to the second digit. A 95% confidence interval is assumed. SW = Shapiro-Wilk.

Δq	μ	σ	SW-test	p-value
ΔH_{bh}	0.57	7.36	0.97	0.002
ΔH_d	2.69	8.05	0.96	0.001
ΔH_g	2.04	5.19	0.95	< 0.001
ΔT	–0.67	4.21	0.92	< 0.001
ΔRH	–1.40	12.26	0.99	0.46
ΔWS_{max}	0.15	3.20	0.98	0.08
ΔWS_{ave}	1.77	6.82	0.90	< 0.001

Table 8. Statistical metrics for the normalized parameter errors Δq for two sites in the solar energy zone-C (SEZ-C). The metrics refer to the monthly mean values of the parameters q in the period 2013–2021. The means μ and standard deviations σ of Δq's are also provided as additional information. All values are rounded to the second digit. A 95% confidence interval is assumed. SW = Shapiro-Wilk.

Δq	μ	σ	SW-test	p-value
ΔH_{bh}	0.41	9.40	0.98	0.73
ΔH_{d}	4.22	5.03	0.95	0.22
ΔH_{g}	3.73	4.99	0.94	0.13
ΔT	–0.93	4.49	0.85	0.003
ΔRH	6.94	10.36	0.97	0.67
ΔWS_{max}	0.86	4.58	0.97	0.61
ΔWS_{ave}	1.91	5.82	0.99	0.96

4. Conclusions

In this study, we addressed the absence of TMYs for Saudi Arabia by deriving them for the first time at thirty sites distributed across most regions of the country. Our primary objective was to generate TMYs tailored for solar energy applications.

To achieve this, hourly mean data for seven parameters (i.e., global, direct, and diffuse horizontal irradiance; air temperature; relative humidity; and maximum and mean wind speed) were retrieved from the RMMM program, covering a 9-year period (2013–2021). The datasets were rigorously checked for missing and outlier values. Instances of 2–3 consecutive missing hourly values within a given month were replaced with the corresponding monthly averages from the same month in other years. Outlier values were excluded from the datasets. This process resulted in a clean and consistent dataset for each site.

The modified Sandia National Laboratories method, based on the Finkelstein–Schafer (FS) statistical procedure, was applied to each site's dataset. Specific weighting factors were assigned to the seven parameters to guide the FS process. This methodology identified the twelve statistically representative TMMs for each site, which were then compiled to construct the corresponding TMY.

To evaluate the accuracy of the generated TMYs, AMY values were computed for each site and compared with the respective TMYs. The normalized differences between each TMY and its corresponding AMY (calculated as the difference divided by the AMY) were within ±10% for all sites, indicating a high level of agreement. Additionally, the Spearman correlation coefficient between the TMY and AMY time series exceeded 0.8 in most cases, further confirming the reliability of the generated TMYs.

Researchers identified three solar energy zones (SEZs) in Saudi Arabia, which were optimized for maximum solar energy capture on tilted surfaces. These SEZs were incorporated into our analysis. Normalized differences and correlation coefficients were calculated for all sites within each SEZ, and the results were consistent with the findings from the individual site analyses.

To further assess the representativeness of the TMYs, the Shapiro–Wilk test (SW-test) was applied to the normalized differences of all seven parameters within each SEZ, treating all sites in a given SEZ as a collective ensemble. For nearly all parameters and SEZs, the SW-test supported the

alternative hypothesis (H_1), indicating statistically-significant normality of the normalized differences at the 95% confidence level.

The generation of TMY datasets for the thirty stations included in the national solar-radiation and meteorological monitoring network established under Saudi Vision 2030 provides a practical foundation for a wide range of energy-system applications. These datasets offer high-quality, long-term, and spatially-resolved climatic information that can be directly used for the design, optimization, and performance assessment of photovoltaic (PV) systems, solar-thermal installations, and hybrid renewable-energy systems across the Kingdom. Beyond system design, the TMYs support techno-economic feasibility studies, grid-integration analyses, and long-term energy-yield forecasting, all of which are essential for large-scale solar deployment under the Vision 2030 renewable-energy targets. The availability of harmonized TMYs for all thirty sites also enables consistent benchmarking of solar-resource variability, facilitates the development of region-specific building-energy simulations, and provides standardized climatic inputs for academic research, engineering consultancies, and governmental planning agencies. In this way, the datasets produced in this study are not only a scientific contribution but also a practical tool supporting national energy-transition strategies and infrastructure development.

In conclusion, the generated TMYs are representative of the climatic conditions across Saudi Arabia. However, to enhance statistical robustness and ensure broader applicability, future efforts should aim to derive TMYs from datasets spanning more than 10 years with higher data completeness.

Author contributions

Conceptualisation: AF, HDK; Data curation: AF, HDK; Methodology: HDK; Formal analysis: AF, HDK; Resources: AF, AL; Funding acquisition: AF; Writing-original draft: HDK; Writing-review & editing: AF, HDK, AL, KAK.

Use of AI tools declaration

The authors declare they have not used AI tools in the creation of this article.

Acknowledgments

The authors would like to acknowledge the support of the Research Center for Aviation and Space Exploration provided through the Deanship of Research (DR) at the King Fahd University of Petroleum and Minerals (KFUPM) for funding this work through the project INAE2303. They are also thankful to the RMMM programme for providing the data for the 30 sites. Finally, the authors would like to thank the anonymous Reviewers for their comments/suggestions that have helped in the improvement of the manuscript's quality.

Conflict of interest

The authors declare no conflicts of interest.

References

1. Hall IJ, Prairie RR, Anderson HE, et al. (1978) *Generation of a Typical Meteorological Year*, United States, 669–671.
2. Crow LW (1970) *Summary Description of Typical Weather Data*, Chicago Midway Airport.
3. Crow LW (1981) Development of Hourly Data for Weather Year for Energy Calculations (WYEC). *ASHRAE J* 23: 37–40.
4. Crow LW (1983) Development of Hourly Data for Weather Year for Energy Calculations (WYEC), Including Solar Data, at 29 Weather Stations throughout the US and 5 Stations in Canada. Available from: https://www.aivc.org/sites/default/files/members_area/medias/pdf/Airbase/airbase_00844.pdf.
5. Perez R, Ineichen P, Seals R, et al. (1990) Modeling Daylight Availability and Irradiance Components from Direct and Global Irradiance. *Sol Energy* 44: 271–289. https://doi.org/10.1016/0038-092X(90)90055-H
6. Klein SA, Beckman WA, Duffie JA (1976) A Design Procedure for Solar Heating Systems. *Sol Energy* 18: 113–127. https://doi.org/10.1016/0038-092X(76)90044-X
7. Benseman RF, Cook FW (1969) Solar Radiation in New Zealand—the Standard Year and Radiation on Inclined Slopes. *NZJ Sci* 12: 696.
8. Wilcox S, Marion W (2008) Users Manual for TMY3 Data Sets; NREL/TP-581-43156; National Renewable Energy Laboratory. Available from: https://www.doe2.com/Download/Weather/TMY3/Users_Manual_for_TMY3_Data_Sets.pdf.
9. Realpe AM, Vernay C, Pitaval S, et al. (2016) Benchmarking of Five Typical Meteorological Year Datasets Dedicated to Concentrated-PV Systems. *Energy Procedia* 97: 108–115. https://doi.org/10.1016/j.egypro.2016.10.031
10. Wald L, Thomas C, Cousin S, et al. (2011) The Project ENDORSE: Exploiting EO Data to Develop Pre-Market Services in Renewable Energy. 25th EnviroInfo Conference "Environmental Informatics", Shaker Verlag, Aachen, Germany, 1: 549–556.
11. Cebecauer T, Suri M (2015) Typical Meteorological Year Data: SolarGIS Approach. *Energy Procedia* 69: 1958–1969. https://doi.org/10.1016/j.egypro.2015.03.195
12. Ohunakin OS, Adaramola MS, Oyewola OM, et al. (2013) Generation of a Typical Meteorological Year for North-East, Nigeria. *Appl Energy* 112: 152–159. https://doi.org/10.1016/j.apenergy.2013.05.072
13. Ohunakin OS, Adaramola MS, Oyewola OM, et al. (2014) A Typical Meteorological Year Generation Based on Nasa Satellite Imagery (GEOS-I) for Sokoto, Nigeria. *Int J Photoenergy* 2014: 468562. https://doi.org/10.1155/2014/468562
14. Oko COC, Ogoloma OB (2011) Generation of a Typical Meteorological Year for Port Harcourt Zone. *J Eng Sci Technol* 6: 216–226.
15. Chan ALS, Chow TT, Fong SKF, et al. (2006) Generation of a Typical Meteorological Year for Hong Kong. *Energy Convers Manage* 47: 87–96. https://doi.org/10.1016/j.enconman.2005.02.010
16. Chan ALS (2011) Developing a Modified Typical Meteorological Year Weather File for Hong Kong Taking into Account the Urban Heat Island Effect. *Build Environ* 46: 2434–2441. https://doi.org/10.1016/j.buildenv.2011.04.038

17. Chan ALS (2016) Generation of Typical Meteorological Years Using Genetic Algorithm for Different Energy Systems. *Renewable Energy* 90: 1–13. https://doi.org/10.1016/j.renene.2015.12.052
18. Ebrahimpour A, Maerefat M (2010) A Method for Generation of Typical Meteorological Year. *Energy Convers Manage* 51: 410–417. https://doi.org/10.1016/j.enconman.2009.10.002
19. Zang H, Xu Q, Du P (2012) Ichiyanagi, K. A Modified Method to Generate Typical Meteorological Years from the Long-Term Weather Database. *Int J Photoenergy* 2012. https://doi.org/10.1155/2012/538279
20. Yang L, Lam JC, Liu J, et al. (2008) Building Energy Simulation Using Multi-Years and Typical Meteorological Years in Different Climates. *Energy Convers Manage* 49: 113–124. https://doi.org/10.1016/j.enconman.2007.05.004
21. Jiang Y (2010) Generation of Typical Meteorological Year for Different Climates of China. *Energy* 35: 1946–1953. https://doi.org/10.1016/j.energy.2010.01.009
22. Sun J, Li Z, Xiao F (2017) Analysis of Typical Meteorological Year Selection for Energy Simulation of Building with Daylight Utilization. *Procedia Eng* 205: 3080–3087. https://doi.org/10.1016/j.proeng.2017.10.303
23. Kim S, Zirkelbach D, Künzel HM, et al. (2017) Development of Test Reference Year Using ISO 15927-4 and the Influence of Climatic Parameters on Building Energy Performance. *Build Environ* 114: 374–386. https://doi.org/10.1016/j.buildenv.2016.12.037
24. Murphy S (2017) The Construction of a Modified Typical Meteorological Year for Photovoltaic Modeling in India. *Renewable Energy* 111: 447–454. https://doi.org/10.1016/j.renene.2017.04.033
25. Farah S, Saman W, Boland J (2018) Development of Robust Meteorological Year Weather Data. Renew. Energy 118: 343–350. https://doi.org/10.1016/j.renene.2017.11.033
26. Janjai S, Deeyai P (2009) Comparison of Methods for Generating Typical Meteorological Year Using Meteorological Data from a Tropical Environment. *Appl Energy* 86: 528–537. https://doi.org/10.1016/j.apenergy.2008.08.008
27. Markou MT, Kambezidis HD, Bartzokas A, et al. (2007) Generation of Daylight Reference Years for Two European Cities with Different Climate: Athens, Greece and Bratislava, Slovakia. *Atmos Res* 86: 315–329. https://doi.org/10.1016/j.atmosres.2007.07.001
28. Sepúlveda C (2014) Comparison of Methodologies for TMY Generation Using15 Years Data for Chillan, Chile. *IOSR J Eng* 4: 25–28.
29. Pissimanis D, Karras G, Notaridou V, et al. (1988) The Generation of a "Typical Meteorological Year" for the City of Athens. *Sol Energy* 40: 405–411. https://doi.org/10.1016/0038-092X(88)90095-3
30. Argiriou A, Lykoudis S, Kontoyiannidis S, et al. (1999) Comparison of Methodologies for TMY Generation Using 20 Years Data for Athens, Greece. *Sol Energy* 66: 33–45. https://doi.org/10.1016/S0038-092X(99)00012-2
31. Petrakis M, Kambezidis HDD, Lykoudis S, et al. (1998) Generation of a "Typical Meteorological Year" for Nicosia, Cyprus. *Renewable Energy* 13: 381–388. https://doi.org/10.1016/S0960-1481(98)00014-7
32. Kalogirou SA (2003) The Energy Subsidisation Policies of Cyprus and Their Effect on Renewable Energy Systems Economics. *Renewable Energy* 28: 1711–1728. https://doi.org/10.1016/S0960-1481(03)00062-4

33. Kambezidis HD, Psiloglou BE, Kaskaoutis DG, et al. (2020) Generation of Typical Meteorological Years for 33 Locations in Greece: Adaptation to the Needs of Various Applications. *Theor Appl Climatol* 141: 1313–1330. https://doi.org/10.1007/s00704-020-03264-7
34. Skeiker K (2004) Generation of a Typical Meteorological Year for Damascus Zone Using the Filkenstein-Schafer Statistical Method. *Energy Convers Manage* 45: 99–112. https://doi.org/10.1016/S0196-8904(03)00106-7
35. Skeiker K (2008) Ghani, B. A. Advanced Software Tool for the Creation of a Typical Meteorological Year. *Energy Convers Manage* 49: 2581–2587. https://doi.org/10.1016/j.enconman.2008.05.013
36. Pusat S, Ekmekçi I, Akkoyunlu MT (2015) Generation of Typical Meteorological Year for Different Climates of Turkey. *Renewable Energy* 75: 144–151. https://doi.org/10.1016/j.renene.2014.09.039
37. Yilmaz S, Ekmekci I (2017) The Generation of Typical Meteorological Year and Climatic Database of Turkey for the Energy Analysis of Buildings. *J Environ Sci Eng* A 6. https://doi.org/10.17265/2162-5298/2017.07.005
38. Bulut H (2004) Typical Solar Radiation Year for Southeastern Anatolia. *Renewable Energy* 29: 1477–1488. https://doi.org/10.1016/j.renene.2004.01.004
39. Mosalam Shaltout MA, Tadros MTY (1994) Typical Solar Radiation Year for Egypt. *Renewable Energy* 4: 387–393. https://doi.org/10.1016/0960-1481(94)90045-0
40. Faiman PD, Feuermann D, Ibbetson P, et al. (2004). The Negev radiation survey. *J Sol Energy Eng* 126: 906–914. https://doi.org/10.1115/1.1756138
41. Fernández MD, López JC, Baeza E, et al. (2015) Generation and Evaluation of Typical Meteorological Year Datasets for Greenhouse and External Conditions on the Mediterranean Coast. *Int J Biometeorol* 59: 1067–1081. https://doi.org/10.1007/s00484-014-0920-7
42. Petrie WR, McClintock M (1978) Determining typical weather for use in solar energy simulations. *Sol Energy* 21: 55–59. https://doi.org/10.1016/0038-092X(78)90116-0
43. Feuermann D, Gordon JM, Zarmi YA (1985) A typical meteorological day (TMD) approach for predicting the long-term performance of solar energy systems. *Sol Energy* 35: 63–69. https://doi.org/10.1016/0038-092X(85)90037-4
44. Gueymard CA (2012) Clear-Sky Irradiance Predictions for Solar Resource Mapping and Large-Scale Applications: Improved Validation Methodology and Detailed Performance Analysis of 18 Broadband Radiative Models. *Sol Energy* 86: 2145–2169. https://doi.org/10.1016/j.solener.2011.11.011
45. Roth M (2017) Updating the ASHRAE Climate Design Data for 2017. *ASHRAE Trans* 123: 80–90.
46. Renné DS (2016) Resource Assessment and Site Selection for Solar Heating and Cooling Systems. *Advances in Solar Heating and Cooling*, 13–41. https://doi.org/10.1016/B978-0-08-100301-5.00002-3
47. Farahat A, Kambezidis HD, Almazroui M, et al. (2021) Solar Potential in Saudi Arabia for Southward-Inclined Flat-Plate Surfaces. *Appl Sci* 11: 4101. https://doi.org/10.3390/app11094101
48. Farahat A, Kambezidis HD, Almazroui M, et al. (2021) Solar Potential in Saudi Arabia for Inclined Flat-Plate Surfaces of Constant Tilt Tracking the Sun. *Appl Sci* 11: 7105. https://doi.org/10.3390/app11157105

49. King Abdullah City for Atomic and Renewable Energy. Renewable Resource Atlas V2: High-Resolution Solar and Wind Measurement Dataset for the Kingdom of Saudi Arabia, 2025. Available from: https://reatlas.energy.gov.sa.
50. Shaik DS, Kant Y, Mitra D, et al. (2019) Impact of Biomass Burning on Regional Aerosol Optical Properties: A Case Study over Northern India. *J Environ Manage* 244: 328–343. https://doi.org/10.1016/j.jenvman.2019.04.025
51. Gueymard CA (2018) A Reevaluation of the Solar Constant Based on a 42-Year Total Solar Irradiance Time Series and a Reconciliation of Spaceborne Observations. *Sol Energy* 168: 2–9. https://doi.org/10.1016/j.solener.2018.04.001
52. Andersen B, Eidorff S, Hallgreen L, et al. (1982) *Danish Test Reference Year TRY: Meteorological Data for HVAC and Energy*, Technical University of Denmark, Department of Civil Engineering. Denmark, 1–64.
53. Festa R, Ratto CF (1993) Proposal of a numerical procedure to select Reference Years. *Sol Energy* 50: 9–17. https://doi.org/10.1016/0038-092X(93)90003-7
54. Li DHW, Yang L, Lam JC (2012) Impact of Climate Change on Energy Use in the Built Environment in Different Climate Zones—A Review. *Energy* 42: 103–112. https://doi.org/10.1016/j.energy.2012.03.044
55. Liu H, Liang J, Liu Y, et al. (2023) A Review of Data-Driven Building Energy Prediction. *Buildings* 13: 532. https://doi.org/10.3390/buildings13020532
56. Paravantis JA, Malefaki S, Nikolakopoulos P, et al. (2025) Statistical and Machine Learning Approaches for Energy Efficient Buildings. *Energy Build* 330: 115309. https://doi.org/10.1016/j.enbuild.2025.115309
57. Badescu V (2008) *Modeling Solar Radiation at the Earth's Surface*, Springer-Verlag Berlin Heidelberg, 517.
58. Gueymard CA, Ruiz-Arias JA (2016) Extensive Worldwide Validation and Climate Sensitivity Analysis of Direct Irradiance Predictions from 1-Min Global Irradiance. *Sol Energy* 128: 1–30. https://doi.org/10.1016/j.solener.2015.10.010
59. Kalogirou SA (2003) Generation of Typical Meteorological Year (TMY-2) for Nicosia, Cyprus. *Renewable Energy* 28: 2317–2334. https://doi.org/10.1016/S0960-1481(03)00131-9
60. Lam JC, Hui SCM, Chan ALS (1996) A Statistical Approach to the Development of a Typical Meteorological Year for Hong Kong. *Archit Sci Rev* 39: 201–209. https://doi.org/10.1080/00038628.1996.9696818

Appendix A: Dataset completeness

The appendix gives additional information on the process of data cleaning. Table A1 contains several columns. Column 1 is the site number. The other columns refer to the seven parameters considered in the study (as averages). Column A shows the annual mean number of observations for the seven parameters at each site. Column B is column A expressed in percent, i.e., $B = 100 \cdot A/(24 \cdot 30 \cdot 365)$, considering an average year to have 24 hours per day, 30 days per month, and 365 days per year. Column C shows the number of months that correspond to the total average number of hourly data per site. Column D has the same meaning as column B but for the months. Column E gives the total number of years that comprise the original dataset of each site. Column F has the same meaning as column B but for the years. Column G shows the number of (cleaned) hourly data in the

dataset of the site; this new dataset was used in the FS statistical procedure. Column H has the same meaning as column B, but the coverage of the clean hourly data is 100% because the new dataset is considered complete for the FS analysis. Columns I, J and K, L have the same meaning as columns G and H, for the months and years, respectively.

Table A1. Completeness of original (real) and quality-checked (clean) data for all seven parameters per site in the period 2013–2021. All numbers in columns B-L are averages across the seven parameters per site and have been rounded to the second decimal digit, wherever necessary. The derived TMYs consist of twelve TMMs selected from the remaining months in the clean databases; they all consist of hourly values. The number of hourly values in each TMY is 7470 in total (February is taken to have 28 days).

Site #	No of real hourly data (A)	% of real hourly data (B)	No of real months (C)	% of real months (D)	No of real years (E)	% of real years (F)	No of clean hourly data (G)	% of clean hourly data (H)	No of clean months (I)	% of clean months (J)	No of clean years (K)	% of clean years (L)
1	62905	23.94	87.37	80.90	7.28	80.90	43698.71	100	60.69	100	5.06	100
2	40178	15.29	55.80	51.67	4.65	51.67	29187.43	100	40.54	100	3.38	100
3	64657	24.60	89.80	83.15	7.48	83.15	49159.57	100	68.28	100	5.69	100
4	60939	23.19	84.64	78.37	7.05	78.37	48041.00	100	66.72	100	5.56	100
5	63649	24.22	88.40	81.85	7.37	81.85	50776.86	100	70.52	100	5.88	100
6	49680	18.90	69.00	63.89	5.75	63.89	38886.57	100	54.01	100	4.50	100
7	63793	24.27	88.60	82.04	7.38	82.04	46784.00	100	64.98	100	5.41	100
8	62738	23.87	87.14	80.68	7.26	80.68	48475.86	100	67.33	100	5.61	100
9	62953	23.95	87.43	80.96	7.29	80.96	41004.29	100	56.95	100	4.75	100
10	44568	16.96	61.90	57.31	5.16	57.31	35064.29	100	48.70	100	4.06	100
11	65881	25.07	91.50	84.72	7.63	84.72	43808.43	100	60.85	100	5.07	100
12	64441	24.52	89.50	82.87	7.46	82.87	51467.29	100	71.48	100	5.96	100
13	61945	23.57	86.03	79.66	7.17	79.66	49208.86	100	68.35	100	5.70	100
14	51144	19.46	71.03	65.77	5.92	65.77	40225.43	100	55.87	100	4.66	100
15	48938	18.62	67.97	62.93	5.66	62.93	34966.71	100	48.56	100	4.05	100
16	65521	24.93	91.00	84.26	7.58	84.26	52258.14	100	72.58	100	6.05	100
17	46802	17.81	65.00	60.19	5.42	60.19	33544.29	100	46.59	100	3.88	100
18	64609	24.58	89.73	83.09	7.48	83.09	51158.14	100	71.05	100	5.92	100
19	33698	12.82	46.80	43.34	3.90	43.34	44056.86	100	61.19	100	5.10	100
20	61081	23.24	84.83	78.55	7.07	78.55	48622.00	100	67.53	100	5.63	100
21	61728	23.49	85.73	79.38	7.14	79.38	48704.43	100	67.65	100	5.64	100
22	65713	25.00	91.27	84.51	7.61	84.51	45309.29	100	62.93	100	5.24	100
23	47473	18.06	65.93	61.05	5.49	61.05	34366.71	100	47.73	100	3.98	100

Continued on next page

Site #	No of real hourly data (A)	% of real hourly data (B)	No of real months (C)	% of real months (D)	No of real years (E)	% of real years (F)	No of clean hourly data (G)	% of clean hourly data (H)	No of clean months (I)	% of clean months (J)	No of clean years (K)	% of clean years (L)
24	66241	25.21	92.00	85.19	7.67	85.19	52537.57	100	72.97	100	6.08	100
25	61753	23.50	85.77	79.41	7.15	79.41	47423.00	100	65.87	100	5.49	100
26	56977	21.68	79.13	73.27	6.59	73.27	45488.14	100	63.18	100	5.26	100
27	52707	20.06	73.20	67.78	6.10	67.78	38830.14	100	53.93	100	4.49	100
28	64921	24.70	90.17	83.49	7.51	83.49	44639.57	100	62.00	100	5.17	100
29	64489	24.54	89.57	82.93	7.46	82.93	51474.00	100	71.49	100	5.96	100
30	64297	24.47	89.30	82.69	7.44	82.69	44056.86	100	61.19	100	5.10	100

Appendix B: Comparison of WFs from this study and other studies

Table B1. Comparison of weighting factors, WF, from this study and other related works.

Parameter	WF from present TMY study	Typical WF ranges in earlier TMY studies	Representative references	Notes
H_{bh}	0.25	0.20–0.40	[1,29,30]	Highest weight in all solar-energy-focused TMY methods.
H_d	0.10	0.05–0.15	[30,60]	Secondary to beam irradiance; important for tilted-surface PV.
H_g	0.15	0.10–0.25	[1,60]	Often grouped with beam/diffuse in a combined solar-radiation weight.
T	0.15	0.10–0.20	[29,30]	Moderate influence on PV efficiency and thermal loads.
RH	0.10	0.05–0.15	[30,60]	Typically low weight; indirect effect on solar availability.
WS_{ave}	0.15	0.10–0.20	[1,30]	Relevant for convective cooling and ventilation.
WS_{max}	0.10	0.05–0.15	[29,30]	Used mainly for structural/operational considerations.
Total	1.00	-	-	All studies normalize weights to unity.