Build Simul
https://doi.org/10.1007/s12273-026-1416-1

Weather files for building simulation in Brazil: A national benchmark

Mario Alves da Silva[1,2] (✉), Giovanni Pernigotto[1], Alessandro Prada[3], Andrea Gasparella[1], Joyce Correna Carlo[2]

1. Free University of Bozen-Bolzano, NOI Techpark - via Bruno Buozzi 1, 39100, Bolzano, Italy
2. Federal University of Viçosa, avenue PH Rolfs, 36570-900 Viçosa, Brazil
3. University of Trento, Via Mesiano 77, 38123, Trento, Italy

Research Article

Abstract

Weather files strongly influence building performance simulation, especially in climatically diverse countries such as Brazil. Yet, the representativeness of typical meteorological years (TMY) for Brazilian applications remains insufficiently assessed. This study aims to evaluate four established TMY approaches and to introduce a performance-based alternative (Brazilian typical meteorological year, BTMY) that incorporates building-specific climatic sensitivity. Using ERA5-Land data for 480 locations, we compiled 15 actual meteorological years (2008–2022) per site and assessed their impact on operative temperature and energy needs in representative residential models. Spearman filtering removed collinear meteorological parameters, and optimized XGBoost and ε-SVR models predicted daily operative temperature with $R^2 = 0.91 \pm 0.03$ (RMSE = 0.64 ± 0.19 °C) and daily energy needs with $R^2 = 0.87 \pm 0.05$ (RMSE = 0.05 ± 0.01 kWh m^{-2} yr^{-1}), identifying dry-bulb temperature as the dominant meteorological parameter (75% importance for operative temperature; 59% for energy needs). Against 15-year simulations, minimum Finkelstein–Schafer and Best Rank I achieved the lowest mean operative-temperature bias (0.15 °C, SD 0.09–0.10 °C), ISO was similar (0.16 ± 0.13 °C), while Pissimanis performed worst (0.23 ± 0.27 °C). BTMY weather files matched the TMY methods and clarified key climatic drivers. Results showed that ERA5-Land data source mattered more than the reference-year method. Finally, the study delivered more than 3000 new weather files to support simulation practice and policy development.

Keywords

weather data
typical years
sensitivity analysis
building performance simulation
machine learning
residential models

Article History

Received: 05 September 2025
Revised: 08 January 2026
Accepted: 14 January 2026

1 Introduction

The advances in computer hardware and software in the last decades enable assessing building performance regarding different metrics by associating geometric characteristics, constructive properties, and climatic conditions to complex mathematical models. By combining these data, the mathematical models can describe the building performance in a transient regime as a function of the climatic conditions given by the simulator. Building performance simulation (BPS) enables the estimation of different building performance metrics, such as indoor environmental quality (IEQ) and energy consumption. Users can create different weather files from the same dataset, based on different methodologies to select the representative year or months (Herrera et al. 2017). The impact of different weather files on building performance has been discussed for more than 40 years (Crawley 1998). However, studies showed that users must be aware of the approach used for weather file development and carefully consider the implications of their choices regarding the adopted weather data (Pernigotto et al. 2014), by testing versions of weather files obtained from different approaches and assessing their impact on simulated building performance.

Different methodologies can be pointed out to characterize the average climatic conditions, which is the primary purpose of most weather files from the last century to the present (Herrera et al. 2017). They usually apply statistical analysis by selecting representative months from different years and creating a typical/average year containing hourly data from a historical series (Herrera et al. 2017). The test reference years (TRY) were initially used to portray the

E-mail: mario.alvesdasilva@unibz.it

Building Thermal, Lighting, and Acoustics Modeling

climatic conditions for BPS applications (Hall et al. 1978). These files describe the average climate, which happened in a selected year, through variables such as dry bulb temperature, wet bulb temperature, dew point temperature, wind direction and wind speed, pressure, relative humidity, cloud cover, and cloud type. Later, several locations benefitted from modifying the average year to average months (typical meteorological years, TMY), where specific changes were made in different situations, such as the proposals by Pissimanis et al. (1988), Festa and Ratto (1993), Layi Fagbenle (1995), and Lund (1995).

Studies pointed out that using different weather files on BPS can lead to different results (Crawley 1998; de Miguel and Bilbao 2005; Pyrgou et al. 2017). In the late 1990s, Crawley (1998) showed the implications of using TRY and TMY for the performance assessment of commercial buildings. de Miguel and Bilbao (2005) presented a new methodology for TRY weather files and discussed the use of different radiation data sources. Pyrgou et al. (2017) mainly discussed the representativeness of weather data and the impacts of microclimates. The overall recommendation was related to the lack of representativeness of TRY weather files and the necessity of TMY with records closer to the long-term.

The efforts to update the weather files methodology, combined with the advances in the computational field, enabled the creation of weather files for several locations across the world. The procedures of collecting and producing weather data, through the combination of measured/observed data with mathematical models and satellite records, allow users to easily access hourly data collections containing records of the main meteorological parameters in a weather file. Databases such as ERA5-reanalysis (Hersbach et al. 2020), ERA5-Land (Muñoz-Sabater et al. 2021), and NASA-POWER (Zhang et al. 2008) provide an extensive source of weather quantities for several locations throughout the world. The databases are frequently used to gather data and create weather files that are later required for BPS assessments (Afshari 2023; Wu et al. 2023).

The possibility of creating weather files for any location on the globe also allows assessing the performance of buildings worldwide. Different studies show approaches based on updating weather files or even creating new procedures to process the meteorological data used for building performance simulation, such as the TMY3 proposed by Wilcox and Marion (2008) or the calibrated TMY by Yuan et al. (2022). Throughout the world, different studies propose using the EN ISO 15927-4 (European Committee for Standardization 2005) for creating new weather files based on the original method or by adjusting weights to represent better the observed historical period, as suggested by Kalamees et al. (2012), Kim et al. (2017), and Pernigotto et al. (2014). Furthermore, the impact of different weather variables on building performance changes for different building typologies, underlining in such a way the need for the creation of adapted weather files. Once users alter the process by understanding the climatic influence under different building typologies in different locations, creating weather files can be adapted to site-specific conditions (Hosseini et al. 2020; Bigtashi et al. 2024; Papakyriakou et al. 2024). However, users must take caution regarding the method used to compile the weather files, since they are highly influenced by the weather variables used to determine the representative year.

Building performance is site dependent. Therefore, it is directly related to the weather conditions at the location used for the simulation. Hosseini et al. (2020) and Kalamees et al. (2012) proposed a similar method for selecting the reference year for each month to generate a new typical meteorological year. Both studies first analyse the impact of meteorological parameters on the energy demand of different building typologies. Then, the most significant meteorological parameters are employed to define the representative years. The studies, however, differ in the sensitivity analysis approach to define parameters and weights. Kalamees et al. (2012) defined the range of variation of each meteorological parameter and then compared their influence on residential and non-residential typologies. Hosseini et al. (2020) used the results of building performance simulations with actual meteorological years (AMYs), and a machine learning method to define the weights by identifying the impact of each meteorological parameter on the building performance. The results indicate a more accurate energy estimation and better correlation to site conditions than generic methods. Papakyriakou et al. (2024) presented an update in the methodology of Hosseini et al. (2020) by summarizing the weight based on the climate zones instead of a different set for each location. In this case, the results were promising, and no performance loss was identified when using the climate zone-based instead of the site-specific weather files.

In the Brazilian context, different studies addressed the use, creation, and modification of weather files. Machado et al. (2019), Bonini et al. (2022), and Tippett et al. (2024) discussed the creation of weather files focusing on solar radiation but without considering BPS applications. Despite different studies dealing with climate and building performance in Brazil, they present a limitation regarding the different Brazilian climatic conditions (e Silva Machado et al. 2024; da Silva et al. 2024) or the building particularities within the territory (Melo et al. 2014; Triana et al. 2015). Some recent studies are related to moisture (Zanoni et al. 2023), thermal comfort (de Sousa et al. 2023), climate zoning (da Silva et al. 2025; Machado et al. 2025), and climate change (Bracht et al. 2024; Krelling et al. 2024), but do not

address the impact of different weather files on BPS. In this context, the originality of this study lies in its scale and building-centred benchmarking design. We provide a national assessment of weather-file representativeness for Brazilian building simulation using a consistent ERA5-Land dataset (2008–2022), covering 480 locations and 15 AMYs per site. Because all candidate weather files are derived from the same underlying database, the comparison isolates the effect of the reference-year selection rule on building-performance outcomes. We also include performance-based weather files because fixed-weight TMY schemes, often centred on dry-bulb temperature, may be insufficient when used alone to represent building response across diverse climates and building envelope characteristics, particularly when considering the specific features of Brazilian social-housing typologies. Therefore, we use machine learning as an interpretable sensitivity step that translates AMY simulation outcomes into explicit, reproducible weights for reference-year selection, following Hosseini et al. (2020) and subsequent refinements (Bigtashi et al. 2024; Papakyriakou et al. 2024). This is implemented through complementary steps: a Spearman-based pre-processing stage to reduce multicollinearity and remove redundant predictors, two tuned models (XGBoost and ε-SVR) applied in parallel to stabilize feature-importance estimates and limit dependence on a single algorithm, and an independent validation stage combining the Mann-Whitney U-test with root mean squard error (RMSE), coefficient of variation of root mean squared error (CVRMSE) and normalized mean bias error (nMBE) by comparing hourly AMY simulations and the derived TMYs. Finally, the analysis spans Tropical (A), Arid (B) and Temperate (C) climates (Köppen-Geiger) and supports a reference-year selection strategy that can be adapted to both site and building specificity across the Brazilian territory. The workflow also produces site-specific and climate-zone performance-based typical years and delivers more than 3000 new weather files to support Brazilian simulation practice and policy applications.

In summary, this paper makes three main contributions: (i) a national benchmark of four TMY methods; (ii) a performance-based Brazilian typical meteorological year (BTMY) approach that translates building response into explicit climatic weights for reference-year selection; and (iii) a large repository of AMY/TMY/BTMY weather files for 480 locations.

2 Methods

Figure 1 shows an overview of the methodology adopted in this study. The method encompasses the process of selecting the locations along with collecting and processing of weather records (Section 2.1). Section 2.2 describes the creation of TMY, while performance-based weather files are described in Section 2.3. Section 2.4 presents the building models and simulation settings. Finally, the overall assessment to define the suitability of different weather files to building performance simulation in Brazil is presented in Section 2.5.

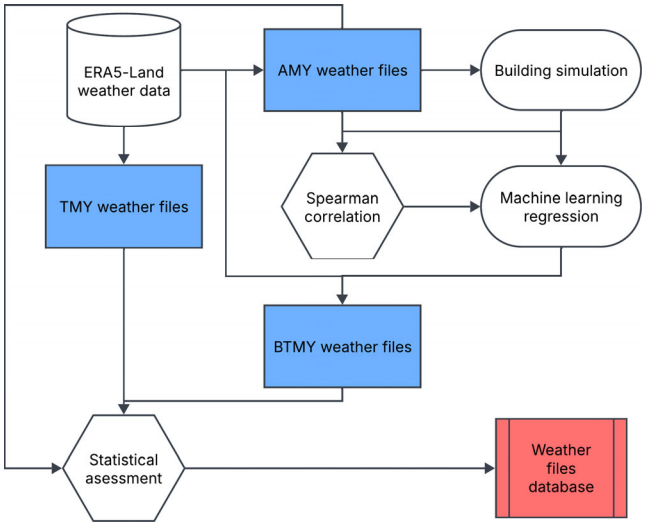

Fig. 1 Summary of the methodological steps adopted in this study (the blue shaded indicates intermediary products and the red shade indicates the main product of this study)

2.1 Locations selection: weather data retrieval, processing, and AMY compilation

There are 5570 municipalities in Brazil, according to the Brazilian Institute of Geography and Statistics (IBGE), distributed in five Major Regions. The North and Central-West comprise the lowest number of municipalities, 450 (8.1%) and 467 (8.4%). The South region has the third position with 1191 (21.4%) municipalities. The Southeast and Northeast regions have the highest number of municipalities, corresponding to more than 60% of the Brazilian municipalities, 1668 (29.9%) and 1794 (32.2%), respectively. Following the geographical distribution and a comprehensive assessment of the climatic conditions of Brazil from 2008 to 2022 (da Silva et al. 2024), this study selected 480 locations as a sample for the subsequent analysis process. The locations not only provide a reliable geographical variability by encompassing different climatic conditions but also provide a comprehensive characterization of the territory in terms of temperature, humidity, and solar radiation (Figure 2).

The ERA5-Land database was adopted as the main source to download all the weather data for the 480 Brazilian municipalities, considering the records from 2008 to 2022. This study compiled all weather files following the EnergyPlus Weather File format (EPW), since it can be employed by different building simulation codes – EnergyPlus (Crawley

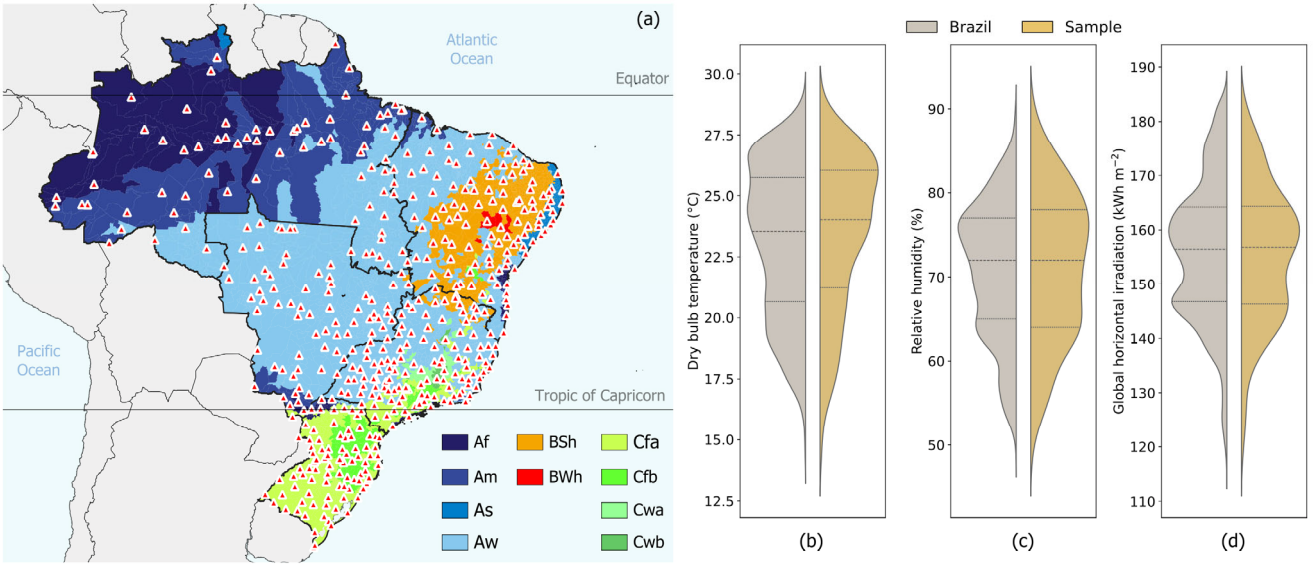

Fig. 2 The 480 locations (red triangles with white outline) used in this study overlapping the Köppen-Geiger classification (a) presented by da Silva et al. (2024), and the distribution of the annual records of dry bulb temperature (b), relative humidity (c), and global horizontal irradiation (d)

et al. 2001), TRNSYS (Beckman et al. 1994), and IES VE (Integrated Environmental Solutions 2024). The format represents a comprehensive database of meteorological variables, but not all are used for building performance simulation, depending on the building simulation code considered by the simulator. This study adopted EnergyPlus 23.2 to perform all the simulations. Therefore, the weather quantities gathering and processing followed the software requirements.

Dry bulb temperature (DBT), dew point temperature (DPT), pressure, precipitation, global horizontal irradiation (GHI), horizontal infrared radiation, wind vector for the U and V components, and albedo for the 15 years were downloaded from the ERA5-Land repository (Muñoz-Sabater et al. 2021). From the downloaded variables, only the albedo from weather files is not used in EnergyPlus. However, the hourly records were downloaded to create the weather files with as much data as possible since future users can use the data in other types of analysis and software. MetPy plugin for Python (May et al. 2022) was utilized to calculate the wind speed and direction from the vectorial components. The plugin was also applied to calculate the relative humidity (RH) as the ratio between the saturation vapor pressure from DPT (Bolton 1980). The Boland–Ridley–Laurent (BRL)-Brazil model was employed to split the GHI on direct normal (DNI) and diffuse horizontal (DHI) hourly irradiation values (Lemos et al. 2017). The final EPW also contained the ground temperatures (Lawrie 2003) and the typical/extreme periods following the procedure described by Rodrigues et al. (2023). Then, for each one of the 480 locations, 15 AMYs were compiled.

After creating the AMY weather files, a primary assessment was carried out to summarize the weather quantities through climatic performance indicators, as described in Bracht et al. (2024). This study used the heating and cooling degree hours, with base temperatures of 16 °C for heating (HDH_{16}) and 26 °C for cooling (CDH_{26}), as proposed by Bracht et al. (2024). The analysis also considered the annual summaries of DNI and DHI, and the maximum, minimum, and average RH. The AMY weather files were also used in EnergyPlus simulations that will be later described in Section 2.4.

2.2 TMY compilation

This section describes the adopted existing weather files compilation methodologies and the comparison between their results and the AMY simulation outcomes. This study used the 15-year data to create TMY weather files following the the EN ISO 15927-4 (European Committee for Standardization 2005), the minimum Finkelstein–Schafer (Layi Fagbenle 1995), the Pissimanis (Pissimanis et al. 1988), and the Best Rank I (Pernigotto et al. 2014) methods.

The EN ISO 15927-4, the minimum Finkelstein–Schafer, and the Best Rank I methods follow the initial steps of the Sandia method (Hall et al. 1978), described as follows:

1) Calculate the daily average $\overline{p}$ of the climatic parameters p (DBT, RH, and GHI), for every month m and year y.
2) List all daily averages $\overline{p}$ from the multiyear records, sorting in increasing order all the averages $\overline{p}$, and calculate the cumulative distribution function (CDF) for each daily record i^{th} as shown in Equation (1):

$$\Phi(p,m,i) = \frac{K(i)}{N+1} \quad (1)$$

where $K(i)$ is the rank order of the i^{th} value of the daily means and N is total number of days for a month over the multiyear series.

3) Assemble all daily averages $\bar{p}$ for every calendar year y, sorting in increasing order all the averages $\bar{p}$, and calculate the CDF for each daily record i^{th} as shown in Equation (2):

$$F(p,y,m,i) = \frac{J(i)}{n+1} \quad (2)$$

where $J(i)$ is the rank order of the i^{th} value of the daily means and n is the number of days of the month for the calendar year.

4) Calculate the Finkelstein–Schafer statistic FS(p, y, m) for each year and climatic parameter using Equation (3):

$$\mathrm{FS}(p,y,m) = \sum_{i=1}^{n} \left| F(p,y,m,i) - \Phi(p,m,i) \right| \quad (3)$$

2.2.1 Best Rank I, EN ISO 15927-4, and minimum Finkelstein–Schafer reference years

After calculating the FS statistic, ranks are assigned to each calendar year based on the increasing order of the statistic, for every month and variable. Then, for every month, the ranks are summed and sorted in increasing order. According to the Best Rank I method, the year with the lowest rank is selected as the reference.

For the EN ISO 15927-4, the three months with the lowest rank are initially selected. Then, the calendar year with the lowest monthly wind speed deviation with respect to the long-term records is chosen as the reference year.

The minimum Finkelstein–Schafer approach, likewise the Best Rank I method, bases the selection of the reference year on the FS values. However, the approach uses the lowest summed FS statistic instead of assigning ranks.

2.2.2 Pissimanis reference years

The Pissimanis method significantly differs from the previous selection process, since it mainly relies on the GHI deviation between each calendar year and the multiyear records. First, the RMSE is calculated for the hourly GHI values. Then, the years with deviations below a certain threshold or a pre-defined minimum value are selected as the reference years. In case the deviation analysis results in more than one candidate year, the FS statistic is calculated for the GHI and DBT variables, and the year with the lowest FS is selected as the reference year for the specific month.

2.3 Performance-based weather files

Hosseini et al. (2020) proposed a systematic approach to create weather files for building performance simulation using a methodology that combined AMY simulations and machine learning as an interpretation tool to determine the appropriate weight for different weather quantities and select the reference years to compile weather files. The method was then improved by including a pre-processing stage to deal with multicollinearity and a sequential machine learning regression model (Bigtashi et al. 2024). Lately, the method evolved to a proposal for climate zone-based weights, instead of unique values for each distinct location (Papakyriakou et al. 2024). This study seeks to combine both approaches. First, the study intends to create performance-based weather files for each location, and, afterwards, the analysis will focus on the creation of climate zone-based weights. The climate zone follows the proposal of e Silva Machado et al. (2024) updated by da Silva et al. (2024), according to the ERA5-Land records. In summary, the performance-based method can be described in four main steps and one additional for the climate zone-based methodology:

1) Compile weather files based on AMYs, define building models and simulation outputs, and run the simulations.
2) Summarize the weather variables and simulation outputs that will be used in the machine learning process according to the desired resolution (daily, weekly, or monthly).
3) Select a machine learning algorithm, create a prediction model based on the weather variables and performance results, and extract the impact of each weather variable on the simulation outputs.
4) Define a methodology to select the reference year for each month using the impact of each weather variable as weights.
5) Climate zone-based: average the site-specific weights (step 3) based on a climate zoning distribution, create new weather files using the definitions from the previous step, and quantify the differences between the location-based method.

For step 2, the weather data records were summarized as the daily average, maximum, minimum, and spread of DBT, DPT, RH, and water vapor pressure (WVP), and daily integral of GHI, DNI, and DHI. The simulation outputs were summarized as the daily average operative temperature and daily total energy needs (cooling and heating). According to the Brazilian standard, NBR 15575 (ABNT 2024), used as the guide for the building performance simulation settings and presented in Section 2.4, two types of simulation are required: (1) free-floating temperature aimed to assess the operative temperature and (2) with ideal systems to assess

space heating/cooling needs. Considering all weather quantities, a pairwise comparison was performed to identify and remove quantities that were strongly correlated, by the means of the Spearman correlation. Therefore, weather quantities with a pairwise correlation equal or higher than 0.5 or equal or less than −0.5 were filtered and the correlation between those variables and the simulation outputs was tested. The one with the lowest correlation with the simulation outputs was removed from the dataset and not used by the machine learning model.

For step 3, this study adopted the eXtreme Gradient Boosting (XGBoost) (Chen and Guestrin 2016) and the epsilon-support vector regression (SVR) (Platt 1999) machine learning algorithms. The choice of algorithms was based on their fast computation and large applicability to building performance related studies, being suited for complex and non-linear problems (Zhou et al. 2019; Mo et al. 2019; Yan et al. 2022; Alawi et al. 2024). As stated by Bigtashi et al. (2024), the use of XGBoost can also be supported "by its regularization parameters that help reduce overfitting", the algorithm also delivers high quality predictions and deals well with multicollinearity. The use of SVR is based on its capacity of producing regression models that are less pruned to outliers and noisy data (Qi et al. 2025). The use of two algorithms also sought to avoid bias from a single machine learning algorithm approach. The results provided by Alawi et al. (2024) also supported the choice of XGBoost and SVR in this study. Among the different classes used in the study, XGBoost provided the best results along with all the other ensemble learning algorithms. SVR also displayed reliable and stable results, outperforming the widely used classes, as artificial neural networks. Additionally, the combination of both machine learning algorithms represents an innovative approach, since there are no records of them being combined as predictors for generating weights used in the creation of weather files.

Before identifying the impact of each variable on the operative temperature and energy needs of the social housing models, an optimization process was implemented to enhance each algorithm's performance. As a principle, an optimization process requires at least one objective function and one parameter, $F(x)$ and x, respectively. Then, it is necessary to define the parameter range, the optimization algorithm, and the stop criteria. The optimization algorithm performs an exploration throughout the search space defined by the bounds of the parameter x and evaluates the improvements on the $F(x)$. The best result corresponds to the lowest $F(x)$, followed by the parameters that lead to the $F(x)$. Table 1 shows the hyperparameters for the XGBoost and SVR algorithms. For the optimization approach present in this study, the Hyperopt library for Python was used (Bergstra et al. 2013) and the algorithms' accuracy, i.e., the percentage of correct predictions, was defined as the objective function. The stop criterion used was the number of evaluations within the optimization process, which was set to 50 evaluations in this study.

The hyperparameter tuning process combined the optimization process and a k-fold cross-validation process to increase the accuracy of the machine learning regression models. The k-fold cross-validation strategy seeks to use the whole database as training and test samples (Rodriguez et al. 2010). The dataset is split in "k" parts where $k-1$ parts will be used as training and the remaining part will be used as test. The process can also be repeated N times, to increase models' accuracy based on the diversity within the different training sample. For this study, the dataset was split into four parts, and the process was repeated twice. Therefore, each simulation output and machine learning algorithm provided a set of eight regression models.

After identifying the best settings for each algorithm, the regression machine learning models were created, and the Permutation Feature Importance method (Pedregosa et al. 2011) was used to calculate the impact of each parameter on the simulation outputs. The importances, ranging from 0 to 1, were averaged considering the XGBoost and SVR results, and adopted as the new weights for the reference year selection process. For the climate zone-based method, the weights were averaged once again but now considering the Brazilian climate zones. Following the definition of the new weights, two sets of reference years were identified, respectively for the simulation of operative temperature in free-floating mode and the analysis of energy needs with ideal load systems. Figure 3 shows the correlation analysis and the machine learning application within the weather file creation process.

The CDFs were calculated for each month of every calendar year and the 15-year records. Given its robustness and solid statistical basis, the Kolmogorov-Smirnov statistic (Massey 1951) was used to calculate the difference between the long-term and each calendar year CDFs using the daily

Table 1 XGBoost and SVR hyperparameters adopted in the optimization process

Algorithm	Hyperparameter	Bounds
XGBoost	Maximum depth	2 to 15
	Learning rate	0.001 to 0.3
	Subsample ratio	0.5 to 1
	Number of estimators	50 to 2000
SVR	Gamma	Scale or auto
	C	1 to 100
	Max. iterations	500 to 1000
	Epsilon	0.001 to 1

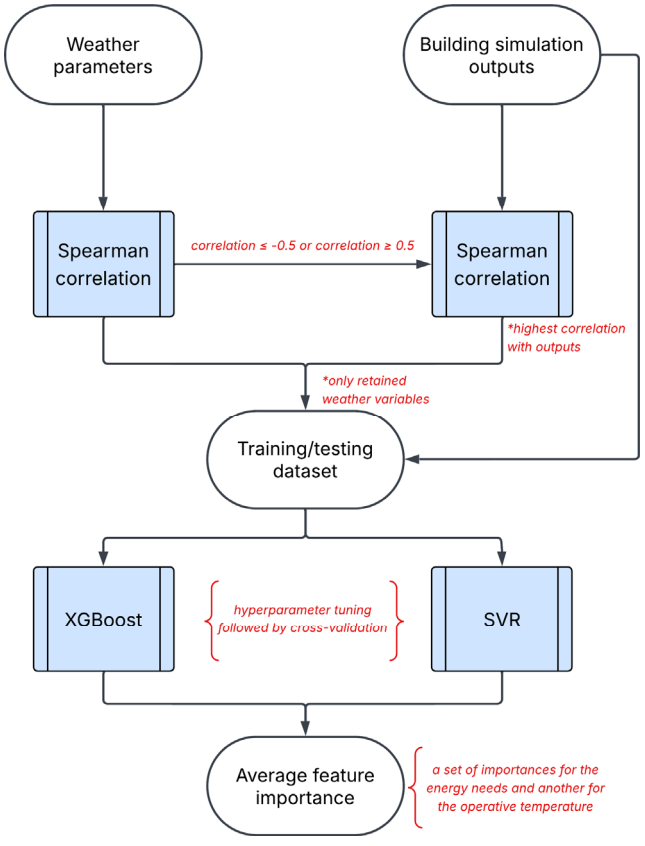

Fig. 3 Proposed methodology based on the correlation analysis and machine learning algorithms

weather quantities for the relevant weather variables identified by the correlation analysis. Then, the test statistics were ranked, and each rank was multiplied by the weights. Finally, the ranks were summed and the one with the lowest rank was defined as the reference. The process, repeated for the different simulation outputs, which are described in Section 2.4, considered the location-based and the climate zone-based weights, i.e., the average of site-specific weights based on the current Brazilian zoning for building performance applications (e Silva Machado et al. 2024; da Silva et al. 2024). For each case, a comparison of the resulting reference years for the operative temperature and for the energy needs pointed out whether the location-based method and the climate-zone-based method required more than a single weather file. After determining this aspect, the created EPW files were used in new simulations with the models and configurations presented in Section 2.4. The approach, called Brazilian Typical Meteorological Year (BTMY), could be implemented with location-based ($BTMY_{SITE}$) or climate zone-based weights ($BTMY_{ZONE}$). The prefix "OT" indicated the reference years based on the operative temperature while the "EUI" indicated the reference years based on energy needs.

2.4 Building archetypes and simulation settings

This study investigated different building archetypes and envelope configurations, since building performance is directly related to building geometry, internal loads, occupancy, and construction components. The residential archetypes described by Triana et al. (2015) and LabEEE et al. (2024) were the basis for the analysis carried out in this study. The residential models are representative typologies based on the social housing designs promoted by the Brazilian government. They follow a similar internal arrangement, encompassing two bedrooms, one bathroom, a shared living room, and kitchen space. The net floor area varies from 43 m^2 to 47 m^2. The building models represent a detached house, a single-story terrace, and a multifamily five-story H-shaped building (Figure 4). Given the complexity of the multifamily typology and the consequent higher demand for computational resources, only the ground floor, the intermediate floor, and the top floor were modelled. This study also modelled only two units per floor, since the building is symmetric.

Schedules for occupancy, lighting, and loads from the Brazilian standard NBR 15575 that describes the minimum performance requirements of residential buildings (ABNT 2024) guided the buildings use configuration. The simulation was repeated in free-floating mode and with an ideal HVAC system. In both cases, the conditioning modes follow the NBR 15575 settings. The natural ventilation model adopts a 19 °C setpoint for opening windows, i.e., ventilation is available for all the occupied hours in which the outdoor temperature is higher than 19 °C and the indoor air temperature is higher than the outdoor. Additionally, the heating setpoint is 21 °C and the cooling setpoint 23 °C (ABNT 2024). The building envelope assumes nine different configurations, following the characteristics shown in Table 2. For the operative temperature, this study adopted the weighted average based on the thermal zones floor area, for each building model, to assign a single value for the whole building. Likewise, the energy needs for cooling and heating were aggregated for the ideal loads model, resulting in the

Table 2 Wall and roof thermal properties (da Silva et al. 2024)

Construction set	Surface	*U*-value (W m^{-2} K^{-1})	Thermal capacity (kJ m^{-2} K^{-1})	Solar absorptance
CS1	Walls	0.41	125	0.3, 0.5, and 0.7
	Roof	0.56	230	0.6
CS2	Walls	2.51	150	0.3, 0.5, and 0.7
	Roof	2.01	21	0.6
CS3	Walls	4.84	220	0.3, 0.5, and 0.7
	Roof	2.41	233	0.6

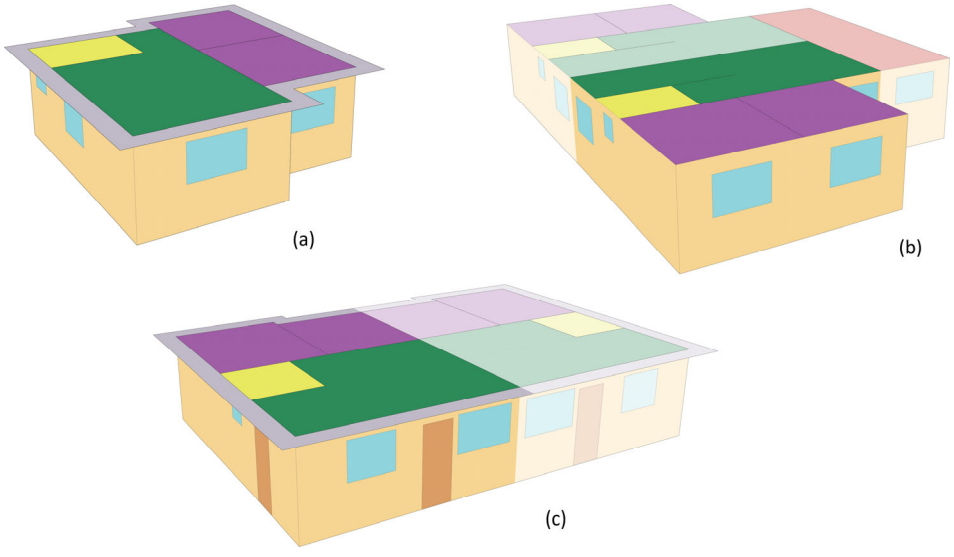

Fig. 4 Detached house (a), multifamily (b), and single-story terrace building geometries (c). The roof color indicates the thermal zones: green indicates the living room, purple indicates the bedrooms, yellow indicates the bathroom, and red indicates the corridor. For the multifamily and single-story terrace models, the low-opacity surfaces indicate the other living units

building envelope configuration and azimuth. Also, the total energy needs. In both cases, the analysis considered each average of all models and building envelope configurations resulted in a single output for each location, containing hourly values for the operative temperature and the total energy needs, representative of all building archetypes used in this study. The average of a single year (8760 hourly records) and the average of a 15-year record (131400 hourly records) were the final assessed outputs.

2.5 Overall assessment

To assess the performance of the typical weather files, new simulations were done using the hourly weather data (Section 2.2) and the building models (Section 2.4). The results were summarized as the average of all models, building envelope configuration, and azimuth. The Mann-Whitney U-test (Mann and Whitney 1947) was used in a goodness-of-fit analysis to compare the hourly distribution of AMY and TMY files. The results were also compared using the RMSE, CVRMSE and nMBE metrics for a practical assessment of deviations. This study adopted the 1 °C threshold for the operative temperature comparison. A CVRMSE of 30% and a nMBE of ±10% were used for energy needs assessment (ASHRAE 2023). The comparison sought to determine which TMY provides the best approximation to the multiyear records.

A BPS annual summary followed the AMY and TMY weather files, considering the different building archetypes and the outcomes for the operative temperature and energy needs throughout the 15-year period. Additionally, after simulating and summarizing the performance-based weather files results, a new comparison was carried out. The Mann-Whitney U-test statistic enabled the assessment of the deviations among the 15-year results, the TMY, and BTMY weather files. A comparison was also carried out using the RMSE, CVRMSE, and nMBE metrics, to determine whether the approaches show a significant difference for the hourly values of operative temperature and energy needs.

3 Results and discussion

3.1 AMY: multiyear summary and archetypes performance assessment

This section presents a brief climatic assessment regarding the air temperature, solar radiation, and relative humidity, since an extensive analysis was already presented in da Silva et al. (2024). The initial assessment of the AMY weather files for the 480 locations showed an average HDH_{16} of 2378 K·h with a standard deviation of 4424 K·h, and a median HDH_{16} of 172 K·h. These values highlight the asymmetrical distribution of the degree hours for heating, emphasized by more than 75% of the locations showing an HDH_{16} below the average. The heating degree hours varied from 0 K·h in 153 locations to a maximum of 24423 K·h in Urubici-SC. The highest HDH_{16} occurs in a location within the 1R bioclimatic zone, characterized as a very cold zone with a severe winter. Regarding the cooling degree hours, the average CDH_{26} was 8141 K·h, varying from 56 K·h in Urubici-SC to 26791 K·h in Piripiri-PI. Likewise, a correlation

between the highest CDH_{26} and the bioclimatic zoning can be made, since it occurred in a very high and dry zone (6B). Therefore, the DBT dependent analysis showed a strong correlation with the Brazilian bioclimatic zoning, with the climatic summary extremes corresponding to the extreme bioclimatic zones (Figure 5).

Regarding the solar radiation summaries, the DNI component showed an average of 698 kWh m^{-2} yr^{-1}, varying from 934 kWh m^{-2} yr^{-1} in São Gabriel da Cachoeira-AM to 2312 kWh m^{-2} yr^{-1} in Ibotirama-BA. The DNI results were directly related to the climatic environment within the locations, since the first is located within the Amazon rainforest region and the second is located within the Semiarid region. The humidity and consequently correlation between nebulosity and precipitation directly resulted in the DNI conditions shown in both locations. Nevertheless, the DHI component with an average 698 kWh m^{-2} yr^{-1} were more influenced by the combination of the latitude and nebulosity within the Brazilian territory. Quaraí-RS, in the South region, had the lowest DHI (507 kWh m^{-2} yr^{-1}) while São Gabriel da Cachoeira-AM, in the North, presented the highest DHI (941 kWh m^{-2} yr^{-1}). The humidity results showed an average RH of 73%, varying from 52% in Barra-BA to 91% in São Gabriel da Cachoeira-AM. The results restate the impact of the Amazon rainforest and the Semiarid over the climatic conditions shown in these locations. Therefore, the results highlight the influence of the most humid and the hottest and dryest environments found in Brazil.

Considering the building archetypes, the variation between the maximum and minimum operative temperature showed an average of 1 °C, with a standard deviation of 0.2 °C. The operative temperature variation presented a uniform distribution across the territory considering both the geographical and the bioclimatic zoning. The energy needs presented an average variation of 28% with a relative deviation of 39%. Likewise, the energy needs variation showed an even distribution for most of the country, except for the South region and, consequently, for the coldest bioclimatic zones. The higher variation, up to 63% for the zone 1R locations, is a direct consequence of the lower energy needs in the region and the numerical influence of slight changes between models, reflecting in higher variations compared to other locations.

From the detached house to the top floor of the multifamily building, the archetypes exhibited an increasing average operative temperature, mainly related to the envelope configuration and the geometry of each model. The energy needs results also demonstrated a high correlation with the envelope boundary conditions. While the top floor of the multifamily building remains as the model with the highest energy needs, the ground floor and intermediary have the lowest values, and the detached house and single-story terrace present intermediary conditions.

3.2 Performance-based weather files assessment

The correlation results mostly displayed positive interactions between the weather quantities (Figure 6). The average

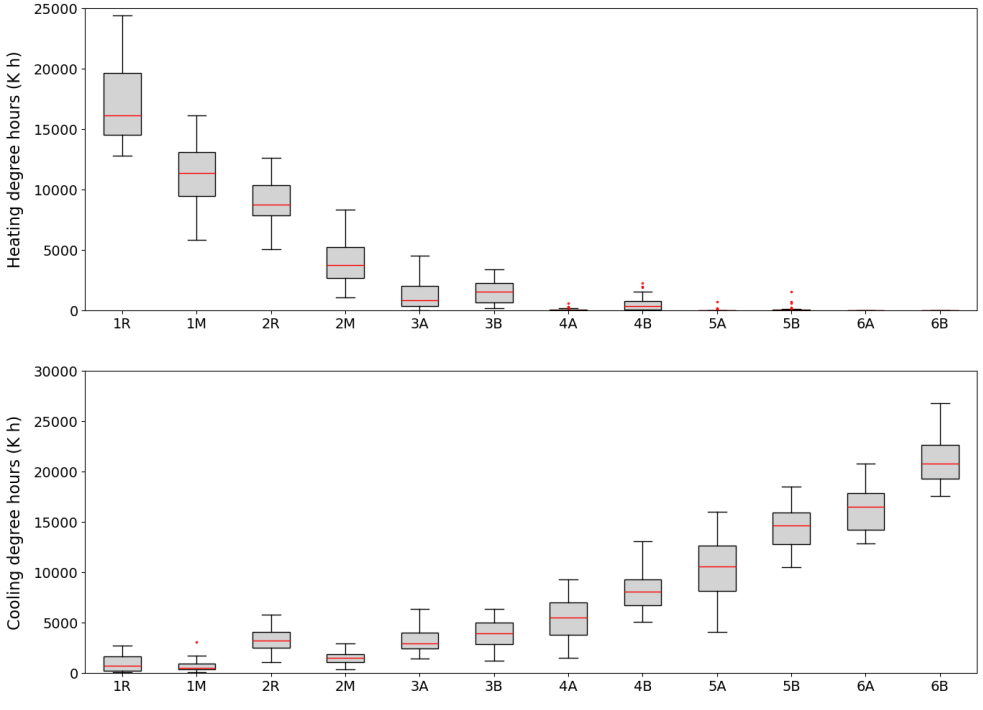

Fig. 5 HDH_{16} and CDH_{26} distribution among the Brazilian bioclimatic zones

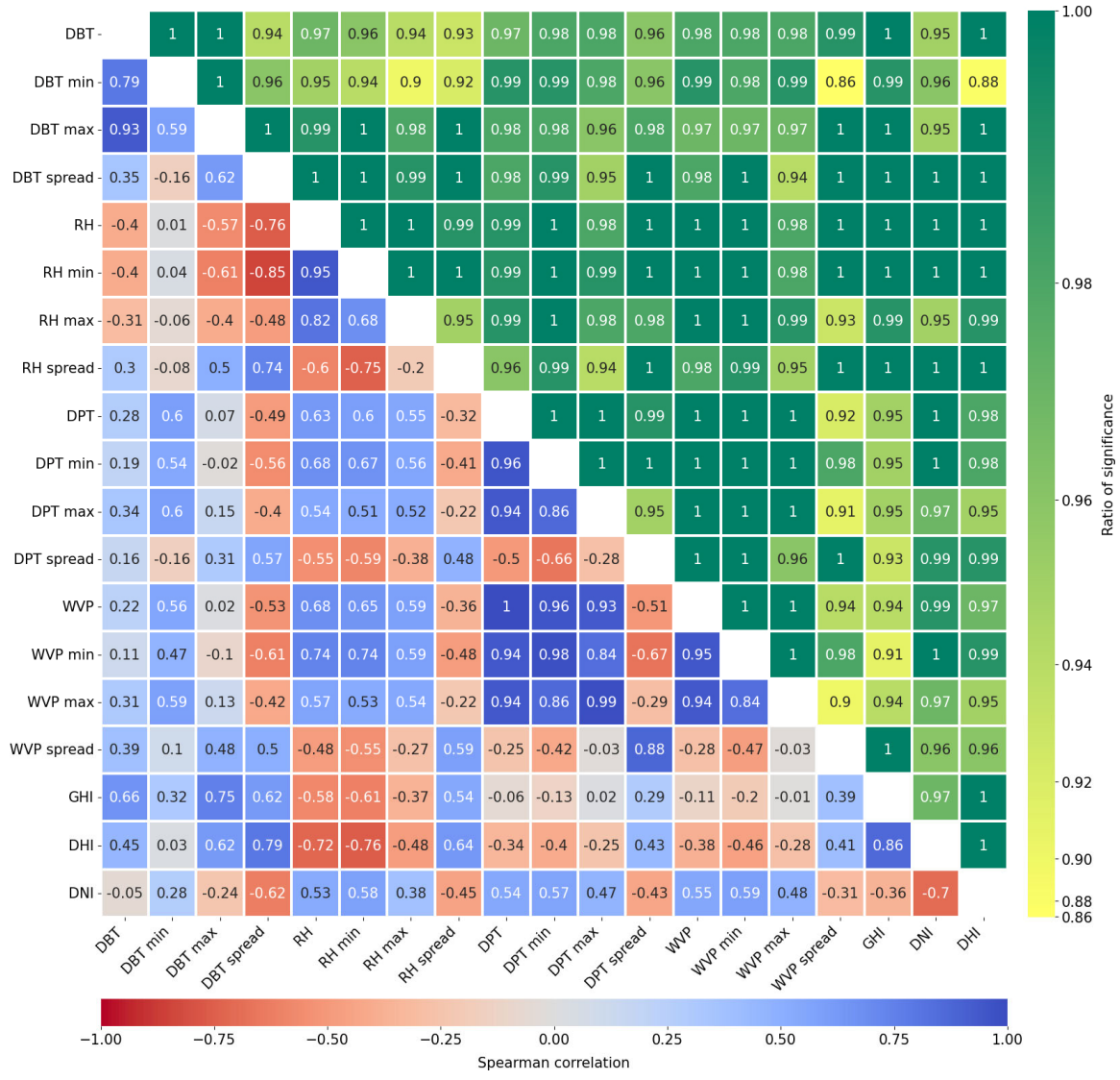

Fig. 6 Average Spearman correlation statistic (lower triangle) and the ratio of statistically significant correlations, i.e., occurrences of p-value ≤0.05 divided by the number of locations (upper triangle)

Spearman statistic showed 98 positive correlations (57%), varying from 0.012 (RH – DBT min) to 0.996 (WVP – DPT). The positive correlations also presented an average of 0.557 with a standard deviation of 0.259. Considering only the most significant positive correlations, i.e., equal or above 0.5, the distribution shows a uniform pattern, with a gradual decrease, that starts with 63 pairwise results equal or higher than 0.05, reaching still 13 occurrences above 0.9, and six that are higher than 95%. For the negative correlations (43%), the average statistic was −0.39, with a standard deviation of 0.206, varying from −0.006 (GHI – WVP max) to −0.847 (RH min – DBT spread). Likewise, the distribution displays a uniform pattern, but with considerably less correlations deemed significant. Most of the statistics falls outside the negative correlation threshold, with only 22 occurrences equal or lower than −0.5, reaching a single occurrence that surpasses the −0.8 mark. DPT and WVP presented the overall strongest correlation, since they are intrinsically physically correlated. Even though the correlation between those variables is known and demonstrated by physical equations, the Spearman test was used to quantify and potentially define their suitability as predictors for the machine learning models. The significance of the Spearman correlation analysis corroborates with the results, since most of the pairwise results showed a p-value within the 0.05 threshold. The ratio of significance, i.e., the number of p-values ≤ 0.05 divided by the number of locations (480), varied from 0.86 to 1, with an average ratio of significance of 0.98 and a standard deviation of 0.03.

After identifying the correlation among the weather quantities, a similar analysis was done with the simulation outputs. The number of weather quantities considered

as input for the machine learning regression model was significantly reduced, based on the Spearman correlation, by performing a pairwise comparison for the weather quantities and by pairing the weather quantities and the simulation outputs. The results showed that all DBT quantities had the strongest correlation with the operative temperature and the energy needs, except for the DBT spread. Apart from DBT, the GHI quantities also displayed a significant correlation, with an average of 0.62. All the remaining weather quantities showed a weak correlation, with DPT spread and DHI presenting the lowest significance. As a result of the Spearman correlation analysis, several weather quantities could be removed from the machine learning model stage. However, the locations resulted in 360 sets of retained weather parameters rather than an expected single set, since the correlation analysis considered nine building envelope configurations, four building orientations, five residential models, and two simulation outputs.

Despite the different sets for each location, the calculation of the ratio of occurrence, i.e., the number of times each variable was retained for the location divided by 360 correlation analysis performed for each location, enabled identifying the dominant variables, as shown in Figure 7. Considering the 360 possibilities for each location, the maximum DBT was the variable with the lowest retention, since it showed a maximum occurrence below 20%, i.e., retained 62 times for the same city. For all the other variables, the correlation analysis resulted in retention in, at least, one location. The average DBT showed the highest occurrence, being retained for all locations but varying from 0.3 to 1.0. It means that, despite being present for all locations, the combination of different simulation outputs, building envelope configurations, orientation, and simulation models resulted in different sets of retained weather parameters for the machine learning stage. Furthermore, since the combinations for a single building model result in 72 possibilities, in neither case—maximum or average DBT—was the set of retained variables the same across the entire set of combinations. This allows us to state that not only the simulation output affected the retained variables, but also the combination of building envelope configurations and the orientation.

The R^2 metric and the RMSE were used to assess the models' accuracy. Regarding the operative temperature models, the average R^2 was 0.91, with a standard deviation of 0.03, and the RMSE resulted in an average of 0.64 °C with a deviation of 0.19 °C. The energy needs models showed an average R^2 of 0.87 with a deviation of 0.05, while the RMSE presented an average of 0.05 kWh m^{-2} yr^{-1} and a deviation of 0.01 kWh m^{-2} yr^{-1}. The results show that, for both situations, the combination of the XGBoost and SVR machine learning algorithms was capable to produce regression models with high accuracy and low error to support the analysis based on the feature importance component (Figure 8).

Given the previously identified significance, the feature importance results for the DBT showed an average of 75% for the operative temperature models and 59% for the energy needs regression models (Figure 9). In both cases, neither of the remaining weather quantities resulted in an average above 10%. The energy needs results showed the DBT min in the second position (8%). The feature importance distribution

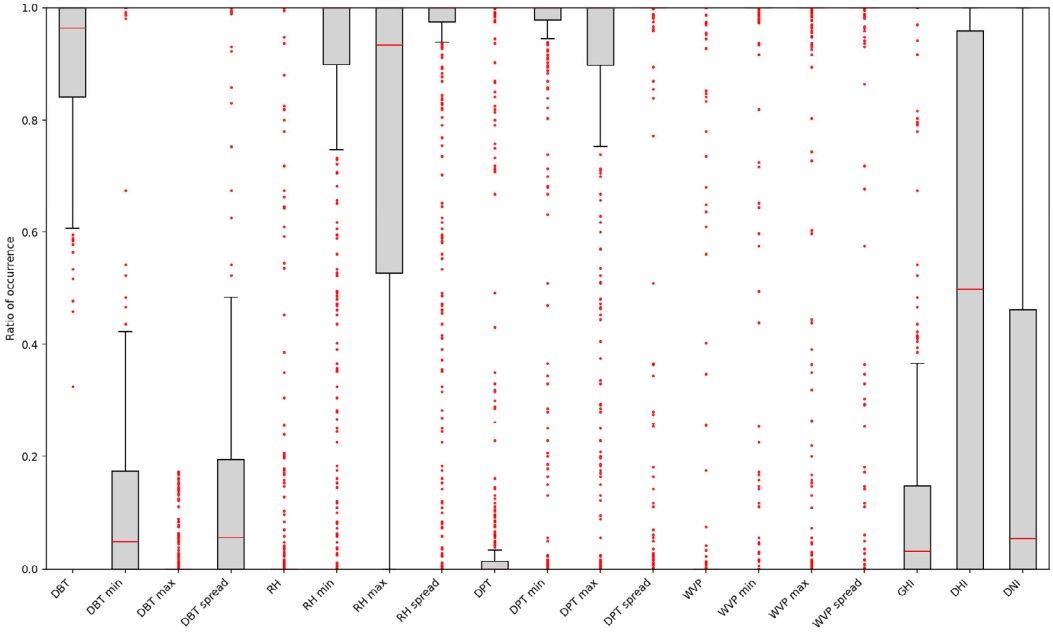

Fig. 7 Occurrence of retained weather parameters within the 480 locations

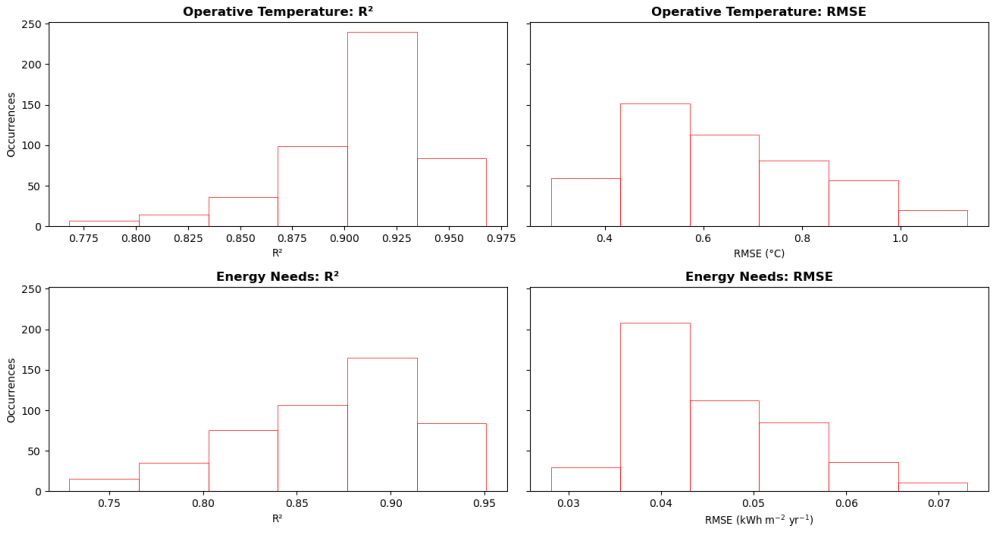

Fig. 8 Average performance results for the machine learning regression models

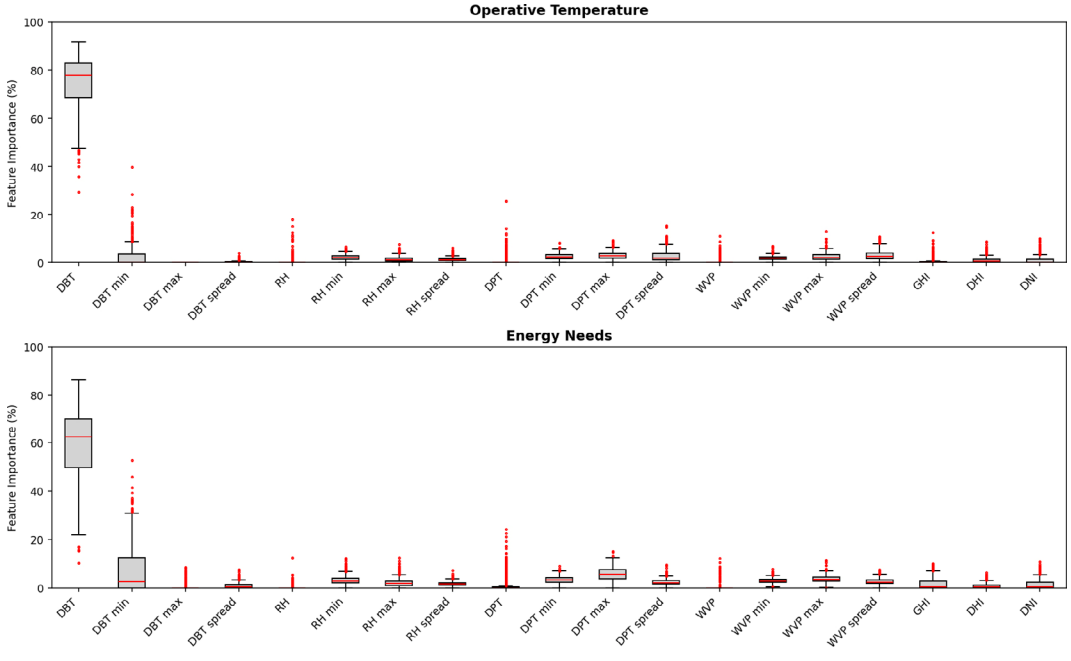

Fig. 9 Distribution of the climatic variables' weights for the reference years selection. The weights for each location and output are provided in the electronic supplementary material, which is availabe in the online version of this paper

for all the weather quantities highlights and restates the significance of the DBT on building performance, given the outstanding superiority, despite some occurrences below 50%. As a direct consequence of the feature importances used as weight to select the reference years, only 47 locations presented a single $BTMY_{SITE}$ (i.e., with the same operative temperature and energy needs-based weights). Therefore, most of the locations showed one weather file for the weights based on the operative temperature and another one based on the energy needs weights.

Following the methodology proposed, the feature importances were grouped according to the bioclimatic zones, as the average importance from the locations within each zone and simulation output (Table 3). DBT displayed an increasing importance for the operative temperature results from zones 1 to 6, where zone 1R presented the lowest importance (58%) and zones 6A and 6B presented the highest importance (82%). Furthermore, the daily average DBT was the only significant weather variable for the operative temperature results (average of 73% and a standard deviation of 7 percentage points). Even though the DPT spread and WVP spread reached an average of 6% for zones 1 and 2, for the remaining zones their importance was below 5%. The remaining weather variables showed an extremely low

Table 3 Weather variables' weights, presented in %, for the $BTMY_{ZONE}$ weather files

		DBT				RH				DPT				WVP						
Zone	Output	Avg	Min	Max	Spread	Avg	Min	Max	Spread	Avg	Min	Max	Spread	Avg	Min	Max	Spread	GHI	DHI	DNI
1M	EUI	59	6	1	*	*	3	2	2	3	3	5	2	*	3	3	3	2	1	2
	OT	67	*	*	*	*	1	*	1	*	3	3	6	*	2	4	5	*	3	3
1R	EUI	53	8	*	1	*	3	2	2	3	3	6	2	*	3	4	2	2	*	2
	OT	58	*	*	*	*	2	*	*	*	4	3	7	*	2	5	7	4	3	3
2M	EUI	64	4	*	*	*	3	2	1	2	3	5	2	*	3	3	2	1	*	2
	OT	71	2	*	*	*	1	*	1	*	3	3	4	*	2	3	4	*	1	2
2R	EUI	65	5	*	*	*	3	2	2	*	3	5	2	1	2	3	2	*	*	1
	OT	63	*	*	*	3	*	*	*	*	4	3	8	*	2	4	7	*	2	3
3A	EUI	53	9	*	*	*	4	3	2	1	3	5	3	*	3	3	2	2	*	2
	OT	68	4	*	*	*	1	*	*	2	3	3	4	*	2	4	4	*	*	1
3B	EUI	57	10	*	1	*	3	2	1	1	3	6	2	*	3	4	2	2	*	2
	OT	77	2	*	*	*	2	1	1	1	2	3	2	*	1	2	3	*	*	*
4A	EUI	60	6	*	*	*	3	2	2	2	3	5	2	*	3	4	2	2	1	2
	OT	77	2	*	*	*	1	*	1	1	2	3	2	*	2	2	2	*	*	*
4B	EUI	58	7	*	1	*	3	2	1	2	3	6	2	*	3	4	3	2	*	1
	OT	74	5	*	*	*	2	1	1	*	2	3	2	*	2	2	3	*	*	*
5A	EUI	58	8	*	*	*	3	2	2	2	3	5	2	*	3	4	2	2	*	1
	OT	79	2	*	*	*	3	2	2	2	2	2	1	*	2	2	1	*	*	*
5B	EUI	59	9	*	*	*	3	2	2	1	3	5	2	*	3	3	2	2	*	1
	OT	76	5	*	*	*	2	1	1	*	2	3	1	*	2	2	2	*	*	*
6A	EUI	57	13	*	1	*	3	3	2	2	3	5	2	*	2	3	2	2	*	*
	OT	82	*	*	*	*	3	3	2	*	2	2	1	*	2	2	2	*	*	*
6B	EUI	57	8	*	*	*	3	2	2	*	3	6	3	*	3	4	3	2	*	2
	OT	82	1	*	*	*	3	2	1	*	1	2	1	*	1	2	2	*	*	*

Note: The sum might not add to 100 given the rounding and values below 1%, marked as "*", in this visualization. The electronic supplementary materials provide the raw data with the specific weights.

importance or 0 as the result for the climate zone-based weights. The importance based on the energy needs model showed a more even distribution, varying from 53% (1R and 3A) to 65% (2R). Likewise, the operative temperature results, the average daily DBT (average of 59% and a standard deviation of 3 percentage points) has the highest influence, but also DBT min (average of 8% and a standard deviation of 2 percentage points) and DPT max (average of 5% and a standard deviation of 0.3 percentage points) presented an average importance of at least 5%. The remaining variables followed the previous pattern with a low or no importance. As regards the creation of the $BTMY_{ZONE}$ weather files, all the 480 locations required a single file for each location, with no occurrences of a $BTMY_{ZONE}$ for the operative temperature and a distinct one for the energy needs as occurred with the $BTMY_{SITE}$ files, independent of the weights applied to the performance-based approach for selecting the reference years.

The feature importance analysis performed for the energy needs and operative temperature separately led to different reference years for 433 locations out of the 480 used in this study. Therefore, it resulted in $EUI.BTMY_{SITE}$ and $OT.BTMY_{SITE}$ weather files. The analyses of the daily aggregation of all 480 locations for the operative temperature show that the weather files have the same average, median, and standard deviation, respectively, 27.2 °C, 28.1 °C, and 3.7 °C (Figure 10). Likewise, the energy needs pointed out that both weather files have a coincident average (0.46 kWh m^{-2} day^{-1}), median (0.49 kWh m^{-2} day^{-1}), and standard deviation (0.21 kWh m^{-2} day^{-1}). Therefore, regarding the central tendency statistics, the weather files present equivalent results, and so, they are equally adequate for BPS applications.

The nMBE pointed out that all locations deviations are within a ±10% range while the CVRMSE pointed that most of them are within the 30% threshold (Figure 10). Likewise,

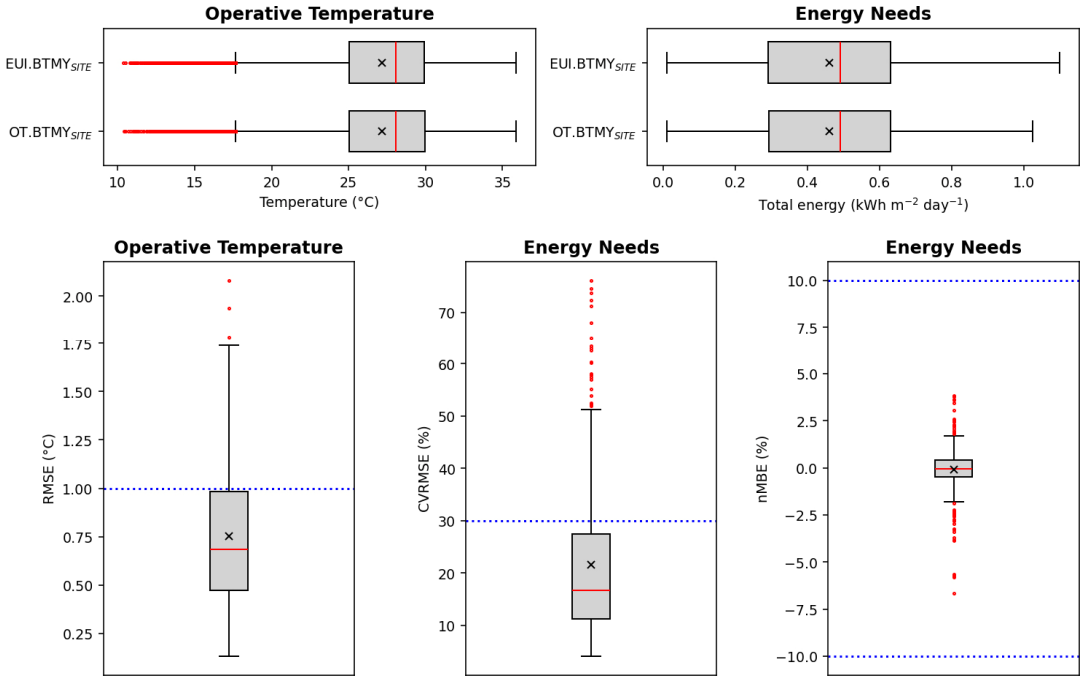

Fig. 10 Distribution of the daily summary of operative temperature and energy needs (first row) and distribution of the hourly deviations of operative temperature and energy needs from OT.BTMY$_{SITE}$ and EUI.BTMY$_{SITE}$ (second row). On the second row, the blue dotted lines indicate the deviation threshold for the operative temperature and energy needs described in Section 2.3

the operative temperature errors presented a similar pattern, with most of the locations below the 1 °C deviation regarding the hourly records. These results allow stating that, despite the feature importance leading to different reference years, the variation within the simulation outputs are small and a single weather file can be used for the locations.

3.3 Traditional and performance-based TMY vs. 15-year record (AMY)

The annual operative temperature and energy needs of all four TMY methods led to simulation outputs close to the AMY records (Figure 11). The minimum Finkelstein–Schafer and Best Rank I showed the best results for the operative temperature, with an absolute average difference of 0.15 °C and a standard deviation of 0.09 °C and 0.10 °C, respectively. The ISO results presented an intermediary performance, with a similar absolute average difference (0.16 °C) and a slightly higher standard deviation (0.13 °C). The Pissimanis had the worst results, with an absolute average difference of 0.23 °C and a standard deviation of 0.27 °C. An in-depth analysis also shows that the Minimum Finkelstein–Schafer and Best Rank I deviation were never outside the ±0.5 °C range, while the ISO shows only four occurrences above 0.5 °C, and the Pissiminis reached 40 locations outside the ±0.5 °C range. For the energy needs, the ISO resulted in the best approximation, the Minimum Finkelstein–Schafer and the Best Rank I methods showed an intermediary variation, while the Pissimanis had the worst approximation with the AMY records, reaching a maximum overestimation of 29.5 kWh m^{-2} yr^{-1} for the top floor of the multifamily model. Considering the average results for all archetypes, the Pissimanis weather files led to the lowest average differences but with the highest deviation, for both the operative temperature and the energy needs. The remaining weather files presented a similar performance. Therefore, the ISO, the minimum Finkelstein–Schafer, and the Best Rank I methods provided a better approximation to the 15-year records.

Regarding the comparison between the existing TMY, BTMY$_{SITE}$, and BTMY$_{ZONE}$ and the AMY weather files, the Mann-Whitney statistic revealed that for the operative temperature most of the locations fail to reject the null hypothesis (Figure 12). Therefore, the results do not provide enough evidence to state a significant difference between the long-term records and the reference years, indicating a possible agreement between the long-term simulation records obtained with the different weather files. The locations in which the hourly distribution displayed a statistically significant difference varied from 2 when using the minimum Finkelstein–Schafer or EUI.BTMY$_{SITE}$ methods to 38 with the Pissimanis approach. The comparison between the hourly distribution of the performance-based weather files (BTMY$_{SITE}$ and BTMY$_{ZONE}$) and the AMY results showed that most of the locations have a similar operative temperature hourly distribution.

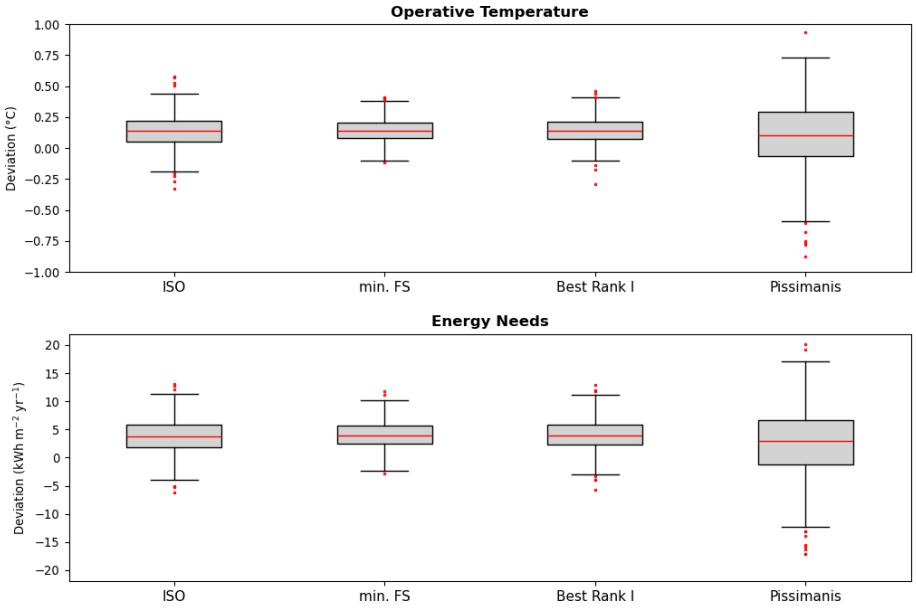

Fig. 11 Annual deviations between the AMY and the TMY weather files

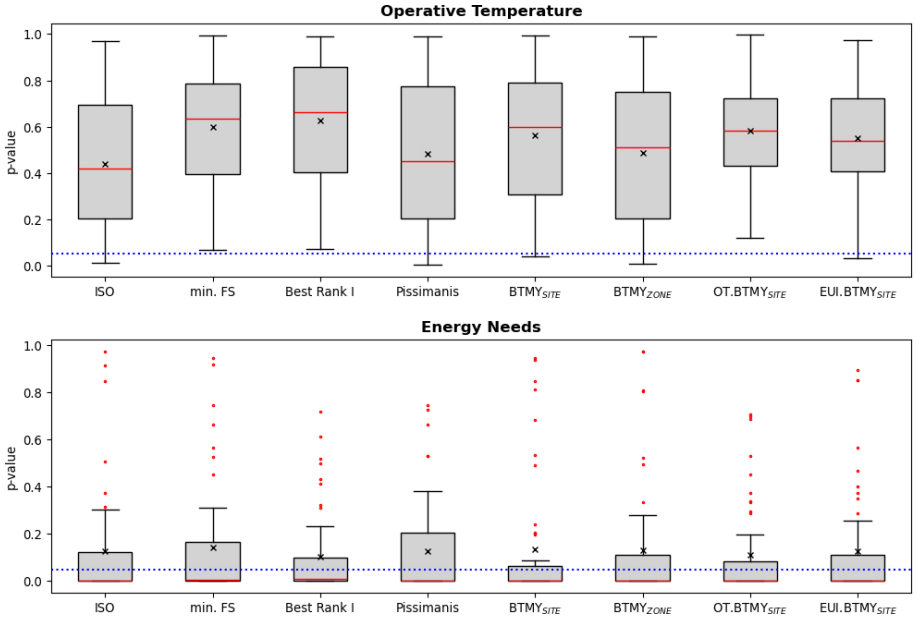

Fig. 12 Summary of the hourly assessment of the distribution for the operative temperature and energy needs according to the Mann-Whitney *p*-value (the blue dotted lines indicate the 0.05 significance level)

Nevertheless, the energy needs results in most of the locations showed a *p*-value ≤ 0.05. The $BTMY_{ZONE}$ weather files presented the lowest energy needs approximation to the long-term records (68% of the locations with a *p*-value ≤ 0.05), while the $BTMY_{SITE}$, ISO, and Pissimanis had the best approximation (74% of the locations with a *p*-value ≤ 0.05).

The deviation results showed most of the locations with an operative deviation higher than 1 °C (Figure 13). The average deviation for the ISO, minimum Finkelstein–Schafer, and Best Rank I methods was 1.1 °C, while the Pissimanis resulted in an average of 1.3 °C. The TMY showed the lowest difference (0.4 °C), while the highest difference varied from 2.2 °C (Best Rank I and minimum Finkelstein–Schafer) to 2.4 °C (ISO). The BTMY weather files, both site-specific ($BTMY_{SITE}$) and climate zone-based ($BTMY_{ZONE}$), demonstrate similarities with the TMY methods for the operative temperature RMSE (average of 1.1 °C). However, the results presented a clear difference for $BTMY_{ZONE}$ as the only weather file displaying the mean and median below the 1 °C threshold. The energy needs deviation analysis showed that more than 50% of the locations presented a CVRMSE within the 30% threshold, while 90% locations presented a nMBE

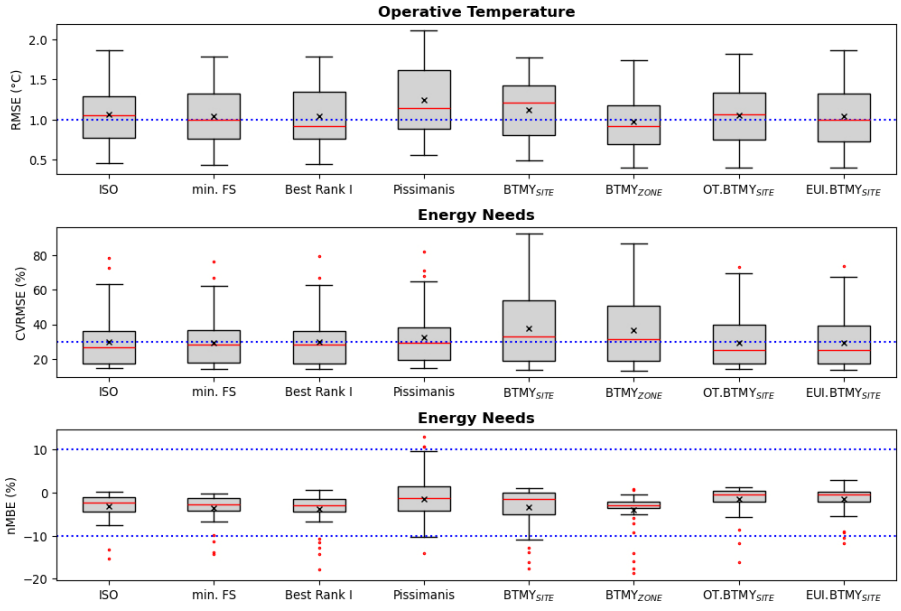

Fig. 13 Summary of the hourly assessment of the deviations for the operative temperature and energy needs (the blue dotted lines indicate the deviation threshold for the operative temperature and energy needs described in Section 2.3)

deviation within the ±10% range. Even though the energy needs resulted in a higher agreement than the operative temperature, CVRMSE reached almost 100% for some locations, while the nMBE presented a minimum value varying from −20.8% (ISO) to −37.5% (Pissimanis).

Furthermore, the comparison between the $BTMY_{SITE}$ and $BTMY_{ZONE}$ showed that by averaging the weights of different locations to define climate zone weights led to a similar behaviour for the operative temperature and energy needs hourly distribution, as observed with the previous results for the *p*-value. The RMSE, CVRMSE, and nMBE metrics for the $BTMY_{SITE}$ and $BTMY_{ZONE}$ also showed that the application of site specific and climate zone-based weights produced similar results. Therefore, the use of $BTMY_{ZONE}$ instead of $BTMY_{SITE}$ can produce reliable results and lower errors, in some cases, including the operative temperature, with the mean and median within the 1 °C threshold.

Although the similarities among the different weather files results, the Pissimanis weather files presented the highest RMSE and nMBE deviations, while the BTMY weather files led to the highest CVRMSE differences between the typical and 15-year simulation results. Among the different weather files created, those based on the existing TMY methods and the $BTMY_{SITE}$ for specific operative temperature and energy needs weights showed the lowest deviations. Despite being a simplification of the site-specific weights, the $BTMY_{ZONE}$ weather files displayed a very similar range to the existing methods and weight-specific approaches. Overall, building performance simulation analysis can use all the weather files to provide good approximation to the long-term records, with Pissimanis presenting the worst overall approximation.

4 Conclusions

This study focused on assessing the impact of different weather file compilation methodologies on the performance of different social housing buildings across 480 locations in Brazil. The simulations covered three archetypes (43–47 m^2), nine envelope configurations, four orientations, and two operating modes (free-floating operative temperature and ideal-load energy needs). The analysis mainly focused on existing weather file compilation methods as the EN ISO 15927-4, minimum Finkelstein–Schafer, Best Rank I, and Pissimanis. For the performance-based approach, the study follows a performance-based workflow that couples building simulation outputs with machine-learning-derived climatic weights. Coupling machine learning algorithms to create prediction models based on weather data as inputs and simulation results as outputs to aid interpreting simulation results and extract the influence of each weather variable as weights represents a significant time reduction in the simulation process. Therefore, the AMYs provided simulation results related to the historical series, and the machine learning algorithm enabled performing a sensitivity analysis to identify the weather quantities that most affect building performance, leading to an effective approach to select reference years. The optimized XGBoost and ε-SVR models achieved R^2 = 0.91 ± 0.03 for daily operative temperature and R^2 = 0.87 ± 0.05 for daily energy needs (RMSE = 0.64 ± 0.19 °C and 0.05 ± 0.01 kWh m^{-2} yr^{-1}), identifying dry-bulb temperature as the dominant driver (75% importance for operative temperature; 59% for energy needs). The methodology adopted in this study also reduced the

uncertainties in the comparison, since all the weather files derived from the same dataset, i.e., ERA5-Land records from 2008 to 2022, compiling 15 AMYs per site. The performance-based weighting selected different reference years for operative temperature and energy needs in 433 of 480 locations (only 47 locations yielded the same year).

The existing TMY methods present in this study share the mean daily DBT as one of the main variables to select the reference years, and the machine learning approach used to determine the new set of weights also pointed out that DBT was the most impactful weather variable. Therefore, the methods share a determining variable in the process of selecting the reference year. Although the methods led to different reference years, the impact of DBT led to the selection of years with a similar profile and quantities for the climatic variable. Despite the BTMY methods being performance-driven, there was no statistically significant improvement in proposing a new set of weather files with site-specific weights or climate zone-based approaches, since they led to similar distribution and deviation from existing methods. However, the results allowed reducing the uncertainty and randomness associated with the climatic quantities and weights employed in the process of defining the reference years. In the 15-year comparison, minimum Finkelstein–Schafer and Best Rank I showed the lowest mean operative-temperature bias (0.15 °C), ISO was similar (0.16 °C), while Pissimanis performed worst (0.23 °C) and overestimated annual energy needs by up to 29.5 kWh m^{-2} yr^{-1} in a multifamily case. In a context of national scale analyses aimed at defining energy policies for the built environment, the choice between traditional TMY methods and the new performance-based one shows limited impact on the simulation outcome. Nevertheless, when the focus is put on specific locations, buildings, and goals (e.g., indoor environmental condition or energy performance), the weather file able to minimize the deviations with respect to the long-term building behaviour should be selected. In this regard, the repository of weather files made available as outcome of the present research is expected to support building simulation practitioners in improving simulation accuracy and the representativeness of its outputs.

Indeed, building performance simulation results are directly dependent on the weather file used during the simulation process. From the different set of comparison between the AMY weather data and the TMY and BTMY methods, the results displayed a similar trend for the Mann-Whitney test. The *p*-value interpretation led to the conclusion of the operative temperature hourly distribution presenting a better approximation to long-term records than the energy needs. For operative temperature, only 2–38 locations showed statistically significant distribution differences depending on the method, whereas energy needs showed *p*-value ≤ 0.05 in most locations (68% for $BTMY_{ZONE}$ and 74% for $BTMY_{SITE}$, ISO, and Pissimanis). Regarding the deviation analysis, the RMSE, CVRMSE, and nMBE results indicated a high occurrence of locations outside the operative temperature and energy needs CVRMSE thresholds, while the nMBE showed most of the locations within the ±10% range. More than 50% of locations were within the 30% CVRMSE threshold, and 90% were within ±10% nMBE; however, minimum nMBE ranged from −20.8% (ISO) to −37.5% (Pissimanis). As a general assessment, the Pissimanis method presented the worst performance, while the remaining methods, including the proposed BTMY, demonstrated a better agreement with the long-term records.

The implication of the results of the extensive analysis to identify the impact of different weather files in building performance simulation for different Brazilian locations can be summarized below:

- Dry-bulb temperature is the most significant climatic parameter in the process of defining the representative year (75% importance for operative temperature; 59% for energy needs).
- The Pissimanis weather files led to the highest building simulation results deviations among the reference year methods used in this study (0.23 °C mean operative-temperature bias; up to 29.5 kWh m^{-2} yr^{-1} energy overestimation; ±0.5 °C exceeded in 40 locations).
- Building performance results showed a small deviation between the different weather files. This indicates that the original weather dataset, in this case the ERA5-Land records, played a more significant role in the building performance outcomes than the method to define the reference years, in this specific study.

5 Limitations and future developments

The decisions made in this work directly influence the results obtained in the comparison of the different weather files, as well as in the process of analysing the feature importance and the compilation of the performance-based weather files. Regarding the building archetypes, even though the analysis considered a wider range of building envelope configurations and building orientations, the simulation settings impact the energy needs and operative temperature outcomes, e.g., schedules, setpoints (e.g., for Ideal Loads and Air Flow Network models). Still, regarding the performance-based weather files, the feature importance was averaged for each location and simulation output, across the different building archetypes, nine building envelope configurations and four simulation outputs. Therefore, an

analysis of each archetype, building envelope and orientation could lead to different reference years and, consequently, to different simulation results.

Additionally, deviations could be generated when adopting different weather data sources, including those from mathematical and physical modelling, satellite images, and ground measurements. As a limiting factor, the investigation considering different sources of weather data can influence the creation of weather files, especially the performance-based weather files, and, consequently, result in different recommendations regarding the suitability of different weather files for BPS applications.

Future developments should include non-residential building typologies and consider the impact of different building simulation codes on building performance simulation outcomes, and consequently on the performance-based approach to defining the reference years. Furthermore, future studies should address extreme events and future climates, since they can significantly impact building performance in terms of operative temperature and energy needs.

Electronic Supplementary Material (ESM): the supplementary material is available in the online version of this article at https://doi.org/10.1007/s12273-026-1416-1.

Data availability

The dataset of weather files used in this study (TMY and BTMY) can be found at https://doi.org/10.5281/zenodo.10546589.

Acknowledgements

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, FAPEMIG PAPG (5.12/2022), CNPq (406426/2022-8), and CAPES PDSE (88881.846099/2023-01). This research was partially developed using the UFV's (Federal University of Viçosa) computer cluster.

Funding note: Open access funding provided by Free University of Bozen-Bolzano within the CRUI-CARE Agreement.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval

This study does not contain any studies with human or animal subjects performed by any of the authors.

Author contribution statement

Mario Alves da Silva: writing—original draft, writing—review & editing, conceptualization, methodology, investigation. Giovanni Pernigotto: supervision, writing—review & editing; Alessandro Prada: writing—review & editing; Andrea Gasparella: supervision, writing—review & editing; Joyce Correna Carlo: supervision, writing—review & editing, funding acquisition, project administration. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

ABNT (2024). NBR 15575: Edificações habitacionais – Desempenho. Parte 1: Requisitos gerais. Brazilian National Standards Organization. (in Portuguese)

Afshari A (2023). Optimization of urban design/retrofit scenarios using a computationally light standalone urban energy/climate model (SUECM) forced by ERA5 data. *Energy and Buildings*, 287: 112991.

Alawi OA, Kamar HM, Yaseen ZM (2024). Optimizing building energy performance predictions: A comparative study of artificial intelligence models. *Journal of Building Engineering*, 88: 109247.

ASHRAE (2023). ASHRAE Guideline14-2023: Measurement of Energy, Demand, and Water Savings. Atlanta, GA, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

Beckman WA, Broman L, Fiksel A, et al. (1994). TRNSYS The most complete solar energy system modeling and simulation software. *Renewable Energy*, 5: 486–488.

Bergstra J, Yamins D, Cox DD (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In: Proceedings of the 30th International Conference on International Conference on Machine Learning.

Bigtashi A, Papakyriakou A, Lee B (2024). Defining generation parameters with an adaptable data-driven approach to construct

typical meteorological year weather files. *Energy and Buildings*, 303: 113781.

Bolton D (1980). The computation of equivalent potential temperature. *Monthly Weather Review*, 108: 1046–1053.

Bonini VRB, Starke AR, Lemos LFL, et al. (2022). Validation of 441 typical meteorological year (TMY) created with INMET/Brazil stations data. In: Proceedings of IX Congresso Brasileiro de Energia Solar.

Bracht MK, Olinger MS, Krelling AF, et al. (2024). Multiple regional climate model projections to assess building thermal performance in Brazil: Understanding the uncertainty. *Journal of Building Engineering*, 88: 109248.

Chen T, Guestrin C (2016). XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA.

Crawley DB (1998). Which weather data should you use for energy simulations of commercial buildings? *ASHRAE Transactions*, 104(2): 498–515.

Crawley DB, Lawrie LK, Winkelmann FC, et al. (2001). EnergyPlus: Creating a new-generation building energy simulation program. *Energy and Buildings*, 33: 319–331.

da Silva MA, Pernigotto G, Gasparella A, et al. (2024). Towards climate, bioclimatism, and building performance—a characterization of the Brazilian territory from 2008 to 2022. *Buildings*, 14: 2568.

da Silva MA, Pernigotto G, Prada A, et al. (2025). Impact of the type of weather files on the outcome of a weather-based climate classification: The case of Brazil. In: Berardi U (Ed.), Multiphysics and Multiscale Building Physics. Singapore: Springer. pp. 258–263.

de Miguel A, Bilbao J (2005). Test reference year generation from meteorological and simulated solar radiation data. *Solar Energy*, 78: 695–703.

de Sousa DLP, de Almeida Nobre JC, dos Santos Borges L, et al. (2023). Utilização dos arquivos climáticos IWEC, SWERA, INMET, TRY e TMY da cidade de belém-PA para análise de desempenho térmico estrutural. *The Journal of Engineering and Exact Sciences*, 9: 15554–1551e.

e Silva Machado RM, Bre F, Mazzaferro L, et al. (2024). Bioclimatic zoning for building performance using tailored clustering method and high-resolution climate data. *Energy and Buildings*, 311: 114157.

European Committee for Standardization (2005). 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.

Festa R, Ratto CF (1993). Proposal of a numerical procedure to select Reference Years. *Solar Energy*, 50: 9–17.

Hall IJ, Prairie RR, Anderson HE, et al. (1978). Generation of a typical meteorological year. Albuquerque, NM, USA: Sandia National Laboratories

Herrera M, Natarajan S, Coley DA, et al. (2017). A review of current and future weather data for building simulation. *Building Services Engineering Research and Technology*, 38: 602–627.

Hersbach H, Bell B, Berrisford P, et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146: 1999–2049.

Hosseini M, Bigtashi A, Lee B (2020). A systematic approach in constructing typical meteorological year weather files using machine learning. *Energy and Buildings*, 226: 110375.

Integrated Environmental Solutions (2024). Virtual Environment. Available at https://www.iesve.com/software/virtual-environment.

Kalamees T, Jylhä K, Tietäväinen H, et al. (2012). Development of weighting factors for climate variables for selecting the energy reference year according to the EN ISO 15927-4 standard. *Energy and Buildings*, 47: 53–60.

Kim S, Zirkelbach D, Künzel HM, et al. (2017). Development of test reference year using ISO 15927-4 and the influence of climatic parameters on building energy performance. *Building and Environment*, 114: 374–386.

Krelling AF, Lamberts R, Malik J, et al. (2024). Defining weather scenarios for simulation-based assessment of thermal resilience of buildings under current and future climates: A case study in Brazil. *Sustainable Cities and Society*, 107: 105460.

LabEEE, Veiga C, Bagio J, et al. (2024). Reference Models for Brazilian Social Housing Pre-2023. Laboratório de Eficiência Energética em Edificações da Universidade Federal de Santa Catarina. Projeto Hab.LabEEE. Available at https://doi.org/10.5281/zenodo.12209191. (in Portuguese)

Lawrie L (2003). EnergyPlus Weather Converter: Subroutine for ground temperature calculation. CalcGroundTemps. F90, Fortran Code.

Layi Fagbenle R (1995). Generation of a test reference year for Ibadan, Nigeria. *Energy Conversion and Management*, 36: 61–63.

Lemos LFL, Starke AR, Boland J, et al. (2017). Assessment of solar radiation components in Brazil using the BRL model. *Renewable Energy*, 108: 569–580.

Lund H (1995). The Design Reference Year Users Manual: A Report of Task 9: Solar Radiation and Pyranometer Studies. Technical University of Denmark.

Machado RD, Bravo G, Starke A, et al (2019). Generation of 441 typical meteorological year from INMET stations - Brazil. In: Proceedings of the ISES Solar World Congress 2019 and IEA SHC International Conference on Solar Heating and Cooling for Buildings and Industry 2019, Santiago, Chile.

Machado RMES, Bre F, Melo AP, et al. (2025). The impact of climate data uncertainty on bioclimatic zoning for building design. *Building and Environment*, 269: 112423.

Mann HB, Whitney DR (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1): 50–60.

Massey FJ, Jr (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46: 68–78.

May RM, Goebbert KH, Thielen JE, et al. (2022). MetPy: A meteorological Python library for data analysis and visualization. *Bulletin of the American Meteorological Society*, 103: E2273–E2284.

Melo AP, Cóstola D, Lamberts R, et al. (2014). Development of surrogate models using artificial neural network for building shell energy labelling. *Energy Policy*, 69: 457–466.

Mo H, Sun H, Liu J, et al. (2019). Developing window behavior models for residential buildings using XGBoost algorithm. *Energy and Buildings*, 205: 109564.

Muñoz-Sabater J, Dutra E, Agustí-Panareda A, et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13: 4349–4383.

Papakyriakou A, Bigtashi A, Lee B (2024). Evaluating the applicability of a machine learning methodology to improve TMY weather file generation for different Canadian climate zones. *Journal of Building Engineering*, 95: 110096.

Pedregosa F, Varoquaux G, Gramfort A, et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830.

Pernigotto G, Prada A, Gasparella A, et al. (2014). Analysis and improvement of the representativeness of EN ISO 15927-4 reference years for building energy simulation. *Journal of Building Performance Simulation*, 7: 391–410.

Pissimanis D, Karras G, Notaridou V, et al. (1988). The generation of a "typical meteorological year" for the city of Athens. *Solar Energy*, 40: 405–411.

Platt JC (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Smola AJ, Bartlett P, Schölkopf B, et al. (Eds), Advances in Large Margin Classifiers. Cambridge, MA, USA: MIT Press.

Pyrgou A, Castaldo VL, Pisello AL, et al. (2017). Differentiating responses of weather files and local climate change to explain variations in building thermal-energy performance simulations. *Solar Energy*, 153: 224–237.

Qi Z, Feng Y, Wang S, et al. (2025). Enhancing hydropower generation Predictions: A comprehensive study of XGBoost and Support Vector Regression models with advanced optimization techniques. *Ain Shams Engineering Journal*, 16: 103206.

Rodrigues E, Fernandes MS, Carvalho D (2023). Future weather generator for building performance research: An open-source morphing tool and an application. *Building and Environment*, 233: 110104.

Rodriguez JD, Perez A, Lozano JA (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32: 569–575.

Tippett A, Gonçalves AR, Pereira EB, et al. (2024). Data source sensitivity in solar radiation typical meteorological year (TMY) for five different regions of Brazil. *Tellus A*: *Dynamic Meteorology and Oceanography*, 76: 29–41.

Triana MA, Lamberts R, Sassi P (2015). Characterisation of representative building typologies for social housing projects in Brazil and its energy performance. *Energy Policy*, 87: 524–541.

Wilcox S, Marion W (2008). Users Manual for TMY3 Data Sets. Technical Report NREL/TP-581-43156. Golden, CO, USA: National Renewable Energy Laboratory.

Wu Y, An J, Gui C, et al. (2023). A global typical meteorological year (TMY) database on ERA5 dataset. *Building Simulation*, 16: 1013–1026.

Yan H, Yan K, Ji G (2022). Optimization and prediction in the early design stage of office buildings using genetic and XGBoost algorithms. *Building and Environment*, 218: 109081.

Yuan J, Huang P, Chai J (2022). Development of a calibrated typical meteorological year weather file in system design of zero-energy building for performance improvements. *Energy*, 259: 125031.

Zanoni V,, Cavalcante C, Moura ABB (2023) Arquivos climáticos de Brasília disponíveis no Climate One Building para simulações higrotérmicas. In: Proceedings of Encontro Nacional de Conforto no Ambiente Construído. (in Portuguese)

Zhang T, Chandler WS, Hoell JM, et al. (2008). A global perspective on renewable energy resources: Nasa's prediction of worldwide energy resources (power) project. In: Goswami DY, Zhao Y (eds), Proceedings of ISES World Congress 2007. Berlin: Springer. pp. 2636–2640.

Zhou X, Zi X, Liang L, et al. (2019). Forecasting performance comparison of two hybrid machine learning models for cooling load of a large-scale commercial building. *Journal of Building Engineering*, 21: 64–73.