sustainability

MDPI

Article

Weather Data Compression for Laboratory Testing of Building Sustainability

Sara Fateri, Ljubomir Jankovic *, Grant Henshaw, Richard Fitton and William Swan

Energy House Labs, University of Salford, Manchester M5 4WT, UK; s.fateri1@salford.ac.uk (S.F.); g.p.henshaw@salford.ac.uk (G.H.); r.fitton@salford.ac.uk (R.F.); w.c.swan@salford.ac.uk (W.S.)
* Correspondence: l.jankovic@salford.ac.uk

Abstract

The research reported in this article is a byproduct of a project on experimental evaluation of the performance of a hempcrete test cell in an environmental chamber under controlled conditions. The work was conducted at Energy House 2.0 at the University of Salford, where two large environmental chambers can generate weather conditions with air temperature from −23.5 °C to +51 °C, and control relative humidity, solar radiation, wind, rain, and snow. However, synthesizing and varying the weather conditions over a long time is impractical and expensive: impractical because it dedicates the laboratory for a single test or a single project, and expensive because it requires significant resources, including the facility, energy and staff costs. For that reason, the need to have a weather data set that is representative of the whole year but using a fraction of annual hours has arisen. This article reports on a weather data compression method that creates a representation of annual weather data with twelve days of hourly data, one hourly data set per month. The Finkelstein–Schafer (FS) method was adapted to identify one representative day for each month by comparing hourly weather distributions. Applied to the Manchester TMYx dataset, the mean squared standardised errors of the daily means and totals of the selected days were 71–91% lower than the value expected for a randomly selected day, and their hourly profiles were closer to the monthly mean diurnal profile than the average day in 56 of 60 month–variable combinations. As many building science laboratories face similar time and cost constraints, this method could be useful for planning shorter weather-based tests of buildings, building components and materials, including sustainable materials and retrofit measures.

Keywords: weather data compression; representative weather data; environmental chamber testing; building performance testing; accelerated experimental testing; dynamic weather simulation; experimental methodology; climate simulation; full-scale building testing; FS method; Finkelstein–Schafer; sustainable building testing

Academic Editor: Papadaki Dimitra

Received: 23 October 2026
Revised: 17 November 2026
Accepted: 23 November 2026
Published: 26 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Large-scale environmental chambers can reproduce a wide range of weather conditions in controlled laboratory environments, allowing climatic variables such as air temperature, relative humidity, solar radiation, wind, and precipitation (rain and snow) to be independently controlled and repeatedly applied. This provides an effective platform for investigating the performance of buildings, building components, and construction materials under controlled weather conditions. Energy House 2.0 at the University of Salford

is an example of such a facility, allowing full-scale buildings to be tested under controlled and repeatable climate conditions [1]. To investigate how a building would perform under real weather conditions, the weather reproduced in the chamber should represent, as closely as possible, the climatic conditions and variations experienced at that location.

However, reproducing the full year's hourly weather sequence under laboratory conditions would require the facility to operate for approximately one year. Such a test would occupy a specialised laboratory facility for a prolonged period and require considerable resources in terms of facility operation, energy consumption, equipment, and staff time. Therefore, a shorter weather dataset is needed to represent the typical climate of a location within a shorter testing period. This leads to the question of which weather conditions should be used to represent the whole year. Selecting a few days at random is unlikely to give a reliable picture of the annual climate, because nothing guarantees that the chosen days reflect the distribution of conditions experienced throughout the year. A method is therefore needed to select a small number of days that best represent the weather conditions throughout the year.

The concept of reducing long-term weather data to a shorter representative dataset is not new and has been used extensively in building performance simulation. Typical Meteorological Years (TMYs), Test Reference Years (TRYs), and other similar representative weather files have been developed for this purpose, reducing long-term weather records covering many years to a single representative year [2–6]. The original motivation was largely practical: rather than simulating building performance using several decades of historical weather data, a much shorter dataset could be used to reduce computational time and resources while retaining the main climatic characteristics of a location [5,7–11]. A representative weather year provides an indication of typical building performance and a consistent basis for comparing different buildings, construction systems, or technologies under the same climatic conditions [2,3,5,9,10,12,13].

Several methods have been developed for weather data compression, and different approaches to selecting representative weather periods have been reported in the literature. These approaches use different statistical criteria and, more recently, have also included simulation-based, data-driven, and optimisation techniques [5–7,14–20]. Among the statistical approaches, approaches based on the Finkelstein–Schafer (FS) statistic have been among the most widely used for selecting representative weather periods and generating typical weather files [9,11,12,14,16,17]. The FS statistic was originally proposed by Finkelstein and Schafer [21] as a goodness-of-fit statistic, measuring the departure of the ordered sample values from a completely specified continuous distribution. It was later adapted in the Sandia method [13] to compare the cumulative distribution function (CDF) of a candidate month with the corresponding long-term CDF for the generation of TMY weather files.

A lower FS value indicates a closer match between the two distributions. When several weather variables are considered, their FS values can be weighted and combined to rank the candidate periods [13]. In the Sandia method, the best-ranked candidates are then screened using persistence criteria before the final selection [13].

Although the FS approach is well established and widely used, later studies have identified some limitations and proposed modifications to address them. The FS statistic is calculated separately for each weather variable, and the final selection depends on the weights assigned to these variables. These weights may vary between different methods and are not necessarily appropriate for all climates or applications [5,9,16,22,23]. In addition, a period that is statistically representative of the weather may not necessarily produce a representative building response [8,9,12,15]. The method is also intended to identify typical conditions and may therefore underrepresent extreme weather events [3,5,24]. Despite these limitations, the FS approach has several advantages. It compares the

distribution of weather variables rather than relying only on their mean values [5,16]. It can also consider several climatic variables through a weighted combination of their individual FS statistics [13,14], providing a relatively simple method for identifying representative weather periods without requiring a building simulation model [5,15]. These advantages, together with its established use in typical weather file generation, have contributed to its continued widespread use [9,11,12,14,16,25,26]. The FS statistic has also been applied in recent tools for generating future weather files under climate change scenarios [24,26].

While reducing long-term weather records to a single representative year provides a practical solution for computer-based building performance simulation, the same timescale is not practical for physical testing. In a simulation, a year of hourly data can be processed in a fraction of real time, whereas an environmental chamber must reproduce the weather in real time. Therefore, further compression of the weather data is needed to make such testing practical while retaining the main characteristics of the annual weather.

Shorter datasets based on representative days have also been explored. Vilasboas et al. [27] adapted the steps of the Sandia method to summarise 20 years of historical weather data into typical meteorological days, including one day for each month, with persistence assessed over consecutive hours and an additional hourly profile criterion, and evaluated them through the forecast electricity production of a solar thermal power plant. The FS statistic has also been used to identify reference days from 30 years of observed weather data for estimating the long-term energy consumption of buildings with reduced computational effort, with weights derived from the correlation of the weather variables with simulated building performance and days selected to span the distribution of weather conditions [28]. Other studies have selected representative days for building and renewable energy system simulation using data-driven and clustering approaches [29,30], and the FS statistic has been applied to hourly heating and cooling needs obtained from long-term building energy simulations to select typical months [15]. These studies evaluated the selected days through simulated system or building responses. In energy system modelling, temporal aggregation methods, such as k-means and k-medoids clustering, are widely used to reduce time series to a small number of representative days [31,32]. These methods differ in how representative periods are identified. K-means clustering may produce representative profiles that do not correspond to actual weather days, while k-medoids selects an observed day based on its similarity to other days. Unlike clustering of daily weather profiles, the FS method compares each candidate day with the hourly weather distribution of its corresponding month. It also considers the distribution of hourly values rather than their order throughout the day. Therefore, two days with the same set of hourly values but different hourly sequences will produce the same FS value. To the authors' knowledge, the use of representative days for environmental chamber testing has received little attention. This application requires each selected period to be an actual, continuous 24-h weather record that can be physically reproduced. Furthermore, a weather-based selection criterion allows representative days to be identified without relying on a building model, which may not be available for new materials or construction systems.

In this study, an approach based on the FS statistic was developed, drawing on the representative weather selection principles of the Sandia method, to reduce a TMY to 12 representative days, one for each calendar month. The adaptation differs technically from the original Sandia method in three respects. First, the Sandia method calculates FS statistics using distributions of daily weather indices across candidate and reference months. When the candidate period is a single day, each daily index provides only one value; therefore, hourly observations were used to construct the daily and monthly empirical distributions. Second, the empirical cumulative distribution function as originally defined

by Finkelstein and Schafer [21] was used instead of the plotting position of the Sandia method, and the weights assigned to the original daily indices were combined for each weather variable. Third, the persistence criteria used in the Sandia method to assess multiday weather sequences were not applied, as such sequences do not exist within a single candidate day. Each candidate day was compared with the hourly weather distributions of its corresponding calendar month. Unlike Vilasboas et al. [27], who selected typical meteorological days from a 20-year weather record using the FS statistic together with persistence and hourly profile criteria, the present study selects days from a typical year based only on hourly weather distributions. The selected days are evaluated using weather statistics, including means, hourly profiles, distributions, distribution tails, threshold exceedances and relationships between variables, as well as the sensitivity of the selection to the FS formulation, rather than the simulated response of an energy system. The selected days are complete, observed 24-h weather records that preserve the original hourly combinations and chronological order of weather variables within each day, although their inter-variable relationships may differ from those of the corresponding month. The method is demonstrated using the TMYx weather file for Manchester, UK, as a case study.

Previous studies have used different approaches to examine how well typical weather datasets represent long-term weather conditions. These approaches have included comparisons of mean and median values [7,9,11,13,14,17], frequency distributions [6,7,11–13], and monthly average and hourly weather profiles with long-term records [14,25,33]. Some studies have also used building or system simulations to compare the resulting performance with that obtained using long-term weather data [2,4,7–9,12,13,15,16,22,23,34]. ISO 15927-4 [6] also highlights the importance of preserving not only the mean values of meteorological variables, but also their frequency distributions and cross-correlations. In the present study, as the selected weather conditions have not yet been applied to a building model, the representativeness of the selected days was examined by comparing their statistical characteristics and hourly weather patterns with those of their corresponding months, together with the overall statistical characteristics of the compressed year and the full TMYx file.

2. Materials and Methods

This section first describes the Sandia method and the FS statistic, followed by its adaptation for representative-day selection and the assessment of the selected days.

2.1. Sandia Method and FS Statistic

The Sandia method selects a Typical Meteorological Month (TMM) for each calendar month from a long-term weather record. The selection is based on four weather variables: dry-bulb temperature, dew-point temperature, wind speed, and global horizontal solar radiation [13]. From these variables, 13 daily indices are calculated: the daily maximum, minimum, mean, and range of dry-bulb temperature, dew-point temperature, and wind speed, together with the daily total global horizontal solar radiation [13]. For each daily index, the values for each calendar month are ordered from lowest to highest, and CDFs are constructed for both the long-term data and each candidate month. In the Sandia method, the CDF is defined as [13]:

$$S_n(x) = \begin{Bmatrix} 0, & x < x_{(1)} \\ \dfrac{k-0.5}{n}, & x_{(k)} \le x \le x_{(k+1)} \\ 1, & x > x_{(n)} \end{Bmatrix} \quad (1)$$

where n is the number of observations and $x_{(k)}$ is the k-th ordered observation. The FS statistic is calculated for each weather index by averaging the absolute differences between the long-term and candidate-month CDFs. A lower FS value indicates that the candidate month has a distribution closer to the long-term distribution. The use of CDFs allows the overall distributions to be compared rather than relying only on the mean and standard deviation, which can be similar even when the distributions are different [13]. The FS statistic is expressed as follows:

$$FS = \frac{1}{n}\sum_{i=1}^{n} \delta_i \quad (2)$$

where n is the number of daily observations in the candidate month and δ_i is the absolute difference between the long-term and candidate-month CDFs at the i-th ordered observation [13,21]:

$$\delta_i = \left|F_{\text{long}}(x_{(i)}) - S_n(x_{(i)})\right| \quad (3)$$

The individual FS values are then weighted and summed to rank the candidate months.

In the original Sandia method, nine of the 13 daily indices are given non-zero weights, as shown in Table 1. Global horizontal solar radiation has the highest total weight, 12/24, while dry-bulb temperature, dew-point temperature, and wind speed each have a total weight of 4/24. The weighting allows greater importance to be given to weather indices that are more relevant to the intended application. Hall et al. noted that the choice of these weighting factors is "not clear cut" and depends on the intended application of the generated typical year [13].

Table 1. Weighting factors used in the original Sandia method [13].

Weather Variable	Daily Index	Weighting Factor
Dry-bulb temperature	Maximum	1/24
	Minimum	1/24
	Mean	2/24
Dew-point temperature	Maximum	1/24
	Minimum	1/24
	Mean	2/24
Wind speed	Maximum	2/24
	Mean	2/24
Global horizontal solar radiation	Daily total	12/24

The weighted sum for each candidate month is calculated as:

$$WFS = \sum_{i=1}^{M} w_i\, FS_i \quad (4)$$

where FS_i is the FS statistic for the i-th daily index, w_i is its weighting factor, and M is the number of daily indices.

The five candidate months with the lowest WFS values are retained for further assessment. The final selection also considers the persistence of mean daily dry-bulb temperature and daily total global solar radiation. This additional step is used because a candidate month can have a CDF close to the long-term distribution while still containing atypically long sequences of warm, cool, or low-radiation days. The final TMM is selected by considering the WFS value together with deviations in the monthly mean and median and the persistence structure [13].

2.2. Representative-Day Selection Method

The FS approach described in Section 2.1 was adapted to select one representative day for each calendar month.

The weather data used in this study were obtained from the TMYx file for Manchester Airport, England (WMO 033340; 53.354° N, 2.275° W; elevation 78.3 m), available from Climate.OneBuilding.org [35]. According to the file header, the TMYx file was generated using the NCEI ISD/ERA5 record for 1934–2023, with 56 years of data available within this period. The file contains 8760 hourly records and consists of twelve typical months, each selected from a different year, as listed in Table 2. The dataset contains dry-bulb temperature (°C), dew-point temperature (°C), relative humidity (%), global horizontal radiation (Wh/m^2), wind speed (m/s), precipitation, and other weather variables. The selected days therefore represent this TMYx file, a typical year composed of months selected from the period 1934–2023 (Table 2), rather than the long-term, present-day or future climate of Manchester. Tests intended to represent a different climatic period would require the selection to be applied to a weather file representing that period. The hourly global horizontal radiation in the EPW file represents the radiant energy received on a horizontal surface during the preceding hour, expressed in Wh/m^2 [36]. Dividing this value by the one-hour interval gives the mean irradiance in W/m^2, with the same numerical value. Daily irradiation totals (Wh/m^2) were obtained by summing the hourly values.

Table 2. Source year of each typical month in the Manchester TMYx file.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Source year	1994	1973	2004	1982	2019	1993	2016	2023	2010	1988	2008	1991

The weather variables and their weighting factors are not fixed in the proposed approach and can be selected based on the purpose of the study. Different combinations of weather variables and weighting factors have been used in representative weather generation, reflecting differences in the intended application and climatic conditions [5,6,9,13,14,16,22]. For example, later TMY methods introduced direct normal irradiance and redistributed the original Sandia weights, while the IWEC method assigned equal weights (40%) to dry-bulb temperature and global radiation [5,14,23,37,38]. More recently, optimisation and data-driven approaches have been used to derive application-specific weighting factors, including genetic algorithms [18], entropy-based weighting [39] and machine learning methods [16,19,20].

In the present study, the four weather variables used in the original Sandia method were retained: dry-bulb temperature, dew-point temperature, wind speed, and global horizontal solar radiation. However, the change from monthly to daily selection required a modification to the indices used in the original method. In the Sandia method, daily indices such as the mean, maximum and minimum are calculated for each day within a candidate month. In the present method, each candidate period is itself a single day. The hourly values were therefore used directly to construct the daily and monthly distributions hour by hour rather than first reducing each day to daily summary indices. For each candidate day, the daily distribution was constructed from its hourly observations, while the monthly distribution was constructed from all hourly observations within the corresponding calendar month, separately for each of the four weather variables. Hourly values were also used directly by Nik [40], who, for each calendar month, compared the quantiles of the hourly air temperature of each year with those of the same month pooled over the 30-year period, and selected the corresponding month from the year with the least absolute difference as the typical month, which is similar to comparing the CDFs with the FS statistic.

As the individual daily indices used in the Sandia method were not calculated separately, their weights were combined for each weather variable. The weights assigned to the maximum, minimum and mean dry-bulb temperature were summed, giving a total weight of 4/24. The same procedure gave weights of 4/24 for dew-point temperature and 4/24 for wind speed, while global horizontal solar radiation retained its original weight of 12/24. These weights were adopted as a reference weighting scheme to investigate the adaptation of the FS approach to representative-day selection rather than to develop an optimised weighting scheme. Although this preserves the relative weights assigned to the four weather variables, the original allocation of weights among their daily indices has no direct counterpart in the hourly distributions, which incorporate the full range of hourly observations rather than weighting the mean and extreme values separately. The adopted weights are therefore not claimed to be optimal for the hourly approach, as the appropriate weighting scheme depends on the intended application [13].

The distributions were expressed as empirical cumulative distribution functions (ECDFs). The daily ECDF was compared with the corresponding monthly ECDF at each hourly observation, and the absolute differences between the two distributions were calculated.

The plotting position $(k - 0.5)/n$ of the Sandia method (Eq. (1)) was not applied directly. Instead, the empirical cumulative distribution function (ECDF) as defined by Finkelstein and Schafer [21] was used, since the aim was to adapt the FS approach itself. The ECDF is defined for any value x, including repeated values, whereas a plotting position is defined through the rank of each observation. The distributions were constructed from the hourly observations as:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\,(X_i \le x) \tag{5}$$

where n is the number of observations used to construct the ECDF, X_i is the i-th observation, and $I(X_i \le x)$ is an indicator function equal to 1 when $X_i \le x$ and 0 otherwise. The monthly ECDF is constructed from the hourly observations of the corresponding month, while the daily ECDF is constructed from those of each candidate day. The number of observations used depends on the weather variable, as explained below.

The FS value for each weather variable was then calculated as the mean absolute difference between the daily and monthly ECDFs:

$$FS_j = \frac{1}{n_d}\sum_{i=1}^{n_d}\left|F_{m,j}(x_i) - F_{d,j}(x_i)\right| \tag{6}$$

where $F_{m,j}(x_i)$ and $F_{d,j}(x_i)$ are the monthly and daily ECDF values, respectively, for weather variable j, and n_d is the number of hourly observations used for the candidate day. A lower FS value indicates a closer match between the daily and monthly distributions. For dry-bulb temperature, dew-point temperature and wind speed, n_d = 24, and the monthly ECDF was calculated from all hourly observations of the month.

Global horizontal solar radiation was treated differently because of the large number of zero values in hourly data corresponding to night-time hours. In the original Sandia method, solar radiation is represented by daily totals, which include the night-time hours but are not affected by their zero values. In the hourly approach, by contrast, night-time zeros enter the ECDFs as separate observations. Because the fraction of night-time hours varies relatively little within a calendar month, these zeros largely coincide in the candidate day and the month and can reduce the differences between their ECDFs. Retaining them may therefore reduce the FS values of radiation and its contribution to the weighted sum, particularly in winter. To avoid this influence and focus the comparison on the

distribution of positive radiation values, only strictly positive GHR observations (GHR > 0) were used to construct the daily and monthly ECDFs. Consequently, the sample sizes for GHR were determined by the number of positive hourly observations in each candidate day and its corresponding month, rather than by the total number of hourly records. The FS statistic for GHR was therefore calculated by averaging the absolute differences between the daily and monthly ECDFs over the positive observations of each candidate day.

Because the FS statistic is a mean over the evaluated observations, the FS values of all variables lie between 0 and 1 irrespective of the number of observations. However, the realised contribution of each variable to the weighted sum depends on its FS value, while its influence on candidate-day rankings also depends on how its FS values vary across candidate days. The prescribed weights therefore represent nominal coefficients rather than fixed percentages of the final weighted FS statistic.

Before the selection, the hourly data should be checked for missing and invalid values, because weather files may contain missing or interpolated values, which the EPW format identifies through data source and uncertainty flags [41] and represents by numerical missing-value codes [36]. The present dataset contained no missing and no negative values. If present, short gaps would be filled by linear interpolation [11], and candidate days with longer gaps would be excluded, as months with longer gaps are excluded in the selection of typical months [37] and days with invalid data in the selection of representative days [30]. Negative radiation values, which are physically invalid, would be set to zero, as is done for flagged night-time values [42]. A candidate day without any positive radiation value would be excluded from the candidates, and in a month without any positive radiation, as at high latitudes in winter, the radiation term would be omitted from the weighted FS statistic; neither occurred in the present dataset.

The FS values for the four weather variables were combined using Equation (4), with weights of 4/24 for dry-bulb temperature, 4/24 for dew-point temperature, 4/24 for wind speed, and 12/24 for global horizontal solar radiation. The day with the lowest weighted FS value in each calendar month was selected as the representative day for that month. One day per calendar month was used to follow the monthly structure of typical and reference years, which are composed of one typical month selected for each calendar month [6,13], and to limit the number of conditions to be reproduced in the chamber. This choice is not claimed to be sufficient for any particular building response quantity. Unlike the Sandia method, no persistence-based screening was applied, as persistence criteria assess sequences of consecutive days within a month and are therefore not applicable to the selection of a single day. The selected day is thus determined only by the weighted FS value, which makes the selection reproducible.

To assess the sensitivity of the representative-day selection to the treatment of tied observations, the FS calculations were repeated using the mid-distribution function, defined as the proportion of observations strictly below a given value plus half the proportion equal to that value. The evaluation points, variable weights, and treatment of zero global horizontal solar radiation values were kept unchanged.

A further sensitivity analysis was carried out to examine the effect of evaluating Equation (6) only at the observations of the candidate day. The FS values were recalculated using all hourly observations of the corresponding month as evaluation points, including repeated values. The ECDF definitions, variable weights, and treatment of zero global horizontal solar radiation values remained unchanged. This changes the weighting of the evaluation points from their frequency in the candidate day to their frequency in the month.

2.3. Assessment of the Selected Representative Days

The representativeness of the selected days was assessed by comparing their statistical characteristics with those of their corresponding months in the original TMYx file [7,9,11,14,16]. These comparisons are in-sample, as the selected days are compared with the same file from which they were selected. They therefore describe the fidelity of the compression of this file, and do not establish its performance for other years, versions of the weather file, or locations. For each month, the daily mean dry-bulb temperature, dew-point temperature, relative humidity and wind speed, the daily total global horizontal radiation, and the daily temperature range were compared with the corresponding monthly averages of these daily values. Relative humidity and daily temperature range were included in the assessment although they were not directly considered in the selection process. The daily temperature range was defined as the difference between the maximum and minimum hourly dry-bulb temperatures of the day.

For each variable, the difference between the selected day and the corresponding monthly average was calculated as:

$$D_{m,j} = X_{s,m,j} - \overline{X}_{m,j} \tag{7}$$

where $X_{s,m,j}$ is the value of variable j for the selected day in month m, and $\overline{X}_{m,j}$ is the mean of the corresponding daily values over all days of that month.

The mean bias error (MBE) and root mean square error (RMSE) were calculated for each variable using the differences across the twelve months:

$$MBE_j = \frac{1}{12}\sum_{m=1}^{12} D_{m,j} \tag{8}$$

$$RMSE_j = \sqrt{\frac{1}{12}\sum_{m=1}^{12} D_{m,j}^2} \tag{9}$$

As the variability of weather conditions differs between months, the differences were also standardised using the standard deviation of the daily values within each month:

$$Z_{m,j} = \frac{D_{m,j}}{\sigma_{m,j}} \tag{10}$$

where $\sigma_{m,j}$ is the population standard deviation of the daily values of variable j within month m (daily means for dry-bulb temperature, dew-point temperature, relative humidity and wind speed, daily totals for global horizontal radiation, and daily ranges for the daily temperature range).

Equation (10) is evaluated only when $\sigma_{m,j} > 0$. If $\sigma_{m,j} = 0$, the standardised error is undefined and is not applicable; that month–variable combination is excluded from summaries of standardised errors, while its unstandardised error is retained. This did not occur in the present dataset.

The MBE and RMSE were also calculated using the standardised differences to allow for comparison between variables with different units and levels of variability. For month–variable combinations with $\sigma_{m,j} > 0$, the mean square of the standardised differences across all days of the month equals 1 by definition. The expected mean squared standardised error of a set of days selected at random, one from each month, is therefore 1, although the corresponding RMSE is not exactly 1. A mean squared standardised error below 1 indicates that the selected days are closer to the monthly averages than expected for a randomly selected day.

To assess how closely the selected days followed the typical diurnal course of the weather conditions, the hourly profiles of five weather variables were compared with

their corresponding monthly mean diurnal profiles. For each variable, the monthly mean diurnal profile was calculated by averaging the values recorded at each hour of the day across all days within the corresponding month. The 10th and 90th percentiles of the hourly values were also calculated for each hour of the day across all days of the month to show the typical range of hourly conditions within each month.

The root mean square deviation (RMSD) between the hourly profile of each selected day and the corresponding monthly mean diurnal profile was calculated as [25]:

$$RMSD_{s,m,j} = \sqrt{\frac{1}{24}\sum_{h=1}^{24}\left(X_{s,m,j,h} - \overline{X}_{m,j,h}\right)^2} \quad (11)$$

where $X_{s,m,j,h}$ is the value of weather variable j at hour h of the selected day in month m, and $\overline{X}_{m,j,h}$ is the mean value of the same variable at that hour across all days of the corresponding month.

To provide a reference for evaluating the selected days, the same RMSD calculation was performed for every candidate day within each month, using the corresponding monthly mean diurnal profile. The mean daily RMSD was then calculated for each month and variable.

The RMSD of each selected day was divided by the mean daily RMSD of its corresponding month:

$$R_{m,j}^{RMSD} = \frac{RMSD_{s,m,j}}{\overline{RMSD}_{m,j}} \quad (12)$$

where $RMSD_{s,m,j}$ is the RMSD of the selected day and $\overline{RMSD}_{m,j}$ is the mean RMSD of all candidate days within month m for weather variable j.

A ratio below 1 means that the selected day lies closer to the monthly mean diurnal profile than the average day of the month. Unlike in the calculation of the FS statistic, all 24 h were used, including the night-time hours when solar radiation is zero, because the actual 24-h profile, and not a distribution, is compared.

To assess how well the twelve selected days collectively represented the annual statistical characteristics of the TMYx dataset, the annual mean and standard deviation (SD) of five weather variables (dry-bulb temperature, dew-point temperature, relative humidity, global horizontal radiation and wind speed) were compared between the original dataset and a compressed dataset. The compressed dataset was constructed by repeating each selected day according to the number of days in its corresponding month (31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30 and 31 days), resulting in 8760 hourly values. The mean and population SD were calculated for each variable using all 8760 hourly values in each dataset, including night-time zero values for solar radiation. The mean and SD values were then compared to assess how well the selected days represented the annual average conditions and variability in the original dataset. To examine the lower and upper parts of the distributions, the 5th, 50th and 95th percentiles of the hourly weather variables were compared, together with the number of hours meeting selected thresholds. For global horizontal radiation, percentiles were calculated using positive hourly values only. The thresholds were selected to examine the representation of cold conditions (dry-bulb temperature below 0 °C), warmer conditions (above 20 °C), high relative humidity (at least 90%), and higher wind speeds (above 8 m/s, corresponding to Beaufort force 5 or higher [43]).

The repetition of the selected days was used solely for calculating month-length-weighted annual weather statistics and does not represent a chronological weather sequence or a prescribed laboratory testing schedule. The twelve selected days provide a set of meteorological conditions for developing an application-specific testing programme,

rather than a fixed 288-h experimental sequence. The order, duration and conditioning requirements of such a programme depend on the objectives of the experiment and the thermal and hygric characteristics of the tested building elements.

In addition, the annual ECDFs of the five weather variables were calculated using all 8760 hourly observations in each dataset, following Equation (5). Night-time zero values of global horizontal radiation were included. The ECDFs of the two datasets were then compared.

To further assess the representativeness of the selected days, the relationships between weather variables were evaluated using Pearson correlation coefficients. Correlations were calculated for all ten unique pairs of the five weather variables: dry-bulb temperature, dew-point temperature, relative humidity, global horizontal radiation and wind speed. For each month, the correlation coefficients were calculated using all hourly observations of the original month and the 24 hourly observations of the corresponding selected day. The absolute difference between the two correlation coefficients for each variable pair was calculated as:

$$\Delta r_{m,jk} = \left| r_{m,jk}^{selected} - r_{m,jk}^{original} \right| \quad (13)$$

where $r_{m,jk}^{selected}$ and $r_{m,jk}^{original}$ are the Pearson correlation coefficients between weather variables j and k for the selected day of month m and for month m, respectively. The absolute differences were averaged across the twelve months for each variable pair and across all 120 month–pair combinations to provide an overall measure of correlation differences. As in the preceding hourly assessments, night-time zero values of global horizontal radiation were retained. Relative humidity was included in this assessment although it was not used in the FS selection criterion.

To interpret these differences, a second reference was used, the mean within-day correlation of the month, calculated as the arithmetic mean of the Pearson coefficients of the 24 hourly values of each day of the month, and the differences of the selected days from it were calculated as in Equation (13). For both references, the absolute difference was also calculated for every day of the month and averaged over the days of each month and then over the twelve months, which gives the difference expected for a day chosen at random (Table A4).

3. Results and Discussion

3.1. Selection of Representative Weather Days

The weighted FS method was applied to the Manchester TMYx dataset following the procedure described in Section 2.2. One day was selected from each month based on the lowest weighted FS value. Table 3 lists the twelve selected days, their daily weather statistics and the corresponding weighted FS values.

Table 3. Selected representative days, their daily weather statistics, and weighted FS values.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Selected day	2	11	14	20	22	22	27	13	7	16	3	25
Daily mean DB (°C)	5.7	5.6	5.7	8.8	12.1	13.9	16.5	16.1	14.2	11.4	7.9	3.8
Daily min DB (°C)	2.2	1.9	2.6	3.0	6.1	11.5	14.0	14.0	11.0	8.5	7.0	1.6
Daily max DB (°C)	8.6	8.0	10.8	14.3	17.0	17.1	19.0	19.0	19.0	14.2	9.0	6.7
Daily mean DP (°C)	3.7	2.7	2.7	3.2	7.1	8.7	13.4	12.1	11.3	8.4	4.5	2.1
Daily min DP (°C)	0.9	−1.2	−0.5	−2.3	3.1	6.8	12.0	10.0	9.0	6.8	1.9	0.7
Daily max DP (°C)	5.7	4.0	4.8	7.2	9.0	10.7	15.0	14.0	13.0	9.9	7.0	5.3
Daily total GHR (Wh/m²)	613	1207	2526	4837	4758	5563	4393	4185	3433	1713	945	631
Daily mean WS (m/s)	6.1	6.1	3.8	3.6	2.6	3.5	3.7	3.3	2.9	3.9	5.5	3.2

Weighted FS	0.090	0.113	0.073	0.051	0.069	0.077	0.074	0.069	0.088	0.082	0.117	0.106

Note: DB = dry-bulb temperature; DP = dew-point temperature; GHR = global horizontal radiation; WS = wind speed.

The weighted FS values ranged from 0.051 (April) to 0.117 (November), with higher values generally found in the winter months. As a low FS value alone does not confirm that a day is representative, the selected days were further assessed in the following sections.

Repeated values were frequent in the hourly weather data, accounting for 94–99% of dry-bulb temperature observations, 96–99% of dew-point temperature observations, 97–99% of wind-speed observations, and 49–75% of positive solar radiation observations within individual months. When the FS calculations were repeated using the mid-distribution function, the same representative days were selected in 10 of the 12 months. In June and September, the originally selected days ranked second under the mid-distribution approach. To assess the effect of these changes, both the original and alternative days were compared with the monthly mean values of five weather variables. In June, the RMS standardised error was 0.47 for the day selected by Equation (6) and 0.60 for the day selected using the mid-distribution method. In September, the corresponding errors were 0.50 and 0.35. The mid-distribution method therefore gave a better result in September, but not in June. This suggests that the alternative treatment of tied values does not necessarily improve the selection.

The FS calculation was repeated using all hourly observations of the month instead of those of the candidate day. This resulted in the selection of day 15 in June and day 12 in September, the same as with the mid-distribution function. In both months, the days selected by Equation (6) were ranked second, with differences in weighted FS values of 0.003 and 0.008, respectively. The selection remained unchanged in the other ten months. The results show that the choice of representative day was largely unaffected by the alternative calculations, except in June and September, where the two highest-ranked days had similar FS values.

3.2. Daily Statistics of the Selected Days

The daily statistics of the twelve selected days were compared with the corresponding monthly averages. Table 4 reports the mean bias error (MBE), root mean square error (RMSE), the corresponding standardised values and the maximum absolute standardised difference for dry-bulb temperature, dew-point temperature, relative humidity, global horizontal radiation, wind speed, and daily temperature range. Although relative humidity was not included in the FS selection criteria, it was also assessed to see how well the selected days represented a variable not used in their selection. Its hourly values were extracted from the original weather dataset for the same 24-h periods. The monthly errors and corresponding standardised errors for all six variables are provided in Tables A1 and A2 in Appendix A, respectively.

Daily temperature range, defined as the difference between the maximum and minimum hourly dry-bulb temperatures within a day, was included as an additional evaluation metric to assess how well the selected days reproduced the typical magnitude of diurnal temperature variation. It was not used as a criterion in the FS-based day selection.

Table 4. Statistical comparison of the selected days with the corresponding monthly averages.

Variable	FS Weights (%)	MBE	RMSE	MBE (z)	RMSE (z)	Max $\|z\|$ (Months)
DB (°C)	16.7	0.44	0.81	0.17	0.30	0.49 (Feb)
DP (°C)	16.7	0.42	0.97	0.14	0.36	0.68 (Jan)
RH (percentage points)	–	−0.15	3.60	−0.02	0.54	0.81 (Nov)

Daily total GHR (Wh/m²)	50	204	409	0.24	0.44	1.08 (Apr)
WS (m/s)	16.7	0.05	0.71	0.02	0.37	0.67 (Feb)
Daily temperature range (°C)	–	−0.03	1.95	−0.01	0.80	1.19 (Nov)

Note: GHR errors refer to daily irradiation totals, calculated as the sum of the 24 hourly EPW values. The daily temperature range is the difference between the maximum and minimum hourly dry-bulb temperatures of the day. z is the standardised difference (Equation (10)). RH MBE and RMSE are expressed in percentage points; standardised statistics are dimensionless. Monthly values are given in Appendix A. The MBE and RMSE are unweighted means over the twelve months.

The results show relatively small differences between the selected days and the monthly averages. Dry-bulb and dew-point temperatures had RMSE values of 0.81 °C and 0.97 °C, corresponding to standardised RMSE values of 0.30 and 0.36, respectively. Wind speed had a standardised RMSE of 0.37, while relative humidity showed a larger value of 0.54. For $\sigma_{m,j} > 0$, the mean of z^2 across all days of a month equals 1. Thus, a random twelve-day set with one day sampled uniformly from each month has an expected mean squared standardised error of 1, although its expected RMSE(z) is not generally exactly 1. For daily means and irradiation totals, the selected set had RMSE(z)² values of approximately 0.09–0.29 for the five variables (Table 4), representing a 71–91% reduction relative to this theoretical reference.

The positive MBE values for dry-bulb temperature, dew-point temperature and global horizontal radiation indicate that the selected days were, on average, slightly warmer, had higher dew-point temperatures, and were sunnier than the monthly averages. For daily temperature range, the MBE was close to zero (−0.03 °C), as positive and negative monthly errors largely cancelled each other out. The range of the selected day was larger than the monthly mean in January–May, September and December, and smaller in June–August, October and November. However, the RMSE was 1.95 °C, approximately 29% of the mean of the monthly mean ranges (6.6 °C). The standardised RMSE was 0.80, higher than those of the other variables (0.30–0.54), indicating that daily temperature range was reproduced less closely relative to its monthly variability. The absolute standardised error exceeded one in April (1.02) and November (1.19).

Global horizontal radiation had a standardised RMSE of 0.44, despite receiving the highest weight (50%) in the FS selection process. This can be explained by the fact that the FS method compares weather distributions, whereas the RMSE measures differences in daily statistics. A higher weight in the selection process therefore does not necessarily lead to a lower RMSE.

The largest standardised differences were found for the daily temperature range in November (−1.19), global horizontal radiation in April (1.08) and the daily temperature range in April (1.02), which were the only cases above 1 among the 72 month–variable combinations (Table A2); among the remaining cases, the largest was that of relative humidity in November (0.81). Overall, these results show that the selected days generally represent the monthly averages well, although some differences remain for individual variables and months.

3.3. Hourly Profiles of the Selected Days

To examine how well the selected days represent hourly weather patterns, dry-bulb temperature and global horizontal radiation were chosen for graphical comparison. Both variables are important for building thermal performance, as they influence heat transfer and solar heat gains. Global horizontal radiation also received the highest weight (50%) in the FS selection process. Figures 1 and 2 compare the 24-h profiles of the selected days with the monthly mean diurnal profile and the 10th–90th percentile range of the

corresponding month, while the agreement across all five weather variables was evaluated statistically (Table 5).

For dry-bulb temperature (Figure 1), the hourly profiles of the selected days remained within the 10th–90th percentile range of their months for virtually all hours. The lowest RMSD ratio for DB was found in July (0.30). In April and May, the selected days had a larger diurnal range than the monthly mean (by 2.8 °C and 2.5 °C; Table A1), with cooler early-morning hours and warmer afternoons. The selected days in January, February, and October were generally warmer than the monthly mean during the afternoon and evening. The November day showed a relatively flat profile above the monthly mean, with a relatively small diurnal temperature range.

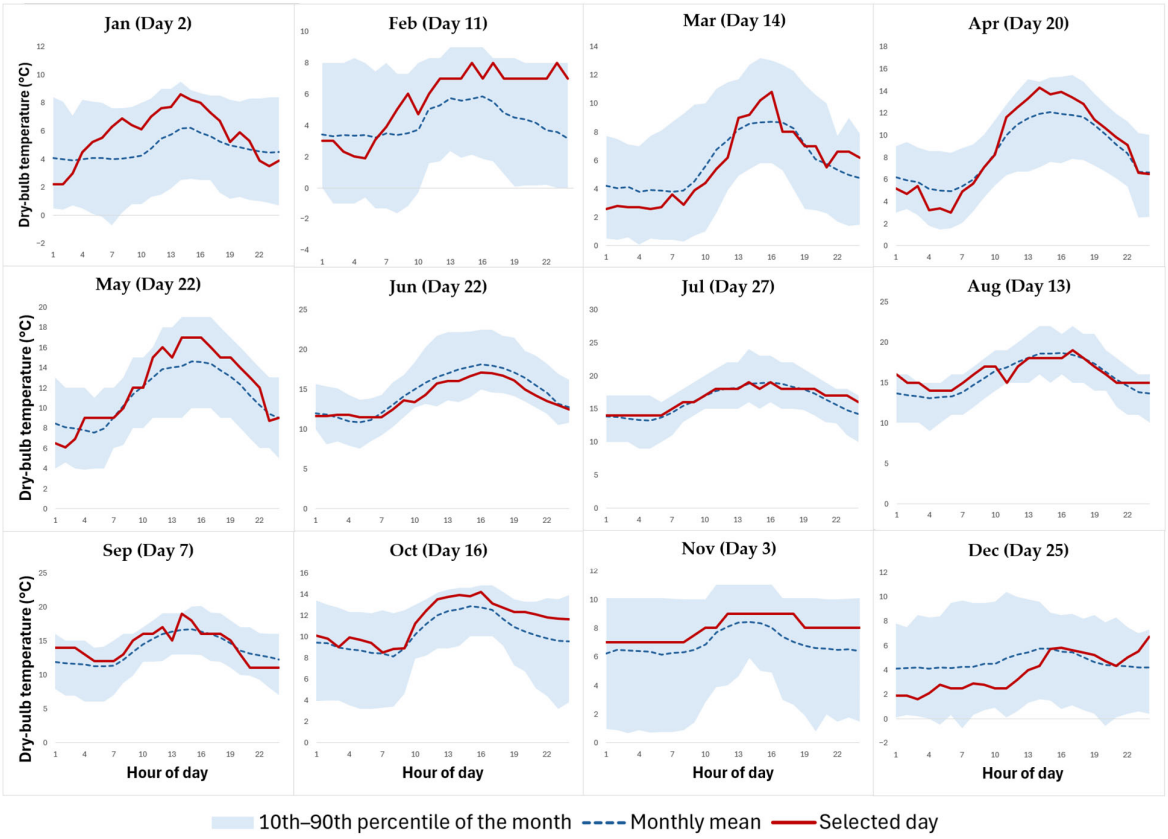


Figure 1. Hourly dry-bulb temperature of the selected days compared with the monthly mean diurnal profile and the 10th–90th percentile range of the corresponding month. Note the different y-axis scales.

For global horizontal radiation (Figure 2), the general pattern of daytime radiation was similar to the monthly mean profile in all months. The profiles of the selected days in January, February, July and October closely followed the monthly mean. In several months, particularly March, April, June, September and December, the selected days had higher midday radiation than the monthly mean, which is consistent with the positive MBE of global horizontal radiation in Table 4. April was the only month in which the midday radiation of the selected day reached the upper bound of the 10th–90th percentile range.

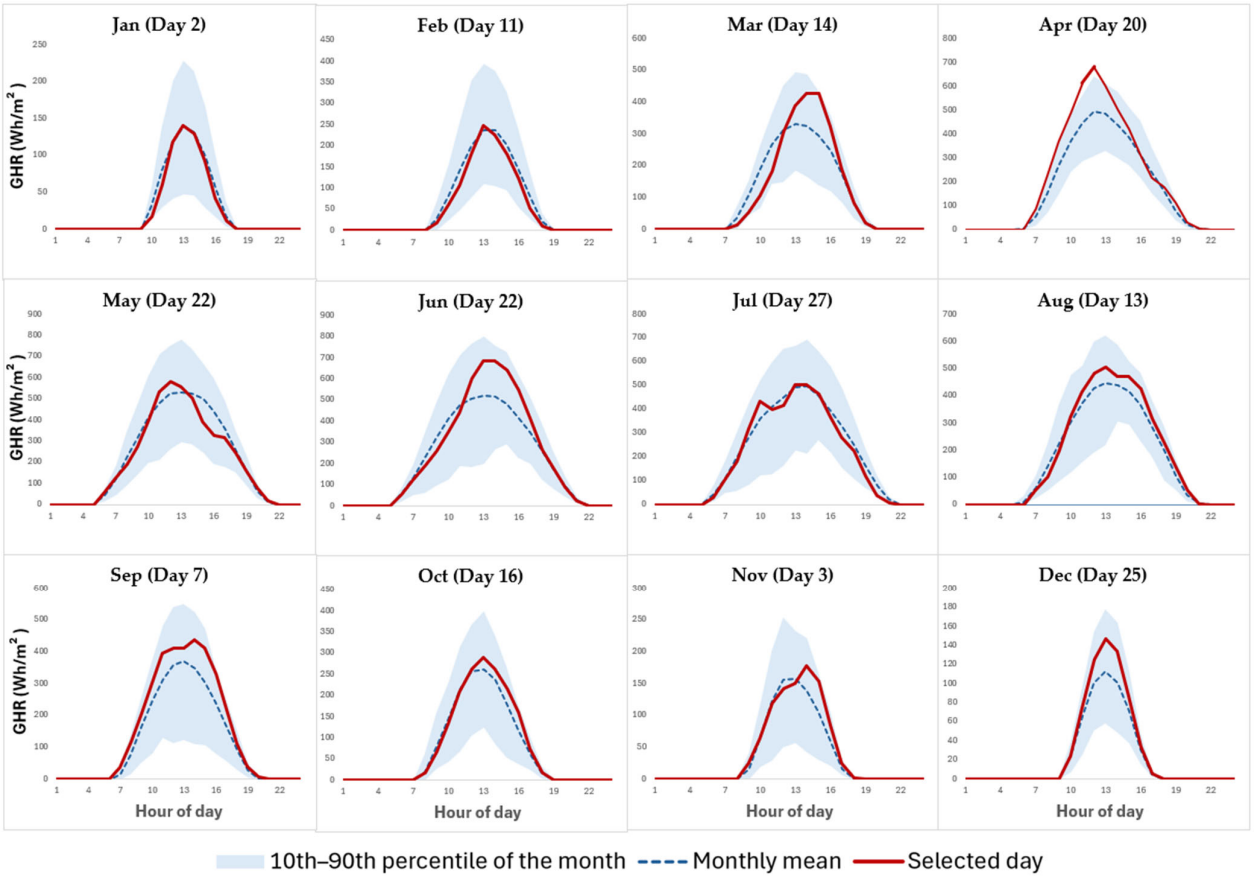


Figure 2. Hourly global horizontal radiation of the selected days compared with the monthly mean diurnal profile and the 10th–90th percentile range of the corresponding month. GHR values represent irradiation accumulated over the preceding hour (Wh/m²). Note the different y-axis scales.

Table 5 presents the ratio of the RMSD of the selected day to the mean RMSD of all candidate days within the corresponding month, for each weather variable. A ratio below 1 indicates that the hourly profile of the selected day is closer to the monthly mean diurnal profile than the average RMSD across all candidate days. The RMSD combines the difference between the daily mean of the selected day and the monthly mean with the difference in the shape of the diurnal course, so part of the advantage of the selected days over the average day reflects the closeness of their daily means. The ratio was below 1 in 56 of the 60 month–variable combinations. These comparisons are not independent, as the weather variables are correlated and all months were selected using the same procedure, and they are therefore reported descriptively. The mean of the monthly ratios was 0.49 for dry-bulb temperature, 0.57 for dew-point temperature, 0.70 for relative humidity, 0.54 for global horizontal radiation and 0.70 for wind speed. The ratio was below 1 in all twelve months for dry-bulb temperature, dew-point temperature, and wind speed. It exceeded 1 for relative humidity in March (1.08), April (1.06) and November (1.10), and for global horizontal radiation in April (1.13). Three of these four cases relate to relative humidity, which was not included in the selection process. The higher ratio for global horizontal radiation in April is consistent with Figure 2 and with the standardised difference of 1.08 reported for global horizontal radiation in April in Table 4.

Table 5. Ratio of the RMSD between the 24-h profile of the selected day and the monthly mean diurnal profile to the mean RMSD of all days of the month, for each month and weather variable. The colours indicate the value of the ratio: dark blue corresponds to a value of 0, red corresponds to a value of 2, and midpoint corresponds to a value of 1. Thus, darker blue indicates lower ratios (closer to the monthly mean diurnal profile), while pink indicates ratios above 1.

Month	DB	DP	RH	GHR	WS
January	0.66	0.79	0.75	0.23	0.80
February	0.71	0.53	0.51	0.30	0.97
March	0.40	0.40	1.08	0.83	0.79
April	0.51	0.84	1.06	1.13	0.51
May	0.56	0.55	0.28	0.42	0.56
June	0.40	0.84	0.77	0.68	0.75
July	0.30	0.47	0.44	0.27	0.68
August	0.48	0.53	0.70	0.42	0.83
September	0.56	0.66	0.69	0.66	0.65
October	0.46	0.38	0.30	0.31	0.69
November	0.37	0.50	1.10	0.50	0.76
December	0.48	0.40	0.70	0.72	0.45
Mean of monthly ratios	0.49	0.57	0.70	0.54	0.70

For dry-bulb temperature, the selected days therefore deviated from the typical diurnal profile by only about half as much as an average day. Even in the four exceptions, the ratios remained close to 1, indicating that the selected days were similar to an average day of the month rather than atypical.

3.4. Annual Characteristics of the Compressed Year

Table 6 compares the annual mean and standard deviation of the five weather variables between the original TMYx dataset and the compressed year, and Figure 3 shows their annual ECDFs. The annual means of the compressed year were generally close to those of the original TMYx dataset, with differences of 0.43 °C for dry-bulb and 0.42 °C for dew-point temperatures, and of 0.10 percentage points for relative humidity and 0.04 m/s for wind speed. The mean global horizontal radiation was 7.5% higher in the compressed year, which is consistent with the positive MBE of global horizontal radiation in Table 4. This corresponds to an annualised difference of approximately 74 kWh/m² in horizontal irradiation; the largest contributions came from April, September and June, partly offset by months with lower radiation totals. The relative errors of the daily irradiation totals of the selected days with respect to the mean daily totals of their months ranged from −12.5% (February) to +26.5% (September) (Table A1). The effect of these differences on solar gains depends on surface orientation, the direct and diffuse components of solar radiation and the characteristics of the building elements. This effect and its implications for the intended laboratory tests were not evaluated in the present study.

Table 6. Annual mean and standard deviation (SD) of hourly weather variables in the original TMYx dataset and the compressed year.

Weather Variable	Mean			SD		
	TMYx	Compressed	Diff.	TMYx	Compressed	Diff. (%)
DB (°C)	9.74	10.17	+0.43 (°C)	5.59	4.84	−13.4
DP (°C)	6.27	6.70	+0.42 (°C)	5.15	4.15	−19.4
RH (%)	79.66	79.56	−0.10 pp	13.14	11.43	−13.0
Hourly GHR (Wh/m²)	112.61	121.06	+8.46	172.96	175.20	+1.3
WS (m/s)	3.97	4.01	+0.04	2.48	2.00	−19.2

Note**:** Annual statistics were calculated using 8760 hourly values for each dataset. Night-time zero values of global horizontal radiation (GHR) were included in these calculations, although they were excluded from the FS calculation during representative-day selection. GHR mean and SD refer to hourly irradiation values, not daily totals. Mean differences are expressed in the units of each variable, with RH differences given in percentage points (pp). SD differences are relative percentage changes from the original TMYx dataset. Annual statistics are weighted by the number of days in each month, whereas the monthly errors summarised in Table 4 are equally weighted. The mean differences in Table 6 therefore need not equal the MBE values in Table 4. Differences and percentage changes were calculated from unrounded values.

The ECDFs of the compressed year generally followed those of the TMYx dataset over the central parts of the distributions, although differences were observed towards the lower and upper ends of some distributions (Figure 3). The standard deviations of dry-bulb temperature, dew-point temperature, relative humidity and wind speed were approximately 13–19% lower in the compressed year (Table 6), indicating a reduction in the variability represented by the selected days. In contrast, the standard deviation of global horizontal radiation, which includes the night-time zero values, differed by only 1.3%; this all-hour value does not describe the daytime variability, and the percentiles of the positive values are compared in Table 7.

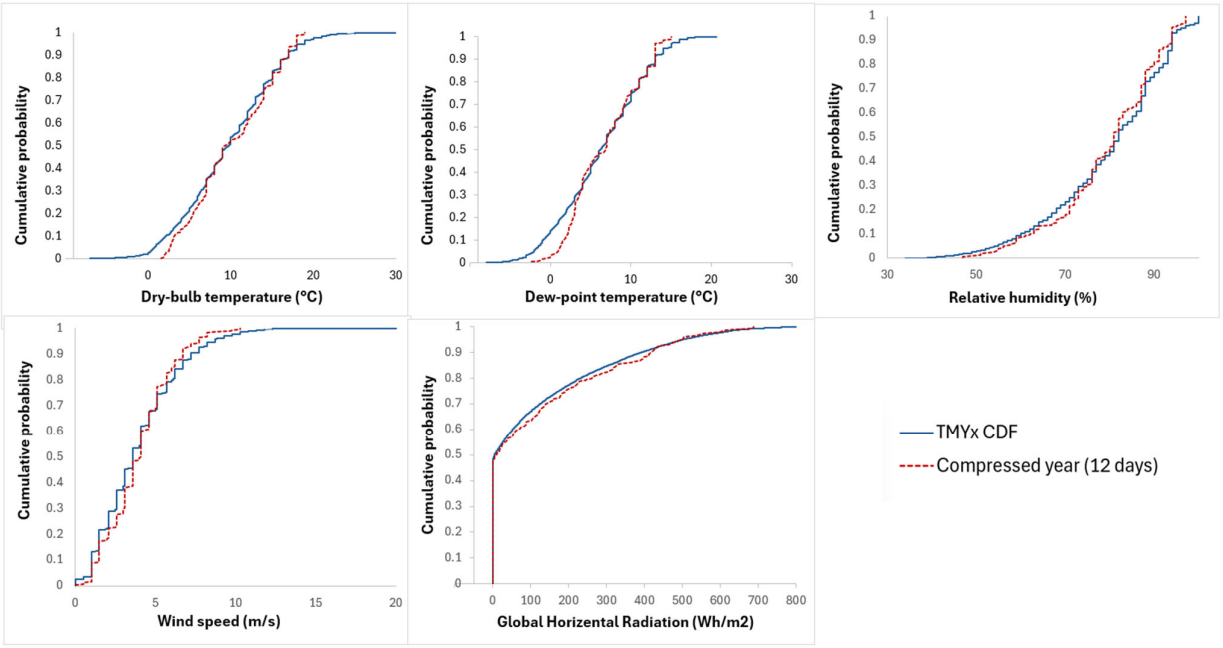


Figure 3. Annual empirical cumulative distribution functions (ECDFs) of hourly dry-bulb temperature, dew-point temperature, relative humidity, wind speed and global horizontal radiation for the original TMYx dataset and the compressed year (12 days). Night-time zero values of global horizontal radiation are included. GHR values represent irradiation accumulated over one-hour intervals (Wh/m^2).

Table 7 provides a further comparison of the hourly weather distributions using selected percentiles and threshold exceedances. The median values were relatively close between the two datasets, although differences were observed at the lower and upper percentiles of some variables. For dry-bulb and dew-point temperature, the 5th percentiles were 1.8 and 3.0 °C higher in the compressed dataset, respectively, while the corresponding 95th percentiles were 1.0 and 1.5 °C lower. The dry-bulb temperature of the

compressed dataset ranged from 1.6 to 19.0 °C, whereas the TMYx file contained 179 h below 0 °C and 204 h above 20 °C, neither of which occurred in the compressed dataset. The number of hours with relative humidity of at least 90% decreased by approximately 17%, and the number of hours with wind speed above 8 m/s by approximately 47%. In contrast, the 95th percentile of global horizontal radiation and the 5th percentile of wind speed remained close to those of the original dataset, whereas the median of the positive values of global horizontal radiation was 12% higher, indicating a shift towards higher values in the central part of the distribution of positive radiation, consistent with the higher annual mean. The 5th–95th percentile range of the positive hourly values, used here as a measure of distributional spread, was slightly narrower in the compressed dataset (573 against 581 Wh/m²); it does not describe the variation in radiation within the day.

Table 7. Selected percentiles and frequencies of specified weather conditions in the TMYx file and the compressed dataset (8760 h each). Percentiles of global horizontal radiation refer to hours with positive values.

Measure	TMYx	Compressed
Dry-bulb temperature P5/P50/P95 (°C)	0.8/9.8/19.0	2.6/9.4/18.0
Dew-point temperature P5/P50/P95 (°C)	−2.0/6.2/14.5	1.0/6.8/13.0
Relative humidity P5/P50/P95 (%)	55/82/96	57/81/94
Wind speed P5/P50/P95 (m/s)	1.0/3.6/8.7	1.0/4.1/7.7
Hourly GHR, positive hours P5/P50/P95 (Wh/m²)	7/162/588	11/182/584
Hours with dry-bulb temperature < 0 °C	179	0
Hours with dry-bulb temperature > 20 °C	204	0
Hours with relative humidity ≥ 90%	2212	1845
Hours with wind speed > 8 m/s	606	320

These results indicate that, although the median values of the weather variables were relatively well represented, some colder, warmer, more humid and windier conditions were underrepresented. Repeating one selected day for each month preserves its within-day variability but does not reproduce the variability between different days of that month, limiting the range of conditions represented in the compressed dataset. The specific conditions omitted also depend on the characteristics of the selected days.

3.5. Relationships Between Variables

The comparison of Pearson correlation coefficients for the ten pairs of weather variables is summarised in Table 8, with detailed results provided in Tables A3 and A4 (Appendix B). The mean absolute difference between the correlations of the selected days and those of the original months was 0.37 (Equation (13)), ranging from 0.12 for relative humidity and global horizontal radiation to 0.49 for dry-bulb temperature and wind speed. This overall mean difference was slightly higher than the corresponding mean across all candidate days (0.33). The signs of the correlations agreed with those of the original months in 93 of the 120 month–pair combinations (78%, compared with 76% across candidate days), including all twelve months for relative humidity and radiation and eleven months for dry-bulb temperature with relative humidity and with radiation. The selected days had stronger absolute correlations than the original months in 96 of the 120 combinations (80%, compared with 71% across candidate days).

Table 8. Summary of the comparison of Pearson correlation coefficients between the selected days and the original months for the ten pairs of weather variables across twelve months. The reported differences are absolute differences between correlation coefficients. The statistics for candidate days were calculated for every day of each month, including the selected day, and averaged across days within each month and then across the twelve months (details in Tables A3 and A4).

Measure	Selected Days	Mean Across All Candidate Days
Mean absolute difference from the correlation of the original month (Equation (13))	0.37	0.33
Mean absolute difference from the mean within-day correlation of the month	0.27	0.28
Month–pair combinations with the same sign as the original month	78% (93 of 120)	76%
Month–pair combinations with a stronger correlation than the original month	80% (96 of 120)	71%

Monthly correlations reflect both within-day and between-day variability, whereas correlations calculated for individual days reflect only within-day variability. Therefore, a day whose correlations differ from those of the month may still exhibit relationships commonly observed within individual days. When compared with the mean within-day correlations of their corresponding months, the selected days showed a mean absolute difference of 0.27, similar to the mean across all candidate days (0.28). The higher proportion of stronger correlations among the selected days (80%, compared with 71% across candidate days) may partly contribute to their larger differences from the monthly correlations; however, this does not establish the cause of these differences.

These findings are consistent with the selection criterion used in this study. Preserving the relationships between weather variables is an aim of reference-year selection in ISO 15927-4 [6] and the Sandia method [13]. However, the FS statistic evaluates the distribution of each weather variable separately, and the individual scores are combined using predefined weights, without explicitly accounting for correlations between variables. This limitation has also been noted in studies of typical year selection methods [5,12]. Therefore, selecting a day that closely represents the individual weather-variable distributions does not necessarily ensure that their correlations are equally well represented. Because the selected days are complete, observed 24-h weather profiles, the actual hourly combinations and temporal alignment of weather variables are preserved. Nevertheless, the extent to which differences in correlation coefficients influence the thermal and hygric responses of building elements remains to be investigated.

3.6. Limitations and Future Research

While the representative-day approach offers a practical means of reducing testing complexity, time and costs, several limitations should be acknowledged.

First, the test material may exhibit long-term thermal and hygric lag effects, which cannot be fully captured by twelve isolated representative days. As a result, the influence of preceding weather conditions may be underestimated. Repeating a selected day over several consecutive cycles until the specimen reaches a periodic response, as routinely done in our laboratory practice, reduces the influence of the initial conditions, but it does not reconstruct the preceding weather history or the storage of heat and moisture over several days. Periodic conditioning to the conditions of a selected day should therefore be distinguished from the preservation of the physical memory of the specimen. The conditioning duration and the criterion for a periodic response, the initial thermal and moisture states of the specimen, the treatment of the transition between the last and first hours of a repeated day, and a comparison between repeated single days and naturally consecutive multiday sequences require further experimental investigation.

A second limitation arises from the very definition of a typical day. Because representative days are selected to reflect average rather than extreme conditions [3,5,33],

extreme weather events are not represented. Consequently, the methodology may not adequately assess building performance under rare but critical conditions. This issue can be addressed by conducting additional testing that explicitly includes extreme weather scenarios.

Furthermore, a single representative day cannot fully capture the variability that occurs within a month. Selecting only one day per month preserves neither the full range of day-to-day weather fluctuations nor the original sequence of weather events. As building performance depends on prior weather exposure, this loss of variability may be significant. Instead of repeating the conditions of a single day over several successive days, applying the representative days of successive months one after another may be used to examine the transient response of the specimen to a change in boundary conditions. However, such a transition is a controlled change between representative conditions rather than a realistic short-term weather sequence, and the form of this transition, such as a step change or a gradual change, would need to be defined for each experiment.

Future research could explore adaptive, data-driven weighting approaches, drawing on feature-importance ranking and attention mechanisms [44], conditional interaction modulation [45], and similarity-adaptive weighting of alternative interaction paths [46]. Although originally developed for biomedical prediction tasks, these approaches could inform the selection of weather variable weights based on their influence on the thermal and hygric response of building elements, while also motivating the development of selection criteria that explicitly account for interactions between weather variables.

It should also be noted that this study focused on a method for selecting representative weather days, while translating the selected days into chamber boundary conditions is beyond its scope. For context, the environmental chambers of Energy House 2.0 can control air temperature from −23.5 °C to +51 °C, with a design requirement to maintain temperatures within ±0.5 °C of the set point, and can generate simulated solar radiation, wind speeds of up to approximately 17 m/s, rain of up to 200 mm/h and snow of up to 250 mm per day, as detailed in [1], where further details of the facility are also provided. Although only global horizontal radiation was used in the selection process, the TMYx file also contains direct normal and diffuse horizontal radiation, which can be used to calculate irradiance on building surfaces of different orientations and inclinations using established methods [47]. However, these calculations were not carried out in the present study. For future chamber implementation, hourly irradiation (Wh/m^2) would be divided by the one-hour interval to obtain mean irradiance (W/m^2); this conversion was not performed in the present study. Developing and validating a protocol for reproducing the selected weather conditions in the chamber would require further research and could form the basis of a separate study.

Finally, the evaluation of representative days was based solely on statistical comparisons of weather data rather than on actual building performance outcomes. While the selected days may accurately reflect key weather characteristics, their effectiveness in predicting building energy use, thermal comfort, or other performance metrics remains to be verified. Future research should therefore compare full-year building simulations with simulations based on representative days scaled to the number of days in each month. Such analyses would provide a clearer understanding of the impact of weather-data simplification on building performance predictions.

4. Conclusions

This study presented a weather data compression method for environmental chamber testing, reducing a typical meteorological year to twelve representative days, one for each calendar month. The Finkelstein–Schafer approach used in the Sandia method was adapted from the month-to-year scale to the day-to-month scale, using hourly

distributions of dry-bulb temperature, dew-point temperature, wind speed, and global horizontal radiation, with the original Sandia weights combined for each variable. The method was demonstrated using the TMYx file for Manchester, UK, and the results are in-sample comparisons with this file.

The assessment showed that the selected days were generally representative of their corresponding months. In terms of daily statistics, the standardised RMSE values for the four variables used in the selection ranged from 0.30 to 0.44, corresponding to mean squared standardised errors 80–91% lower than the value expected for a randomly selected day. The standardised differences were within one standard deviation of the monthly average for all assessed variables and months, except for global horizontal radiation in April and the daily temperature range in April and November. In terms of hourly profiles, the selected days were closer to the monthly mean diurnal profile than the average day of the month in 56 of the 60 month–variable combinations. Relative humidity, which was not included in the selection process, was also reasonably represented, whereas the magnitude of the diurnal temperature range was reproduced less closely (standardised RMSE of 0.80). Mean hourly global horizontal irradiation was 7.5% higher in the compressed year than in the original TMYx dataset. These results were obtained using the original Sandia weighting factors, without any adjustment or optimisation, indicating that the adapted FS approach can identify representative days even with a general-purpose weighting scheme.

When combined into a compressed year, the twelve selected days reproduced the annual mean conditions relatively closely, except for global horizontal radiation, whose annual mean was 7.5% higher. However, the compressed year did not retain the full variability of the original TMYx dataset. The standard deviations of dry-bulb temperature, dew-point temperature, relative humidity and wind speed were approximately 13–19% lower, and some colder, warmer, higher dew-point temperature, and windier conditions were underrepresented. Repeating a single selected day for each month eliminates day-to-day variability within that month, while the specific conditions omitted depend on the characteristics of the selected days. These omissions are not explicitly imposed by the FS selection criterion. The selected days provide a basis for investigating building-element performance under a reduced set of representative monthly weather conditions, although their ability to reproduce building responses under the full weather record remains to be evaluated. Where specific conditions are important to the testing objectives, additional days could be selected to complement the twelve representative days.

The proposed method reduces the length of a weather data sequence representing a typical year from 365 to 12 days, making representative weather testing more practical in environmental chambers. This is, however, a nominal reduction in the number of weather days. The actual test duration is typically longer, depending on the experimental requirements, such as conditioning periods, repeated cycles, transitions between days and additional extreme-weather tests. The savings in testing time, cost and energy are therefore anticipated benefits that remain to be quantified. The method is also flexible, as both the weather variables and their weighting factors can be adapted to the purpose of the study. For example, greater weight can be given to the variables that most influence the performance under investigation, and additional variables, such as rainfall for hygrothermal testing, can be included in the selection. The duration of each test could also be adjusted depending on the purpose of the experiment and the laboratory facilities. For example, the weather conditions of a selected day could be repeated for several days to reduce the influence of the initial conditions of the specimen, although such repetition does not reproduce the effect of preceding weather history.

Several directions could be considered for future research. First, the effect of alternative weighting schemes, including those used in later TMY methods and ISO 15927-4,

could be investigated [6,22,37,38]. Recent optimisation and data-driven approaches could also be explored as alternatives for selecting representative days [16,18–20,39]. Second, the representativeness of the selected days could be evaluated in terms of building response by comparing building performance simulated using the compressed year with that obtained using the full TMYx dataset. Such comparisons could also be used to identify weighting factors that best reproduce the annual building response for a given application. Third, the compressed year could be applied in the environmental chambers of Energy House 2.0 to test the hempcrete test cell under representative weather conditions. As the present study was limited to a single case, the Manchester TMYx file, the applicability of the method to other climates and weather datasets remains a hypothesis. It could be tested by applying the method to independent weather datasets and contrasting climates and by assessing the stability of the selected days across them. Future research could also investigate selecting more than one representative day per month to better capture day-to-day weather variability. The acceptable level of error depends on the building response quantity of interest and was not defined in this study. Therefore, the adequacy of selecting one day per month for assessing building performance has not yet been established. Future studies could compare alternatives, such as one, two or three representative days per month or short consecutive periods, based on the thermal and hygric responses of building elements, to determine an appropriate number of representative days for a specified accuracy requirement. The FS criterion does not explicitly account for relationships between weather variables as discussed in Section 3.5. Future research could investigate multi-objective selection methods that consider both the representativeness of individual weather-variable distributions and the preservation of inter-variable correlations, while retaining complete, observed 24-h profiles. Such methods could also be evaluated against the thermal and hygric responses of building elements to determine whether improved correlation preservation leads to more representative laboratory results.

Overall, this study provides a practical approach to selecting representative weather conditions for environmental chamber testing, with the potential to reduce testing time and costs while supporting the evaluation of building performance under typical climatic conditions. This method could be particularly useful for projects involving new and retrofitted buildings, where the performance of sustainable building materials and construction methods could be evaluated under controlled laboratory conditions.

Author Contributions: Conceptualization, S.F. and L.J.; data curation, S.F.; formal analysis, S.F.; funding acquisition, W.S.; investigation, S.F.; methodology, S.F.; resources, L.J., G.H., R.F. and W.S.; supervision, L.J.; writing—original draft preparation, S.F. and L.J.; writing—review and editing, L.J., G.H., R.F. and W.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was carried out as part of the Retrofit Advanced Materials and Manufacturing Centre (RAMMIC) project supported by the UK Research and Innovation Local Innovation Partnerships Fund [grant number UKRI5673].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CDF	Cumulative Distribution Function
DB	Dry-Bulb Temperature
DP	Dew-Point Temperature
ECDF	Empirical Cumulative Distribution Function
EPW	EnergyPlus Weather
FS	Finkelstein–Schafer
GHR	Global Horizontal Radiation
IWEC	International Weather for Energy Calculations
MBE	Mean Bias Error
RH	Relative Humidity
RMSD	Root Mean Square Deviation
RMSE	Root Mean Square Error
SD	Standard Deviation
TMM	Typical Meteorological Month
TMY	Typical Meteorological Year
TMYx	TMYx data set of Climate.OneBuilding.Org, based on more recent weather data than earlier TMY data sets
TRY	Test Reference Year
WFS	Weighted FS Value
WS	Wind Speed

Appendix A

This appendix gives the monthly errors of the selected days with respect to the mean of the daily values of their months (Table A1) and the same errors expressed in standard deviations of the daily values of each month (Table A2). The annual summaries of these errors are given in Table 4.

Table A1. Monthly error of the selected day: value of the selected day minus the mean of the daily values of its month (dry-bulb temperature, DB; dew-point temperature, DP; relative humidity, RH; wind speed, WS; global horizontal radiation, GHR, as the daily total; and daily temperature range).

Month	DB (°C)	DP (°C)	RH (Percentage Points)	WS (m/s)	Daily Total GHR (Wh/m²)	GHR (%)	Daily Temperature Range (°C)
Jan	0.98	1.64	3.33	0.71	−63	−9.3	1.52
Feb	1.41	0.92	−2.72	1.64	−172	−12.5	1.00
Mar	−0.18	0.68	5.00	−0.63	134	5.6	1.64
Apr	0.28	−0.10	−2.48	−0.22	939	24.1	2.83
May	0.87	0.90	−1.05	−0.47	−219	−4.4	2.50
Jun	−0.65	−1.57	−4.62	0.24	571	11.4	−2.79
Jul	0.37	0.95	2.44	−0.04	−159	−3.5	−2.38
Aug	0.43	−0.17	−3.56	0.29	346	9.0	−2.07
Sep	0.50	1.30	4.19	−0.82	720	26.5	0.48
Oct	1.12	1.04	−0.45	−0.32	118	7.4	−0.61
Nov	1.00	−0.10	−5.66	0.90	112	13.5	−2.69
Dec	−0.84	−0.45	3.81	−0.71	119	23.3	0.17
MBE	0.44	0.42	−0.15	0.05	204	7.6	−0.03
RMSE	0.81	0.97	3.60	0.71	409	14.7	1.95

Note: Positive values: the selected day is higher than the average day of its month. The daily temperature range is the maximum minus the minimum of the hourly dry-bulb temperature of the day. MBE: mean of the twelve-monthly errors; RMSE: root mean square of the twelve-monthly errors (Equations (8) and (9)). GHR (%) is the error of the daily irradiation total of the selected day relative to the mean daily total of its month. For GHR (%), the MBE and RMSE give equal weight to each month and therefore differ from the relative difference in the annual mean in Table 6.

Table A2. Monthly error expressed in standard deviations of the daily values of the month, z = error/σ (Equation (10)). z is dimensionless. σ is the population standard deviation of the daily values of the month (daily means for DB, DP, RH and WS; the daily total for GHR; the daily range for the temperature range), calculated over all days of the month. Where $\sigma = 0$, standardised errors are not applicable and excluded from standardised summaries.

Month	DB	DP	RH	WS	GHR	Daily Temperature Range
January	0.41	0.68	0.47	0.35	−0.18	0.81
February	0.49	0.28	−0.43	0.67	−0.25	0.42
March	−0.06	0.21	0.80	−0.28	0.17	0.78
April	0.13	−0.04	−0.40	−0.13	1.08	1.02
May	0.32	0.30	−0.11	−0.43	−0.14	0.90
June	−0.30	−0.63	−0.67	0.22	0.33	−0.95
July	0.14	0.38	0.39	−0.03	−0.11	−0.96
August	0.20	−0.11	−0.60	0.34	0.32	−0.99
September	0.22	0.47	0.55	−0.51	0.64	0.22
October	0.39	0.30	−0.08	−0.16	0.17	−0.23
November	0.35	−0.03	−0.81	0.52	0.28	−1.19
December	−0.24	−0.13	0.62	−0.26	0.56	0.08
MBE (z)	0.17	0.14	−0.02	0.02	0.24	−0.01
RMSE (z)	0.30	0.36	0.54	0.37	0.44	0.80
max $\lvert z \rvert$	0.49	0.68	0.81	0.67	1.08	1.19
Month of max $\lvert z \rvert$	February	January	November	February	April	November
Months with $\lvert z \rvert > 1$	0	0	0	0	1	2

Note: Scale of the errors of the daily temperature range relative to its reference: the monthly mean daily range was between 4.7 and 8.8 °C (mean of the twelve months 6.6 °C); the monthly errors ranged from −57% (November) to +33% (April) of the monthly mean range, and the RMSE of 1.95 °C corresponds to 29% of the mean of the monthly mean ranges.

Appendix B

This appendix gives the correlations between the pairs of weather variables for the original months and the selected days (Table A3) and their comparison with the average over all days of the month (Table A4).

Table A3. Pearson correlations between the pairs of weather variables for the original months (all hourly values of the month) and for the selected days (24 hourly values), and the mean absolute difference between them. GHR includes night-time zero values.

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean	Mean Absolute Difference
DB–DP	Original month	0.83	0.89	0.76	0.53	0.53	0.60	0.65	0.61	0.67	0.86	0.89	0.94	0.73	
	Selected day	0.61	0.78	−0.26	0.66	0.83	−0.72	0.20	−0.33	0.49	0.93	0.61	0.80	0.38	0.43
DB–RH	Original month	−0.15	0.06	−0.22	−0.53	−0.55	−0.61	−0.59	−0.75	−0.39	−0.11	0.02	−0.08	−0.32	
	Selected day	−0.60	−0.57	−0.87	−0.60	−0.90	−0.97	−0.90	−0.89	−0.85	−0.93	0.16	−0.71	−0.72	0.42
DB–GHR	Original month	0.12	0.03	0.34	0.46	0.55	0.56	0.47	0.55	0.49	0.19	0.15	−0.05	0.32	
	Selected day	0.66	0.42	0.74	0.65	0.77	0.77	0.69	0.78	0.84	0.66	0.68	0.08	0.64	0.32
DB–WS	Original month	0.57	0.26	0.48	0.29	0.41	0.19	0.23	0.45	0.31	0.36	0.37	0.47	0.36	
	Selected day	−0.17	0.77	0.75	0.77	0.90	0.93	0.76	0.73	0.71	0.80	−0.35	0.73	0.61	0.49

DP–RH	Original month	0.41	0.48	0.45	0.41	0.40	0.26	0.22	0.06	0.41	0.41	0.46	0.24	0.35	
	Selected day	0.22	0.03	0.65	0.18	−0.53	0.87	0.25	0.73	0.03	−0.73	0.88	−0.15	0.20	0.47
DP–GHR	Original month	−0.15	−0.16	−0.07	−0.12	−0.06	−0.02	−0.11	0.04	−0.02	−0.11	−0.05	−0.11	−0.08	
	Selected day	0.17	0.39	−0.28	0.15	0.52	−0.36	−0.13	−0.14	0.24	0.59	0.03	−0.34	0.07	0.31
DP–WS	Original month	0.31	0.13	0.33	−0.03	0.13	−0.05	−0.06	0.03	0.09	0.19	0.16	0.32	0.13	
	Selected day	−0.02	0.46	−0.05	0.33	0.64	−0.64	0.42	−0.25	0.48	0.76	−0.48	0.33	0.17	0.40
RH–GHR	Original month	−0.44	−0.40	−0.55	−0.58	−0.65	−0.68	−0.72	−0.66	−0.62	−0.53	−0.40	−0.27	−0.54	
	Selected day	−0.61	−0.14	−0.66	−0.68	−0.78	−0.66	−0.73	−0.64	−0.82	−0.64	−0.36	−0.53	−0.60	0.12
RH–WS	Original month	−0.37	−0.22	−0.21	−0.35	−0.34	−0.30	−0.36	−0.54	−0.28	−0.27	−0.37	−0.38	−0.33	
	Selected day	0.14	−0.59	−0.60	−0.73	−0.88	−0.90	−0.56	−0.64	−0.54	−0.73	−0.39	−0.81	−0.60	0.35
GHR–WS	Original month	0.09	0.06	0.17	0.25	0.43	0.23	0.40	0.55	0.19	0.08	0.20	−0.01	0.22	
	Selected day	0.22	0.45	0.40	0.48	0.71	0.79	0.39	0.74	0.70	0.59	−0.33	0.56	0.47	0.35
All pairs															0.37

Table A4. Mean absolute difference between the correlations of the selected days and two references, compared with the average over all days of the month. The references are the correlation of all hourly values of the original month (Equation (13)) and the mean within-day correlation of the month. All values are averaged over the twelve months.

Pair	Selected Days vs. Original Month	Average over All Days vs. Original Month	Selected Days vs. Mean Within-Day	Average over All Days vs. Mean Within-Day
DB–DP	0.43	0.38	0.34	0.34
DB–RH	0.42	0.40	0.22	0.27
DB–GHR	0.32	0.30	0.12	0.17
DB–WS	0.49	0.35	0.34	0.27
DP–RH	0.47	0.40	0.40	0.39
DP–GHR	0.31	0.35	0.27	0.30
DP–WS	0.40	0.36	0.38	0.33
RH–GHR	0.12	0.20	0.13	0.20
RH–WS	0.35	0.30	0.23	0.26
GHR–WS	0.35	0.30	0.23	0.26
All pairs	0.37	0.33	0.27	0.28

References

1. Fitton, R.; Henshaw, G.; Swan, W.; Roberts, B.; Le Hunte, T.; Zhang, X.; Jankovic, L. Energy House 2.0: The Design, Build, Commissioning, Performance & Early Findings of a Large-Scale Building Physics Test Facility. *Build. Serv. Eng. Res. Technol.* **2026**, *47*, 247–262. https://doi.org/10.1177/01436244261419278.
2. Crawley, D.B. Which Weather Data Should You Use for Energy Simulations of Commercial Buildings? *ASHRAE Trans.* **1998**, *104*, 498–515.
3. Crawley, D.; Lawrie, L. Should We Be Using Just 'Typical' Weather Data in Building Performance Simulation? In Proceedings of the 16th Conference of IBPSA (Building Simulation 2019), Rome, Italy, 2–4 September 2019; pp. 4801–4808.
4. Crawley, D.; Lawrie, L. Rethinking the TMY: Is the 'Typical' Meteorological Year Best for Building Performance Simulation? In Proceedings of the BS2015: 14th Conference of International Building Performance Simulation Association, Hyderabad, India, 7 December 2015; IBPSA: Hyderabad, India.

5. Rady, M.; Muhammad, M.K.I.; Shahid, S. Evolving Typical Meteorological Year (TMY) Data for Building Energy Simulation: A Comprehensive Review of Methods, Challenges, and Future Directions. *Adv. Build. Energy Res.* **2025**, *19*, 269–299. https://doi.org/10.1080/17512549.2025.2457649.
6. *BS EN ISO 15927-4:2005*; Hygrothermal Performance of Buildings—Calculation and Presentation of Climatic Data—Part 4: Hourly Data for Assessing the Annual Energy Use for Heating and Cooling. British Standards Institution: London, UK, 2005.
7. Zeng, Z.; Kim, J.-H. (Jeannie); Muehleisen, R.T. An Update to the Sandia Method for Creating Typical Meteorological Years from a Limited Pool of Calendar Years. *Energy* **2025**, *320*, 135143. https://doi.org/10.1016/j.energy.2025.135143.
8. Vandemeulebroucke, I.; Caluwaerts, S.; Van Den Bossche, N. Decision Framework to Select Moisture Reference Years for Hygrothermal Simulations. *Build. Environ.* **2022**, *218*, 109080. https://doi.org/10.1016/j.buildenv.2022.109080.
9. Pernigotto, G.; Prada, A.; Cóstola, D.; Gasparella, A.; Hensen, J.L.M. Multi-Year and Reference Year Weather Data for Building Energy Labelling in North Italy Climates. *Energy Build.* **2014**, *72*, 62–72. https://doi.org/10.1016/j.enbuild.2013.12.012.
10. Barnaby, C.S.; Crawley, D.B. Weather and Climate in Building Performance Simulation. In *Building Performance Simulation for Design and Operation*; Hensen, J.L.M., Lamberts, R., Eds.; Routledge: Abingdon, UK, 2019; pp. 191–220, ISBN 978-0-429-40229-6.
11. Lee, K.; Yoo, H.; Levermore, G.J. Generation of Typical Weather Data Using the ISO Test Reference Year (TRY) Method for Major Cities of South Korea. *Build. Environ.* **2010**, *45*, 956–963. https://doi.org/10.1016/j.buildenv.2009.10.002.
12. Azimi, M.; Baltazar, J.-C. A Comparative Analysis of Different Weather Datasets for Future Proofing Building Performance Analysis. In Proceedings of the SimBuild 2024: Eleventh National Conference of IBPSA-USA, Denver, CO, USA, 21 May 2024; IBPSA-USA: Denver, CO, USA.
13. Hall, I.J.; Prairie, R.R.; Anderson, H.E.; Boes, E.C. *Generation of Typical Meteorological Year for 26 SOLMET Stations*; Sandia Laboratories: Albuquerque, NM, USA, 1978.
14. Chan, A.L.S.; Chow, T.T.; Fong, S.K.F.; Lin, J.Z. Generation of a Typical Meteorological Year for Hong Kong. *Energy Convers. Manag.* **2006**, *47*, 87–96. https://doi.org/10.1016/j.enconman.2005.02.010.
15. Seyed Salehi, S.S.; Kalamees, T.; Kurnitski, J.; Thalfeldt, M. New Typical Meteorological Year Generation Method Based on Long-Term Building Energy Simulations. *Build. Environ.* **2024**, *256*, 111504. https://doi.org/10.1016/j.buildenv.2024.111504.
16. Lall, S.; Rajasekar, E.; Arya, D.S.; Natarajan, S. Data-Driven Approach to Generate Test Reference Year Weather Files for Building Energy Simulations. *J. Build. Eng.* **2025**, *111*, 113218. https://doi.org/10.1016/j.jobe.2025.113218.
17. Oko, C.O.C.; Ogoloma, O.B. Generation Of a Typical Meteorological Year for Port Harcourt Zone. *J. Eng. Sci. Technol.* **2011**, *6*, 204–214.
18. Chan, A.L.S. Generation of Typical Meteorological Years Using Genetic Algorithm for Different Energy Systems. *Renew. Energy* **2016**, *90*, 1–13. https://doi.org/10.1016/j.renene.2015.12.052.
19. Hosseini, M.; Bigtashi, A.; Lee, B. A Systematic Approach in Constructing Typical Meteorological Year Weather Files Using Machine Learning. *Energy Build.* **2020**, *226*, 110375. https://doi.org/10.1016/j.enbuild.2020.110375.
20. Zhang, W.; Li, H.; Wang, M.; Lv, W.; Huang, J.; Yang, L. Enhancing Typical Meteorological Year Generation for Diverse Energy Systems: A Hybrid Sandia-Machine Learning Approach. *Renew. Energy* **2024**, *225*, 120369. https://doi.org/10.1016/j.renene.2024.120369.
21. Finkelstein, J.M.; Schafer, R.E. Improved Goodness-of-Fit Tests. *Biometrika* **1971**, *58*, 641–645. https://doi.org/10.1093/biomet/58.3.641.
22. Kalamees, T.; Jylhä, K.; Tietäväinen, H.; Jokisalo, J.; Ilomets, S.; Hyvönen, R.; Saku, S. Development of Weighting Factors for Climate Variables for Selecting the Energy Reference Year According to the EN ISO 15927-4 Standard. *Energy Build.* **2012**, *47*, 53–60. https://doi.org/10.1016/j.enbuild.2011.11.031.
23. Seo, D.; Huang, J.; Krarti, M. Evaluation of Typical Weather Year Selection Approaches for Energy Analysis of Buildings. *ASHRAE Trans.* **2009**, *115*, 654–667.
24. Xie, H.; Eames, M.; De Grussa, Z.; Mylona, A.; Challenor, P. Creating State-of-the-Art Weather Files to Enable a Climate-Resilient Built Environment in the UK. *Build. Serv. Eng. Res. Technol.* **2026**, *47*, 165–177. https://doi.org/10.1177/01436244261423730.
25. Pissimanis, D.; Karras, G.; Notaridou, V.; Gavra, K. The Generation of a "Typical Meteorological Year" for the City of Athens. *Sol. Energy* **1988**, *40*, 405–411. https://doi.org/10.1016/0038-092X(88)90095-3.
26. Aliabadi, A.A.; McLeod, R.M. The Vatic Weather File Generator (VWFG v1.0.0). *J. Build. Eng.* **2023**, *67*, 105966. https://doi.org/10.1016/j.jobe.2023.105966.
27. Vilasboas, I.F.; Da Silva, J.A.M.; Venturini, O.J. On the Summarization of Meteorological Data for Solar Thermal Power Generation Forecast. *Energies* **2023**, *16*, 3297. https://doi.org/10.3390/en16073297.

28. Zou, Y.; Chen, Z.; Lou, S.; Huang, Y.; Xia, D.; Cao, Y.; Li, H.; Lun, I.Y.F. Accelerating Long-Term Building Energy Performance Simulation with a Reference Day Method. *Build. Simul.* **2024**, *17*, 2331–2353. https://doi.org/10.1007/s12273-024-1190-x.
29. Chen, Y.; Bhattacharya, S.; Pang, Z.; Sivaraman, D.; Huang, S.; Vrabie, D. Assessing Building Energy Performance via Selection of Representative Simulation Days. In Proceedings of the Building Simulation 2019: 16th Conference of IBPSA, Rome, Italy, 2–4 September 2019; pp. 4825–4832.
30. Hassan, M.A.; Khalil, A.; Abubakr, M. Selection Methodology of Representative Meteorological Days for Assessment of Renewable Energy Systems. *Renew. Energy* **2021**, *177*, 34–51. https://doi.org/10.1016/j.renene.2021.05.124.
31. Hoffmann, M.; Kotzur, L.; Stolten, D.; Robinius, M. A Review on Time Series Aggregation Methods for Energy System Models. *Energies* **2020**, *13*, 641. https://doi.org/10.3390/en13030641.
32. Teichgraeber, H.; Brandt, A.R. Clustering Methods to Find Representative Periods for the Optimization of Energy Systems: An Initial Framework and Comparison. *Appl. Energy* **2019**, *239*, 1283–1293. https://doi.org/10.1016/j.apenergy.2019.02.012.
33. Melro, B.; Coelho, G.B.A.; Baltazar, L.G. Hygrothermal Simulation of Buildings: A Study on the Different Types of Weather Files. In Proceedings of the World Summit: Civil Engineering-Architecture-Urban Planning Congress (CAUSummit 2024), Antalya, Turkiye, 2 September 2024; Pensoft Publishers: Antalya, Turkiye; Volume 7, pp. 118–127.
34. Salonvaara, M.; Sedlbauer, K.; Holm, A.; Pazera, M. Effect of Selected Weather Year for Hygrothermal Analyses. In *Proceedings of the 11th International Conference on Thermal Performance of the Exterior Envelopes of Whole Buildings (Buildings XI)*; ASHRAE: Clearwater, FL, USA, 2010.
35. Crawley, D.B.; Lawrie, L.K. Development of Global Typical Meteorological Years (TMYx). Available online: https://climate.onebuilding.org/WMO_Region_6_Europe/GBR_United_Kingdom/index.html#IDENG_England (accessed on 5 May 2026).
36. U.S. Department of Energy EnergyPlus Weather File (EPW) Data Dictionary. Available online: https://bigladdersoftware.com/epx/docs/9-0/auxiliary-programs/energyplus-weather-file-epw-data-dictionary.html (accessed on 8 November 2026).
37. Marion, W.; Urban, K. *User's Manual for TMY2s: Typical Meteorological Years Derived from the 1961–1990 National Solar Radiation Data Base*; National Renewable Energy Laboratory: Golden, CO, USA, 1995.
38. Wilcox, S.; Marion, W. *Users Manual for TMY3 Data Sets*; National Renewable Energy Laboratory: Golden, CO, USA, 2008.
39. Li, H.; Huang, J.; Hu, Y.; Wang, S.; Liu, J.; Yang, L. A New TMY Generation Method Based on the Entropy-Based TOPSIS Theory for Different Climatic Zones in China. *Energy* **2021**, *231*, 120723. https://doi.org/10.1016/j.energy.2021.120723.
40. Nik, V.M. Making Energy Simulation Easier for Future Climate—Synthesizing Typical and Extreme Weather Data Sets out of Regional Climate Models (RCMs). *Appl. Energy* **2016**, *177*, 204–226. https://doi.org/10.1016/j.apenergy.2016.05.107.
41. Crawley, D.B.; Hand, J.W.; Lawrie, L.K. Improving the Weather Information Available to Simulation Programs. In *Proceedings of the Building Simulation '99*; IBPSA: Kyoto, Japan, 1999.
42. Ayoub, M. Generation of Typical Weather Datasets for Daylight Simulations Using the Enhanced Simplified Sandia Method. *J. Build. Eng.* **2025**, *111*, 113223. https://doi.org/10.1016/j.jobe.2025.113223.
43. MeteoSwiss. *Table of Medium Wind Speeds According to Beaufort Wind Scale*; Federal Office of Meteorology and Climatology MeteoSwiss: Zurich-Airport, Switzerland, 2023.
44. Deng, Y.; Chen, M.; Gao, P.; Luo, R.; Li, Z.; Yao, Y. Identifying Potential Ligand-Receptor Interactions by Integrating LSTM Network and the Attention Mechanism for Cell-Cell Communication Prediction. *J. Transl. Med.* **2026**, *24*, 942. https://doi.org/10.1186/s12967-026-08033-0.
45. Zhou, Z.; Chen, M. CIM-DTA: Conditional Interaction Modulation for Cold-Start Drug–Target Affinity Prediction. *IEEE Trans. Comput. Biol. Bioinforma.* **2026**, 1–11. https://doi.org/10.1109/TCBBIO.2026.3741529.
46. Zhou, Z.; Chen, M.; Zhou, W.; Xiao, Q. Similarity-Adaptive Interaction Routing for Drug–Target Affinity Prediction. *J. Biomed. Inform.* **2026**, *182*, 105100. https://doi.org/10.1016/j.jbi.2026.105100.
47. Duffie, J.A.; Beckman, W.A. *Solar Engineering of Thermal Process*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006; ISBN 978-0-471-69867-8.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.