UNIVERSIDADE FEDERAL DE VIÇOSA

Multi-scale analysis of weather data for building performance assessment in Brazil

Mario Alves da Silva
Doctor Scientiae

VIÇOSA - MINAS GERAIS
2024

MARIO ALVES DA SILVA

Multi-scale analysis of weather data for building performance assessment in Brazil

Thesis submitted to the Architecture and Urban Planning Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Joyce Correna Carlo

VIÇOSA - MINAS GERAIS
2024

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
S586m
2024

Silva, Mario Alves da, 1996-
Multi-scale analysis of weather data for building performance assessment in Brazil / Mario Alves da Silva. – Viçosa, MG, 2024.
1 dissertação eletrônica (130 f.): il. (algumas color.).

Texto em inglês.
Inclui apêndices.
Orientador: Joyce Correna Carlo.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Arquitetura e Urbanismo, 2024.
Referências bibliográficas: f. 108-117.
DOI: https://doi.org/10.47328/ufvbbt.2025.024
Modo de acesso: World Wide Web.

1. Edifícios - Engenharia ambiental. 2. Arquitetura e clima. 3. Meteorologia - Métodos estatísticos. 4. Aprendizado do computador. I. Carlo, Joyce Correna, 1973-. II. Universidade Federal de Viçosa. Departamento de Arquitetura e Urbanismo. Programa de Pós-Graduação em Arquitetura e Urbanismo. III. Título.

CDD 22. ed. 696

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

MARIO ALVES DA SILVA

Multi-scale analysis of weather data for building performance assessment in Brazil

Thesis submitted to the Architecture and Urban Planning Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: November 29, 2024.

Assent:

Mario Alves da Silva
Author

Joyce Correna Carlo
Adviser

ACKNOWLEDGMENTS

To God.
To my mom, Ana Cristina, my dad, Edilson, my sister, Mariana, and my grandmother Cidinha.
To my friends Flor, Sofia, Tiffany, Mell, Gustavo, and Gregorio Borelli.
To my lab friends from LATECAE, especially Caio and Matheus.
To my advisor, Joyce Carlo, and my co-advisors, Giovanni Pernigotto and Andrea Gasparella.
To the Federal University of Viçosa, for the opportunity to complete the postgraduate course.
To the UFV's computer cluster.
To the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) for funding my exchange period at Free University of Bozen-Bolzano in Italy – Finance Code 88881.846099/2023-01.
To the Conselho Nacional de Desenvolvimento Científico and Tecnológico (CNPq) – Finance Code 406426/2022-8.
To the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) – Finance Code 5.12/2022.
This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

ABSTRACT

SILVA, Mario Alves da, D.Sc., Universidade Federal de Viçosa, November, 2024. **Multi-scale analysis of weather data for building performance assessment in Brazil**. Adviser: Joyce Correna Carlo.

The choice of weather data is fundamental for assessing building performance in an accurate and representative way. Typical weather files usually apply a statistical approach to select months representative of current climatic conditions, but they do not encompass site-specific characteristics for different locations worldwide.

This work analyses the potential of a multi-scale analysis for building performance assessment from different resolutions of weather data, in a comprehensive geographical territory, given the size of Brazil. It presents an overview of weather data and its application on building performance analysis and a general procedure to retrieve and process weather data, and different approaches to compile weather files for building performance assessment. The study also provides an extensive analysis of the Brazilian territory, presenting a climatic profile and trends for the entire territory. The analysis focuses on a climatic and bioclimatic summary, and on building performance simulations for representative cities according to the Brazilian bioclimatic zoning. Then, it compares the records from ERA5-Land and INMET to quantify the differences and present the impact on building performance analysis. The study proposes a new weather file compilation method for Brazil and applies statistical tests to determine whether the new approach delivers better results than the existing TMY methods. The procedure encompasses correlation and sensitivity analysis based on machine learning models to propose a performance-based method.

The initial analysis of the Brazilian territory showed predominantly a temperature increase based on the 2008-2022 records, with some locations reaching more than 1 °C. However, the bioclimatic approach based on Givoni's chart showed that ventilation strategies are still the most effective approach instead of HVAC systems. Following the comparison between high resolution spatial data and weather stations from Brazil, some locations present insufficient years for a multi-year analysis and some municipalities show a significant variation not only of weather data, but also of the building performance results. Finally, the analysis of new weather files for Brazil allowed concluding that creating typical year weather files based on a performance approach delivers the best outcomes, since they are closer to the historical records.

Keywords: climate analysis; building performance simulation; weather files; machine learning

RESUMO

SILVA, Mario Alves da, D.Sc., Universidade Federal de Viçosa, novembro de 2024. **Análise multiescala de dados climáticos para avaliação de desempenho de edificações no Brasil**. Orientadora: Joyce Correna Carlo.

A escolha de dados meteorológicos é fundamental para avaliar o desempenho de edifícios de forma precisa e representativa. Arquivos climáticos típicos geralmente aplicam uma abordagem estatística para selecionar meses representativos das condições climáticas atuais, mas não abrangem características específicas do local para diferentes locais no mundo.

Este trabalho analisa o potencial de uma análise multiescala para avaliação de desempenho de edifícios a partir de diferentes resoluções de dados meteorológicos, em um território geográfico abrangente, dado o tamanho do Brasil. O trabalho apresenta uma visão geral dos dados meteorológicos e sua aplicação na análise de desempenho de edifícios e um procedimento geral para coletar e processar dados meteorológicos, e diferentes abordagens para compilar arquivos climáticos para avaliação de desempenho de edifícios. O estudo também fornece uma análise extensa do território brasileiro, apresentando um perfil climático e tendências para todo o território. A análise se concentra em um resumo climático e bioclimático e em simulações de desempenho para cidades representativas de acordo com o zoneamento bioclimático brasileiro. Em seguida, o trabalho compara os registros da base de dados *ERA5-Land* e do INMET para quantificar as diferenças e apresentar o impacto na análise de desempenho de edifícios. O estudo propõe um novo método de compilação de arquivos climáticos para o Brasil e aplica testes estatísticos para determinar se a nova abordagem fornece melhores resultados do que os métodos TMY existentes. O procedimento abrange análise de correlação e sensibilidade com base em modelos de aprendizado de máquina para propor um método baseado em desempenho.

A análise inicial do território brasileiro mostrou predominantemente um aumento de temperatura com base nos registros de 2008-2022, com alguns locais atingindo mais de 1 °C. No entanto, a abordagem bioclimática com base no diagrama de Givoni mostrou que as estratégias de ventilação ainda são a abordagem mais eficaz em vez dos sistemas HVAC. Após a comparação entre dados espaciais de alta resolução e estações meteorológicas do Brasil, alguns locais apresentam anos insuficientes para uma análise multianual e alguns municípios mostram uma variação significativa não apenas dos dados meteorológicos, mas

também dos resultados de desempenho do edifício. Finalmente, a análise de novos arquivos climáticos para o Brasil permitiu concluir que a criação de arquivos climáticos típicos com base em uma abordagem de desempenho fornece os melhores resultados, uma vez que estão mais próximos dos registros históricos.

Palavras-chave: análise climática; simulação de desempenho de edificações; arquivos climáticos; aprendizado de máquina

ACRONYMS AND ABBREVIATIONS

AMY	Actual Meteorological Year
BRL	Boland–Ridley–Laurent model
BTMY	Brazilian Typical Meteorological Year
CDD	Cooling degree day
CDH	Cooling degree hour
CgT	Thermal load
CgTa	Thermal load for heating
CgTr	Thermal load for cooling
CVRMSE	Coefficient of variation of the root mean squared error
DBT	Dry bulb temperature
DHI	Diffuse horizontal irradiation
DNI	Direct normal irradiation
DPT	Dew point temperature
GHI	Global horizontal irradiation
HDD	Heating degree day
HDH	Heating degree hour
HVAC	Heating, ventilation, and air conditioning system
INMET	National Institute of Meteorology from Brazil
MBE	Mean bias error
NVP	Natural ventilation potential
PHFT	Percentage of hours in a certain range of operative temperature
PPT	Precipitation
PRES	Atmospheric pressure
RH	Relative humidity
RMSE	Root mean squared error
TMY	Typical Meteorological Year
Top	Operative temperature
TRY	Test Reference Year
U_{high}	Building envelope with high U-value
U_{low}	Building envelope with low U-value
U_{medium}	Building envelope with medium U-value

WBT	Wet bulb temperature
WD	Wind direction
WS	Wind speed
WVP	Water vapor pressure

TABLE OF CONTENTS

CHAPTER 1: Weather data for building performance assessment – an overview

The advances in 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 but always rely on the climatic aspects given by the simulator. Then, meteorological data is essential in building performance simulation since the results are significantly climate-dependent.

In the past, most weather data was based on records from ground weather stations, which usually recorded the dry bulb temperature, dew point temperature, pressure, relative humidity, wind speed and direction, solar radiation, and precipitation. During a building performance assessment, we usually intend to quantify building performance based on the weather for a generic location inside the administrative boundary that also covers the ground-station. However, ground stations are limited by their immediate surrounding conditions; thus, the data can provide inaccurate results when used in building performance assessment for generic locations, leaving the following question: should we stop using ground station data and rely only on other data sources?

The advances in hardware and software allowed improvements in several fields, including significant approaches to model climate conditions to generate past, present, and future weather data. Climate models are capable of simulating the interaction between the atmosphere, land surface, ocean, and sea ice by using mathematical models that can provide results for any location worldwide. Beyond the possibility of having weather data for any location worldwide, the climate databased that uses modeled data also has an advantage by providing a complete record without any gaps. The database completeness allows the creation of weather files for building performance simulations without any interpolation method.

The European Centre for Medium-Range Weather Forecasts (ECMWF) provides reliable climate datasets based on mathematical and physical models that account for atmospheric interactions and the resulting impact on the Earth's surface. ERA5 and ERA5-Land are datasets that contain different weather variables at a high resolution and are used in different studies (1,2,3). The ERA5 database provides a global database with a 31 km resolution, and the ERA5-Land dataset provides a refined model with a 9 km resolution (4,5). The database also provides data in high temporal resolution, with projections going back to 1940 for ERA5 and 1950 for the ERA5-Land database. From a building performance assessment perspective, ERA5-Land can provide more adequate weather data for a generic building performance

assessment since it can reproduce the mesoclimatic scale of a city.

Weather files are a collection of data capable of representing the meteorological condition of a given place. Then, a weather file is necessary for each new place to describe its conditions. Over the years, different methods to compile meteorological data into weather files were developed: files containing hourly data regarding variables such as temperature, radiation, pressure, wind, and humidity. These methods usually apply statistical analysis by selecting representative months from different years and creating a typical/average year containing hourly data from a historical series (6). Different studies show approaches based on updating weather files or even creating new procedures to process the meteorological data used for building performance simulation (7,8,9,10,11,12,13,14,15,16).

The Test Reference Years (TRY) were initially used to assess the meteorological conditions (17). These files describe the average climatic conditions 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, the method was modified and applied to several locations, where specific changes were made in different situations (18,19,20,21,10,22).

The Sandia method enabled the creation of a typical meteorological year based on analyzing a historical record by applying cumulative distribution functions and the Finkelstein-Schafer (FS) statistic (23). The method was then used to create the Typical Meteorological Year (TMY), where the reference year was selected based on Dry Bulb and Dew Point Temperature, Wind Velocity, and Global Radiation. For each weather parameter, a weight was assigned, representing the importance of the parameter on the TMY creation. The following versions, TMY2 and TMY3, updated the weights and added the Beam Radiation as a significant parameter for selecting the reference years (9,24). They were developed for USA locations, as the meteorological parameters and weights were defined based on the characteristics of the US climate.

The ISO 15927-4 (25) also uses the Sandia method to define the reference years but does not apply any weight to the meteorological parameters. The method, then, provides a general approach for defining the TMY weather file based on the Dry Bulb Temperature, Relative Humidity, Global Radiation, and Wind Velocity. The capability of generalization of the EN ISO 15927-4 allowed the mass production of weather files for any location with hourly data of the meteorological parameters for more than one year. Using the approach based on ISO 15927-4 (25), the Climate.OneBuilding.Org platform makes weather files available for

different locations worldwide (26). The availability of weather files for almost any location on the platform enables users to assess the performance of their building and can increase the proposition of more efficient buildings. However, using a generic approach to obtain weather files for different locations, users can reach performance values that are not close to reality since the most influential meteorological parameters can vary for each location. The same analysis can be made for most Typical Meteorological Year (TMY) generation methods since they usually do not apply sensitivity analysis methods to define meteorological parameter influences and the corresponding numerical importance.

Throughout the world, different studies propose using the ISO 15927-4 for creating new weather files based on the original method or by adjusting weights to represent better the observed historical period (16,10,11,27). Kalamees et al. (16) used the ISO 15927-4 method but added a sensitivity analysis for defining the most significant meteorological parameters on building energy demand in Finland. Given the general approach adopted by EN ISO 15927-4, its applicability was also successful in different locations from South Korea (27). The approach approximated the historical record for all the analyzed locations based on statistical indicators such as R^2 and the Pearson correlation.

Pernigotto et al. (10) addressed ISO 15927-4 modifications on the reference year selection method. The study focused on two main aspects: the reference year's final selection and using weights associated with the meteorological parameters. The results showed that the current procedure for the final selection of the reference year of ISO 15927-4, based on the wind velocity, does not provide the best approach. They also pointed out that using weights to design weather files focused on building performance indicators, such as heating and cooling energy demand, can provide results that best describe the relationship between building performance and meteorological parameters.

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 (7). Several climatic conditions can be found worldwide, so defining a single set of meteorological parameters and weights to select representative years is impossible. In the Brazilian scenario, given the climatic diversity (28), it is improbable that using a generic approach such as the ISO 15297-4 (25) or the TMY3 (9) users will be able to obtain building performance results with a high confidence level for all the locations. Then, adjustments must be made to adequate the existing methods to local aspects.

Simulations enable estimating building performance regarding different metrics, such

as thermal comfort and energy consumption. However, the building's performance directly depends on the weather data used. The user can create weather files for the same weather, depending on the methodology used to define the representative year or months. Then, the results can be highly influenced by the methodology adopted. Weather files are fundamental for Building Performance Simulation (BPS) since they can virtually emulate climatic conditions of a specific location and be used to obtain the performance of a building in the defined condition.

Since the last century, studies pointed out that using different weather files on BPS can lead to different results (29,30). By testing different formats of weather files and their impact on building performance, studies showed that users must consider the implications of their particular choices on which weather file is used to assess building performance. For instance, users must be aware of the approach used to select the representative months that will be used to obtain the 8760 hourly data that is present on the weather files they intend to use. 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.

From the proposition of Test Reference Years (TRY) in the 70s to methods such as TMY3 and IWEC2 (9,31), studies focused on enhancing the representativeness of the obtained weather files by applying new statistical methods to determining the most influential years, capable of better describing the average condition of the selected timeframe (6). The efforts to update the weather files methodology, combined with the advances in the computational field, enabled the creation of weather files for any location. 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 principal meteorological parameters in a weather file.

The possibility of creating methodologies that allow the generation of weather files for any location on the globe also represents a possibility to assess the performance of a building in any location on Earth. Despite the advances and possibilities above, users must take caution since the following question remains: Are the weather file generation methods used the most adequate and represent a good fit for any location worldwide?

Many studies provide new approaches regarding the definition of the most significant variables and their weights in the selection process (32,10,16,33). Despite providing analyses on different locations, most of these studies share similar aspects. They usually define representative meteorological parameters and weights, create new weather files based on these combinations, run simulations, and compare the results with simulations from Actual

Meteorological Year (AMY) weather files. The weather file that provides the closest result to the historical average is the best TMY for the location.

Building performance is site-dependent. Therefore, it is directly related to the weather conditions at the location used for the simulation. Hosseini, Bigtashi, and Lee (34) and Kalamees et al. (16) proposed a similar method for selecting the reference year for each month to generate a new typical meteorological year. Both studies first analyze the impact of meteorological parameters on the energy demand of different building typologies. Once they define the most significant meteorological parameters, they use them with their respective weights to define the representative years. The studies, however, differ in the sensitivity analysis approach to define the parameters and weights. Kalamees et al. (16) define the range of variation of each meteorological parameter and then compare their influence on residential and non-residential typologies. Hosseini, Bigtashi, and Lee (34) used the results of building performance simulations with actual meteorological years (AMYs), and for defining the weights, they used a machine learning method to identify the impact of each meteorological parameter on the building performance. The results indicate potential energy saving and better correlation to site conditions than generic methods (7).

The climate change scenario raised a significant question in the building performance simulation field: Are Typical (average) Meteorological Years satisfactory for designing efficient buildings for today and the future? Different studies approach the necessity of using not only typical weather files but also assessing building performance based on extreme scenarios for heating, cooling, and even humidity (12,35,36,37).

Usually based on the Sandia method and the creation procedure of TMY, the studies induced a slight change in the selection of the reference year that completely modified the final weather file (12,35,36,37). During the TMY creation, after the FS statistic calculation, users must increasingly order the FS values to rank the years with the lower difference from the historical data. The approach used by Crawley and Lawrie (37), Gasparella et al. (12), and Pernigotto, Prada, and Gasparella (35) rank the FS values in the decreasing order, so the years that have the highest differences are in the first positions. Then, each study proposes specific selection methods to define which years should be part of the extreme reference years. Ji et al. (36) provide a similar approach but focus on the heat extremes since the study uses weather files that provide closer data on heat wave events.

The extreme weather events can be derived from a historical record, as previously described, or they can be related to weather changes in future scenarios. The Intergovernmental

Panel on Climate Change (IPCC) provides scientific assessments on the effects of climate change on Earth's climate. The different reports focus on a specific historical record and the resulting projections for greenhouse gas emissions and temperature increase. Since weather files are based on hourly records of different weather variables, it's possible to use the IPCC projections to create weather files based on extreme events derived from future climate scenarios. Rodrigues, Fernandes, and Carvalho (38) created the *Future Weather Generator* tool that creates future weather files based on the last IPCC report. The tool requires a weather file with data within the 1985-2014 range, and then it produces the future data based on different models for the 2050 and 2080 timeframes (39,40,41).

1.1. The Brazilian scenario

The Brazilian National Institute of Meteorology (INMET) provides weather data from automatic weather station (AWS) records for more than 400 locations and weather data from analogic stations for several other locations. The AWS records date from 2000 to the present day and mainly comprise air temperature, relative humidity, dew point temperature, pressure, relative humidity, solar radiation, wind speed and direction, and precipitation on an hourly basis. The records provided by INMET can be directly used to compile weather files since they have the main components required to create a weather file. However, they can present gaps that affect the weather data quality and the reliability of simulation results.

The Climate.OneBuilding.Org repository represents the most extensive weather files database for Brazil. The database has 1867 weather files for Brazil in three different compilation approaches, but they share the same database for most climatic variables. There are 17 TRY files using weather records from before 2000 representing a single year, and 411 INMET weather files that used the TMY method to select the representative years for reference months based on records from 2000 to 2010 (23). Finally, the TMYx weather files have three possible variations: 2004-2018, 2007-2021, or complete record. For the first temporal series, there are 201 weather files available, and they are based on INMET’s records. The 2007-2021 files are available for 578 locations, and the INMET data was combined with solar radiation from ERA5. Finally, 660 weather files used all the historical records from the INMET’s database.

Different studies discussed the impact of different weather files on building performance assessment in Brazil, focusing on moisture (42), climate zoning impacts (43), building labeling (44), and thermal comfort (45). All the studies discussed the impacts of different weather files, pointing out the preferred weather file over the others. However, none focused on proposing a

new compilation method or determining the most reliable weather file. Machado et al. (32) propose a new typical weather file for Brazil based on the Sandia method but focusing on radiation parameters since the studies intend to estimate better the Brazilian territory's capacity for generating solar energy. Despite this advance, the study does not focus on the building performance. It employs the usual approach of defining the new typical weather file based on an analysis of weather files created by different combinations of weights. Then, the study points out a new combination of weights to obtain the best TMY but only focuses on the potential for solar energy generation.

The Brazilian territory is extensive and encompasses different climatic conditions (28) Thus, applying a weather file creation method based on generic or even different climatic conditions seems inadequate and can lead to results that do not represent the real influence of weather on Brazilian building typologies. The available methods consider different meteorological parameters and weight criteria to select the most representative data and generate the required weather file for building performance simulations. However, selecting those parameters and weights is usually based on standard building typologies that are tested or climate conditions that are different from Brazil (9,20,18,11,16). According to the Köppen classification, the Brazilian territory encompasses different climates majorly represented by tropical, subtropical, and semi-arid climates, where annual mean dry bulb temperatures can go from 13.5°C in São Joaquim-SC (28.29 S, 49.93 W) to 28.6°C in Bom Jesus-PI (9.075 S, 44.36 W) (46,28). As for the building typologies, different studies can point out the typical building characteristics for the Brazilian territory (47,48,49). Thus, a single approach based on a fixed weight set for the territory can lead to biased results.

This study's novelty relies on providing an extensive climate analysis for the Brazilian territory from a building-centered perspective, supported by a quantitative and qualitative assessment of weather records from 2008 to 2021. We intend to quantify the similarities between the ERA5-Land and INMET’s data by determining the impact of the surrounding microclimatic conditions on building performance. We also propose a new compilation method for the Brazilian weather files based on a sensitivity analysis approach. The new method, specific to the Brazilian territory, intends to identify the most significant meteorological parameters on the performance of different building typologies and performance indicators. Then, the parameters' significance value will be used to generate new weather files based on the Sandia method. The approach will follow the methodology proposed by Hosseini, Bigtashi, and Lee (34). We will address different locations and building typologies to encompass the

heterogeneity of the Brazilian territory. Finally, we will create a new weather file database for Brazil containing all the weather files created in this study.

1.2. Hypothesis, objectives, and scope

Based on the presented scenario, this research proposes the following hypothesis: **A proper characterization of Brazil, given the geographical extension of the territory and the climatic heterogeneity, must rely on a thorough climatic comprehension of the whole territory and the consequences on the built environment, i.e., a multi-scale analysis.** Following the hypothesis, we intend to answer the following questions:

1. How much can microclimatic (automatic weather station) and mesoclimatic data affect the results of building performance assessment?
2. How can a building performance-related method based on sensitivity analysis influence the proposition of new weather files for the whole Brazilian territory?

This research proposes a multi-scale analysis of weather data used in building performance-related analysis, from outdoor environmental characterization to indoor analysis supported by machine learning, sensitivity analysis, building performance simulation, and statistical tests. The research encompasses the following specific objectives to support the primary objective:

1. Create a climatic profile for the Brazilian territory;
2. Quantify the differences between the Brazilian automatic weather stations and ERA5-Land data;
3. Determine the impact of using different typical meteorological years for building performance assessment;
4. Propose and validate a new methodology to select representative months for TMY in Brazil.

1.3. Overall methodology for creating weather files

As the main topic of this thesis, weather files require two main elements: hourly weather records for different weather variables and a compilation methodology to select the reference years, when dealing with typical meteorological years (TMY). The following subsections will present the process of retrieving and processing weather data from two main sources, the ERA5-Land database and the Brazilian meteorological institute (INMET). The subsections will also

present a step-by-step description of the reference years selection procedure of different TMY methodologies.

1.3.1. Retrieving and processing weather records

The ERA5-Land represents a worldwide database with weather records for almost all inland locations within a 0.1 ° resolution (approximately 9 km) (4). The modeling procedure is based on the ERA5-Reanalysis (5) component with a higher focus on the land components to increase the data resolution and provide a more refined grid. The modeled records are obtained from mathematical and physical calculations that emulate the interactions between the different atmospheric components, such as radiation, precipitation, and temperature. Therefore, the modeled results keep the correlation with real conditions. The records from INMET derive ground stations scattered across the Brazilian territory and provide measured data for different weather variables.

Throughout this thesis, all weather files follow the EnergyPlus Weather File (EPW) compilation mode, since EnergyPlus will be used for the simulations. The EPW format comprises hourly records of several weather variables and can be used in different building performance simulation software that are in compliance with the BESTEST evaluation from ASHRAE 140, such as TRNSYS (50) and IES VE (51). Despite several weather information, many weather variables are not currently used in EnergyPlus (52) as shown in Table 1.1.

Table 1.1. Weather data present in the EPW format.

Weather data	Is it used in EnergyPlus calculations?
Dry bulb temperature	Yes
Dew point temperature	Yes
Relative humidity	Yes
Atmospheric pressure	Yes
Extraterrestrial horizontal radiation	Yes
Extraterrestrial direct normal radiation	No
Horizontal infrared radiation intensity	Yes
Global horizontal radiation	No
Direct normal radiation	Yes
Diffuse horizontal radiation	Yes
Global horizontal illuminance	No
Direct normal illuminance	No
Diffuse horizontal illuminance	No
Zenith luminance	No
Wind direction	Yes
Wind speed	Yes
Total sky cover	Yes
Opaque sky cover	Yes
Visibility	No
Ceiling height	No
Present weather observation	Yes
Present weather codes	Yes
Precipitable water	No
Aerosol optical depth	No
Snow depth	Yes
Days since last snowfall	No
Albedo	No
Liquid precipitation depth	Yes
Liquid precipitation quantity	No

Dry bulb temperature (DBT), dew point temperature (DPT), pressure, precipitation,

global horizontal irradiation (GHI), horizontal infrared radiation (IR), wind vector for the U and V components, and albedo records for the 15 years were downloaded from the ERA5-Land, used in chapters two to four. As an advantage, ERA5-Land records do not have gaps, therefore they reduce the uncertainty when dealing with missing values of weather variables. From the downloaded variables, only the albedo from weather files is not used in EnergyPlus. However, the data was 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. The MetPy plugin for Python (53) was used to calculate the wind speed (WS), direction (WD), and relative humidity (RH).

Different from modeled data provided by the ERA5-Land database, the INMET automatic weather stations (AWS) already provide the RH, WS, and WD records, but do not measure the IR component, neither the albedo. The INMET data can also have missing records, and this topic will be discussed in Chapter 3. For the INMET compilation, the procedure described in the EnergyPlus Engineering Reference manual (54) was used to calculate the horizontal infrared radiation and the albedo data was left blank, since it does not interfere with the EnergyPlus simulation.

For both ERA5-Land and INMET data, the solar position variables were calculated according to Meeus (55), the BRL-Brazil model (56) was used to separate the GHI into direct normal (DNI) and diffuse (DHI), and the illuminance variables followed the Pérez model (57). For the EPW compilation, the ground temperatures and the typical/extreme periods were determined following the procedure described by Rodrigues et al. (38).

1.3.2. Weather file compilation

The creation of an Actual Meteorological Year (AMY) is enabled after compiling all the data from Section 1.3.2. As a weather file that contains the records from a single year for all the 8760 hourly records, AMY represents a real condition.

For building simulation analysis, however, a simulation with an AMY can be misleading, since it does not represent and average condition and a multi-year analysis, with multiple AMYs can have a high computational cost. In this context, Typical Meteorological Years (TMY) are usually employed in building performance simulations, since they are usually compiled by selecting monthly records that are closer to a multi-year series.

There are several methodologies to select the reference months of TMY and the following subsections will describe the selection process according to the ISO 15927-4 (25), minimum Finkelstein-Schafer (22), Pissimanis (21), and Best Rank I methods (10). The process

of selecting the reference year for each month usually relies on using some statistical analysis to compare the long-term records and the individual years.

The ISO 15927-4, the minimum Finkelstein-Schafer, and the Best Rank I methods follow the initial steps of the Sandia method (23), described as follows:

1. Calculate the daily average $\bar{p}$ of the climatic parameters p (DBR, RH, and GHI), for every month m and year y;

2. Put together all daily averages $\bar{p}$ from the multi-year records, sorting in increasing order all the averages $\bar{p}$, and calculate the cumulative distribution function for each daily record i^{th} as shown in Equation (1.1):

$$\Phi(p, m, i) = \frac{K(i)}{N+1} \tag{1.1}$$

where $K(i)$ is the rank order of the ith value of the daily means and N is total number of days for a month over the multi-year series.

3. Put together all daily averages $\bar{p}$ for every calendar year y, sorting in increasing order all the averages $\bar{p}$, and calculate the cumulative distribution function for each daily record i^{th} as shown in Equation (1.2):

$$F(p, y, m, i) = \frac{J(i)}{n+1} \tag{1.2}$$

where $J(i)$ is the rank order of the ith 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 (1.3):

$$FS(p, y, m) = \sum_{i=1}^{n} |F(p, y, m, i) - \Phi(p, m, i)| \tag{1.3}$$

1.3.2.1. Best Rank I, ISO 15927-4, and minimum Finkelstein-Schafer reference years

After calculating the FS statistic, for every month and variable, the values are sorted in

increasing order and ranks are assigned. Then, for every month the ranks are summed and sorted in increasing order. For the Best Rank I method, the month with the lowest rank is selected as the reference year for the month.

For the ISO 15927-4, pick the three months with the lowest rank and select the one with the lowest deviation for the monthly mean wind speed between the calendar year and the multi-year is selected as the reference year.

The minimum Finkelstein-Schafer method is similar to the Best Rank I method, however, instead of assigning ranks, the method selects as reference the year with the lowest summed FS values.

1.3.2.2. 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 multi-year records. First, the root mean squared error (RMSE) is calculated for the hourly GHI values. Then, the years with deviations below a certain threshold or with the 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.

Then, for each method (ISO, Best Rank I, minimum Finkelstein-Schafer, and Pissimanis) a reference year is selected for each month, resulting in complete year composed of 12 representative years.

1.4. Thesis structure

This section presents an overview of this thesis structure, distributed along five chapters. Chapter one presents an overview of weather data and its application to building performance analysis, providing a worldwide scenario and the Brazilian context. The first chapter also presents the hypothesis, objective, and scope of this research. Addressing the weather file subject, Chapter one also provides the methodology used in chapter three and four to process the raw weather data and compile the different weather files used in this study.

Chapter two is a transcription of the paper “Towards Climate, Bioclimatism, and Building Performance – A Characterization of the Brazilian Territory from 2008 to 2022” published in the Buildings MDPI journal. The study provides a comprehensive characterization of the Brazilian territory, analyzing main climatic features for 5,567 Brazilian municipalities.

Chapter three focus on the comparison of modeled weather data, such as the one provided by the ERA5 repository and the measured data from the Brazilian meteorological network. The approach quantified the difference on the data regarding raw data assessment and building performance simulation applications. For the building performance assessment, the analysis followed the Brazilian standard for the minimum performance requirements of residential buildings – NBR 15575.

Chapter four addresses the suitability of different weather files to building performance simulation, considering 480 municipalities in Brazil and the records from 2008 to 2022. The approach focused on initially compiling AMY weather files and TMY based in known methodologies. Then, the best TMY was identified for each location. After, a performance-based weather file was proposed and the applicability of new weights on the reference years selection process was tested against the 15-year simulation outputs and the best TMY. Therefore, the best weather file for each location was determined.

Chapter five presents a conclusion based on the overall results of the previous chapters. The chapter provides a discussion of the hypothesis and objectives from section 1.2, considering all the results obtained. The final chapter also presents the limitations and recommendations for future research.

CHAPTER 2: A multi-year characterization of the Brazilian territory

Chapter two is a transcription of the paper “Towards Climate, Bioclimatism, and Building Performance – A Characterization of the Brazilian Territory from 2008 to 2022” published in the journal Buildings from MDPI, co-authored by Giovanni Pernigotto, Andrea Gasparella, and Joyce C. Carlo.

This chapter intended to characterize the Brazilian territory based on the annual average of dry bulb temperature, relative humidity, solar radiation, wind speed, and precipitation downloaded from the ERA5-Land database. We considered a 15-year record (2008-2022) to analyze Brazil’s climatic profile and update the Köppen-Geiger classification of 5567 Brazilian municipalities. We also used the dry bulb temperature and relative humidity values from ERA5-Land to provide a monthly bioclimatic analysis of the Brazilian territory using the Givoni bioclimatic chart and assessed the natural ventilation potential. We also analyzed the heating and cooling demands using the degree days indicator.

Finally, we identified representative locations according to the Brazilian bioclimatic zoning criteria and the Kolmogorov-Smirnov test to provide a building performance characterization of the territory.

2.1. Methods

2.1.1. Weather data pre-processing

First, all the Brazilian municipalities with data within the ERA5-Land database were identified by setting coordinates as close as possible to the urban area. Data were retrieved for 5567 Brazilian municipalities, i.e., 99.9% of them, except for Fernando de Noronha-PB, Itaparica-BA, and Madre de Deus-BA, which did not have any of their territory within the ERA5-Land database. Then, using the ERA5-Land Monthly Aggregated data available on the Google Earth Engine (GEE) catalog, this study gathered the records from 2008 to 2022 for the dry bulb temperature (DBT), dew point temperature (DPT), global horizontal solar radiation (GHI), precipitation (PPT), and the vectorial components of wind (U and V). Daily aggregated data were also downloaded, but only for DBT and DPT, to provide a more detailed bioclimatic analysis, as described in Section 2.1.3.

After downloading the climatic variables, the MetPy library for Python was used to estimate the relative humidity (RH) from DBT and DPT and to calculate the wind speed (WS) and direction.

2.1.2. Climatic approach

After pre-processing all the weather variables, the 15-year averages were calculated based on the annual averages from monthly DBT, RH, PPT, and GHI records. The 15-year minimum and maximum averages for DBT, RH, and PPT were also calculated based on each year's minimum and maximum monthly records. Following the 15-year analysis of the average, maximum, and minimum records, the cumulative average was used to quantify the difference between 2008 and 2022, and the Mann–Kendall statistic test with a significance level of 5% was used to identify the trends. The choice for the Mann–Kendall test was based on the vast application of this statistical test to identify trends in meteorological records (58,59). The test identifies monotonic trends, i.e., explicit increase, decrease, or no trend, considering a linear relation between the time series and the weather variables.

The GHI records were also analyzed in terms of the annual average from monthly integrals to compare with the records from the Brazilian Solar Atlas (60). This Atlas provides solar radiation data for the entire Brazilian territory with the same resolution as ERA5-Land. It incorporates mathematical models of prediction validated by solar radiation measurements in Brazil spanning from the late 1990s to 2017.

Monthly values of DBT and PPT were used for the climatic characterization update of the Köppen–Geiger classification for the Brazilian territory (61,62) Results were compared with the classification by Alvares et al. (28), which considered weather data records from Brazilian and international sources from 1950 to 1990. In order to facilitate comparison, the results of Alvares et al. (28) were simplified by assigning a single classification for each municipality.

2.1.3. Bioclimatic approach

This approach focused on the bioclimatic characterization of the Brazilian territory, mainly supported by Givoni's bioclimatic chart (63,64,65). For the Brazilian context, Givoni's bioclimatic chart can reach 12 groups, encompassing design strategies and a comfort zone. Groups A, B, and C are related to heating strategies. They represent space heating with active HVAC systems, passive solar heating, and thermal mass combined with solar heating. Group D represents the thermal comfort zone. Groups E, F, and G are related to ventilation strategies and are divided into daytime ventilation, night cooling with thermal mass, and night cooling with thermal mass and evaporative cooling. Group H requires thermal mass and evaporative cooling, group I requires only thermal mass for cooling, and group J only evaporative cooling. Group K requires humidification, and group L represents an active HVAC system for space cooling.

The daily records of DBT and RH described in Section 2.1.1 were used as input since the monthly records can mask daily patterns and oversimplify strategies' applications. The monthly requirements were quantified for each year and the results summarized as the 15-year average for each municipality. For the final monthly requirements, all the strategies with an occurrence of at least seven days, consecutive or not, were considered. Then, the results were compared with the bioclimatic strategies for summer and winter as recommended by the NBR 15220:3 (66), still considering the bioclimatic zoning from Refs. (66,63) since the new proposal does not make bioclimatic recommendations based on the Givoni methodology (67).

In the Brazilian context, ventilation is an important passive design strategy according to different studies (63,65,68,69) since it has the potential to reduce energy demand and increase thermal comfort. Using the daily records, the Givoni bioclimatic analysis was complemented by calculating the daily Natural Ventilation Potential (NVP) and the shading necessity for all the municipalities. Chen et al. (70) and Sakiyama et al. (71) described the NVP indicator as the number of hours in which the outdoor conditions favor using ventilation strategies to cool the indoor environment instead of active systems. The indicator quantifies the potential based on

the DBT, WS, and wet-bulb temperature (WBT). The DBT, WBT, and WS were set according to the ASHRAE 55 parameters (72), using the upper limits of the 80% thermal comfort acceptance level for the DBT threshold and the 16.8 °C limit for the WBT one. For the WS threshold, the same settings as Sakiyama et al. (71) were adopted. Finally, the thresholds were combined, and the average number of days that could use natural ventilation as a cooling strategy was quantified based on the 15-year records.

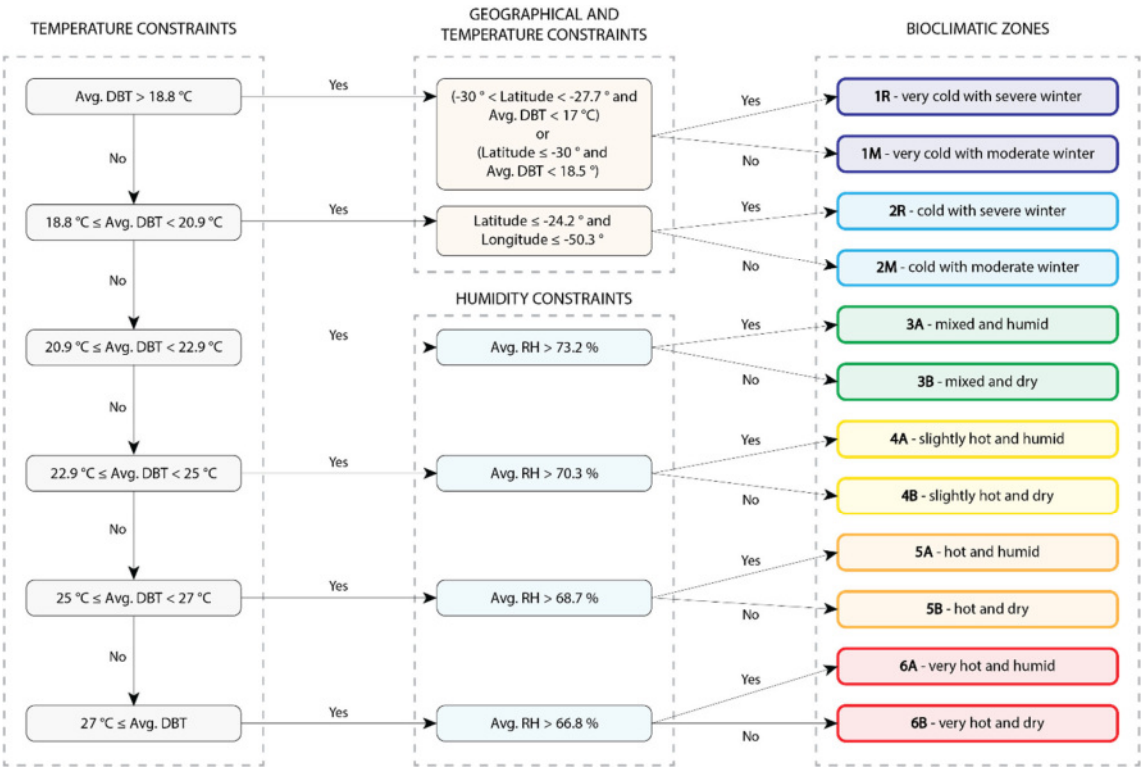
Givoni's chart was used to determine whether passive and active cooling and heating systems were necessary. To bolster this assessment, the average heating degree days (HDD) and cooling degree days (CDD) were calculated for the past 15 years using daily data. The HDD was calculated with an 18 °C base temperature, while the CDD was calculated using base temperatures of 24–26 °C. Using different base temperatures for CDD allowed us to understand the impact of a 1 °C variation on cooling requirements and to identify regions with the highest cooling demands.

2.1.4. Spatial criteria for the climatic, bioclimatic and building performance simulation analysis

All the climatic and bioclimatic analyses from Sections 2.1.2 and 2.1.3 were performed using the ERA5-Land monthly and daily data, focusing on three territorial distributions. First, a general characterization was provided considering all 5567 municipalities, creating a climatic profile for Brazil. Then, the results were analyzed based on the five Brazilian regions (Figure 2.1), showing the overall and predominant characteristics. Finally, the results were analyzed based on the new bioclimatic zones for Brazil (Figure 2.2) using the criteria defined by Silva Machado et al. (67) to determine the climate zone of each municipality based on the average from 15-year DBT and RH records from the ERA5-Land database.

Figure 2.1. The five administrative regions of Brazil and their percentage of municipalities.

Figure 2.2. Criteria for defining bioclimatic zones in Brazil. Adapted from Silva Machado et al. (67).



Silva Machado et al. (67) used the data from 298 locations for the building performance simulation and the proposition of the new Brazilian bioclimatic zones. For the remaining 5272 locations, they used annual average DBT data from the ERA5-Land database and machine learning to predict the annual average RH. Since this present study focuses on the ERA5-Land database, the criteria summarized in Figure 2 were adopted to classify the Brazilian locations using only the ERA5-Land data. Since the analysis based on the zones that resulted from Silva Machado et al. (67) could lead to inconsistencies due to different climate data sources, i.e., locations with DBT and RH that do not respect the limits for each zone, a reclassification was performed using only the ERA5-Land records.

The proposal for a new bioclimatic zoning for Brazil mainly defines the zones based on DBT and RH records. However, it does not define representative locations to summarize the characteristics of each zone or the ideal building envelope configuration to enhance thermal and energy performance (67). Pernigotto et al. (73) proposed a climate cluster methodology based on hierarchical clustering and a method to define the representative locations based on the Kolmogorov–Smirnov statistical test. Since a recent Brazilian bioclimatic zoning proposal will be used as the guide for the new Brazilian standard, the methodology from Pernigotto et

al. (73) was adopted only to define the representative locations for each bioclimatic zone.

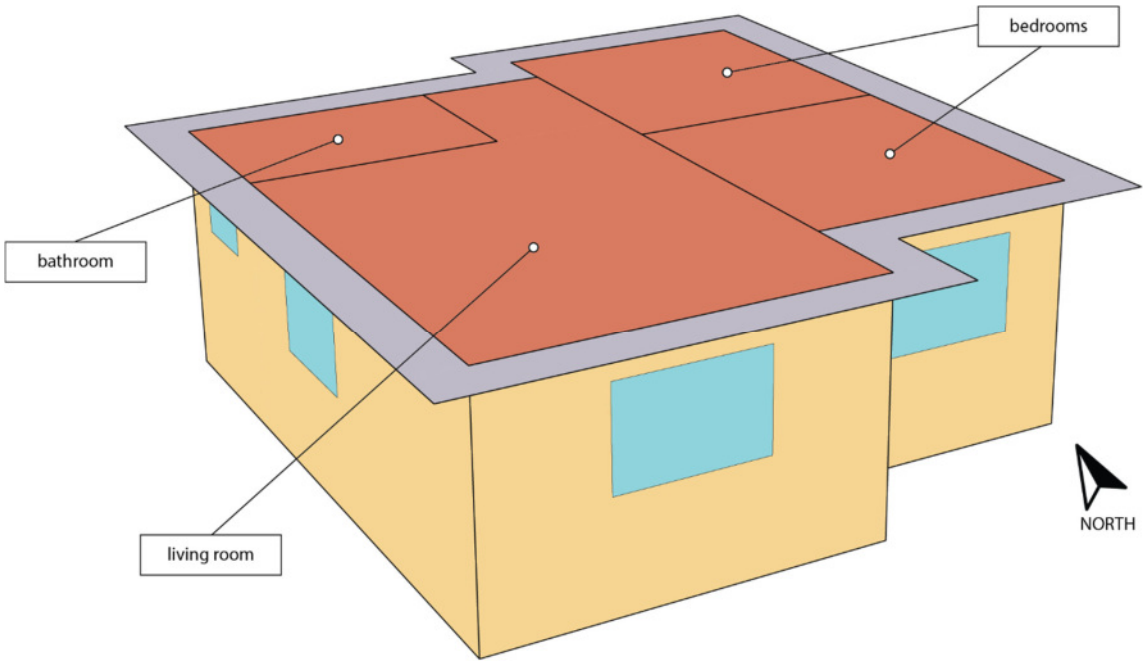
The representative locations were originally defined based on the monthly averages and spreads of DBT, GHI, and water vapor pressure (73). However, the Brazilian bioclimatic zoning proposition already concluded that the average DBT and RH are the most significant weather parameters, except for zones 1R, 1M, 2R, and 2M, where DBT and heating degree days with a 14 °C base temperature (HDD_{14}) were used to define the final zone (67). Based on those findings, first, each municipality's average DBT, HDD_{14} (zones 1 and 2), and RH (zones 3 to 6) were calculated to determine their classification. Then, each location's and group's monthly average (DBT and HDD_{14} for zones 1 and 2, and DBT and RH for zones 3 to 6 were compared using the Kolmogorov–Smirnov statistical test. Individual ranks were assigned to DBT, HDD_{14}, and RH. Finally, the assigned ranks were summed up and the locations with the lowest ranks were selected to represent each climatic zone. The adopted statistical test measures the maximum difference between two distributions, so a lower rank indicates the closer location to the group's average. Therefore, the representative locations from zones 1 and 2 were chosen based on their proximity to monthly DBT and HDD_{14}, while the remaining ones were chosen according to their proximity to monthly records of DBT and RH.

After defining the representative locations, ERA5-Land hourly data were downloaded from the GEE repository to create a 15-year series of weather files for each of the 12 locations. The process described in Section 2.1.1 was implemented to obtain the RH and WS, with the same Python library to calculate the wind direction (WD). GHI was also separated into direct normal (DNI) and diffuse (DHI), using the BRL-Brazil model since it delivers better results for the Brazilian territory (56). Then, each year was compiled for each location as a weather file, representing an actual meteorological year (AMY).

2.1.5. Building performance simulation approach

EnergyPlus 23.2 was used to run the building simulations. Geometry and configurations were prepared according to the software requirements and the standard for the minimum performance requirements for the residential sector in Brazil (NBR 15575) (74). The geometry proposed by Triana et al. (47) and Veiga et al. (48) of a single-family house encompassing two bedrooms, a coupled living room and kitchen, and a bathroom was used in this study (Figure 2.3).

Figure 2.3. Single-family model.



Window's properties (U-value of 5.7 W/m2.K and SHGC of 0.87), schedules, and internal gains were set according to NBR 15575, but different settings were adopted for the internal constructions and building envelope (Table 2.1).

Table 2.1. Building envelope configuration.

Construction set	Surface	U-value (W/m².K)	Thermal capacity (kJ/m²K)	Solar absorptance
U_{low}	Walls	0.41	125	0.3, 0.5, and 0.7
	Roof	0.56	230	0.6
U_{medium}	Walls	2.51	150	0.3, 0.5, and 0.7
	Roof	2.01	21	0.6
U_{high}	Walls	4.84	220	0.3, 0.5, and 0.7
	Roof	2.41	233	0.6

For each bedroom, the standard defines two people with an activity level of 81 W and an occupation period from 10 pm to 8 am. The bedrooms have a lighting power density of 5 W/m2 and two operational periods (6 am to 8 am and 10 pm to midnight). For the living room, the standard defines four people, and, since the space was split into two thermal zones, an occupancy density of 0.187 ppl/m2 was used to account for the correct distribution between them. The living room has an activity level of 108 W, half of the occupation from 2 pm to 6 pm, and a full occupation from 6 pm to 10 pm. The living room also has a lighting power density of 5 W/m2, available from 4 pm to 10 pm. The standard defines an electrical equipment power for the living room of 120 W, available from 2 pm to 10 pm. For this study, a density of 5.6 W/m2 was modeled, considering the two thermal zones representing the living room. The schedules are the same for the entire year since the standard does not account for different operational profiles during the weekends or holidays.

Two conditioning modes were simulated: one naturally ventilated model with a 19 °C setpoint during the occupied hours and one that accounts for the building thermal load using an ideal HVAC system with a 21 °C heating setpoint and a 23 °C cooling setpoint. Four building azimuths (N, E, W, and S) were considered for both conditioning models and all the building envelope configurations.

Following the analysis procedure described in the NBR 15575, the percentage of hours within an operative temperature range (PHFT) was calculated as the performance indicator for the naturally ventilated buildings. The operative temperature (Top) ranges are defined according to the average outdoor DBT (Table 2.2).

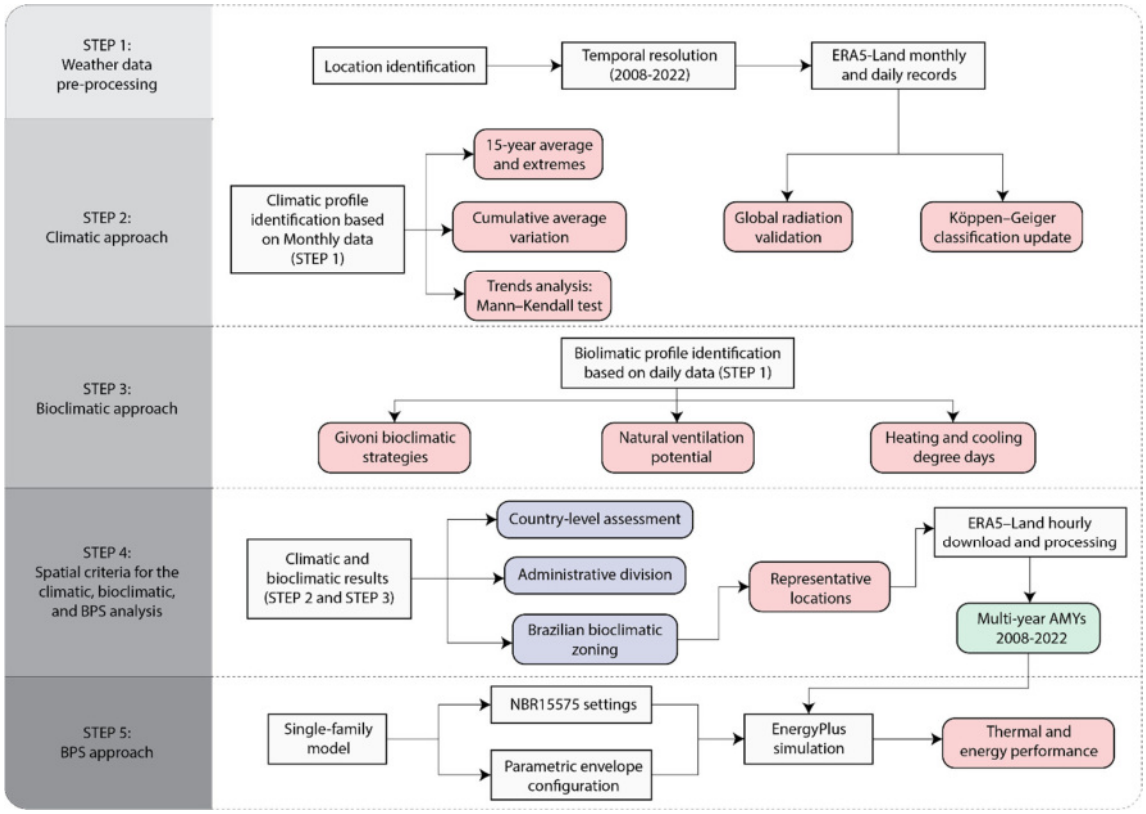
Table 2.2. Operative temperature range for the *PHFT* calculation.

Annual Avg. *DBT* (°C)	Operative temperature range	Bioclimatic zone
Avg. *DBT* < 25 °C	18 °C < *Top* < 26 °C	1R, 1M, 2R, 2M, 3A, 3B, 4A, and 4B
25 °C ≤ Avg. *DBT* < 27 °C	*Top* < 28 °C	5A and 5B
Avg. *DBT* ≥ 27 °C	*Top* < 30 °C	6A and 6B

Then, the average of all thermal zones was estimated to obtain the model's PHFT. For the model with the ideal HVAC system, the heating (CgTa) and cooling demand (CgTr) were determined, expressed in kWh/year. The CgTa indicator is only required for locations with an average DBT below 25 °C and accounts for heating demands in hourly Top ≤ 18 °C records. A similar rule applies to the CgTr indicator, which only accounts for the cooling demands in times that exceed the upper limits from Table 2.2. After calculating the annual thermal loads, the demands of all zones were summed up and the results were analyzed for the single-family house. The multi-year results for all representative locations were analyzed to identify the impact of different construction sets on the detached house performance throughout the 15 years as the average of the building's azimuth.

The methodology proposed for this study can be summarized as a 15-year analysis based on climate and meteorological variables (DBT, RH, PPT, and GHI), an updated Köppen–Geiger classification, and a bioclimatic approach focusing on NVP, degree days, and the Givoni strategies. Finally, focusing on the Brazilian bioclimatic zoning, building performance was assessed in representative locations, as summarized in Figure 2.4.

Figure 2.4. Summary of the methodology used in this study. The red shading represents the analysis outputs, the blue shading indicates the spatial resolution, and the green shading shows the products made available for building performance simulation studies.

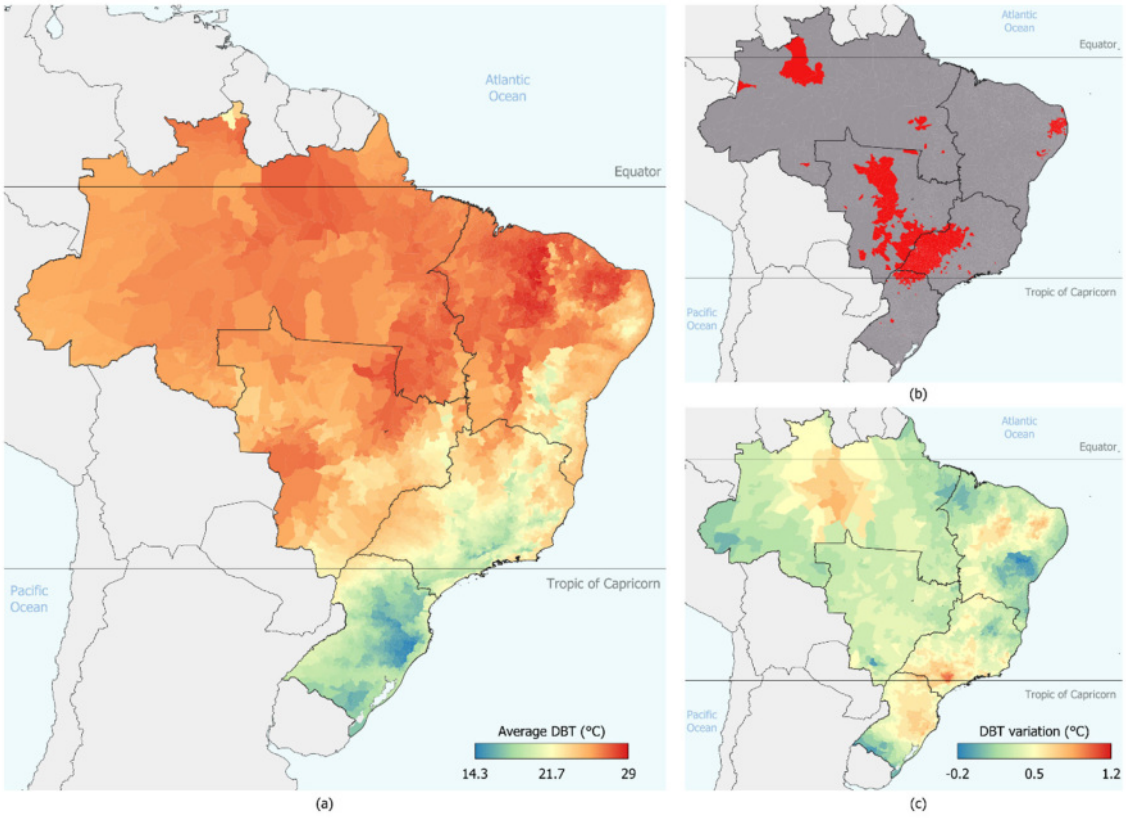


2.2. Results

2.2.1. Climate, meteorological, and bioclimatic data following the administrative boundaries

The annual average dry bulb temperature (DBT) over the years varied from 14.3 °C to 29 °C, with an average value of 23.2 °C and a standard deviation of 3.1 °C (Figure 2.5a). The Mann–Kendall results showed that 15 % of the municipalities (812) had an increasing trend throughout the 15-year annual records (Figure 2.5b). The north region showed the lowest increase (11 municipalities), and the southeast region had the highest occurrence (416 municipalities). The results showed that the southeast region encompasses more than 50% of the municipalities, with an increasing trend, while the other regions had between 12.1% and 15.9%, except for the north region (1.4%).

Figure 2.5. Annual dry bulb temperature (a), Mann–Kendall summary (b), and temperature variation (c) for 5567 Brazilian municipalities based on records from 2008 to 2022. For the Mann–Kendall summary, the gray area indicates no trend, while red indicates an increasing trend.



Over the 15-year record, the mean variation was 0.5 °C, representing a significant temperature increase in 3288 municipalities, approximately 60 % of the Brazilian municipalities (Figure 2.5c). Out of them, 44 municipalities showed temperature variations from 1 °C to 1.2 °C. Except for two municipalities in the state of Paraná, all the municipalities with temperature variations above 1 °C were in the state of São Paulo. Those results, supported by the Mann–Kendall analysis, state that the southeast region had the highest DBT variations in Brazilian territory. The highest variations in the north and northeast regions represented a 0.9 °C difference between 2008 and 2022. Therefore, it can be concluded that most of the Brazilian territory has annual averages of at least 22 °C and a temperature increase between 0.4 °C and 0.7 °C.

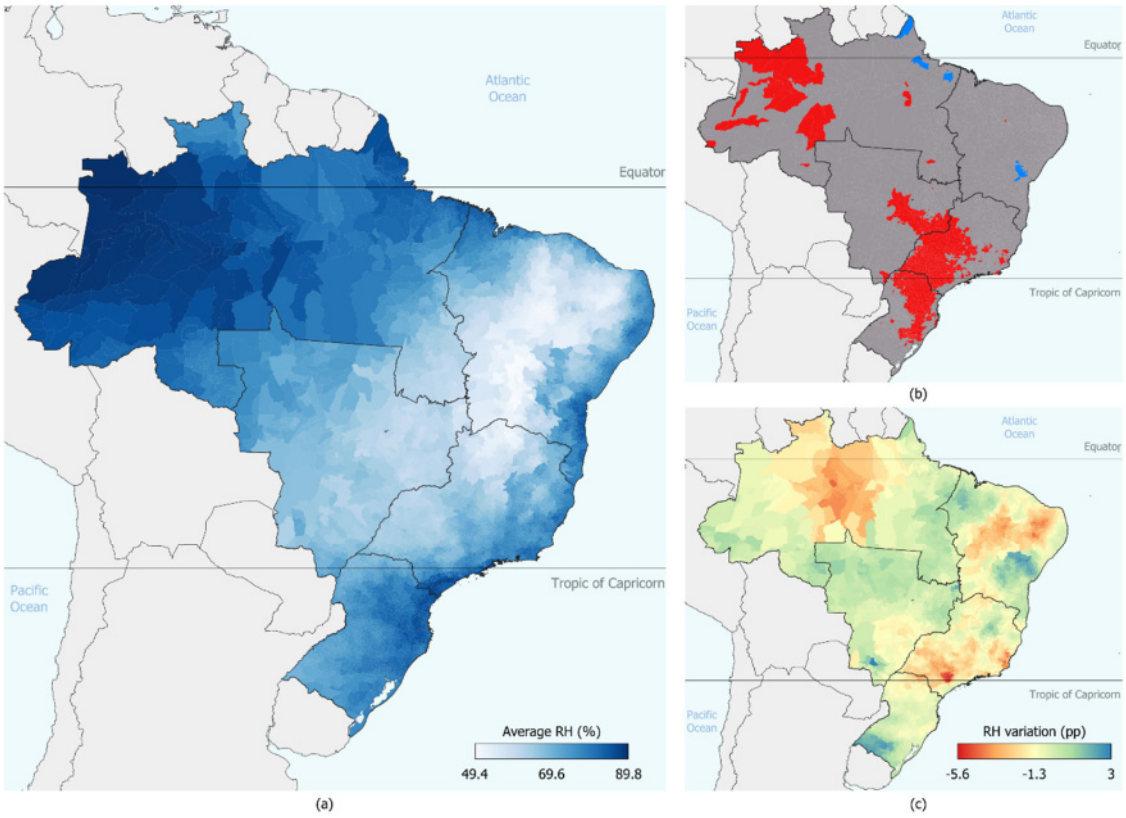
The average of the maximum and minimum annual DBT results were mainly related to the altitude and the latitude since the locations close to the Equator and with lower altitudes had

the highest records, while the locations close to and below the Tropic of Capricorn and with higher altitudes had the lowest ones (Appendix A—Figure A1a,c). The lower average maximum DBT mostly occurred in the high-altitude regions, represented by the blue-shaded region in Figure A1a. Regarding the trends (Appendix A—Figure A1b,d), the Mann–Kendall results showed that all regions presented an increasing trend for the maximum DBT (Appendix A—Figure A1b).

The combined records of the average maximum and minimum DBT resulted in an average spread of 6.3 °C with a standard deviation of 2.5 °C for the Brazilian territory. The north region had the lowest spread (3.7 °C), followed by the northeast (4.2 °C), the central–west (5.5 °C), the southeast (6.7 °C), and the south (10.3 °C). Then, the southeast region had results closer to the country average, and the south showed the highest difference.

The Brazilian territory showed an average relative humidity (RH) of 70.7 %, with the lowest average value of 50.7 % and the highest relative humidity of 89.9 % considering the monthly time series of 15 years (Figure 2.6a). The annual relative humidity occurrences above 80 % happened in regions with dense vegetation, especially the Amazon rainforest. The lowest RH encompasses the Semi-Arid hot climate, which also shows temperature variation above 0.4 °C. The central–west region had the lowest average RH (65.2 %), and the south region had the highest average (76.4 %). Regarding the regions' distribution, the south had a standard deviation of 4.1 percentage points (pp), the lowest among the five regions, and the northeast had the highest (8.9 pp). The highest variation in the northeast region is probably related to the environmental differences between the coastal and continental areas. Figure 6a shows that the coastal regions have higher RH, while the continental area, especially the Semi-Arid region, has lower RH.

Figure 2.6. Annual average relative humidity (a), Mann–Kendall summary (b), and relative humidity variation in percentage points (c). For the Mann–Kendall summary, the gray area indicates no trend, red indicates a decreasing trend, and blue indicates an increasing trend.



The variation in the relative humidity over the 15 years starts from a 5.6 percentage point (pp) decrease to a 3 pp increase (Figure 2.6c). The combined results of the Mann–Kendall test (Figure 2.6b) and the cumulative average also show overlaps in the southeast region, representing not only a linear RH reduction over time but in an overview of the territory. The RH variation in the southeast, northeast, and north regions matches the municipalities with the highest temperature increase.

The highest RH corresponded to the cities in the north region, surrounded by the Amazon rainforest. On the other hand, the lowest RH values corresponded to the SemiArid region, with the highest occurrences in the northeast region. Most of the Brazilian territory presented a maximum reduction of 2 pp. This same reduction represents most of the north, northeast, southeast, and south regions. The central–west region results showed that almost 60 % of its territory had no variations or an increase below 2 pp.

The distribution of the maximum and minimum RH records showed that most of Brazil

presents humid and dry conditions, except for the Semi-Arid region which, despite showing a significant variation between the maximum and minimum records, mainly presented the lowest values. On the contrary, the Amazon rainforest region had the highest values. The Mann–Kendall results also show many municipalities with trends to average maximum RH decrease, especially in the north and central–west regions (Appendix A—Figure A2). Then, following the same pattern as the average DBT and RH, the region defined by the Cerrado and Caatinga biomes (including the Semi-Arid region) had the lowest minimum RH values.

The combined Mann–Kendall results for the average maximum and minimum RH show that most municipalities with a trend presented a reduction in the RH spread but did not become drier, with an exception for some municipalities in the north (three municipalities), southeast (seventy-six municipalities), and south (fifty-four municipalities). The municipalities of the south with a trend to become drier are the majority, according to the Mann–Kendall results (61.4 %).

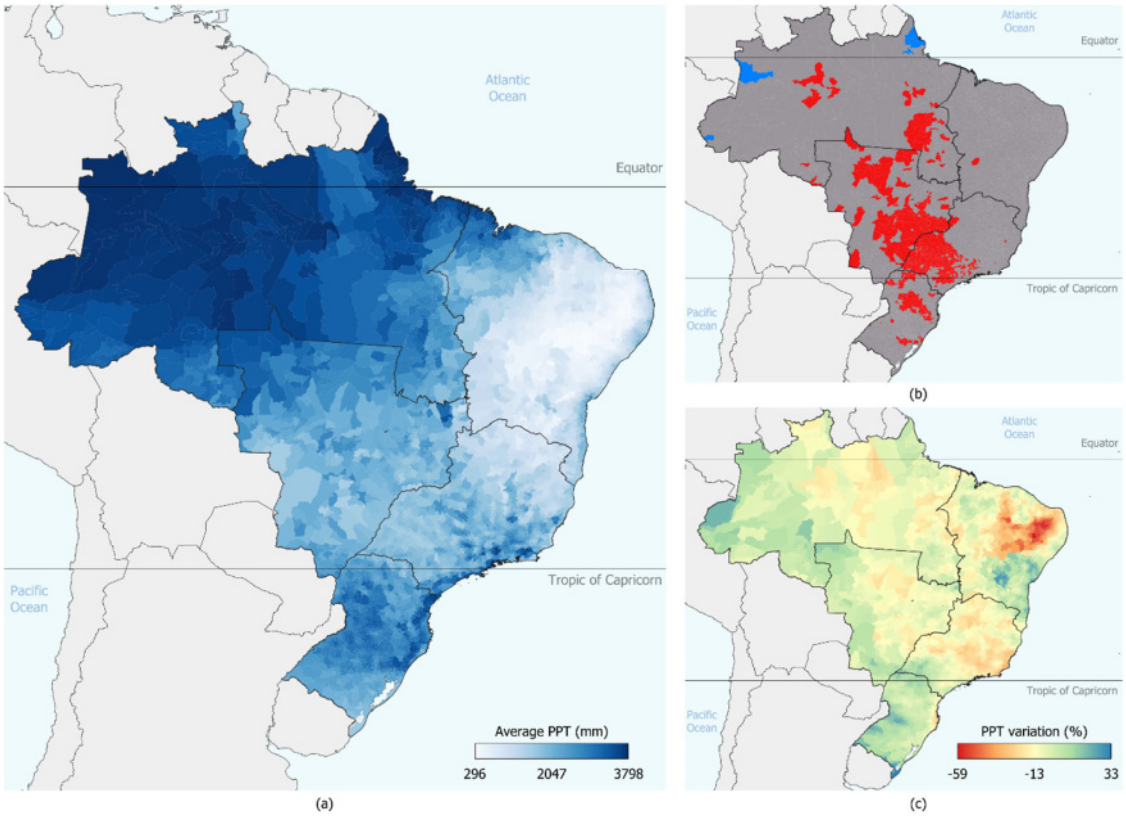
Regarding the RH spread, the average in the Brazilian territory was 25.7 pp with a relative deviation of 44.8 %. Again, the southeast region had the results closest to the Brazilian territory average with the same average spread (25.7 pp). The central–west region had the highest spread (43.6 pp), and the south region showed the lowest spread (17.2 pp). The northeast and the north regions presented intermediary values, respectively, 24.9 pp and 32.4 pp.

Over the 15 years considered in this study, the Brazilian municipalities presented precipitation levels varying from 296 mm to 3798 mm (Figure 2.7a). The average annual precipitation was 1302 mm, with a standard deviation of 500 mm. The results show that despite a high variation in the annual precipitation levels, most of the territory is located on an interval that comprehends levels ranging from 1000 mm to 2000 mm. Municipalities in the Amazon rainforest region had higher precipitation volumes. The lowest precipitation followed the same pattern of the highest temperature, temperature variation, lowest relative humidity, and relative humidity variation. The Semi-Arid region showed the lowest precipitation levels, averaging 692 mm.

The Mann–Kendall results show a significant decrease in the average annual PPT for the central–west and southeast regions and some occurrences in the other regions (Figure 7b). The north region was the only one presenting a linear increase trend for PPT. Regarding the precipitation (PPT) variation over the 15 years, most of the Brazilian territory had a change in the precipitation levels, varying from a reduction of 59 % to an increase of 33 % (Figure 2.7c).

The highest reduction in precipitation occurred in the northeast and some municipalities of the southeast regions. The cities with increased precipitation levels showed a stable annual temperature without significant variations but slightly increased humidity levels.

Figure 2.7. Annual average precipitation (a), Mann–Kendall summary (b), and precipitation variation (c). For the Mann–Kendall summary, the gray area indicates no trend, red indicates a decreasing trend, and blue indicates an increasing trend.



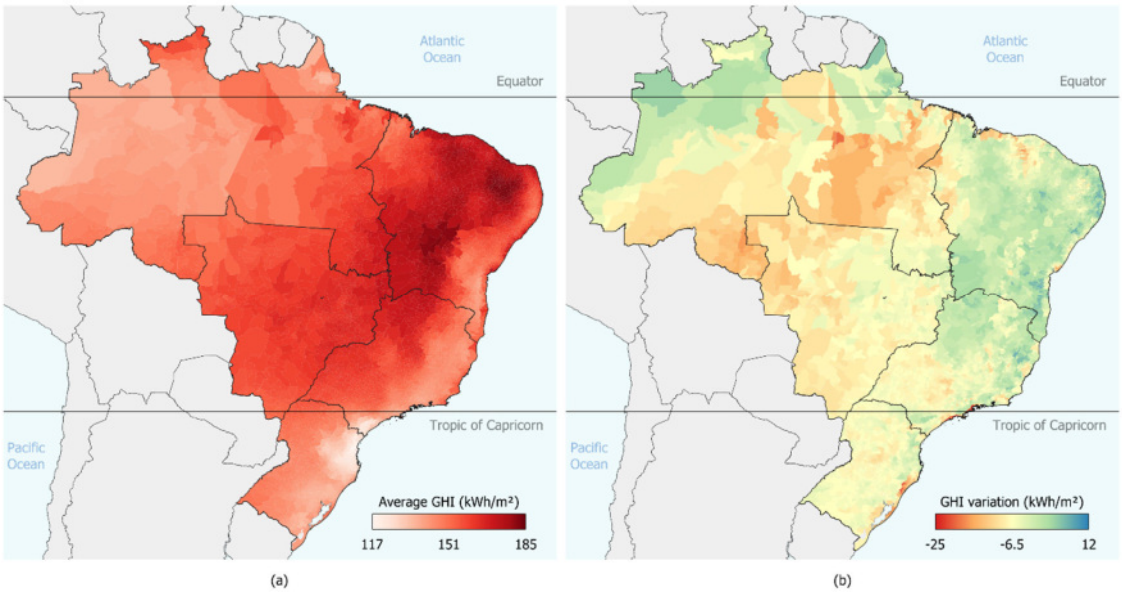
The average minimum and maximum PPT summarized the seasonality in the Brazilian territory (Appendix A—Figure A3a,c). The Amazon rainforest region, followed by the south region, had the highest minimum PPT with averages of 19.3 mm and 38.8 mm, respectively. The results showed that 5556 locations, representing approximately 99.8 % of the Brazilian municipalities, have dry months (below 100 mm), varying from 0.4 mm to 99 mm. The remaining 12 municipalities, with all months within the threshold, varied from 100 mm to 142 mm. The average maximum PPT results showed that most locations had a maximum precipitation between 200 mm and 400 mm. The average maximum PPT varied from 82.2 mm to 729 mm. Thus, 23 locations can be classified as dry throughout the year since the precipitation levels are below 100 mm. The cities with dry conditions during the year are within

the Semi-Arid region. As previously mentioned, they had the lowest precipitation levels, followed by low relative humidity and higher DBT.

The Mann–Kendall results for the average maximum and minimum point to a larger decreased trend for both quantities and a slight trend for PPT increase over the territory (Appendix A—Figure A3b,d). The results show that the decrease trends prevail in regions with high annual average PPT and higher averages for the maximum and minimum quantities. For the increase trends, the situation is similar for the maximum PPT but, for the minimum records, the results show occurrences in regions with the lowest records.

The comparison between ERA5-Land and the Brazilian Solar Atlas annual average from monthly records resulted in an average difference of −3 kWh/m² varying from −25 kWh/m² to 12 kWh/m². The difference also presented a standard deviation of 4.2 kWh/m² and a RMSE of 5.1 kWh/m², proving that the ERA5-Land is a reliable source for an annual analysis based on the average of monthly integrals (Figure 2.8b). The average results by region showed that the northeast records from ERA5-Land best matched the Brazilian Solar Atlas with an average underestimation of 0.1 kWh/m². The north region had the highest difference, with an overestimated 6.6 kWh/m². The ERA5-Land records overestimated the average GHI for the remaining regions based on the monthly integrals varying from 3.1 kWh/m² to 5.6 kWh/m². The results also indicated that 1579 municipalities (28.3 %) had an average GHI difference higher than the RMSE. The central–west region had the lowest number of municipalities above the RMSE threshold, while the southeast region had the highest, encompassing 226 and 426, respectively. A total of 71.6 % of the Brazilian municipalities showed a difference within the RMSE threshold, therefore showing a high correlation between ERA5-Land and the Brazilian Solar Atlas. Only 191 (12.1 %) presented absolute values above 10 kWh/m² from the locations with higher differences.

Figure 2.8. Annual average GHI (a) and GHI difference from the Brazilian Solar Atlas (b).



The annual average from monthly integrals of GHI varied from 117 kWh/m² to 185 kWh/m², with an average value of 156 kWh/m² and a standard deviation of 13 kWh/m² (Figure 2.8a). Regarding the psychrometric relations, the highest values occurred in the cities with the lowest relative humidity since these variables can be directly correlated with the cloud cover that impacts the amount of radiation the land surface receives. The same happens in the opposite direction since the Amazon rainforest regions had high humidity levels, increasing the precipitation and cloud cover, directly affecting the region's radiation values. The Brazilian territory had most municipalities (2834) with radiation levels between 150 kWh/m² and 170 kWh/m², representing 47.3 % of the territory. The north and south regions had 68.4 % and 74.1 % of their territory with radiation levels between 150 kWh/m² and 170 kWh/m². The central–west and southeast regions had 96.7 % and 61.2 % of their territory with radiation levels between 150 kWh/m² and 170 kWh/m². The northeast had 53.6 % of its territory with radiation levels above 170 kWh/m².

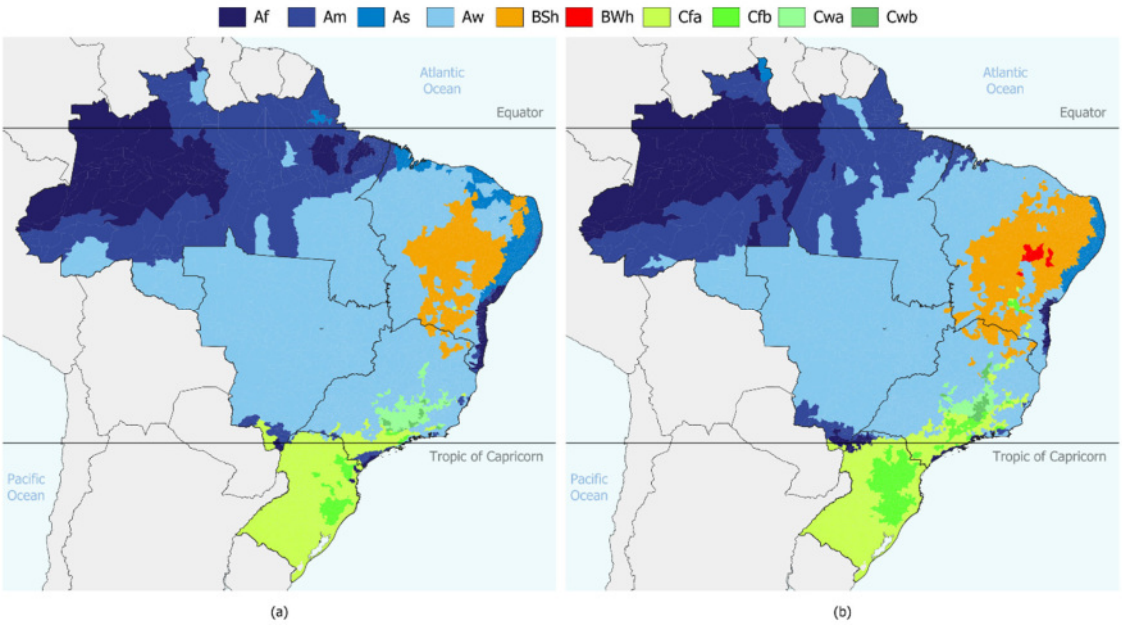
The results from the classification using data from 1950 to 1990 and 2008 to 2022 (Table 2.3) showed that the tropical climate area decreased by 3.1 pp, the dry climate increased by 2.6 pp, and the subtropical climate increased by 0.5 pp. Thus, the Brazilian experienced a shift towards hotter and drier conditions.

Table 2.3. Summary of the Köppen–Geiger classification coverage (percentage) according to records from 1950 to 1990 (A) and 2008 to 2022 (B) and the variation in percentage points.

	Af	**Am**	**As**	**Aw**	**BSh**	**BWh**	**Cfa**	**Cfb**	**Cwa**	**Cwb**
A	14.3	22.3	2.7	45.8	6	-	7	0.7	1.1	0.1
B	16.3	20.5	1	44.2	8.3	0.3	6.3	2.2	0.6	0.3
Var	2	-1.8	-1.7	-1.6	2.3	0.3	-0.7	1.5	-0.5	0.2

Figure 2.9 shows that the classification using the ERA5-Land database differed from Alvares et al. (28) on the territory occupied by each climate group and the classification of some municipalities in the Hot Desert-Arid group (BWh). Since the Köppen–Geiger classification is based on air temperature and precipitation, the regions most affected by the temperature and precipitation variation had the most significant changes in the climate classification. Considering the 15-year data, the tropical climates, represented by group A, occupy the highest portion of the Brazilian territory (82 %). The subtropical climate, represented by group C, occupies the second position, with 9.4 % of the territory. Group B, representing the dry climate, covers 8.6 % of the Brazilian territory.

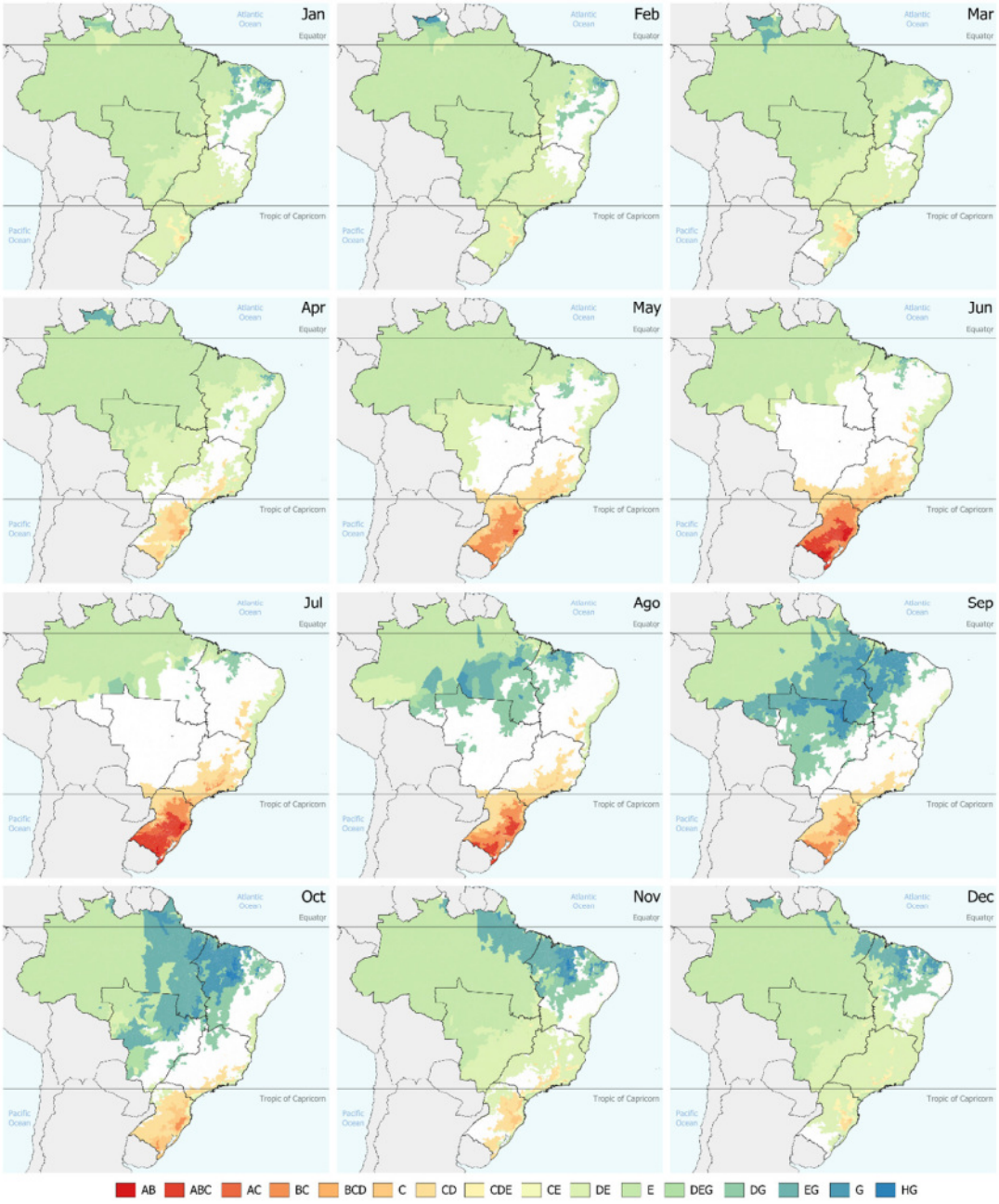
Figure 2.9. Köppen–Geiger classification of 5567 Brazilian municipalities based on records from 1950 to 1990 (a) and 2008 to 2022 (b).



The results for the first trimester indicate that design strategies that rely only on ventilation (E and G) are the majority in reaching thermal comfort in most of Brazil (Figure

2.10). The onset of autumn increases the need for heating solutions in the southern region and parts of the southeast. The temperature reduction generally increases the municipalities' thermal comfort, particularly in the southeast and south regions. The winter solstice represents another turning point for the bioclimatic design requirements in Brazil. The third trimester shows the highest number of municipalities with heating demand. However, thermal comfort has the highest territorial occurrence, since the heating demand comes from small municipalities, and the thermal comfort is achieved in municipalities that occupy larger territories. The last trimester closes the annual cycle and shows an increasing demand for ventilation strategies while a significant reduction for heating strategies.

Figure 10. Monthly bioclimatic strategies distribution over the Brazilian territory. A—artificial heating, B—solar heating, C—thermal mass combined with solar heating, D—thermal comfort, E—daytime ventilation, G—night cooling with thermal mass and evaporative cooling, and H—thermal mass and evaporative cooling. The white areas indicate that only thermal comfort "D" reached the 7-day threshold.

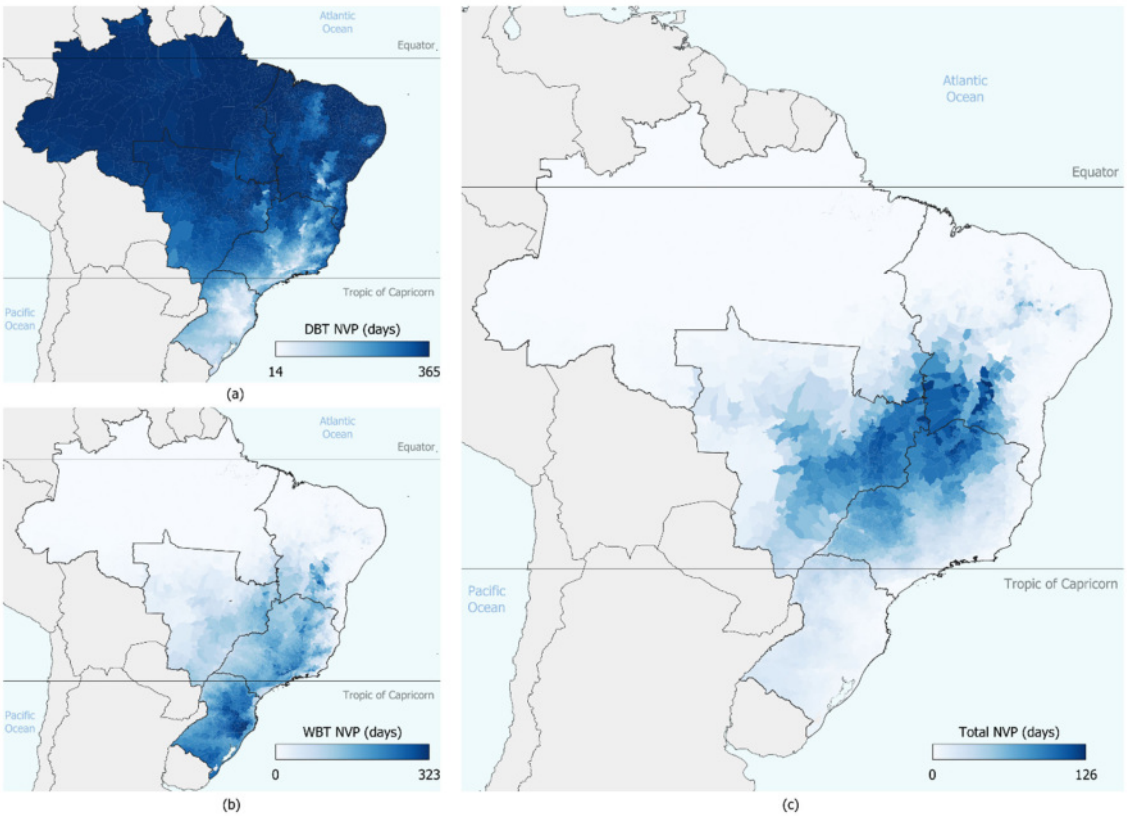


The NVP based on the DBT records reinforced the thermal comfort occurrence over the Brazilian territory as they were based on the adaptive model with an 80 % acceptance from

ASHRAE 55 (Figure 2.11a). The Brazilian territory comprised an average NVP of 274 days, and 75 % of the year within thermal comfort according to the outdoor conditions.

Urupema-SC had the lowest thermal comfort occurrence and, consequently, the lowest potential based only on DBT records (14 days). On the other hand, 695 municipalities (12.5 %) of the north and northeast regions presented an average potential for all days of the year. The south region had the lowest average DBT NVP (160 days), and the north region had the highest average potential (360 days). The central–west and northeast regions also presented a potential above 300 days, and the southeast region—an average of 235 days.

Figure 11. Average annual results for the NVP considering the DBT (a), WBT (b), and the combination of DBT, WBT, and WS (c).



While the temperature threshold allowed for a high ventilation potential in most of the Brazilian territory (Figure 2.11a), the WBT threshold, which is related to humidity levels, resulted in a lower potential for natural ventilation (Figure 2.11b). On average, humidity constraints limited ventilation for 88 days, with a standard deviation of 79 days, affecting 1171 municipalities (21%). The WBT analysis revealed that the regions with the highest potential for the DBT threshold had the lowest potential for the humidity constraints, and vice versa for the

regions with the lowest potential for the temperature constraints. The north region had the lowest potential, averaging 7 days, and the south region had the highest potential of 179 days. For the WS threshold, all municipalities showed a potential for all days of the year, since none exceeded the wind speed upper limits.

The final assessment combined the temperature, humidity, and wind speed threshold (Figure 2.11c), resulting in an average of 22 days suitable for natural ventilation across Brazil with a high standard deviation (27 days). A total of 1317 municipalities did not present natural ventilation potential for the entire year, mainly encompassing the north and northeast regions and some coastal municipalities in the southeast and the south regions. The central–west region had the highest average (57 days), while the north had the lowest potential (6 days).

The heating and cooling analysis from Givoni was complemented by quantifying the average annual degree days along the 15-year records. The territorial distribution in Table 2.4 shows that the south region had the most even distribution for the heating demand, with a relative deviation of 51 %. The north region had a more even distribution for the cooling demand, with an average relative deviation of 24 %.

Table 2.4. Average and standard deviation for the five Brazilian regions' cooling and heating degree days.

	HDD_{18} [K d]	**CDD_{24} [K d]**	**CDD_{25} [K d]**	**CDD_{26} [K d]**
Central-West	12 ± 27	622 ± 330	395 ± 249	228 ± 167
North	0 ± 0	1070 ± 194	725 ± 171	426 ± 132
Northeast	0 ± 3	875 ± 467	395 ± 389	228 ± 297
South	425 ± 213	123 ± 118	68 ± 74	35 ± 42
Southeast	69 ± 84	220 ± 200	123 ± 129	62 ± 74

The HDD_{18} (Appendix A—Figure A4a) reinforced the heating demand by the highest average of 219 K d in the south, a significant demand in the southeast region (84 K d), and a smaller portion in the central–west region (27 K d). The CDD results indicated that, in addition to passive cooling strategies, artificial methods are also necessary for several locations in Brazil, with maximum demand from 1246 K d of CDD_{26} to 1970 K d of CDD_{24}. The results also revealed that the north region had the highest demand, followed by the northeast, central–west, southeast, and south regions. The step analysis (Appendix A—Figure A4b–d) showed a considerable reduction for each region from 24 °C to 26 °C base temperature. For all the

regions, the results showed a significant cooling demand depending on the setpoint defined for the degree days analysis.

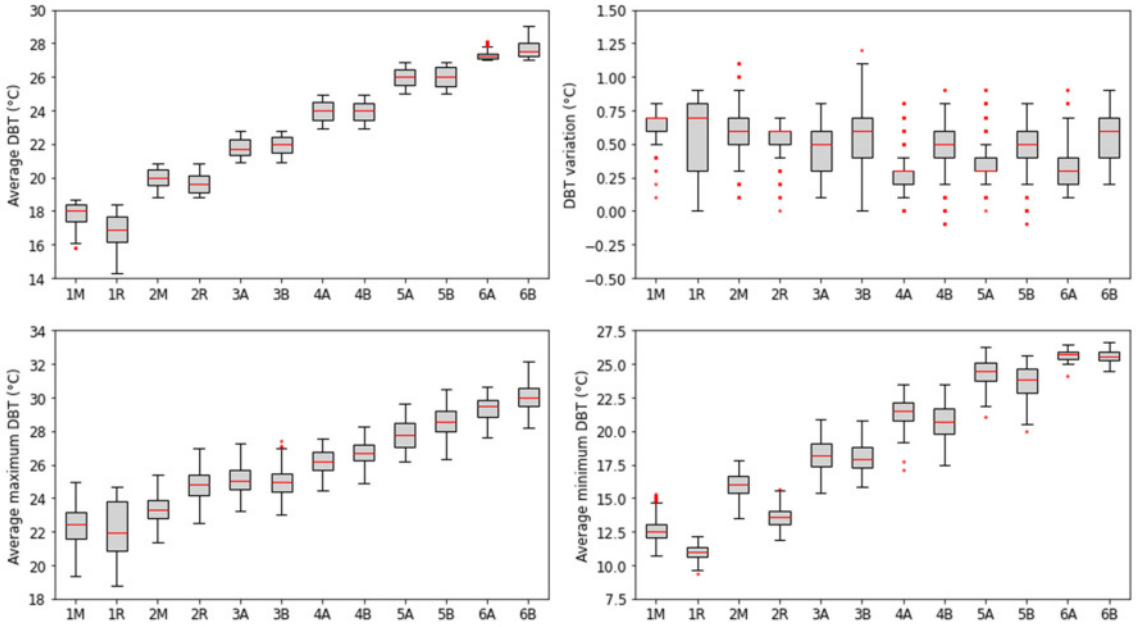
2.2.2. Bioclimatic zones assessment

The reclassification of the Brazilian municipalities based only on ERA5-Land data showed that 539 municipalities (9.6 %) presented a different classification from Silva Machado et al. (67). Most of the changes were influenced by the RH records from the ERA5-Land database, resulting in changes within the same macrozone but migrating from a humid (A) to a dry (B) condition.

2.2.2.1. Climate and meteorological assessment

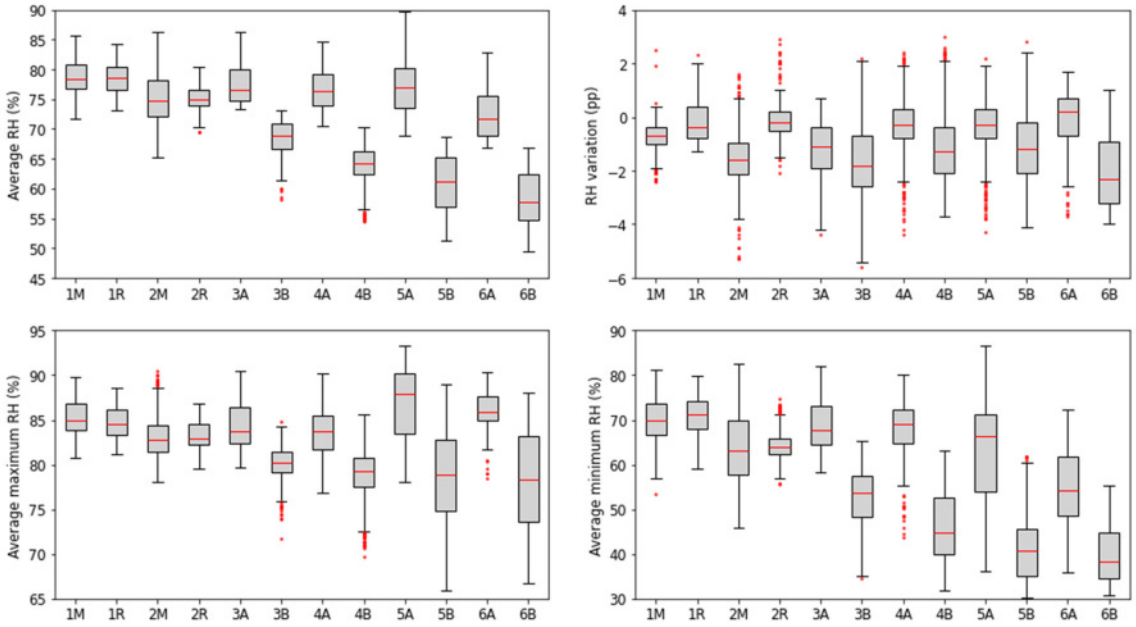
Considering the 15-year annual average for each municipality, Figure 2.12 showed that the average DBT among zones varied from 16.7 °C (1R) to 27.6 °C (6B) with a standard deviation varying from 0.3 °C (6A) to 0.9 °C (1R). The DBT cumulative average variation (variation between the 2008 and 2022 average) showed that zones 4A, 5A, and 6A had the lowest average variation (0.3 °C), while zone 1M had the highest variation among the bioclimatic zones (0.7 °C). Despite varying from 0.3 °C to 0.7 °C among zones, the standard deviation was low for all twelve zones, ranging between 0.1 °C and 0.2 °C. Following the average DBT distribution, the average maximum and minimum DBT showed a similar distribution among the twelve zones. Zone 1R had the lowest averages for the minimum (11.2 °C) and maximum DBT (21.6 °C), while zone 6B had the highest maximum (30.1 °C) and zone 6A had the highest minimum (25.7 °C). Those results showed that the zones had a homogenous distribution regarding the average DBT, as stated by Silva Machado et al. (67), and the average maximum and minimum DBT.

Figure 2.12. Distribution of DBT 15-year average, 15-year variation, 15-year average minimum, and 15-year average maximum among the Brazilian bioclimatic zones following the annual records. The red line indicates the median while the red dots the outliers.



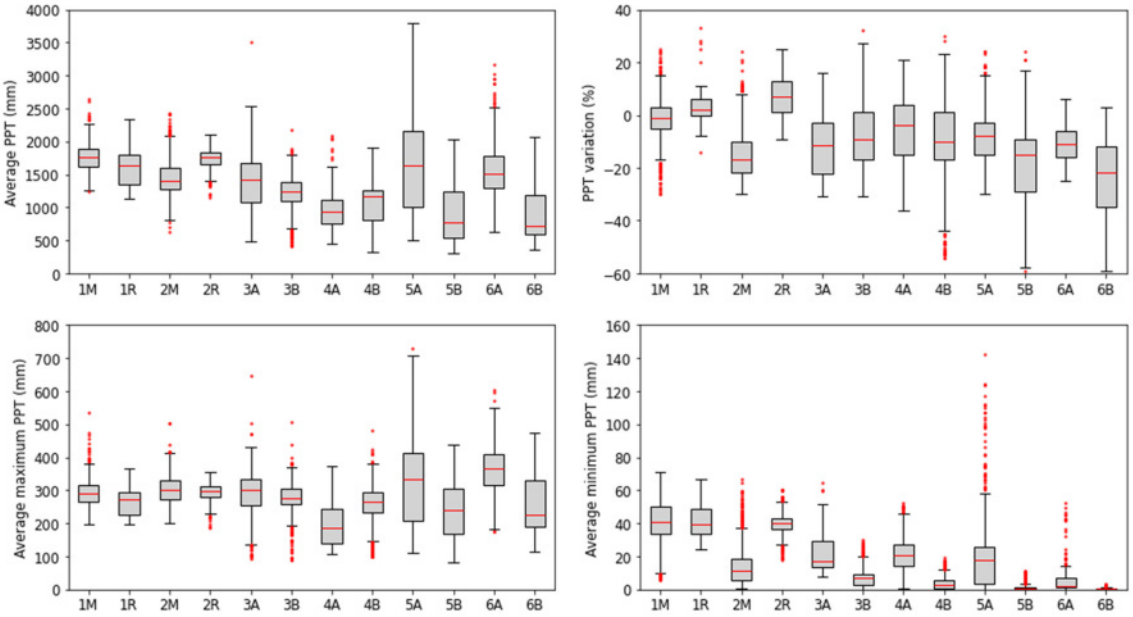
Regarding the annual average RH, the Brazilian bioclimatic zones show a clear separation between humid and dry locations (Figure 2.13), as described in the method presented in Section 2.1.1 and corroborating with the climatic analysis from Section 3.1.1. Zones 1 and 2 showed overlaps between the M and R classes, differing from the other bioclimatic zones. Zone 6B showed the lowest average (58.4 %), and zone 1R had the highest average RH (79.3 %). The driest zone (6B) also had the highest average reduction (2 pp), while zones 6A and 2R had the lowest average reduction (0.1 pp). The standard deviation also shows that zone 5A had the highest variation dispersion (4.8 pp), and zone 2R had the highest homogeneity (2.1 pp). The average maximum RH varied from 78 % (6B) to 86.8 % (5A), while the average minimum varied from 39.7 % also in zone 6B to 71.4 % (1R). Zones 6B and 5A presented the highest dispersion for the average maximum and minimum, respectively, presenting standard deviations of 5.5 pp and 11.2 pp. Zone 2R had the lowest deviation for the average maximum and minimum with standard deviations of 1.6 pp and 3.7 pp.

Figure 2.13. Distribution of RH 15-year average, 15-year variation, 15-year average minimum, and 15-year average maximum among the Brazilian bioclimatic zones following the annual records. The red line indicates the median while the red dots the outliers.



The annual average PPT indicated that all zones have overlaps (Figure 2.14), different from the previous results for the average DBT and RH. The annual average PPT varied from 897 mm for zone 6B to 1760 mm in zone 1M. The results corroborate the rules to define the Brazilian bioclimatic zones, especially for zone 6B, which has the lowest RH levels and directly impacts the PPT records. Regarding the PPT variation, most zones showed an average reduction varying from 1.4 % (1M) to 24.4 % (6B). Zones 1R and 2R were the only ones showing an average increase in precipitation levels, respectively, of 1 % and 7.1 %. Zones 5B and 6B also showed the highest deviations for the average PPT distribution, mainly representing a reduction over the 15-year records and reaching almost 60 %. The average maximum PPT was highly homogeneous among the zones, varying from 197 mm (4A) to 367 mm (6A). Zone 5A also shows the highest average maximum PPT (729 mm), and zone 5B shows the lowest average maximum (82 mm). For the average minimum PPT, zone 6B had the lowest average, with no precipitation, and zones 1M, 1R, and 2R had the highest levels (40 mm). The standard deviation also pointed to a high variation for zone 5A (21 mm) and no variation for zone 6B. The average maximum and minimum analysis depicted the monthly precipitation profile for the bioclimatic zones, showing that zones with the lowest humidity levels also had lower precipitation throughout the 15-year records.

Figure 2.14. Distribution of PPT 15-year average, 15-year variation, 15-year average minimum, and 15-year average maximum among the Brazilian bioclimatic zones following the annual records. The red line indicates the median while the red dots the outliers.



The updated Köppen–Geiger classification for Brazil showed that zones 2M, 3A, 3B, and 4A had the highest variability among the classes, with climates from classes A, B, and C. The results also pointed to regions with lower diversity with only two classes for zones 1R, 2R, and 6B. Despite the different classes within zones, it is possible to distinguish a prevailing class for zone 1M (Cfa = 52.2 %), 1R (Cfb = 69.6 %), 2M (Cfa = 64.3 %), 2R (Cfa = 99.5 %), 3B (Aw = 59.8 %), 4B (Aw = 73.7 %), 5B (Aw = 52.6 %), 6A (Aw = 84.8 %), and 6B (BSh = 52.3 %). For zone 3A, the prevailing classes were Aw (44.5 %) and Cfa (30.5 %). Zone 4A had classes Aw (47.1 %) and As (29.2 %) as the predominant. Finally, zone 5A presented a more even distribution between classes Aw (36.8 %), As (26.7 %), and Am (23.6 %). This pattern resulted from the method used for bioclimatic zoning, which combines both climatic and performance parameters, resulting in a diverse distribution within zones.

2.2.2.2. Bioclimatological assessment

The findings for January to March indicated that daytime ventilation could improve thermal comfort across all zones. The results also showed that some municipalities from zones 1M, 1R, 2M, and 3A also required thermal mass associated with solar heating, while the zones with the highest temperatures, zones 5 and 6, also required strategy G of thermal mass and

evaporative cooling. Zone 4B had the highest number of locations where no strategy met the 7-day threshold apart from thermal comfort.

For April, all of the bioclimatic zones required daytime ventilation, and zones 1 and 2 kept the requirement for heating strategies. Zone 1R also showed a requirement for solar heating (B). May showed a similar pattern but required artificial heating for zone 1R and solar heating combined with thermal mass for zones 3 and 4. May also showed no demand for ventilation strategies from zones 1 and 2R. June followed the requirements of May, but artificial heating was added to zones 1M and 2R, and ventilation strategies were removed from zone 2M. The coldest zones, represented by groups 1 and 2, also showed a reduction in thermal comfort and an increase for zones 4 to 6. Zone 3B also increased for April and May, but June presented a decrease.

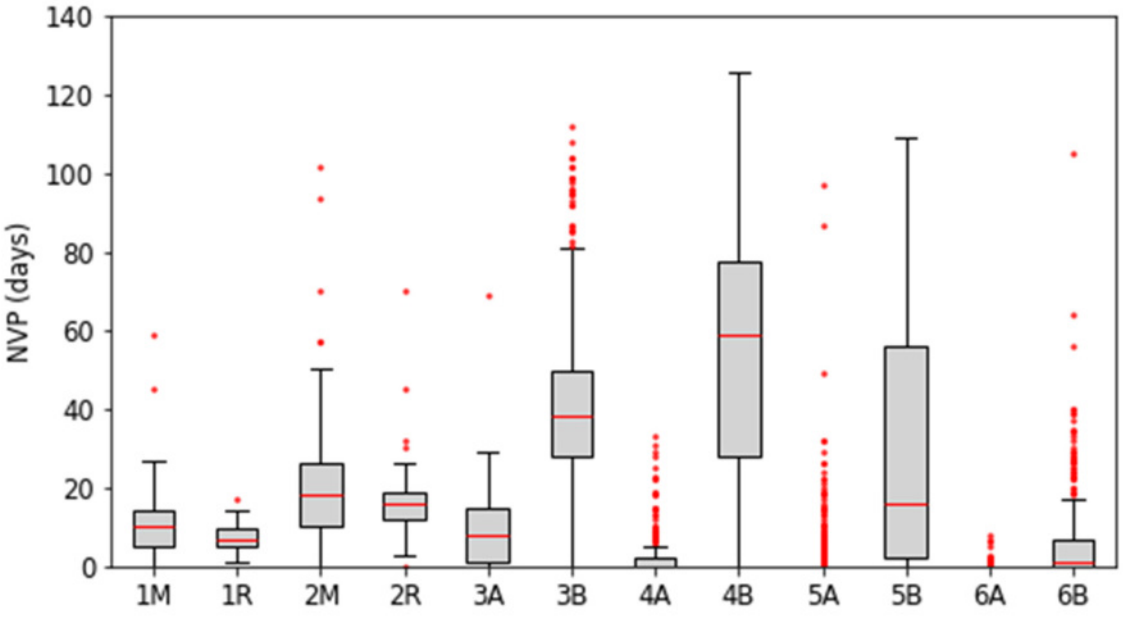
July has the same recommendations as June, except for the combination of thermal mass with solar heating strategies and thermal comfort (CD) for zone 2R. For the CD combination, zone 2R increased from a single location in June to 299 municipalities in July. August, however, starts showing thermal comfort conditions in zone 1M and a general thermal comfort increase in all the other zones except for zone 1R. The results also show a reduction in the demand for heating strategies in all zones, especially for zones 1 and 2. For September, most municipalities did not require artificial heating (A), except for UrupemaSC. The results also showed a demand for strategy H (thermal mass with evaporative cooling) for seven municipalities of zone 6B.

The results showed that October required heating strategies from zones 1M to 3A, mainly defined by thermal mass combined with solar heating (C), except for 27.1 % of the municipalities from zone 1R that still required strategy B. October also showed increased demand for daytime ventilation from zones 2M to 6B. November showed a similar pattern to October but required daytime ventilation for 76.3 % of the municipalities from zone 1R. Finally, December showed that all bioclimatic zones required daytime ventilation (E) for most of their municipalities, which mainly combined thermal mass with solar heating (C) for zones 1 and 2. For zones 3 to 6, daytime ventilation was mainly combined with strategy G or with days within thermal comfort (D).

This paper provides a detailed analysis confirming the need for those strategies and reinforcing the ventilation potential through the NVP results. The results showed that locations with dry classification “B” of zones 3 to 6 had the best outcomes for the NVP (Figure 2.15). Additionally, the colder municipalities within zone 2 had a higher potential than those in zone

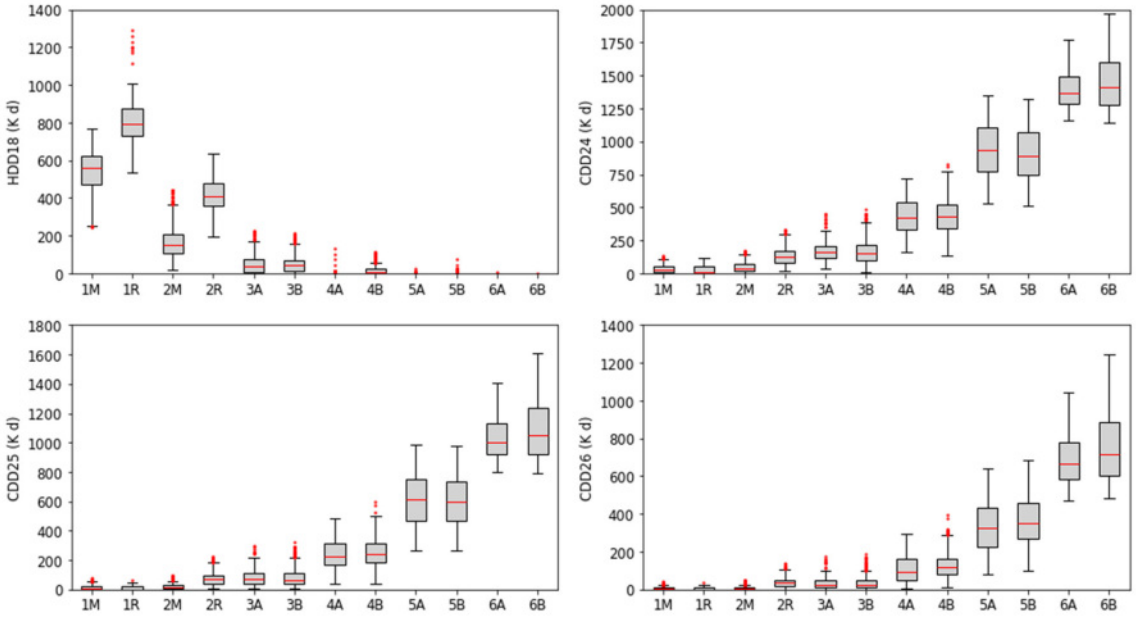
1.

Figure 15. Total natural ventilation potential combining the DBT, WBT, and WS thresholds based on the 15-year average among the Brazilian bioclimatic zones. The red line indicates the median while the red dots the outliers.



Zones 1R, 1M, and 2R have the highest heating demand following the bioclimatic distribution (Figure 2.16). This study found an average greater than zero in zones 3 and 4 but significantly lower than the demand from the coldest zones, ranging between 2 K d and 52 K d of HDD_{18}. The cooling demand increased from zone 1 to 6, following the daily average DBT that delimitates each of the 12 zones. The results also demonstrated that zones 1M to 4B had the highest reduction from CDD_{26} to CDD_{24}, with an average of 79 %. Zones 5A to 6B had a lower average reduction but still showed an average cooling demand reduction ranging from 49 % to 66 %.

Figure 2.16. Distribution of HDD_{18}, CDD_{24}, CDD_{25}, and CDD_{26} 15-year average among the Brazilian bioclimatic zones following the annual records. The red line indicates the median while the red dots the outliers.



2.2.2.3. Representative locations and building performance simulation

Using the Brazilian bioclimatic zoning methodology presented in Section 2.1.2 and the ERA5-Land daily records, the 5567 Brazilian municipalities used in this study were classified to identify the representative locations (Table 2.5). After applying the Kolmogorov–Smirnov test, 12 representative locations were identified with an annual average DBT varying from 16.7 °C in São Marcos-RS (1R) to 27.7 °C in Bela Vista do Piauí-PI (6B). Colônia Leopoldina-AL (4A) and Pio XII-MA (6A) had the lowest DBT variation throughout the 15-year records (0.2 °C), while Agudos-SP (3B), Bela Vista do Piauí-PI (6B), Ibicaré-SC (1M), and São Marcos-RS (1R) had the highest DBT variation (0.8 °C). Olímpia-SP (4B) showed a 0.7 °C variation between 2008 and 2022 and was the only one with an "increase" result for the Mann–Kendall trend analysis. Bela Vista do Piauí-PI (6B) showed the lowest average RH (52.9 %), and Fonte BoaAM (5A) had the highest average RH (88.9 %). Regarding the 15-year variation, Muqui-ES (3A) had the highest RH decrease (3.6 pp), while Chiapetta-RS (2R) and Pio XII-MA (6A) showed an increase in the RH, respectively, of 0.2 pp and 1.1 pp. The Mann–Kendall results also showed that Agudos-SP (3B), Fonte Boa-AM (5A), Três Barras-SC (1M), Olímpia-SP (4B), and São Marcos-RS (1R) had a "decrease" trend for the RH records. Therefore, the results

indicated that Olímpia-SP (4B) was the only location with significant DBT and RH Mann–Kendall analysis results.

The heating demand was highest in zones 1 and 2, while zones 3 and 4 had significantly lower demands. The cooling demand was lowest in zones 1M to 2M, followed by zones 2R to 3B. Zone 4 locations had a higher demand for cooling, and zones 5 and 6 showed the highest cooling demand at 24 °C base temperature.

Table 2.5. Representative locations for each Brazilian bioclimatic zone.

Zone	Municipality	Lat (°)	Long (°)	HDD_{18} (K d)	CDD_{24} (K d)
1M	Três Barras-SC	−26.11	−50.32	666	13
1R	São Marcos-RS	−28.97	−51.07	805	16
2M	Monteiro Lobato-SP	−22.95	−45.84	171	33
2R	Chiapetta-RS	−27.92	−53.94	430	143
3A	Muqui-ES	−20.95	−41.35	26	169
3B	Agudos-SP	−22.47	−48.99	59	190
4A	Colônia Leopoldina-AL	−8.92	−35.72	-	368
4B	Olímpia-SP	−20.74	−48.91	17	443
5A	Fonte Boa-AM	−2.52	−66.09	-	975
5B	Queimada Nova-PI	−8.57	−41.41	-	835
6A	Pio XII-MA	−3.89	−45.18	-	1424
6B	Bela Vista do Piauí-PI	−7.99	−41.87	-	1442

The PHFT results showed that the construction set with the lowest U-value (U_{low}) delivered the best average PHFT, independent of the solar absorptance (Figure 17), for the representative locations for the cold zones (1M, 1R, 2M, and 2R). From zones 3 to 6, the most insulated construction set gradually delivered intermediary to worst PHFT, while U_{medium} (intermediary U-value) with a low solar absorptance had the highest averages. For zones 3 and 6, U_{high}, with a solar absorptance of 0.3, also delivered good PHFT results, and U_{medium}, with a solar absorptance of 0.7, had the lowest PHFT levels.

Figure 2.17. Annual PHFT results. Blue represents U_{low}, green represents U_{medium}, and orange represents U_{high}. The dotted line indicates a solar absorptance of 0.3, the solid line indicates a solar absorptance of 0.5, and the dashed line indicates a solar absorptance of 0.7.

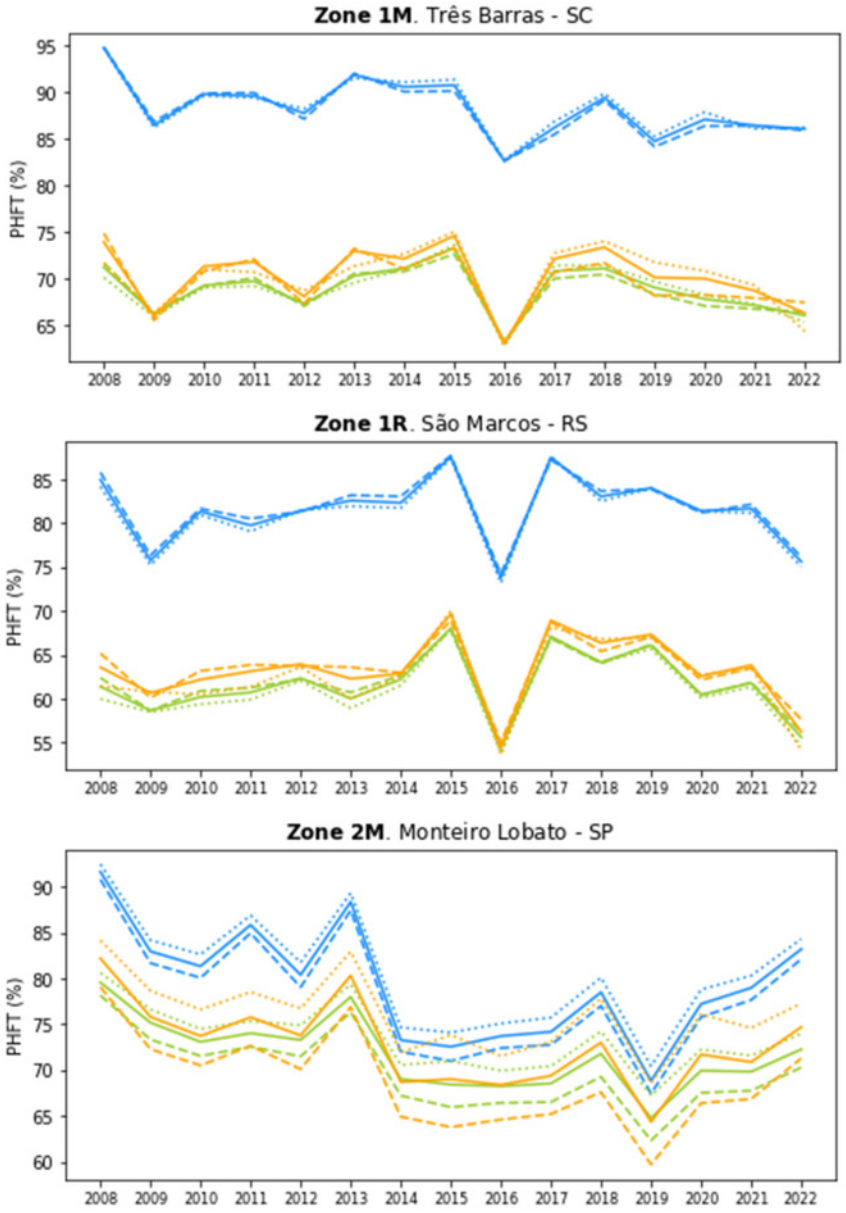


Figure 2.17. *Cont.*

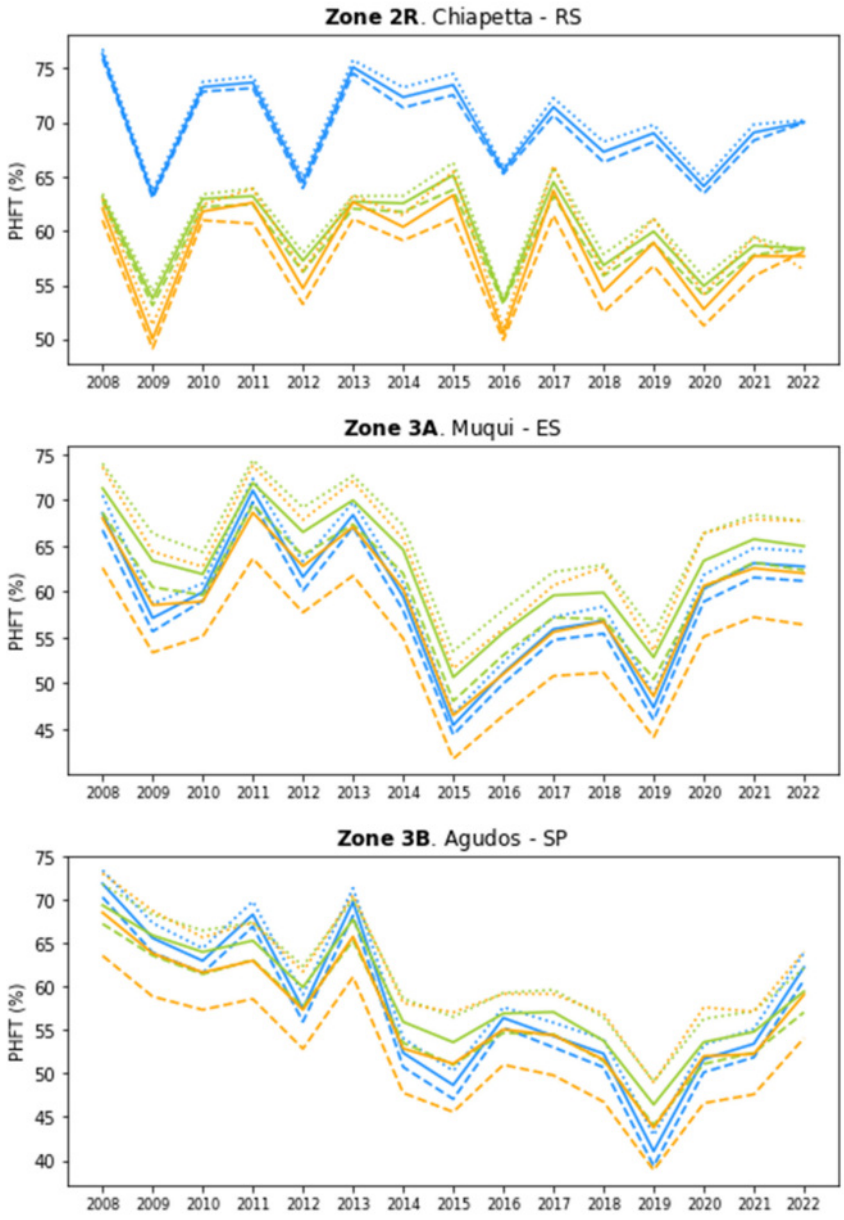


Figure 2.17. *Cont.*

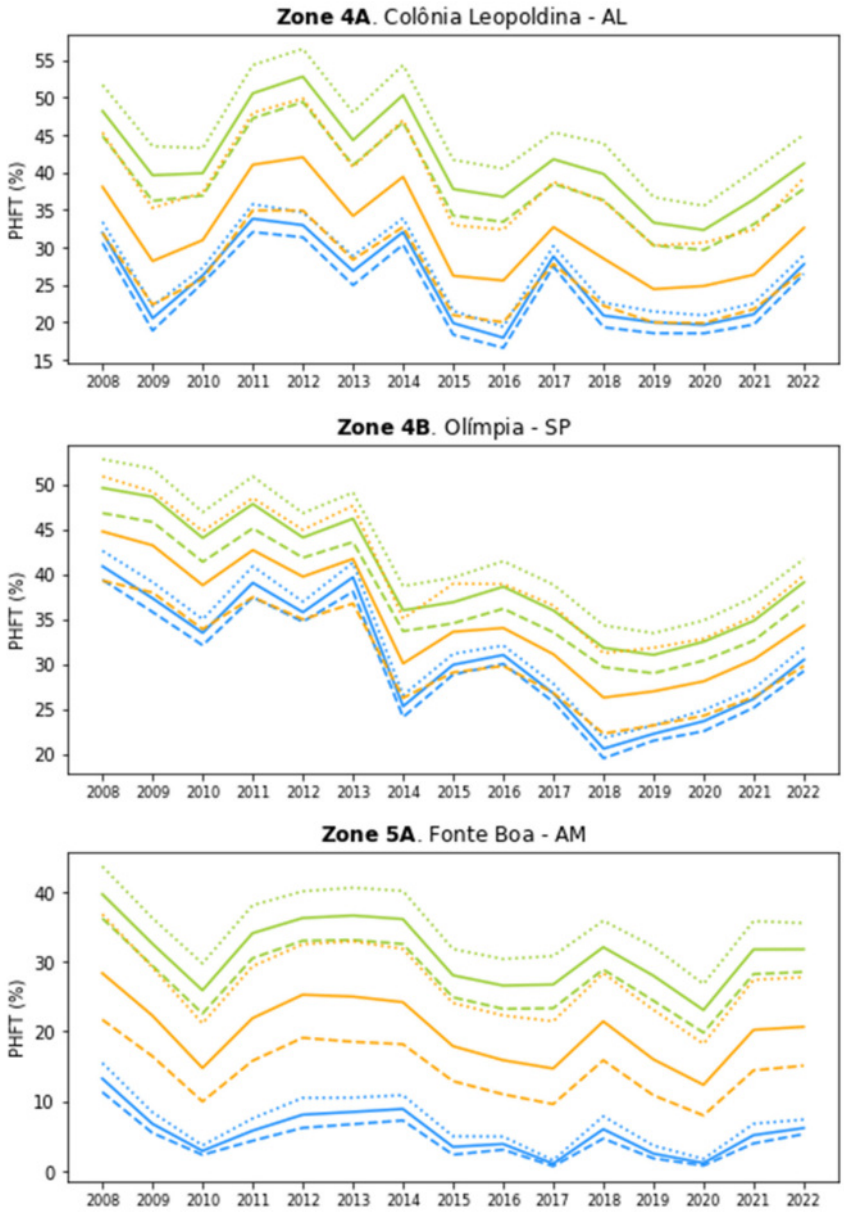
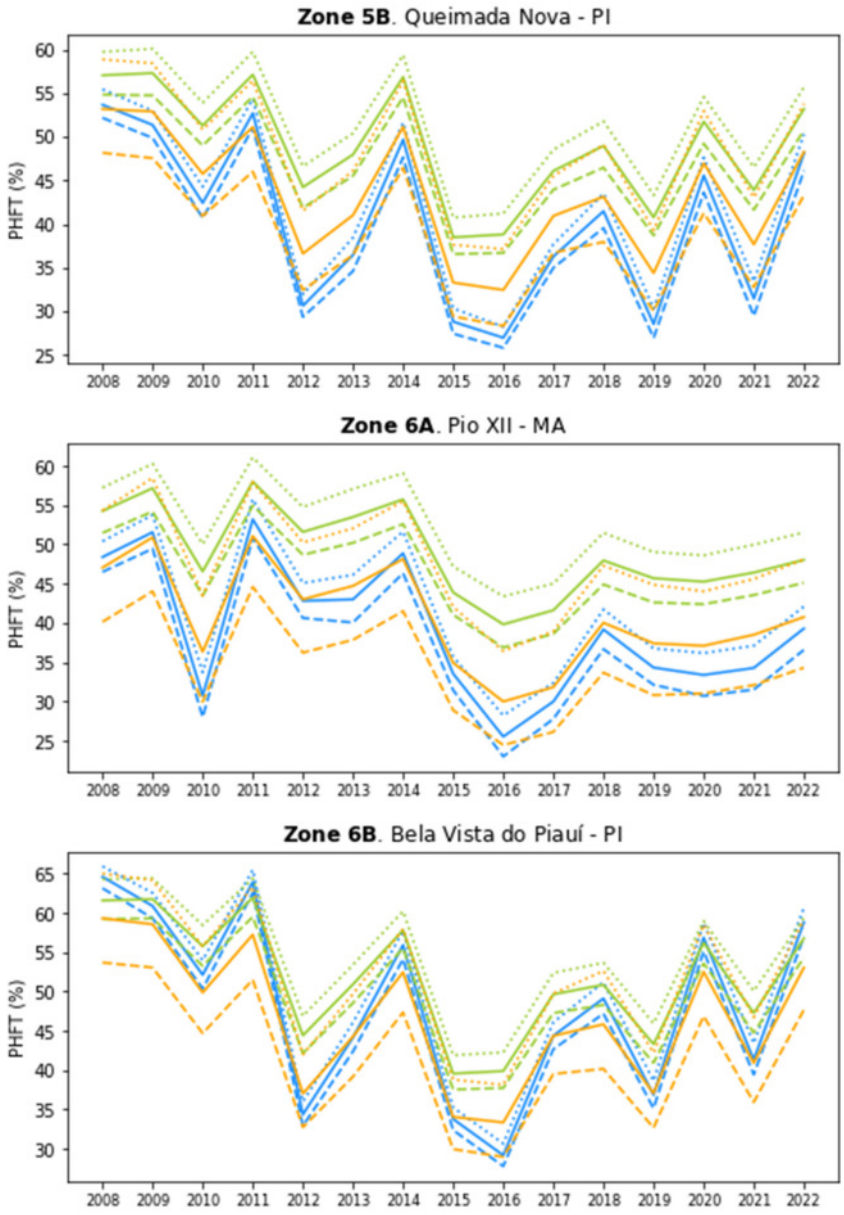


Figure 2.17. *Cont.*



Considering the average results from the 15-year simulation outputs for each construction set and solar absorptance combination, Três Barras-SC (1M) had the best results with an 88.4 % PHFT for U_{low} with low solar absorptance, and Fonte Boa-AM (5A), for U_{low} with the highest absorptance, had the worst results (PHFT of 4.4 %). Fonte Boa-AM (5A) had the greatest increase in PHFT, rising by 30.8 percentage points when transitioning from U_{low} with a solar absorptance of 0.7 (PHFT of 4.4 %) to U_{medium} with a solar absorptance of 0.3 (PHFT of 35.2 %). On the other hand, Agudos-SP (5A) showed the lowest variation, increasing

from 52 % for U_{high} with a solar absorptance of 0.7 to 61.7% with the same construction set, but with a solar absorptance of 0.3.

The analysis by construction set revealed that highly insulated configurations performed best in zones 1 and 2, ranging from 69.9 % to 88.2 %. On average, an increase of 16.6 percentage points was found in the representative locations of colder climates. For zones 3 to 6, the construction set with an intermediary U-value delivered the best results. The average PHFT from zones 3 and 4 varied from 39.9 % to 62.9 %, with an average improvement of 13.9 percentage points. The hottest zones, 5 and 6, had the lowest PHFT levels, from 31.5 % to 51.9 %, but they had the highest improvement in PHFT levels, with an average of 18.6 percentage points.

Considering a multi-year analysis based on the cumulative average, all the locations resulted in a reduction in the PHFT when comparing 2008 and 2022, independently of the building envelope configuration, except São Marcos-RS, whose cumulative average varied of 1 pp with U_{medium} and U_{high} with low absorptance and 0.2 pp with U_{medium} with an intermediary absorptance. Bela Vista do Piauí-PI (6B), a hot and dry climate, had the highest variation with a reduction of 6.6 pp. The most insulated building envelope resulted in the highest average reduction of PHFT (9.6 pp), and the intermediary components (U_{medium}) had the lowest reduction (6.7 pp).

The insulated construction (U_{low}) delivered the best performance regardless of solar absorptance, removing all the heating demand for the representative locations from zones 2M, 3, and 4 (Figure 2.18). In the remaining locations, the heating demand significantly decreased for all the years.

Figure 18. Annual results for the CgTa. Blue represents U_{low}, green represents U_{medium}, and orange represents U_{high}. The dotted line indicates a solar absorptance of 0.3, the solid line indicates a solar absorptance of 0.5, and the dashed line indicates a solar absorptance of 0.7.

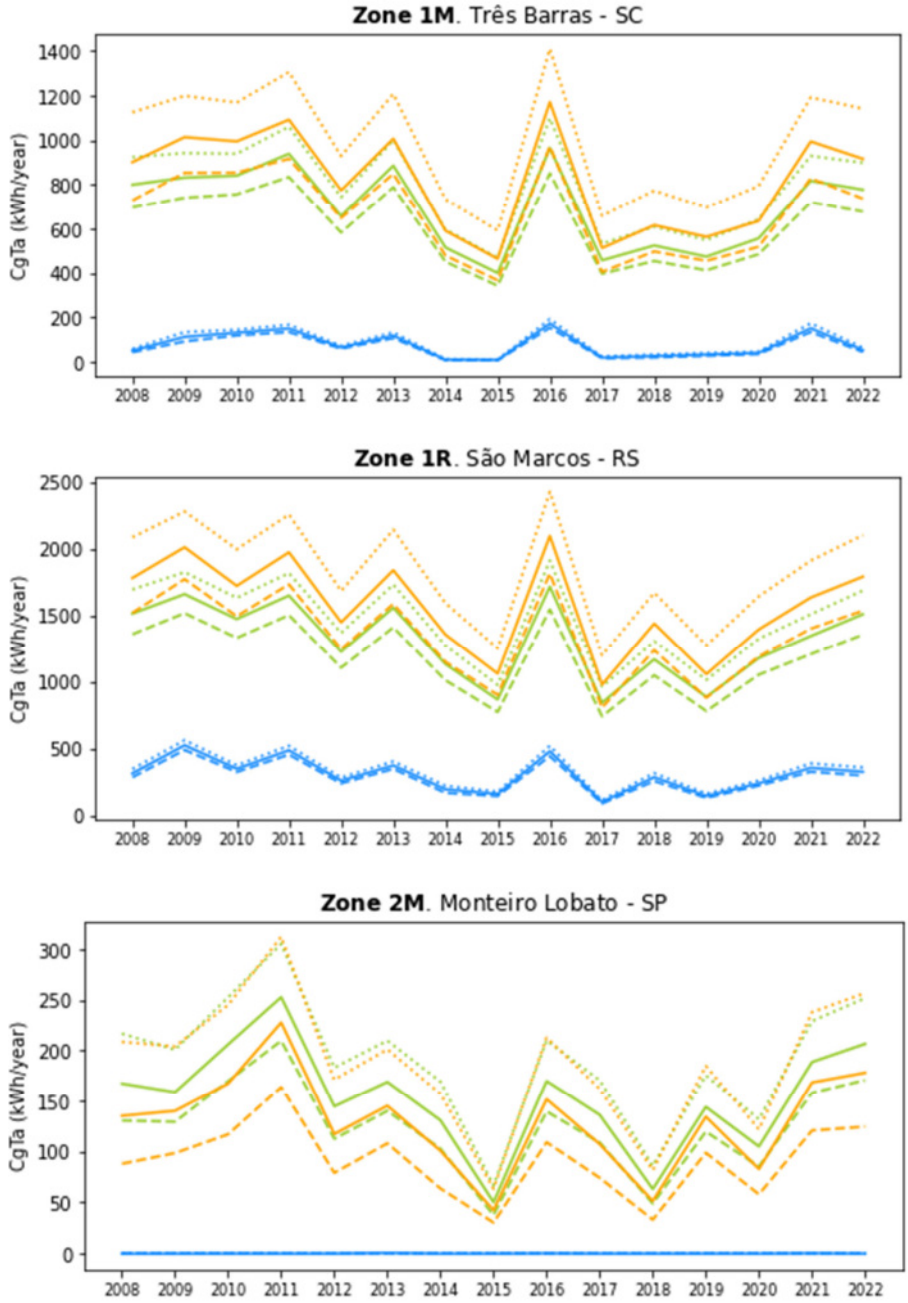


Figure 2.18. *Cont.*

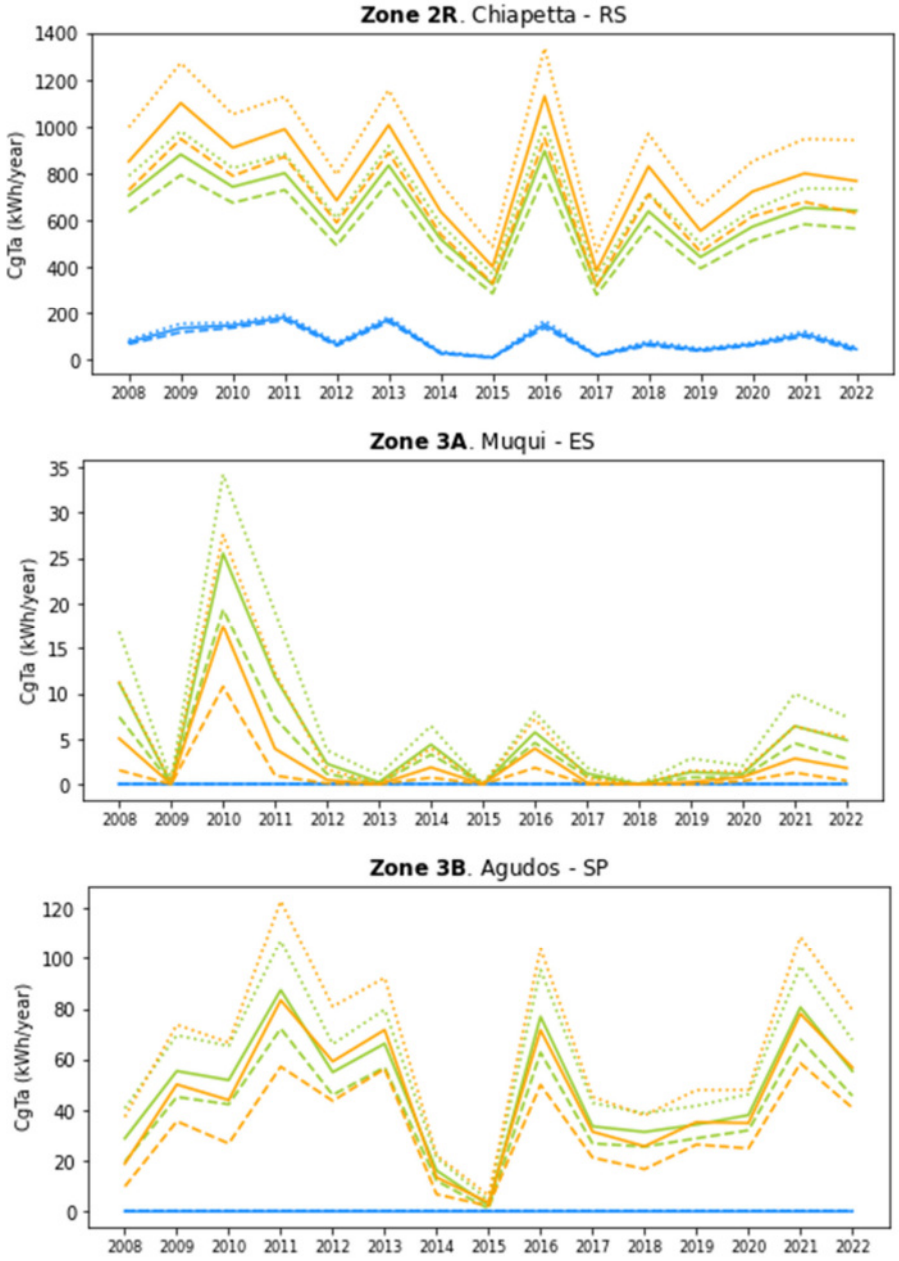
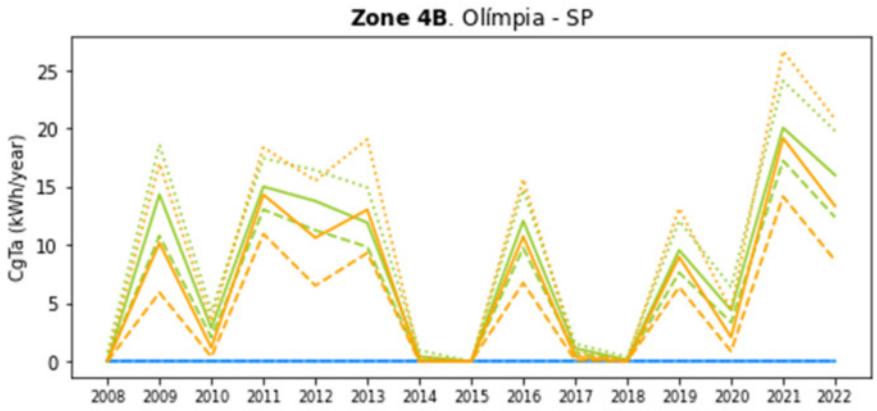


Figure 2.18. *Cont.*



The 15-year average results showed that São Marcos-RS (1R) had the highest heating demand (1833.6 kWh/year) when combining Ulow with the lowest solar absorptance but also presented the highest absolute reduction (1553.4 kWh/year) for the combination of Ulow with a high solar absorptance. Among the other seven locations, Muqui-ES (3A) had the lowest average demand (2.8 kWh/year) and accomplished a 100 % reduction when using Ulow. As expected, the construction set with a high U-value presented the highest heating demand, while the most insulated resulted in a lower or even no demand for heating. In the climatic data analysis from Section 3.1.1, it was found that there was a predominant heating process in Brazil's 15-year series. As a result, the heating demand over this simulation period decreased the cumulative average, even for the building envelope configurations that allow for a higher heating exchange (U_{medium} and U_{high}).

The representative locations from zones 3 and 4 also showed a low heating demand. The 15-year average HDD18 from Figure 16 showed that for zones 3, the average value is 51 K d with an average maximum of 221 K d. Zone 4's average value is 9 K d with an average maximum of 122 K d. The combined analysis of HDD18 and simulation results points out a necessity of a review of the NBR 15575 to adjust the threshold for CgTa calculation.

The 15-year summary (Figure 2.19) indicated that Pio XII-MA (6A) had the highest cooling demand when combining high U-value and solar absorptance (13,070.7 kWh/year). However, it also showed a significant reduction in cooling demand when using the insulated building envelope (U_{low}) with a low absorptance (4171.3 kWh/year). São Marcos-RS had the lowest cooling demand (1214.3 kWh/year) when using the intermediary U-value construction with high solar absorptance and achieved a 73 % reduction using a high insulated construction with low absorptance. The results from the average construction set showed that the ones with

high U-value resulted in the highest demand, while the highly insulated construction presented the lowest demand, with an average reduction of 29 %. Colônia Leopoldina-AL (5A) and Pio XII-MA (6A) showed an exception since the intermediary building envelope with low absorptance delivered the best results.

Figure 2.19. Annual results for the CgTr. Blue represents U_{low}, green represents U_{medium}, and orange represents U_{high}. The dotted line indicates a solar absorptance of 0.3, the solid line indicates a solar absorptance of 0.5, and the dashed line indicates a solar absorptance of 0.7.

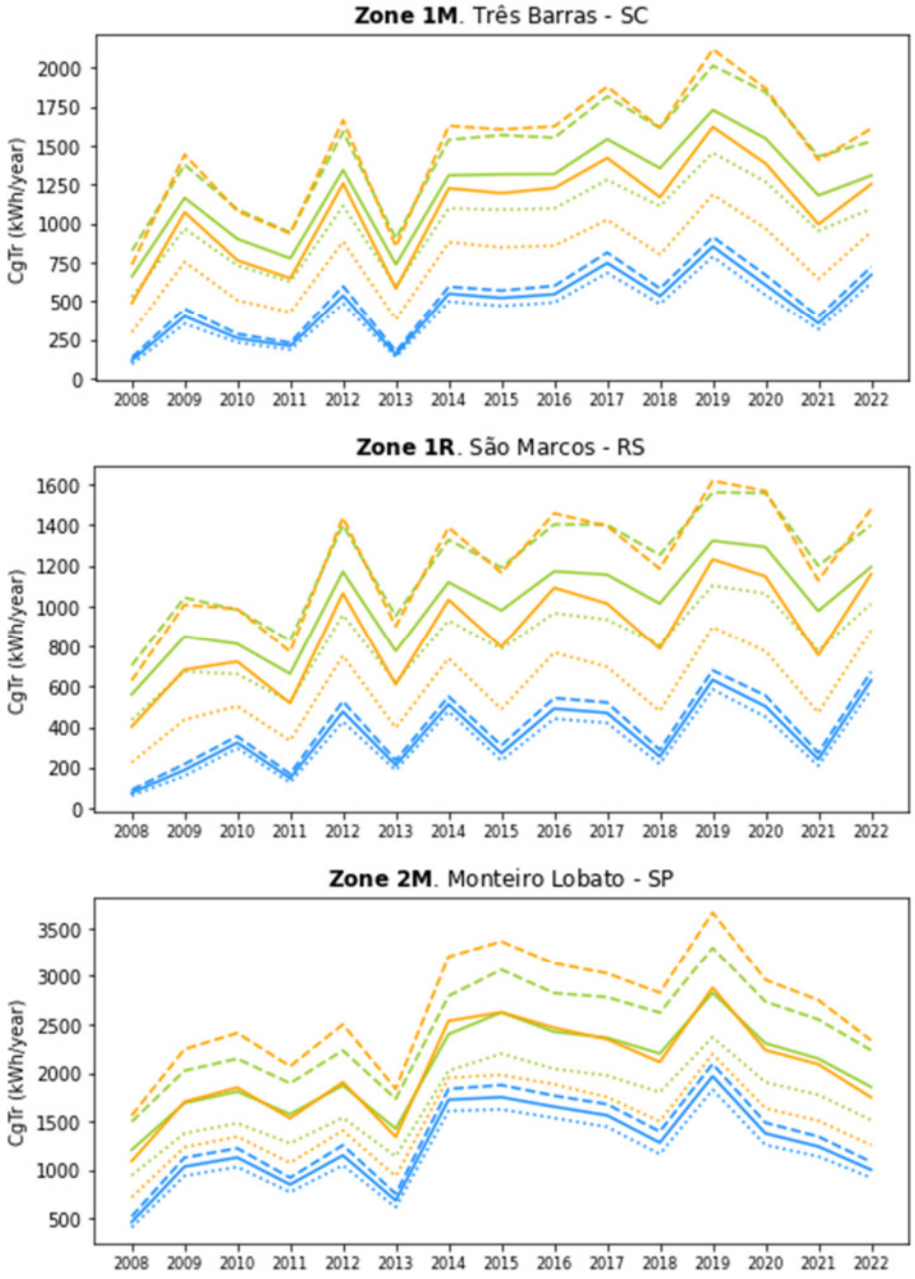


Figure 2.19. *Cont.*

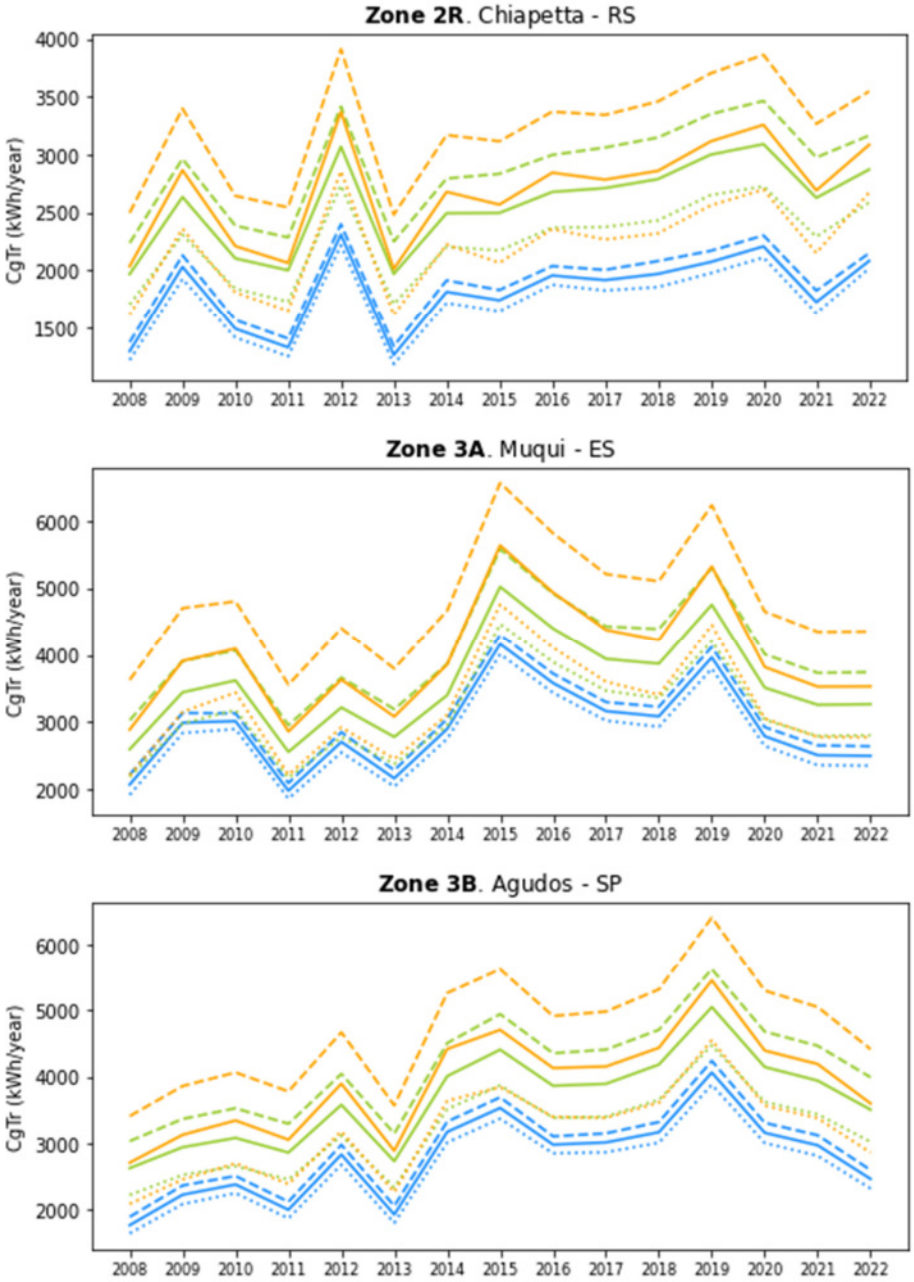


Figure 2.19. *Cont.*

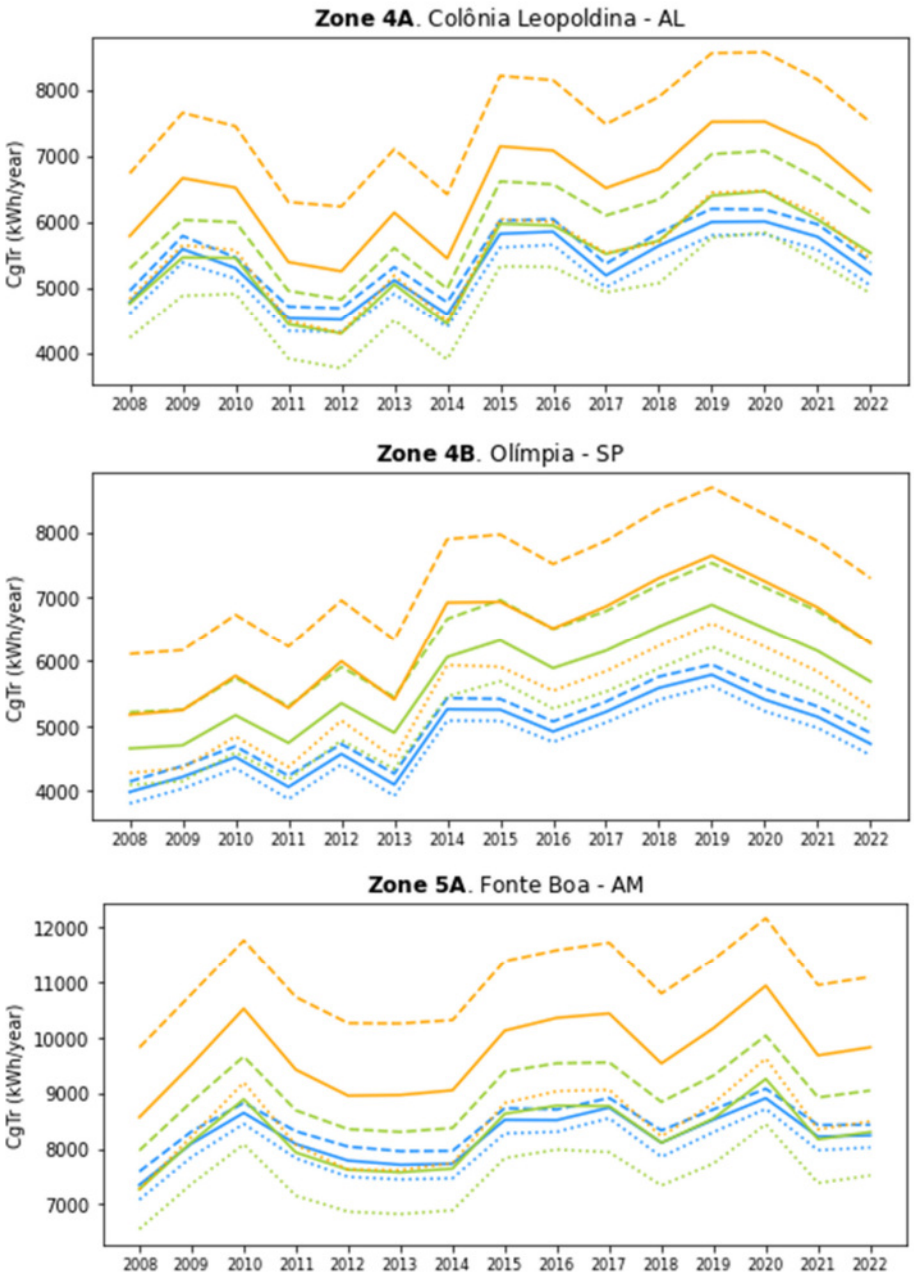
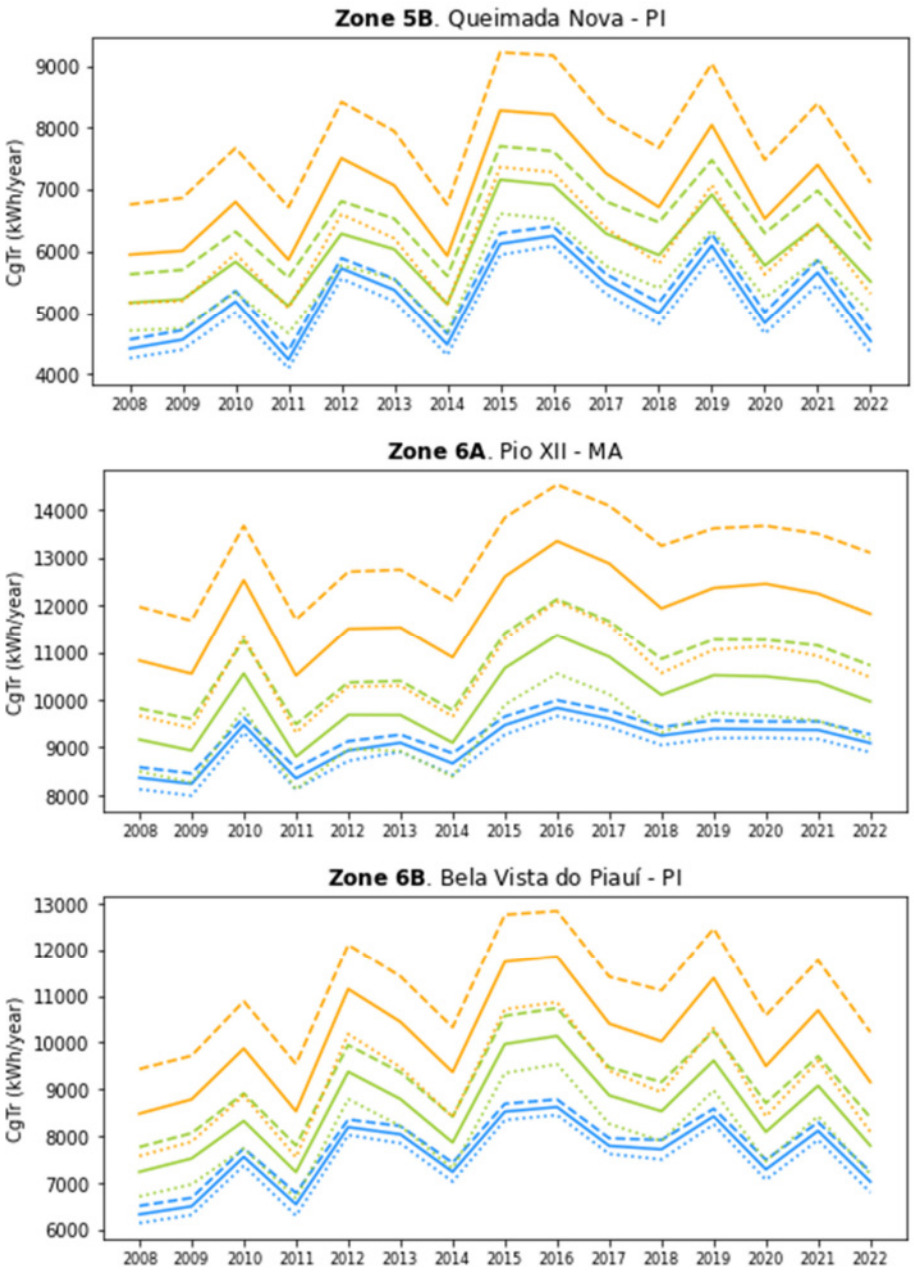


Figure 2.19. *Cont.*



Regarding cooling demand, all locations presented an increasing trend, unlike the PHFT, which mostly indicated a decreasing scenario with some increases based on the building envelope configuration. The multi-year analysis also showed that Bela Vista do Piauí-PI (6B) also had the highest variation for the cooling energy demand, with an average increase of 415.5 kWh/year on the cumulative average. On the opposite side, Chiapetta-RS (2R), with a cold climate, had the lowest increase (218 kWh/year). The building envelope also influenced the

magnitude of the increase, as the highly insulated configuration indicated the lowest average increase (738 kWh/year), while the building envelope with a high U-value had the highest average increase in the building demand (966.4 kWh/year).

2.3. Discussion

Different studies report the effects of climate change and exogenous climatic events on meaningful variations in temperature, humidity, and precipitation in the Brazilian context (75,76,77). Therefore, the presented results are supported by previous research. The highest temperatures are directly linked with four Brazilian biomes: the Amazon rainforest, the Caatinga, and Cerrado—generally characterized as Savanna—and the Pantanal, a tropical wetland. Those regions are also characterized by the highest occurrences of wildfires in the Brazilian territory. The results also present concerning aspects, especially regarding the temperature increase since they are above the threshold to reduce climate change effects (78).

Nevertheless, relative humidity and precipitation show variations throughout the 15-year records and DBT maximum and minimum records show a trend to an annual amplitude increase. The effects of El Niño and La Niña can explain the significant variations in temperature and especially precipitation, resulting in the extreme conditions shown for the northeast region. For the last weather parameter (GHI), the ERA5-Land records proved to be a reliable source for GHI since the difference based on the annual average from monthly integrals is low. However, this study does not analyze the impact of different resolutions (daily and hourly). Thus, it is essential to quantify the differences in building performance and determine the ideal source for each location.

The results from Givoni's chart indicate that passive design strategies are sufficient for most of Brazil to achieve thermal comfort, even with artificial heating in the southern region during the coldest months. This study's approach, which considers a daily resolution and the 7-day threshold, allowed for a detailed evaluation of the recommended design strategies compared to the results provided by the NBR 15220:3 (66). The monthly bioclimatic results confirmed that ventilation is a fundamental strategy to achieve thermal comfort and mitigate heat gains according to outdoor weather characteristics, as previously shown in Silveira et al. (65), Loche et al. (69), Roriz et al. (63), and the NBR 15220 (66). The NVP results supported Givoni's assessment and portrayed a more detailed analysis focusing on the different approaches to quantify the ventilation potential as a cooling strategy. The results also show that the latitude and temperature distribution over the territory were significant factors in

determining the natural ventilation potential based on the DBT records. Despite Givoni's results not requiring artificial cooling, according to the threshold of 7 days, different studies point out the need for HVAC systems in Brazil (67,79,80). It is important to note that the analysis with Givoni's chart considered only the outdoor environment, and the outcomes can vary when analyzing building simulation results due to the outdoor conditions and the building's thermal properties, internal gains, and operation.

In general, when looking at the average PHFT results, colder climates near or below the Tropic of Capricorn achieved their best performance when using construction components with high insulation, regardless of the solar absorptance. On the other hand, warmer climates achieved their best performance with construction components with an intermediary U-value and low solar absorptance. Despite specific occurrences of high PHFT for construction sets with high U-value and low solar absorptance, the average results indicated that U_{high} delivered the worst average results. The third construction set (U_{high}), with the lowest solar absorptance, generally generated the highest heating demand by promoting a higher heating exchange with the outdoors than U_{low} with insufficient heat gains, compared to higher solar absorptances, to compensate for the losses. Cooling demand results followed a similar trend to the heating demand, with the highly insulated model reaching the lowest demand, while the construction set with the highest U-value (U_{high}) had the highest requirement for cooling. The exceptions are Colônia Leopoldina-AL (4A), Fonte Boa-AM (5A), and Pio XII-MA (6A). The best outcomes for these three locations required an intermediary U-value, but they still had a low solar absorptance. The results are likely related to high temperatures and high humidity conditions. The cooling requirements followed the bioclimatic zones classification, showing a lower demand for the colder zones (1M and 1R) and a higher demand for the hottest locations (6A and 6B). The results reinforce the correlation between high and no-insulation components by accentuating the separation between the construction set's demands according to the increase in the annual average DBT and, consequently, the bioclimatic class.

Finally, according to the analysis procedure from NBR 15575 (74), locations with an annual average DBT below 25 °C required the analysis of heating and cooling demand, but Colônia Leopoldina-AL (4A) did not present any heating demand for the 15-year results. This result could indicate the necessity of a review of the standard if other locations of the same zone follow the same pattern. Furthermore, the low heating demand for zones 3 and 4 can also raise the question of the necessity of heating demand assessment for the locations within those zones.

The limitations of this study can be summarized as follows: (1) climatic analysis can rely on monthly and annual records, but a bioclimatic approach requires daily or hourly records to provide a better characterization and more reliable results; (2) using a bioclimatic zone to characterize the territory is more adequate than using the geographical or the Köppen–Geiger distribution, but future studies should identify locations that represent the extremes of each zone to provide a wider characterization; (3) building performance is the last step of characterizing the territory since it encompasses the effects of different climatic data, but future studies should use multiple building geometries.

The primary contribution of this study is a replicable methodology that relies on multi-year records from the worldwide ERA5-Land database, a bioclimatic analysis using Givoni's method, the NVP and degree day indicator, and finally, building performance analysis. Therefore, despite being used in Brazil as a case study, the methodology can be replicated for any location worldwide.

2.4. Main findings

In this chapter a multi-scale analysis for the Brazilian territory was used to present a climatic profile of the country, based weather records for 5567 municipalities from 2008 to 2022.

The climatic analysis showed that most of the territory faces an increase in the average DBT and a reduction in the relative humidity and precipitation. These results directly affected the Köppen-Geiger classification, leading to a higher occurrence of hot climates and the introduction of the BWh classification in the Semi-Arid region. The bioclimatic analysis, combined with the NVP assessment showed that the natural ventilation is the most important bioclimatic strategy for Brazil and it can increase the thermal comfort condition throughout the territory.

The identification of representative locations, based on the Brazilian bioclimatic zoning allowed quantifying the building performance using different building envelope configurations. Moreover, the results showed that using highly insulated building envelopes can significantly reduce the heating demand for colder climates and that building envelope configurations with high U-value deliver the worst performance.

CHAPTER 3: Modeled and measured weather data – INMET vs ERA5-Land

This chapter discusses the use of modeled data from the ERA5-Land database and measured weather data from automatic weather stations from the Brazilian meteorological institute (INMET).

In the first part of the chapter the municipalities with weather records within the 2008-2022 are identified and the gap lengths are quantified for the key weather parameters used in the analysis. Attention was given to the locations with less missing data and the chosen locations were compared against the ERA5-Land records to quantify the differences. The next step consisted in testing different approaches to fill INMET's gaps, based on three approaches: linear interpolation, regression model supported by a machine learning algorithm, and Multivariate Imputation by Chained Equations (MICE). The approaches were compared using the coefficient of determination (R^2) to determine the best methodology for each variable and location.

After identifying the best gap filling approach, the analysis focused in creating the weather files with the ERA5-Land and INMET's records. Then, a residential building with four different orientations and nine building envelope configurations was used to quantify the differences between the weather files from the different data sources. The comparison was based on the performance assessment procedure from the NBR 15575. Therefore, it focused on the operative temperature and thermal load of the building.

3.1. Methods

The following subsections describe the process of comparing INMET's and ERA5-Land's data to quantify their differences from data gathering to building performance assessment.

3.1.1. Location selection

The comparison focused on the assessment of different weather data sources for building performance simulation from automatic weather stations' records from the INMET and ERA5-Land data. First, all automatic weather stations (AWS) catalogued by INMET were identified (566 locations) for the selection of the ones with records available from 2008-2022. It resulted in a sample of 225 locations with records for the 15-year period (Figure 3.1). For some locations, an adjustment of the coordinates used as input in ERA5-Land approximated the urban areas, because the AWS are eventually in distant locations that can represent particular microclimatic conditions.

Figure 3.1. The initial AWS population (black dots) and final sample with 225 (white center).

Within the five Brazilian main regions, 86 locations are from the Southeast region, 49 are from the Northeast region, 42 are from the Central-West region, 39 are from the South region, and 9 are from the North region. The locations cover all the nine Köppen-Geiger climates that occur in Brazil (28). The Southeast region with 38 % of the sample had eight from the nine climates, but mostly the tropical savanna climate with dry-winters (Aw). The Northeast region with 22 % of the locations had six climates, but mostly the tropical savanna climate with dry winters or dry summers (Aw and As). The Central-West had 19 % of the locations, mostly defined by the Aw climate. The South region locations had 17 % of the locations, mostly defined by a humid subtropical climate (Cfa). Finally, the North region had 4 % of the locations and was defined by tropical climates. In general, the tropical savanna climate with dry winter (Aw) had the highest occurrence (120 locations), followed by the humid subtropical climate (45

locations). The remaining climates had less than 20 occurrences (As, BSh, Cwa, Af, Am, Cwb, and Cfb), varying from 19 to 4 occurrences.

3.1.2. Raw data analysis: bias correction and gap filling

This section describes the process of gathering the weather data and comparing the recorded weather variables from ERA5-Land and INMET. Raw data analysis was the procedure of evaluating INMET's data without any modification on its content. Since ERA5-Land is a widely used and validated database, INMET's quality assessment was compared to the ERA5-Land database (5) by quantifying the differences between the records for different weather variables.

Unlike the ERA5-Land database, INMET's automatic weather stations (AWS) can present missing values on their records. The first analysis identified how many missing values each location had for each year and variable. Therefore, all years with more than 72 consecutive records missing (3 days) and gaps that exceed 40 % were removed. The locations with at least five eligible years remained after applying the filters. Then, the download of the hourly records from the ERA5-Land database intended to gather data for the correlation with the INMET's records. Finally, the non-missing values were compared with the corresponding hourly values from the ERA5-Land using the average (DBT, DPT, RH, PRES, WS, and WD), the integral (GHI and PPT), the standard deviation, the root mean square error (RMSE), the coefficient of determination (R^2), and the Spearman correlation.

After quantifying the extension of INMET's gaps, filtering the eligible years, and comparing the databases, the next step was testing three techniques for filling the gaps: linear interpolation, a prediction model based on a machine learning algorithm, and a Multivariate Imputation with Chain-Equation (MICE). The Extreme Gradient Boosting (XGBoost) machine learning algorithm was the prediction model as it can handle large datasets in less time without compromising the results (81). Then, for each weather variable, the estimators were set as the day, hour, and the corresponding variable from the ERA5-Land database. Therefore, the approach seeks to correct bias from the ERA5-Land to provide the correct data for the INMET stations.

The MICE method combines machine learning with an imputation strategy predicting the missing values based on the non-missing ones on the same dataset (82). The XGBoost algorithm was also the predictor. In both situations, an optimization improved the algorithm performance by adjusting the configurations to increase the accuracy of the prediction model.

Following the comparison of the three methods, the approach with the highest R^2 filled the gaps for each weather variable. Regarding DPT and RH variables, only the gaps for the variable with the lowest normalized error were filled, while the remaining had the missing values calculated with the adequate psychrometric correlations.

3.1.3. Building performance: weather data compilation and simulation settings

After analyzing and processing INMET's data, the resulting database and the ERA5-Land records enabled the creation of Actual Meteorological Years (AMYs) for the eligible locations according to the method described in Chapter 1 – Section 1.3.

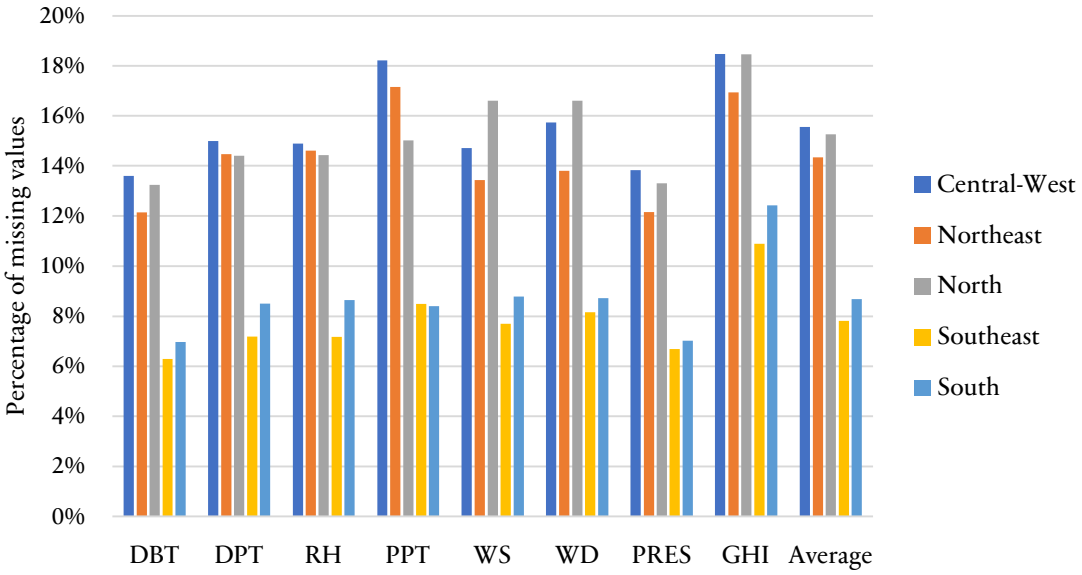
The building modeling and envelope configurations followed the same process described in Chapter 2 – Section 2.1.5. Finally, the study compared the performance of both databases applied to building performance simulations of a detached house using the settings from the Brazilian standard for the minimum performance requirements of residential buildings (NBR 15575) (74). As done in Chapter 2, the assessment also used the Brazilian bioclimatic zoning (67) to guide the analysis process.

3.2. Results and discussion

3.2.1. Weather data: raw data, interpolation results, and multi-year assessment

The results showed that average gap varied from 9 % for DBT records to 14 % for GHI, with relative deviations above 100 %, indicating a high variability despite the low average gap result (Figure 3.2 and Appendix B – Figure B1). For all records, the results also showed extreme situations, with the worst scenario (entire year missing) and the best scenario (entire year available). A clear split in time for the AWS records resulted from the COVID-19 pandemic (2008-2019). Before the pandemic, the stations presented lower gaps and from 2020 until 2022 the gaps increased significantly, with the worst results occurring in 2021. The missing data and the lockdown procedures that prevented maintenance in the AWS affected the data measurement in several locations.

Figure 3.2. Number of missing records summarized by region.



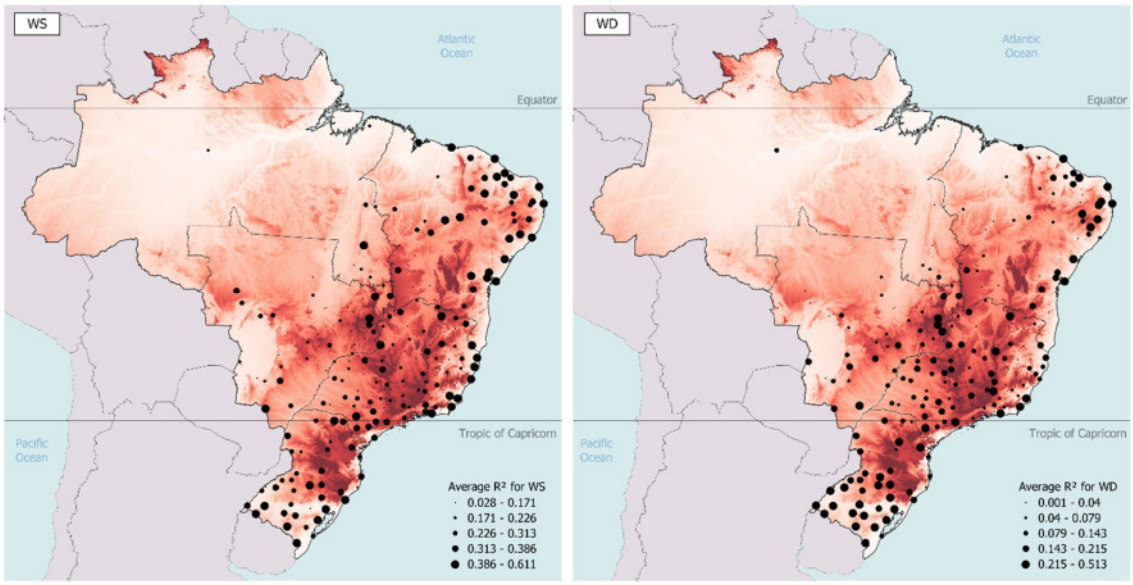
Following the analysis of missing values, the results pointed the locations with 14 and 15-year data, within the 40 % threshold, that represented 53.3 % of the initial sample. Only Campos do Jordão - SP had less than five eligible years and it was excluded from the following analyses. Despite the 40 % threshold results showing a significant number of locations with 14 and 15-year data, the assessment of consecutive missing records showed that 11 locations had less than five eligible years within the 3-day threshold, and only three locations had 15 eligible years. The combination of the two criteria discarded 14 locations from the original sample (225 locations). The remaining 211 locations had an average of nine eligible years, varying from five (16 locations) to 15 years (3 locations).

The eligible years occurrence varied from 88 locations in 2021 to 161 locations in 2009, with an average of 141 locations per year with a standard deviation of 12 from 2008 to 2020. The years of 2021 and 2022, as previously stated, presented the worst results, with 88 and 90 locations, restating the impact of the COVID-19 pandemic on the records and weather data quality. Figure 3.3 shows the distribution of eligible years across the Brazilian territory.

Figure 3.3. Eligible years distribution.

The comparison between the hourly records from the AWS eligible years and the ERA5-Land database, showed that pressure records had the highest correlation between databases, showing a correlation above 90 % and also lower deviations. PPT records showed the highest differences between databases. Despite the significant correlation (above 80 %) and low standard deviation, the temperature records still present a high difference for DBT and DPT with an average RMSE of 1.9 °C and 1.7 °C. The resolution of the data, associated with the data gathering or modeling method explain the differences between databases. The AWS site conditions are also impacting, since the parameters with smaller R^2, i.e., lower correlation, are the most affected by the immediate surroundings (Figure 3.4 and Appendix B – Figure B2).

Figure 3.4. R^2 distribution for wind speed (left) and wind direction (right). The red shading represents a digital elevation model for Brazil, varying from 0 to above 1000 meters.



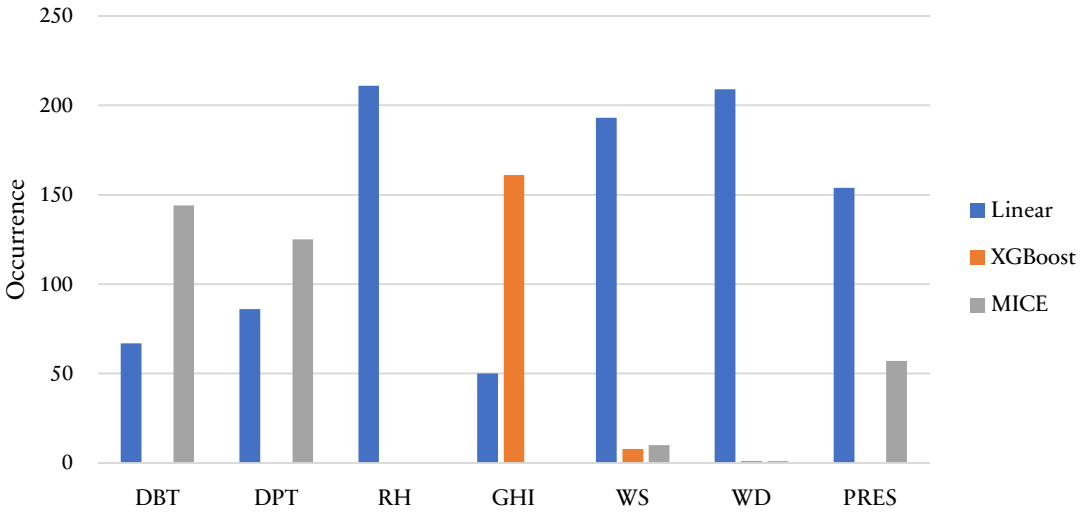
The comparison between the INMET and ERA5-Land records mostly showed that the modeled records provide higher values when considering the mean and median (Table 3.1). The pressure records have the best approximation with the lowest nMBE for both the Mean and Median comparison. Nevertheless, the wind related variables presented the highest deviation. The results also showed that ERA5-Land overestimates the DBT, DPT, RH, GHI, and WS records when considering the mean value. For the remaining weather parameters (WD, PRES, and PPT), the results showed lower values for the ERA5-Land data. The analysis showed the overall difference of the weather records and allowed concluding that only analyzing the mean and median is not enough to fully characterize the correlation between the databases. The mean and median approach provides an initial assessment, but mask the complete distribution and correlation between datasets, misleading the results, especially for the precipitation data.

Table 3.1. Statistical summary of INMET's and ERA5-Land's comparison for the eligible years. The MBE and nMBE were calculated based on the mean and median.

Weather variable	Mean		Median	
	MBE	**nMBE**	**MBE**	**nMBE**
DBT (°C)	-0.11	-0.49%	-0.24	-1.06%
DPT (°C)	-0.30	-1.81%	-0.26	-1.53%
RH (%)	-0.35	-0.49%	0.79	1.06%
GHI (kWh/m²)	-19.47	-4.58%	-29.74	-7.35%
WS (m/s)	-0.34	-16.68%	-0.45	-25.07%
WD (°)	18.67	12.29%	23.17	17.15%
PRES (Pa)	-17.05	-0.02%	-15.17	-0.02%
PPT (mm)	35.78	2.91%	0.00	-

Figure 3.5 indicated that filling the gaps required different interpolation methods for the same location, depending on the weather parameter. After identifying the best method for each weather parameter, the coefficient of determination (R^2) results showed an average above 0.95 and a deviation below 0.02 for the recommended techniques for DBT, DPT, RH, GHI, and PRES. The wind direction average R^2 was of 0.70 and the relative deviation was of 0.9. Those records presented the worst results, with a R^2 of 0.39 and a relative deviation of 0.16. The variation between the maximum and minimum R^2 for each variable restate the good performance of the interpolation techniques over DBT, DPT, RH, GHI, and PRES records with a minimum always above 0.9. For the wind components, however, the R^2 variation intervals were of 0.8 to 0.89 for WD and of 0.36 to 0.89 for WS. Thus, it enables identifying locations where the interpolation techniques provide satisfying results. Moreover, we used the best technique for each weather variable and location to fill the gaps and obtain the 8760 hourly records, even for the variables with a low R^2, since they are required for the simulation stage in EnergyPlus.

Figure 3.5. Occurrence of interpolation techniques according to the weather variable.



Only RH records had unanimous results, since the analysis indicated the linear interpolation as the best gap filling technique. Thus, we did not fill the DPT with the imputation techniques and used the MetPy library to calculate the variable following the psychrometric relationships between the DBT and RH. MICE is the recommended approach for 144 locations and the linear interpolation was for the remaining 67 municipalities. For the GHI records, the XGBoost approach outperformed the linear interpolation for 161 locations while pressure and wind required the use of the linear interpolation method for most of the locations. In general, the results showed that only linear interpolation was recommended for at least a single variable for all locations, while the MICE appeared in the second position, and XGBoost outperformed the other approaches just for the radiation records.

3.2.2. Building performance simulation assessment

The PHFT for both INMET and ERA5-Land simulation pointed out that the intermediary U-value for the overall building envelope (U_{medium}), followed by U_{low} achieved the best operative temperature conditions while the highest U-value (U_{high}) led to the worst performance. An exception occurred in 24 municipalities where U_{high} with a solar absorptance of 0.3 delivered a slightly better PHFT than the others. From those locations, U_{high} with a solar absorptance of 0.3 delivers the best PHFT for Itapira-SP and Resende-RJ for both INMET and ERA5-Land.

Considering the INMET results, the average PHFT was 50.2 %, from 11.5 % in Goiás Velho-GO to 89.2 % in Diamantina-MG, while ERA5-Land's average PHFT was slightly

superior, of 52.4 %, from 13.3 % in Manaus-AM to 84.7 % in Passa Quatro-MG. In general, INMET and ERA5-Land provided similar PHFT, with a deviation ranging from -4.5 percentage points (pp) in Água Boa-MT to 6.3 pp in Teófilo Otoni-MG. Considering the Brazilian bioclimatic zoning, ERA5-Land also presents a better performance with a significant increase in acceptable range of operative temperature for zone 6A (Table 3.3).

Table 3.3. Average PHFT according to the Brazilian bioclimatic zoning.

Bioclimatic zone	**PHFT [%]**		
	INMET	**ERA5-Land**	**Difference [pp]**
1M	73.6	75.4	-1.8
1R	60.0	63.6	-3.6
2M	76.0	77.3	-1.3
2R	57.8	60.0	-2.2
3A	58.2	66.1	-7.9
3B	60.6	62.5	-1.9
4A	41.9	41.3	0.6
4B	36.9	38.2	-1.3
5A	30.1	34.0	-3.9
5B	40.3	38.9	1.4
6A	49.6	65.1	-15.6
6B	46.4	52.2	-5.8

From the 211 municipalities in the initial sample, 158 satisfy the Avg. DBT criteria for the heating thermal load analysis (CgTa), i.e., they have an Avg. DBT below 25 °C, as shown in chapter 2 – Table 2.2. However, the majority of these locations demands none or a considerably low value for heating, independent of the weather data source. Therefore, the results expand the analysis provided in chapter 2, where a modification in the CgTa calculation requirement from NBR 15575 was suggested, and provided a wider analysis that corroborates with the previous findings. Considering the bioclimatic zones limits, shown in chapter 2, the present results certify the exempt of CgTa analysis for zone 4's locations. Then, the Avg. DBT criteria for the need of a heating thermal load analysis can be lowered to locations with Avg. DBT below 22.9 °C. Further studies can even address the topic to reduce even more the DBT limit, since most of zone 3 locations have a low heating demand.

Table 3.4. Frequency of heating demand in the studied locations.

Interval [kWh/year]	Number of locations	
	INMET	ERA5-Land
CgTa < 10	84	86
10 ≤ Cgta < 50	19	19
50 ≤ Cgta < 100	11	10
100 ≤ Cgta < 1000	38	37
Cgta ≥ 1000	6	6

As stated, most of the locations did present either a low heating demand or no demand at all. The most insulated building envelope delivered the best results (U_{low}) while the U_{high} had the worst performance, resulting in higher heating demands. For both INMET and ERA5-Land, the highest heating demand occurred in São José dos Ausentes-RS, 3147 kWh/year and 1669 kWh/m². After all, São José dos Ausentes-RS is a small municipality and the 9 km grid of ERA5-Land can easily diverge from the microclimatic conditions that affect the measured data from INMET stations. The heating demand deviation of the remaining locations varied from an overestimation of 293 kWh/year for the INMET records of Canguçu-RS against an overestimation of 256 kWh/year for the ERA5-Land records of São Gabriel-RS.

Following the bioclimatic zoning, as previously stated, zone 3 had an average CgTa below 25 kWh/year while zone 4 CgTa was not significant . Zones 1 and 2 showed the highest heating demand in alignment with the criteria that defined the bioclimatic zonesand considered them the colder zones among all in Brazil, especially zone 1R that encompasses the coldest locations. Therefore, locations within zone 1R have a significantly higher heating demand.

The cooling demand results (CgTr) pointed that both INMET and ERA5-Land have the lowest demands when using the insulated building envelope (U_{low}) with low solar absorptance. The INMET data resulted in an average CgTr of 4657 kWh/year, while ERA5-Land resulted in an average cooling demand of 4246 kWh/year. The INMET results varied from 29 kWh/year in São José dos Ausentes-RS to 9481 kWh/year in Manaus-AM. The cooling load for São José dos Ausentes-RS (169.93 kWh/year) was also the lowest with ERA5-Land data, but the maximum was for São Luís-MA (9918.33 kWh/year). The low CgTr for São José dos Ausentes-RS, the average PHFT, and the high heating demand, for both databases, points out a significant occurrence of operative temperature below the desired PHFT range caused by the severe cold condition in this location. On the opposite position, Manaus-AM and São Luís-MA mainly

present operative temperature above the desired PHFT range, since the cooling load is significantly higher and the average PHFT is lower than 15 %. INMET and ERA5-Land presented a deviation varying from -3484 kWh/year (Alegre-ES) to 3850 kWh/year (Água Boa-MT).

Despite the significant range of cooling loads, the average difference led to an overestimation of the cooling loads produced by INMET climate data by 411 kWh/year. Across the 12 bioclimatic zones, the average difference between INMET and ERA5-Land cooling loads indicates an overestimation for the results using measured weather data. The colder climates, zones 1 and 2, have the lowest average overestimations, while the hottest and humid zone indicate higher discrepancies between the weather data sources (Table 3.5).

Table 3.5. Average CgTr according to the Brazilian bioclimatic zoning.

Bioclimatic zone	CgTr [kWh/year]			
	INMET	**ERA5-Land**	**Difference**	**Deviation**
1M	1156.08	955.61	200.47	17 %
1R	999.58	962.30	37.29	4 %
2M	1707.81	1582.68	125.13	7 %
2R	2875.42	2513.39	362.03	13 %
3A	3714.53	2836.02	878.51	24 %
3B	3192.96	2939.68	253.29	8 %
4A	5454.21	5299.67	154.54	3 %
4B	5613.02	5194.45	418.57	7 %
5A	7822.82	7347.44	475.38	6 %
5B	6775.34	6515.64	259.71	4 %
6A	7540.55	5317.85	2222.69	29 %
6B	7185.14	6135.52	1049.62	15 %

In general, the simulation results indicated that an insulated building envelope significantly reduces the heating and cooling loads. Therefore, it increases occurrence of hours within the acceptable range of operative temperature. The simulation results also illustrate the correlation and, especially, the differences between the INMET and ERA5-Land data records. In general, INMET produced worse results than ERA5-Land and this result raises the discussion on two major studies in Brazil: the bioclimatic zoning (67) and the leading studies for the

current NBR 15575 (83). In both cases, INMET data was used as the primary weather data source and the findings presented in this study show that the use of another weather database could influence the results.

3.3. Main findings

Chapter 3 focused on the analysis of different weather data sources, mainly characterized as a modeled and a measured weather database. The main goal was to quantify the differences and draw possible implications related to the choice of different weather databases. Initially, the analysis focused on quality control of the measured data, identifying the gaps and filtering eligible years for simulation stage. The initial sample of 225 locations was reduced to 211 and the number of eligible years varied from 5 to 15 to provide a multi-year analysis.

Since the measured data presented missing records, the process of filling gaps was tested with three different approaches (linear interpolation, machine learning regression, and multivariate imputation) that were particularly selected for each location and weather variable, depending on a previous analysis to determine their accuracy using the coefficient of determination. All methods were used, being the linear interpolation the approach with the highest occurrence.

A building simulation analysis for a residential model with different building envelope configurations enabled a full comparison of the datasets. Overall, the results show that an insulated building envelope provides the best outcomes for both INMET and ERA5-Land results, considering both operative temperature (PHFT) and thermal loads (CgTa and CgTr). The study disclosed a considerable difference between the simulation results from the different data sources, mainly represented by a superior performance of ERA5-Land. It also reinforced the necessity of revision of heating load analysis of NBR 15575. As the main conclusion, the choice of the weather data source can directly influence the results of building performance simulation processes. Therefore, they can impact the outcomes of standards and regulations that are building performance related.

CHAPTER 4: Weather files for building performance simulation

This chapter discusses the outcomes of using different weather files to building performance simulation analysis of three residential buildings throughout 480 municipalities in Brazil.

Initially, a multi-year building performance simulation considering different building envelope configurations and building azimuths was performed. The focus was providing a multi-year profile of different climatic conditions across the Brazilian territory quantified according to the NBR 15575. Therefore, primarily focusing on naturally ventilated buildings and conditioned spaces, according to operative temperature and thermal loads for cooling and heating.

In addition, typical meteorological years were created for all locations, based on four different methodologies. The approach focused on calculating the deviation between the TMY and the 15-year records to determine which existing method provides the best approximation to the long-term records. The next step followed the development of performance-based weather files. The combination of a correlation analysis and a machine learning algorithm help establishing new weights for the adapted Sandia method.

Finally, the performance-based weather files were tested against the 15-year records and the best TMY. The Kolmogorov-Smirnov test provided the deviation for the hourly distribution. The analysis also assessed the differences using the root mean squared error (RMSE), the coefficient of variation of the root mean squared error (CVRMSE), and the normalized mean biased error (NMBE).

4.1. Methods

4.1.1. Locations' selection: weather data retrieval and processing

There are 5570 municipalities in Brazil, according to the Brazilian Institute of Geography and Statistics (IBGE), distributed in five 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. The ERA5-Land database was used to download all the weather data for 480 Brazilian municipalities (Figure 4.1), considering the records from 2008 to 2022. As a result, 15 EPWs were obtained for each one of the 480 municipalities, following the procedure described in Chapter 1 – Section 1.3. For each municipality, four TMY were also created, following the ISO 15927-4 (25), the minimum Finkelstein-Schafer (22), the Pissimanis (21), and the Best Rank I methods (10) described in Chapter 1 – Section 1.3.2.

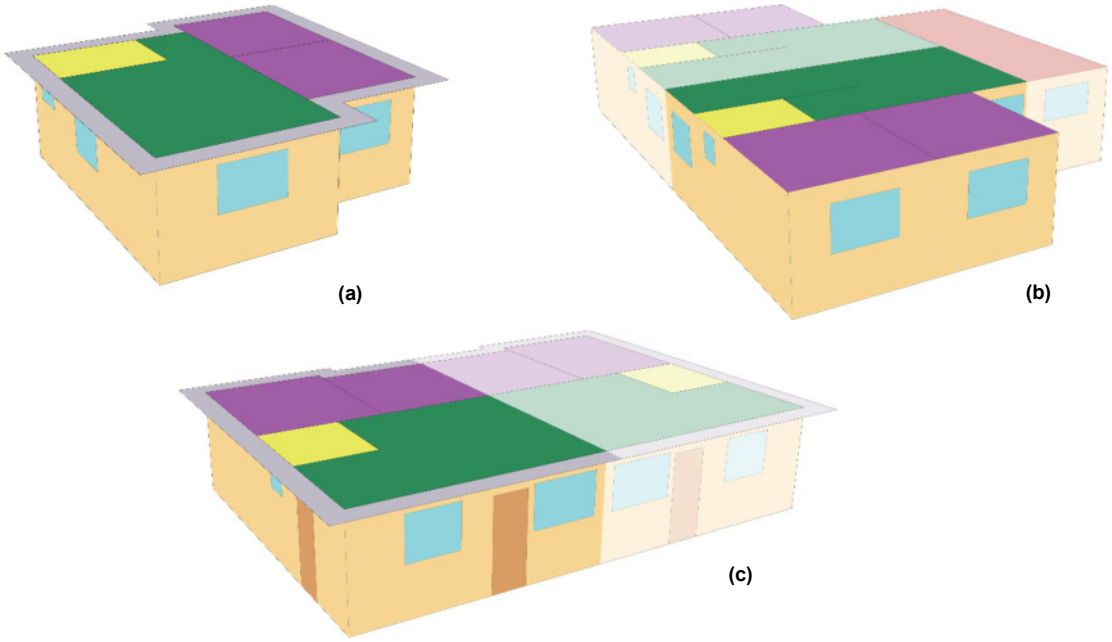
Figure 4.1. The 480 locations (red dots) used in this study and the remaining municipalities (low-opacity black dots).

4.1.2. Building models

Since building performance is directly related to building geometry, internal loads, occupancy, and other patterns, this study required different building typologies. The analysis focused on three typologies of residential buildings based on Triana et al. (47) and Veiga et al. (48). The residential models are representative typologies based on the social housing projects promoted by the Brazilian government. The models follow a similar internal separation, encompassing two bedrooms, one bathroom, and a shared living room and kitchen space. The

net floor area varies from 43.24 m² to 46.45 m². The building models represent a detached house, a single-story terrace, and a multifamily five-story H-shaped building (Figure 4.2).

Figure 4.2. 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 configurations, loads, and use profiles adopted the same values presented in Chapter 2 – Section 2.1.5. The housing units for the single-family models were fully modeled. However, given the complexity of the five-story typology and the consequent higher demand for computational resources, the ground floor, the intermediate floor, and the top floor were simulated as separated geometries. Only half of the building was modeled since both sides are symmetrical and entirely isolated by a central corridor in the shortest axis. All the surfaces that were supposed to be adjacent to a housing unit on the same floor (corridor internal wall) or on a different floor (housing unit and corridor ceiling and/or floor) were set as adiabatic. Independently of the building models, all of the results dealt with an average performance of a single unit.

4.1.3. Performance-based weather files

Hosseini, Bigtashi, and Lee (34) propose a performance-based weather file by

combining BPS results from multi-year weather data series and a machine learning approach. Following the Sandia method approach, the study summarizes the hourly results as daily quantities of cooling and heating demands, and the daily summaries for DBT (mean, maximum, and minimum), DPT (mean, maximum, and minimum), RH (mean, maximum, and minimum), GHI, DNI, and DHI. Then, they used a Random Forest machine learning to fit the data and return the impact of each weather variable on the heating and cooling demands. The method is summarized in four steps, described as follows:

- **Step 1:** create AMY files, define building models and simulation outputs, and simulate;
- **Step 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);
- **Step 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;
- **Step 4:** define a methodology to select the reference year for each month using the impact of each weather variable as weights.

4.1.3.1. Weights definition

For each location, the daily average, maximum, minimum, and spread of DBT, DPT, RH, and water vapor pressure (WVP), daily average and maximum wind speed, and daily integral of GHI summarized the weather data records. The required simulation outputs were the daily average operative temperature and daily thermal loads, energy demand for cooling and for heating. After obtaining the daily quantities, a correlation analysis identified the weather variables strongly correlated to the BPS results for selection.

The Spearman correlation more suited to analyze weather variables and their impact on BPS results, as it quantifies the non-linear correlation between the weather variables. In summary, the process involved first calculating the Spearman correlation, filtering all weather variables pairwise correlations outside the -0.5 to 0.5 range, and finally excluding the weather variable with the lowest correlation with the simulation outputs. Therefore, the process focused on removing redundances that could influence the machine learning model, since weather variables highly correlated could mask the feature importance analysis.

This study followed the approach of Hosseini, Bigtashi, and Lee (34) but used two machine learning algorithms, eXtreme Gradient Boosting (XGBoost) (81) and support vector regression (SVR) (84), to skip bias from a single approach. The choice was based on the applicability of the algorithms in building related problems, their different solving processes, and fast computation.

Machine learning models' performance directly depends on their internal settings and the database used for training to avoid underfitting and overfitting. Overcoming this aspect is fundamental to achieve reliable results, capable of precisely capturing the dataset correlations. Therefore, in this study, an optimization strategy enhanced the algorithms performance and a multiple testing approach to ensure that different training/testing splits of the dataset produce good performances.

The optimization process consisted in using the Hyperopt library for Python (85) to search across the machine learning algorithms parameters' search space for the best parameters combination (Table 4.1). After identifying the best combination, a k-fold cross validation quantified the machine learning algorithm performance across the entire dataset, in different combinations of training and testing. The percentage of correct predictions measured the accuracy of both simulations.

Table 4.1. XGBoost and SVR hyperparameters used in the optimization process.

Algorithm	Hyperparameter	Description	Bounds
XGBoost	Maximum depth	Represents the maximum depth of a tree	2 to 15
	Learning rate	The ratio of errors that new trees will correct	0.001 to 0.3
	Subsample ratio	The ratio of the training data randomly sampled to create each new tree	0.5 to 1
	Number of estimators	Number of trees created by the model	50 to 2000
SVR	Max. number of iterations	Maximum number of iterations possible for the model	500 to 1000
	Gamma	Defines how much a single training subject influences other models	scale or auto
	Epsilon	Tolerance boundary for the predictions	1e-3 to 1
	C	Regularization parameter that controls the tradeoff for the regression model	1 to 100

Once the optimization ensured the models' performance, the Permutation Feature Importance method calculated the impact of each parameter on the simulation outputs (86). The machine learning models previously described produced a set of weights for each building model and simulation output for each municipality. To proceed with the weather file creation process, the average for each simulation output and municipality summarized its weights.

4.1.3.2. Brazilian typical meteorological year

The Kolmogorov-Smirnov test compared the daily summaries of the independent weather variables of each calendar month of the long-term records (15-years) against each one of the 15 years. For each year, the test statistic was calculated to assign a rank from 1 (best) to 15 (worst). The ranks were then multiplied by the weights obtained from the machine learning process and the process repeated for all weather variables. Afterwards, the ranks were summed

up for the selection of the year with the lowest rank as the representative for the given month. The mentioned process applied to all calendar months and both simulation outputs used the resulting weights of each municipality, as described in Section 4.1.3.1. These weather files were named as Brazilian Typical Meteorological Year (BTMY). Depending on the simulation output from which the weights derived, the weather files can also receive the "To" prefix to the weather files with reference years based on the operative temperature and/or the "CgT" to the weather files with reference years based on thermal loads.

Using the same building models and configurations adopted for the AMYs and TMYs analysis, the BTMY weather files enabled accessing the building performance and quantifying the hourly outputs of operative temperature and thermal loads.

4.1.4. Analysis methodology

First, the climatic variables from multi-year records (AMY) provided a characterization of the 480 municipalities used in this study. Then, the method applied by Bracht et al. (80) was used to compare the 15-year records and the four existing TMY methods to identify the sensitivity and quantify the uncertainty on weather files. The approach focuses on the cooling and heating degree hours (K h), direct normal and diffuse horizontal irradiation per year (kWh/m²), and on the maximum, minimum and mean annual relative humidity. For the degree hours calculation, the base temperature was 26 °C for cooling (CDH_{26}) and 16 °C for heating (HDH_{16}). Since the approach only used the climatic variables, it provides a summary of the outdoor condition variations according to the TMY used.

The comparison between the multi-year records, the TMYs and BTMY weather files used the average of all residential models, building envelope and orientation, for each municipality. The procedure focused on quantifying the differences in the hourly results, regarding the distribution and the error. The Kolmogorov-Smirnov test was used as a “goodness of fit” technique, since it quantifies the maximum difference when comparing two distributions (87). The procedure identified the existing TMY method that better approximates the 15-year records, including the four TMY based on existing methods and the BMTY proposed. Then, the TMYs were tested against the long-term hourly results to quantify the differences and determine which methodology produced better results.

For the locations which showed different weather file formats for the operative temperature and thermal loads, the deviations between the methods and the 15-year series guided the selection of a single format. The root mean squared error (RMSE) of the operative

temperature, with a threshold of 1 °C, and the coefficient of variation of the root mean squared error (CVRMSE) of the thermal load outputs, with a 30 % threshold (88).

4.2. Results and discussion

4.2.1. Multi-year and TMY summary

The initial assessment on the 15-year records of DBT, for the 480 municipalities, resulted in an average of 23.5 °C, varying from 14.5 °C in Urubici-SC to 28.5 °C in Piripiri-PI. The RH records showed an average of 71 %, varying from 49 % (Barra-BA) to 90 % (São Gabriel da Cachoeira-AM). The solar radiation results showed an average of 156 kWh/m², going from 118 kWh/m² (Joinville-SC) to 183 kWh/m² (Patos-PB). The climatic summary supports the reliability of the sample used in this study, since the average and extremes are similar to the extremes and average values reported for the whole country in Chapter 2.

When considering the ISO, minimum Finkelstein-Schafer, and Best Rank I weather files, the combination of the variation of the 480 municipalities revealed that HDH, CDH, DNI, and RHmin described an underestimation of the 15-year records. Nevertheless, Pissimanis' weather file presented mostly an overestimation. The average and standard deviation indicated Pissimanis as the responsible for best average performance, but with considerably higher deviations (Table 4.2 and Appendix C – Figure C1).

Table 4.2. Summary of weather variables difference by TMY method. The best values are indicated in blue and the worst in red. (ISO = A, minimum Finkelstein-Schafer = B, Pissimanis = C, and Best Rank I = D)

Weather variables	Average				Standard deviation			
	A	B	C	D	A	B	C	D
HDH	-6.28%	-7.91%	0.63%	-9.18%	39.99%	39.65%	61.06%	31.29%
CDH	-7.31%	-7.04%	1.34%	-7.42%	12.20%	9.59%	21.62%	10.15%
DNI	-0.64%	-0.65%	0.24%	-0.57%	2.17%	2.01%	5.53%	1.81%
DHI	0.81%	0.70%	-0.13%	0.72%	1.44%	1.40%	2.88%	1.26%
RHmax	0.00%	0.02%	0.06%	0.02%	0.40%	0.37%	0.42%	0.37%
RHmin	-0.31%	-0.63%	-2.79%	-0.50%	11.44%	10.86%	13.85%	10.81%
RHmean	0.29%	0.24%	0.01%	0.20%	0.98%	0.82%	2.21%	0.90%

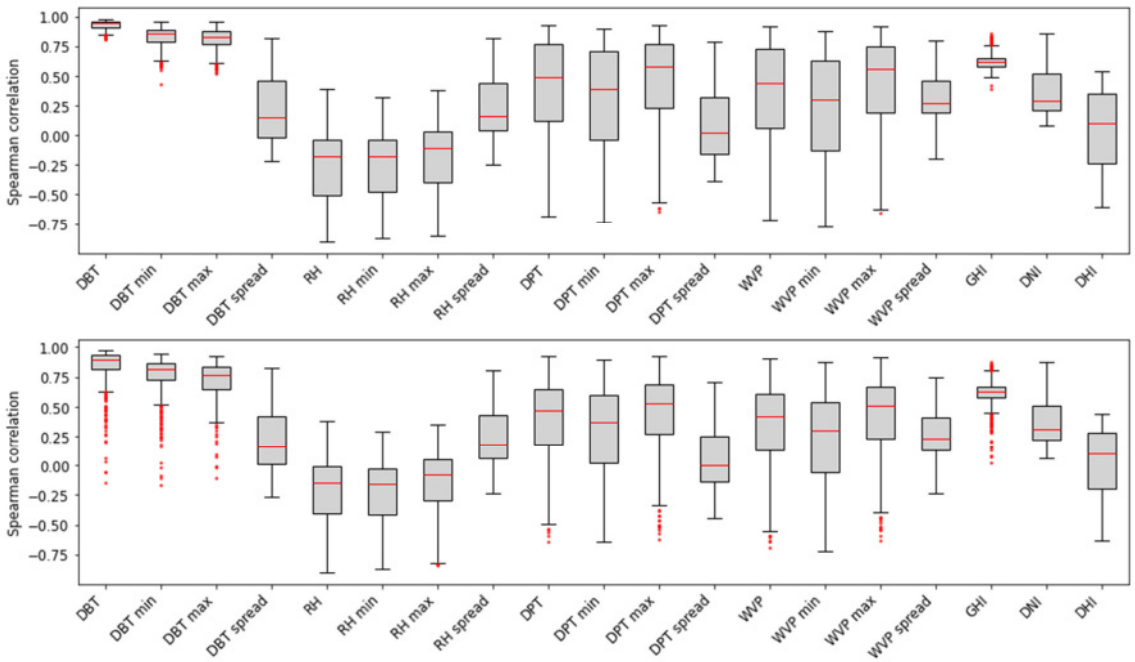
The HDH and CDH pointed to a higher standard deviation and average for all TMY.

However, the difference between AMY and TMY mostly occurred for locations with values close or equal to 0 that still presented a low value for the TMY data, but capable of causing high disturbances in the average and standard deviation. The remaining variables pointed to a lower average and standard deviations, except for the RHmin that presents the same pattern as the degree hour results.

4.2.2. Performance based weather files: correlation and machine learning results

The combined correlation between weather variables (Appendix – Figure C2) and between each weather variable and the simulation outputs (Figure 4.3) showed DBT as the variable with the highest correlation with the simulation outputs. Therefore, the variable was used as the main estimator for the XGBoost and SVR regression models. The average daily, daily maximum and daily minimum DBT showed the strongest and positive correlation for both outputs. The daily DBT presented the highest average with a positive correlation of 0.93 for the operative temperature and 0.84 for the thermal loads. Therefore, the operative temperature and thermal loads variation followed their variation. The remaining variables had a variation in both directions, being direct and inversely correlated to the simulation outputs, but showing average correlations between -0.5 and 0.5, except for the GHI.

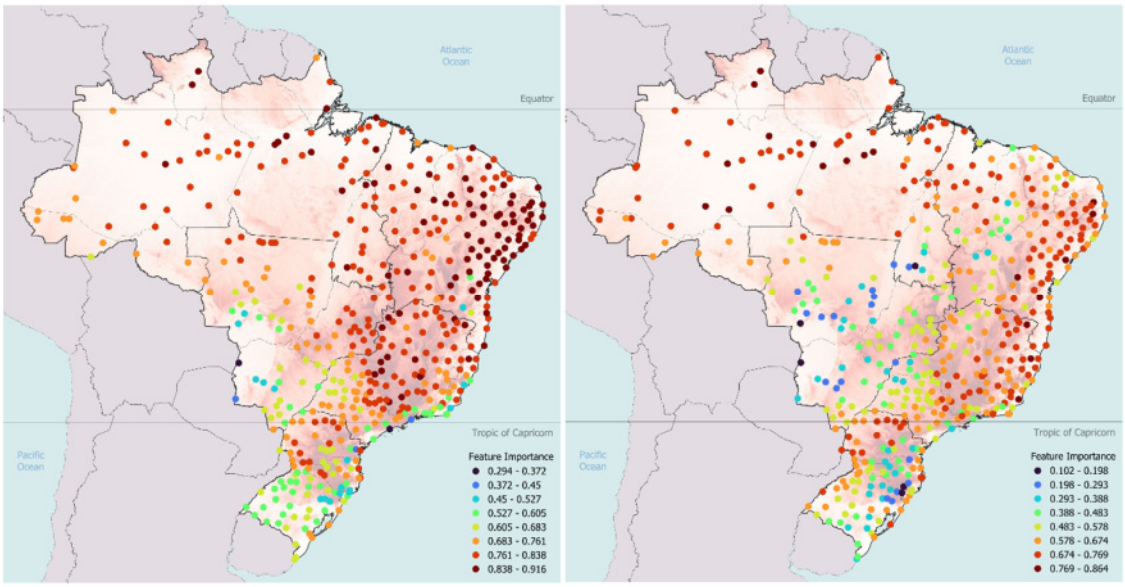
Figure 4.3. Correlation results distribution between each weather variable and the simulation outputs: operative temperature (top) and thermal loads (bottom).



The machine learning models achieved an average R^2 of 0.91 with a standard deviation of 0.03 for the operative temperature and an average R^2 of 0.87 with a standard deviation of 0.05 for the thermal loads. The operative temperature model presented an average RMSE of 0.6 °C with a RMSE deviation of 0.2 °C and the ideal loads model showed an average RMSE of 0.05 kWh/m².day with a deviation of 0.01 kWh/m².day. The feature importance results show that both machine learning models could provide reliable predictions for daily quantities.

The daily average was the only variable which always show some importance for all locations and simulation output. For the operative temperature, the DBT varied from 29 % in Corumbá-MS to 92 % in Lençóis-BA (Figure 4.4). Considering the thermal load results, the daily DBT varied from 10 % in São José dos Ausentes-RS to 86 % in Areia-PB (Figure 4.4). The daily average DBT was the most impactful weather variable for both operative temperature and thermal load models (Figure C3). Despite some presenting some variation, the remaining variables showed significantly lower feature importance values, except for DBT min that reached 53 % for the thermal load model of Corumbá-MS.

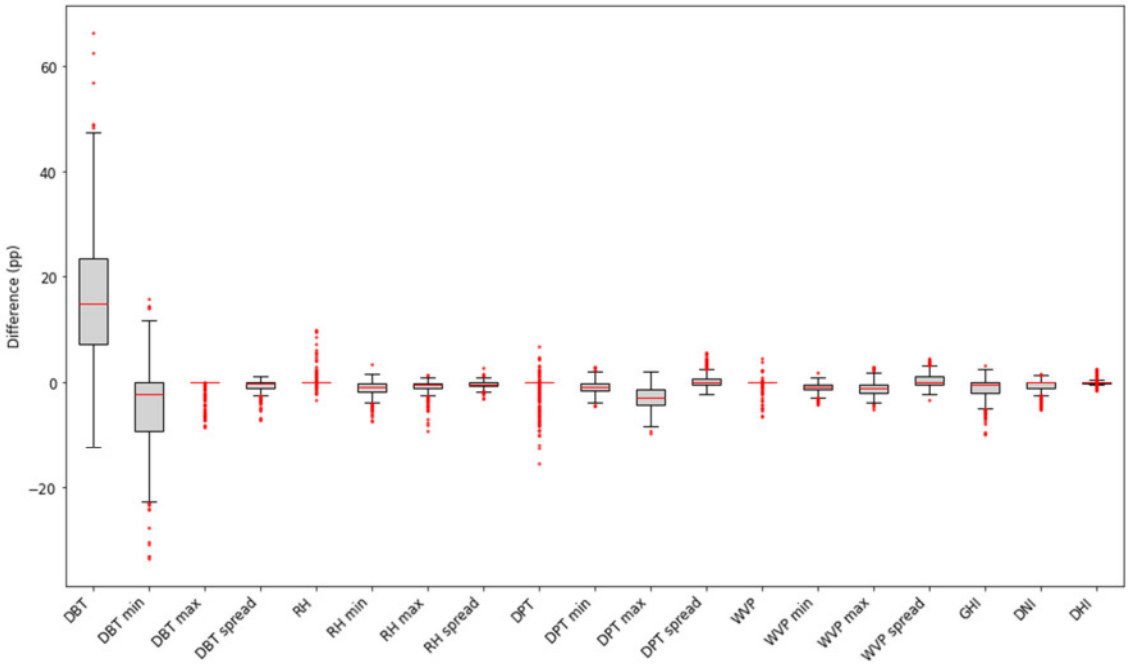
Figure 4.4. Geographical distribution of the daily average DBT feature importance based on the operative temperature (left) and thermal loads (right) machine learning model.



Gurupi-TO resulted in the highest difference for the daily average DBT (66 pp). For the municipality, the daily DBT has the highest impact on the operative temperature model (82 %), but the thermal load model showed a higher influence of the minimum daily DBT (36 %) while the remaining importance was scattered across the other weather variables. The daily average

DBT, as the most influential weather variable, also had the highest MBE indicating that the feature importance based on the operative temperature results present an average importance 16 pp higher than the thermal load models. The daily DBT difference also reached the highest standard deviation (12 pp). The remaining variables resulted in lower differences, with an MBE varying from -5 pp (DBT min) to 0.004 (WVP spread). Figure 4.5 presents the distribution of the differences between the operative temperature and the thermal load feature importance and allows highlighting the variability of the daily DBT while most of the remaining variables showed a lower and more stable difference between models.

Figure 4.5. Distribution of the difference between the operative temperature and the thermal load-based feature importance for each weather variable.



The daily maximum DBT was never used for the operative temperature models while the WVP was the least used variable for the thermal load machine learning model, as shown in Table 4.3. Apart from the daily average DBT which scored the highest importance for both models, all the remaining variables showed an average importance below 4 % for the operative temperature models, and below 8 % for the thermal load models. Therefore, the reference years for the 480 municipalities were mainly defined by the daily DBT records.

Table 4.3. Feature importance summary for the operative temperature and thermal load models.

Variable	Operative temperature					Thermal loads				
	Mean	Std Dev	Min	Max	Occ.	Mean	Std Dev	Min	Max	Occ.
DBT	0.746	0.110	0.294	0.916	480	0.587	0.146	0.102	0.864	480
DBT min	0.026	0.049	0	0.399	187	0.076	0.097	0	0.527	334
DBT max	0	0	0	0	0	0.006	0.017	0	0.084	81
DBT spread	0.002	0.005	0	0.038	200	0.009	0.012	0	0.075	336
RH	0.004	0.018	0	0.177	63	0.002	0.009	0	0.123	67
RH min	0.020	0.011	0	0.064	450	0.030	0.019	0	0.121	465
RH max	0.013	0.011	0	0.075	433	0.021	0.019	0	0.124	453
RH spread	0.012	0.008	0	0.059	450	0.016	0.009	0	0.072	464
DPT	0.008	0.025	0	0.258	84	0.015	0.037	0	0.241	136
DPT min	0.022	0.014	0	0.081	429	0.031	0.017	0	0.090	449
DPT max	0.026	0.016	0	0.089	429	0.054	0.029	0	0.150	453
DPT spread	0.027	0.026	0	0.152	413	0.023	0.016	0	0.095	439
WVP	0.003	0.011	0	0.110	43	0.004	0.015	0	0.121	44
WVP min	0.017	0.009	0	0.066	450	0.027	0.012	0	0.076	457
WVP max	0.024	0.017	0	0.128	440	0.035	0.018	0	0.115	460
WVP spread	0.029	0.023	0	0.106	408	0.024	0.015	0	0.073	422
GHI	0.005	0.012	0	0.122	180	0.016	0.022	0	0.101	323
DNI	0.009	0.017	0	0.098	234	0.015	0.023	0	0.108	274
DHI	0.008	0.013	0	0.087	286	0.008	0.011	0	0.062	276

The reference year selection process resulted in 47 municipalities which required a single BTMY weather file and 433 locations that needed a BTMY for the operative temperature analysis and another for the thermal loads assessment. Considering each calendar month, none of the years can be defined as representative for the majority of the locations (Table C1 and

Table C2). When selecting the reference year for each month, some years presented a higher occurrence throughout the 480 municipalities and the two performance-based models. The results present a similar pattern when the overall selection of each year is considered, disregarding the occurrence for a particular calendar month. Despite some year presenting a higher occurrence, none reached the majority. These results allow concluding that the reference years present an even distribution throughout the 480 municipalities.

4.2.3. Long-term records vs TMY weather files

The hourly distribution analysis, based on the Kolmogorov-Smirnov statistic, pointed out 49 municipalities requiring the same TMY for both operative temperature and thermal loads analysis. For the remaining 431 locations, the error analysis allowed selecting a single weather file for the locations. The performance-based method was indicated as the best weather file for 349 locations, and the remaining TMY appear as the best choice for 131 locations, varying from 13 municipalities (Pissimanis) to 45 municipalities (Best Rank I). The CgT-BTMY and To-BTMY weather files presented a higher occurrence than the BTMY weather file, indicating the need of different reference years for the operative temperature and thermal loads analysis.

The hourly weather data led to situations where thermal load weather files approximated better the 15-year operative temperature results, and operative temperature-based weather files provided closer thermal load results to the 15-year average. Figure 4.6 shows that the best TMY method do not follow a geographical distribution pattern, since all methods are scattered throughout the Brazilian territory, except for ISO, minimum Finkelstein-Schafer, Best Rank I, and Pissimanis that occur less in the North and Central-West regions.

Figure 4.6. Geographical distribution of the TMYs methods.

The RMSE showed that 30 % of the municipalities are within the operative temperature threshold with an average minimum of 0.46 °C. Almost 50 % of the locations are withing the CVRMSE threshold for the thermal load analysis, with an error of 47 %. BTMY methods had the lowest average error. The operative temperature results showed a 1.23 °C error against 1.44 °C from Pissimanis method. The same occurs for the thermal loads results, since BTMY led to an average CVRMSE of 35 % and Pissimanis, of 39 %.

Considering the thresholds determined in this study, the average error and the individual value for several locations are outside the 1 °C RMSE for the operative temperature and the 30 % CVRMSE for the thermal loads. However, these results still represent an improvement,

since the proposed approach showed the best overall performance with the lowest error. It should be reinforced that the entire process of comparing the distribution and errors considers the average of all models, orientations, and building envelope configurations. Therefore, the results show an average condition of the building models.

4.3. Main findings

Chapter 4 discussed the use of different TMY methods and proposed a performance-based weather file for the Brazilian context.

The weather indicator comparison showed that ISO, minimum Finkelstein-Schafer, and Best Rank I have similar results, while Pissimanis' shows higher variations.

The Spearman correlation and machine learning procedure showed that the daily average DBT has the most significant impact on both operative temperature and thermal loads regression models, and so, it became the estimator with the highest weights for the whole process. Thus, the reference years for the BTMY weather files were mainly selected based on the daily average DBT distribution. The reference years presented an even distribution throughout the territory, since none of the years was selected as representative for the majority of the municipalities.

The application of the weights resulted in 47 municipalities that had the same reference year for both operative temperature and thermal loads models. Thus, the majority of the locations required two BTMY for the simulations process.

The comparison between the existing TMY methods and the BTMY weather files and the BTMY method showed that the majority of the locations presented a better hourly distribution of the results and lower errors when using BTMY weather files.

CHAPTER 5: Final considerations

5.1. Conclusions

This research discussed a climatic multi-scale analysis and its influence over the Brazilian territory and building performance simulation. The research is focused on an extensive analysis of Brazil, aiming to investigate the following hypothesis: **A proper characterization of Brazil, given the geographical extension of the territory and the climatic heterogeneity, must rely on a thorough climatic comprehension of the whole territory and the consequences on the built environment, i.e., a multi-scale analysis.** The findings of this research showed that a single approach can be used for initial characterization of a territory, but only an extensive analysis offers sufficient and reliable results to wider conclusions, allowing generalizations. Therefore, the proposed hypothesis is accepted.

The main objective is distributed throughout the chapters, since it corresponds to the proposition of the multi-scale analysis, supported by statistical methods, machine learning, and building performance simulation, and each chapter presents its own methodology that encompasses the different aspects of a multi-scale analysis. All specific objectives were accomplished with this research. Chapter 1 encompasses the analysis that satisfies the first specific objective, chapter 3 satisfies the second specific objective, and chapter 4 answers to the third and fourth specific objective.

5.1.1. A multi-year characterization of the Brazilian territory

Chapter 2 proposed a climatic, bioclimatic, and building performance analysis of the Brazilian territory based on ERA5-Land data from 2008 to 2022. In most Brazilian cities, the average DBT ranged from 22 °C to 26 °C. However, nearly 50 % of the country experienced an average temperature of at least 26 °C. In most areas, the average DBT increased between 0.4 °C and 0.8 °C. RH levels were fairly consistent throughout the territory. The precipitation levels decreased to less than 50 % in the regions with the highest temperature increase. The updated Köppen–Geiger classification introduced the Hot Desert classification “BWh” in the Semi-Arid region of Brazil, showing that the 15-year temperature increase and precipitation decrease significantly impacted the Brazilian territory.

The bioclimatic analysis showed that passive design strategies are still effective in providing thermal comfort, despite the heating process faced by the Brazilian territory. In general, ventilation strategies are Brazil’s predominant approach to reaching thermal comfort,

and the NVP results corroborated the potential of natural ventilation throughout the territory. However, the results of cooling degree days showed several locations requiring cooling strategies. Therefore, combining Givoni and degree days delivered more suitable results for the Brazilian context.

The results of building performance simulation indicated that colder zones achieve a higher percentage of hours with an operative temperature within 18 °C and 26 °C (PHFT), lower cooling demands (CgTr), and even no heating demands (CgTa) when using a construction set with low solar absorptance and U-value. The combination of PHFT, CgTa, and CgTr results showed that, in general, using a construction set with high insulation and low solar absorptance produces the best outcomes, particularly for thermal assessment for zones 1 and 2, and for all zones in terms of heating and cooling demands. However, users should weigh the trade-off between the PHFT and conditioning demand for warm and hot zones when employing a construction set with an intermediary U-value, as it leads to improved PHFT levels. Based on multi-year trend analysis, the simulations indicated a highly insulated building envelope as the cause of reduced variation over the 15 years. Following the results and discussion, the main conclusions of this study are as follows:

- Following worldwide climate change trends, the Brazilian territory showed significant temperature variation for 2008–2022, followed by a reduction in the relative humidity and precipitation levels, based on the ERA5-Land records. For the annual average GHI from monthly integral records, the results also showed that ERA5-Land records are similar to the Brazilian Solar Atlas database;
- The bioclimatic analysis, as depicted in Givoni's chart, indicates the potential of ventilation to achieve thermal comfort in Brazilian territory, supported by the NVP indicator results. However, Givoni's chart was not enough to characterize the cooling requirement, and the degree day analysis showed a significant cooling demand for several municipalities in Brazil;
- Finally, for a complete characterization of the territory, an analysis from the outdoor environment to building performance simulation from a multi-year series allowed us to identify the impacts of the weather variations on building performance by showing that, generally, highly insulated constructions with a low solar absorptance deliver the best performance even in naturally ventilated dwellings.

5.1.2. Modeled and measured weather data – INMET vs ERA5-Land

Chapter 3 focused on quantifying the differences between the weather data from ERA5-Land and INMET. Therefore, this chapter provides answers for the first question presented in chapter 1.

The initial sample encompassed 225 locations. The ground station assessment revealed that 14 locations presented a high number of missing hourly records (greater than 3 days). Considering the constraints defined in this research, these locations could not generate a 5-year series and were discarded from further analyses. Overall, the COVID-19 pandemic affected the measured data, since the lockdown regulations prevented the maintenance of the INMET stations, resulting in higher gaps from 2020 to 2022.

Considering the hours without missing values it is possible to identify that air temperature, relative humidity, and atmospheric pressure have the strongest correlation between ERA5-Land and INMET. Nevertheless, the wind variables and precipitation showed the highest divergence between the data sources. These results can be correlated with the topographical conditions and the particularities of both data sources, since INMET data capture the immediate surrounding of the weather stations, i.e., microclimatic conditions, while the ERA5-Land data comprises a 9 km grid. Still on the quality of INMET records, the investigation of different gap filling techniques showed that the linear interpolation can be used for most variables. Furthermore, the MICE method also provided good results, especially for DBT and DPT. The regression method, based on the XGBoost machine learning algorithm was the least frequent method, but the most recommend for the solar radiation data.

Building performance simulation results showed that ERA5-Land and INMET data can lead to different results, despite agreeing on the best building envelope configuration. PHFT and CgTr results showed that the bigger differences occurred in humid zones. Thus, combining these finding with the discussion presented in chapter 2, where the bioclimatic reclassification was affected mostly by humidity records, allow stating that the humidity records from the databases directly influenced the building performance and resulted in bigger variations. The heating demand (CgTa) also showed divergencies between INMET and ERA5-Land, but mostly allowed a critical view of the setpoint temperature for the CgTa analysis. The extensive analysis can support the definition of a lower setpoint, since several locations that required the analysis of this indicator do not show a heating demand, independently of the database.

Finally, the results show that depending on the database the differences can lead to

significant variations in the results, affecting the proposition of new standards, recommendations or even guidelines, as it happens in Brazil. Therefore, future analysis should focus in quantifying the divergences in different applications, since the recent studies provide robust methodologies that are useful and necessary for Brazil, despite mainly considering a single climate database.

5.1.3. Weather files for building performance simulation

As for the second question presented in chapter 1, chapter 4 focused on the analysis of different weather files for building performance simulation in Brazil, including the proposition of a performance-based method.

Coupling machine learning algorithms to obtain new weights for the different weather variables involved in the weather file generation represents a significant time reduction in the simulation process. AMYs provided simulation results related to the historical series, and the machine learning algorithm allowed per sensitivity analysis to identify the weather quantities that most affect building performance.

Dry bulb temperature was the most significant variable for the reference years, with an average importance greater than 80 %. Despite dry bulb temperature having the primary effect across most locations, there are differences when selecting the reference years for operative temperature and thermal load models.

The results indicate that for 49 locations, the reference years do not change between performance indicators. The statistical analyses indicate differences between the newly created BTMY and the existing TMY files for all locations. The RMSE and CVRMSE metrics reinforced the relevance of creating new Brazilian Typical Meteorological Years, since the BTMY method presented lower errors. Therefore, it produced results closer to the 15-year average. Despite requiring different weather files due to the different reference years between simulation outputs, the statistical analysis allowed reducing the weather files to a single one, for each location, independent of the simulation output.

Finally, the weather files showed an even distribution across the Brazilian territory, but the BTMY method was the most present, since it was the best approximation method for 349 municipalities. The creation of performance-based weather files for different Brazilian locations based on a sensitivity analysis approach that resulted in site-specific weather files for each location was justified by the results found in this study. The results provided a new database that indicates the best weather file for each location and its use is encouraged since

they better approximate the historical records.

5.2. Further developments

For each chapter of this thesis, further developments have been considered.

As regards chapter one, focusing on a multi-scale analysis of the Brazilian territory based on climatic, bioclimatic, and building performance simulation analysis, further developments should consider the three main aspects discussed in the chapter. For the climatic analysis, locations with the most significant changes should be identified and analyzed in higher resolutions, such as the hourly distribution. The bioclimatic analysis should include not only outdoor climatic conditions, but simulation processes coupled to the analysis to understand the real impact of the strategies and their efficiency for the indoor environment. As regards the identification of representative climates, the assessment should include climate change metrics and building performance outputs. For the building model used, considering the variety of the Brazilian building stock, further studies should focus on different building geometries and non-residential building typologies.

As regards the comparison between modeled data from the ERA5-Land database and the measured data from the INMET automatic weather stations, future studies should include different weather data sources and account for the microclimatic effects and its impact on building performance simulation. More building geometries, as more residential and non-residential typologies, should enhance the approach.

The extensive analysis of weather files and its application for building performance simulation in the Brazilian context discussed in chapter four quantified the impact of different building simulation software in the simulation outputs and resulting weights for the feature importance analysis. Further developments should consider summarizing the weights based on the Brazilian bioclimatic zoning, but also analyzing the impact of different building simulation codes on the outputs and weights. Finally, the weather file analysis should consider not only historical records through AMY and TMY, but also extreme events based on historical data and future events from climate change scenarios.

BIBLIOGRAPHY

[1] WU, Y. et al. A global typical meteorological year (TMY) database on ERA5 dataset. **Building Simulation**, 1 jun. 2023.

[2] BABAR, B.; GRAVERSEN, R.; BOSTRÖM, T. Solar radiation estimation at high latitudes: Assessment of the CMSAF databases, ASR and ERA5. **Solar Energy**, v. 182, p. 397–411, 1 abr. 2019.

[3] AFSHARI, A. Optimization of urban design/retrofit scenarios using a computationally light standalone urban energy/climate model (SUECM) forced by ERA5 data. **Energy and Buildings**, v. 287, p. 112991, 15 maio 2023.

[4] MUÑOZ-SABATER, J. et al. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. **Earth System Science Data**, v. 13, n. 9, p. 4349–4383, 2021.

[5] HERSBACH, H. et al. The ERA5 global reanalysis. **Quarterly Journal of the Royal Meteorological Society**, v. 146, n. 730, p. 1999–2049, 2020.

[6] HERRERA, M. et al. A review of current and future weather data for building simulation. **Building Services Engineering Research and Technology**, v. 38, n. 5, p. 602–627, 2017.

[7] HOSSEINI, M.; BIGTASHI, A.; LEE, B. A systematic approach in constructing typical meteorological year weather files using machine learning. **Energy and Buildings**, v. 226, p. 110375, 2020.

[8] PYRGOU, A. et al. Differentiating responses of weather files and local climate change to explain variations in building thermal-energy performance simulations. **Solar Energy**, v. 153, p. 224–237, 2017.

[9] WILCOX, S.; MARION, W. Users manual for TMY3 data sets. **Renewable Energy**, n. May, p. 51, 2008.

[10] PERNIGOTTO, G. et al. Analysis and improvement of the representativeness of EN ISO 15927-4 reference years for building energy simulation. **Journal of Building Performance Simulation**, v. 7, n. 6, p. 391–410, nov. 2014.

[11] KIM, S. et al. Development of test reference year using ISO 15927-4 and the influence of climatic parameters on building energy performance. **Building and Environment**, v. 114, p. 374–386, 1 mar. 2017.

[12] GASPARELLA, A. et al. **Extreme weather data in building performance simulation**. Proceedings of Building Simulation 2021: 17th Conference of IBPSA. **Anais**...KU Leuven, 19 jul. 2022.

[13] BHANDARI, M.; SHRESTHA, S.; NEW, J. Evaluation of weather datasets for building energy simulation. **Energy and Buildings**, v. 49, p. 109–118, 2012.

[14] COSTANZO, V. et al. Updated typical weather years for the energy simulation of buildings in mediterranean climate. A case study for sicily. **Energies**, v. 13, n. 15, 1 ago. 2020.

[15] YUAN, J.; HUANG, P.; CHAI, J. Development of a calibrated typical meteorological year weather file in system design of zero-energy building for performance improvements. **Energy**, v. 259, 15 nov. 2022.

[16] KALAMEES, T. et al. Development of weighting factors for climate variables for selecting the energy reference year according to the en ISO 15927-4 standard. **Energy and Buildings**, v. 47, p. 53–60, 2012.

[17] NCDC - NATIONAL CLIMATIC DATA CENTER. **Test Reference Year (TRY)–Tape Reference Manual TD-9706**. National Climatic Data Center Asheville, NC, USA, , 1976.

[18] EAMES, M. E.; RAMALLO GONZALEZ, A. P.; WOOD, M. J. An update of the UKs test reference year: the implications of a revised climate on building design. **Building Services Engineering Research and Technology**, v. 37, n. 3, p. 316–333, maio 2016.

[19] LUND, H. The Design Reference Year user's manual, Thermal Insulation Laboratory. **Technical University of Denmark: Lyngby**, 1995.

[20] BILBAO, J. et al. Test Reference Year Generation and Evaluation Methods in the Continental Mediterranean Area. **Journal of Applied Meteorology**, v. 43, n. 2, p. 390–400, 2004.

[21] PISSIMANIS, D. et al. The generation of a "typical meteorological year" for the city of Athens. **Solar Energy**, v. 40, n. 5, p. 405–411, 1988.

[22] LAYI FAGBENLE, R. Generation of a test reference year for Ibadan, Nigeria. **Energy Conversion and Management**, v. 36, n. 1, p. 61–63, 1995.

[23] HALL, I. J. et al. **Generation of a typical meteorological year**. United States: 1978. Disponível em: <https://www.osti.gov/biblio/7013202>

[24] MARION, W.; URBAN, K. User's Manual for TMY2s. **NREL**, 1995.

[25] EUROPEAN COMMITTEE FOR STANDARDIZATION - CEN. **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**. Brussels, 2005.

[26] LAWRIE, L. K.; CRAWLEY, D. B. **Development of Global Typical Meteorological Years (TMYx)**. Disponível em: <http://climate.onebuilding.org>.

[27] KIM, Y.; JANG, H.-K.; YU, K.-H. Study on Extension of Standard Meteorological Data for Cities in South Korea Using ISO 15927-4. **Atmosphere**, v. 8, n. 11, 2017.

[28] ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2013.

[29] CRAWLEY, D. B. Which weather data should you use for energy simulations of commercial buildings? **ASHRAE Transactions**, v. 104, n. 2, p. 498–515, 1998.

[30] CRAWLEY, D. B.; HUANG, Y. J.; BERKELEY, L. Does It Matter Which Weather Data You Use in Energy Simulations ? **Building Energy Simulation User News**, v. 18, n. 1, p. 25–31, 1997.

[31] HUANG, Y. J. et al. Development of 3012 IWEC2 weather files for international locations (RP-1477). **ASHRAE Transactions**, v. 120, p. 340+, 2014.

[32] MACHADO, R. D. et al. **Generation of 441 typical meteorological year from INMET stations - Brazil**. Proceedings of the ISES Solar World Congress 2019 and IEA SHC International Conference on Solar Heating and Cooling for Buildings and Industry 2019. **Anais**...International Solar Energy Society, 2020.

[33] SU, F. et al. An evaluation of the effects of various parameter weights on typical meteorological years used for building energy simulation. **Building Simulation**, v. 2, n. 1, p. 19–28, 2009.

[34] HOSSEINI, M.; BIGTASHI, A.; LEE, B. A systematic approach in constructing typical meteorological year weather files using machine learning. **Energy and Buildings**, v. 226, 1 nov. 2020.

[35] PERNIGOTTO, G.; PRADA, A.; GASPARELLA, A. Extreme reference years for building energy performance simulation. **Journal of Building Performance Simulation**, v. 13, n. 2, p. 152–166, 3 mar. 2020.

[36] JI, L. et al. Evaluating approaches of selecting extreme hot years for assessing building

overheating conditions during heatwaves. **Energy and Buildings**, v. 254, p. 111610, 2022.

[37] CRAWLEY, D. B.; LAWRIE, L. K. **Should we be using just "typical" weather data in building performance simulation?** Building Simulation Conference Proceedings. **Anais**...International Building Performance Simulation Association, 2019.

[38] RODRIGUES, E.; FERNANDES, M. S.; CARVALHO, D. Future weather generator for building performance research: An open-source morphing tool and an application. **Building and Environment**, v. 233, p. 110104, 1 abr. 2023.

[39] DÖSCHER, R. et al. The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. **Geoscientific Model Development**, v. 15, n. 7, p. 2973–3020, 2022.

[40] VOLDOIRE, A. et al. Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1. **Journal of Advances in Modeling Earth Systems**, v. 11, n. 7, p. 2177–2213, 2019.

[41] WU, T. et al. BCC-CSM2-HR: a high-resolution version of the Beijing Climate Center Climate System Model. **Geoscientific Model Development**, v. 14, n. 5, p. 2977–3006, 2021.

[42] VANDA, Z.; CAVALCANTE, C.; BARROSO, M. A. B. Arquivos climáticos de Brasília disponíveis no Climate One Building para simulações higrotérmicas. **ENCONTRO NACIONAL DE CONFORTO NO AMBIENTE CONSTRUÍDO**, v. 17, n. 1, p. 1–8, fev. 2023.

[43] AMORIM, A. C.; CARLO, J. C. Análise das propostas de revisão do zoneamento bioclimático brasileiro: estudo de caso de Colatina, ES. **Ambiente Construído**, v. 17, n. 1, p. 373–391, mar. 2017.

[44] SILVA, M.; CARLO, J. **Impacto do uso de diferentes arquivos climáticos na**

classificação de eficiência energética da envoltória de uma edificação segundo a INI-R: o caso de Viçosa-MG. VII Congresso Latino-Americano de Simulação de Edifícios. **Anais**...Florianópolis, SC: Imaginar o Brasil, 2023.

[45] SOUSA, D. L. P. DE et al. 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**, v. 9, n. 3, p. 15554– 01e, fev. 2023.

[46] INSTITUTO NACIONAL DE METEOROLOGIA DO BRASIL – INMET. **Normais Climatológicas do Brasil (1991/2020)**. Brasília, DF, 2022.

[47] TRIANA, M. A.; LAMBERTS, R.; SASSI, P. Characterisation of representative building typologies for social housing projects in Brazil and its energy performance. **Energy Policy**, v. 87, p. 524–541, 2015.

[48] VEIGA, C. Z. M. et al. **Relatório técnico: Criação dos Modelos de Referência de HIS**. Projeto hab.labeee, 2023. Disponível em: <https://hablabeee.ufsc.br/resultados>

[49] MELO, A. P. et al. Development of surrogate models using artificial neural network for building shell energy labelling. **Energy Policy**, v. 69, p. 457–466, 2014.

[50] BECKMAN, W. A. et al. TRNSYS The most complete solar energy system modeling and simulation software. **Renewable Energy**, v. 5, n. 1–4, p. 486–488, 1994.

[51] INTEGRATED ENVRIONMENTAL SOLUTIONS. **Virtual Environment**. , 2024.

[52] U.S. DEPARTMENT OF ENERGY - DOE. **EnergyPlus™ Version 23.2.0 Documentation: Auxiliary Programs**. [s.l: s.n.].

[53] MAY, R. M. et al. MetPy: A Meteorological Python Library for Data Analysis and Visualization. **Bulletin of the American Meteorological Society**, v. 103, n. 10, p.

E2273–E2284, 2022.

[54] U.S. DEPARTMENT OF ENERGY - DOE. **EnergyPlus™ Version 23.2.0 Documentation: Engineering Reference**. [s.l: s.n.].

[55] MEEUS, J. H. **Astronomical Algorithms**. 2. ed. Richmond: Willmann-Bell, Inc., 1998.

[56] LEMOS, L. F. L. et al. Assessment of solar radiation components in Brazil using the BRL model. **Renewable Energy**, v. 108, p. 569–580, 2017.

[57] PEREZ, R. et al. Modeling daylight availability and irradiance components from direct and global irradiance. **Solar Energy**, v. 44, n. 5, p. 271–289, 1 jan. 1990.

[58] OLIVEIRA, P. T.; SANTOS E SILVA, C. M.; LIMA, K. C. Climatology and trend analysis of extreme precipitation in subregions of Northeast Brazil. **Theoretical and Applied Climatology**, v. 130, n. 1, p. 77–90, 2017.

[59] COSTA, R. L. et al. Analysis of climate extremes indices over northeast Brazil from 1961 to 2014. **Weather and Climate Extremes**, v. 28, p. 100254, 1 jun. 2020.

[60] PEREIRA, E. B. et al. **Atlas brasileiro de energia solar**. 2.ed. ed. São José dos Campos: INPE, 2017.

[61] GEIGER, R. **Überarbeitete Neuausgabe von Geiger R., Köppen-Geiger/K-lima der Erde**. GothaKlett-Perthes, , 1961.

[62] KÖPPEN, W. Das geographische System der Klimate. Em: KÖPPEN, W.; GEIGER, R. (Eds.). **Handbuch der Klimatologie**. Berlin: Gebrüder Bornträger, 1936. p. 1–44.

[63] RORIZ, M.; GHISI, E.; LAMBERTS, R. **Bioclimatic zoning of Brazil: a proposal based on the Givoni and Mahoney methods**. (S. S. Szokolay, Ed.)PLEA ’99: Sustaining the future - energy, ecology, architecture. **Anais**...Brisbane, Australia: 1999.

[64] GIVONI, B. Comfort, climate analysis and building design guidelines. **Energy and Buildings**, v. 18, n. 1, p. 11–23, 1992.

[65] SILVEIRA, F. M.; LABAKI, L. C. **Use of natural ventilation in reducing building energy consumption in single-family housing in Brazil**. 2012 International Conference on Renewable Energies for Developing Countries (REDEC). **Anais**...2012.

[66] ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS - ABNT. **NBR 15220-3: Desempenho térmico de edificações - Parte 3 - Zoneamento bioclimático brasileiro e diretrizes construtivas para habitações unifamiliares de interesse social**. Rio de Janeiro: ABNT, 2005.

[67] E SILVA MACHADO, R. M. et al. Bioclimatic zoning for building performance using tailored clustering method and high-resolution climate data. **Energy and Buildings**, v. 311, 15 maio 2024.

[68] MORAIS, J. M. DA S. C.; LABAKI, L. C. CFD como ferramenta para simular ventilação natural interna por ação dos ventos: estudos de caso em tipologias verticais do "Programa Minha Casa, Minha Vida". **Ambiente Construído**, v. 17, 2017.

[69] LOCHE, I. et al. Balcony design to improve natural ventilation and energy performance in high-rise mixed-mode office buildings. **Building and Environment**, v. 258, p. 111636, 2024.

[70] CHEN, Y.; TONG, Z.; MALKAWI, A. Investigating natural ventilation potentials across the globe: Regional and climatic variations. **Building and Environment**, v. 122, p. 386–396, 1 set. 2017.

[71] SAKIYAMA, N. R. M. et al. Natural ventilation potential from weather analyses and building simulation. **Energy and Buildings**, v. 231, p. 110596, 15 jan. 2021.

[72] ANSI/ASHRAE. **STANDARD 55 – THERMAL ENVIRONMENTAL CONDITIONS FOR HUMAN OCCUPANCY**. Atlanta: [s.n.].

[73] PERNIGOTTO, G. et al. **Clustering of European climates and representative climate identification for building energy simulation analyses**. Proceedings of the 16th International IBPSA Conference (Building Simulation 2019). **Anais**...International Building Performance Simulation Association (IBPSA), set. 2019. Disponível em: <http://buildingsimulation2019.org/>

[74] ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS - ABNT. **NBR 15575: Edificações habitacionais - Desempenho**. Rio de Janeiro: [s.n.].

[75] COSTA, M. DA S. et al. Rainfall extremes and drought in Northeast Brazil and its relationship with El Niño–Southern Oscillation. **International Journal of Climatology**, v. 41, n. S1, 30 jan. 2021.

[76] LOPES, A. B. et al. Multiyear La Niña effects on the precipitation in South America. **International Journal of Climatology**, v. 42, n. 16, p. 9567–9582, 30 dez. 2022.

[77] GOZZO, L. F. et al. Climatology and Trend of Severe Drought Events in the State of Sao Paulo, Brazil, during the 20th Century. **Atmosphere**, v. 10, n. 4, p. 190, 9 abr. 2019.

[78] CALVIN, K. et al. **IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.** [s.l: s.n.].

[79] MAZZAFERRO, L. et al. Do we need building performance data to propose a climatic zoning for building energy efficiency regulations? **Energy and Buildings**, v. 225, p. 110303, 2020.

[80] BRACHT, M. K. et al. Multiple regional climate model projections to assess building

thermal performance in Brazil: Understanding the uncertainty. **Journal of Building Engineering**, v. 88, p. 109248, 2024.

[81] CHEN, T.; GUESTRIN, C. **XGBoost: A Scalable Tree Boosting System**. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. **Anais**...: KDD '16.ACM, ago. 2016. Disponível em: <http://dx.doi.org/10.1145/2939672.2939785>

[82] VAN BUUREN, S.; GROOTHUIS-OUDSHOORN, K. mice: Multivariate Imputation by Chained Equations in R. **Journal of Statistical Software**, v. 45, n. 3, p. 1–67, 2011.

[83] KRELLING, A. F. et al. A thermal performance standard for residential buildings in warm climates: Lessons learned in Brazil. **Energy and Buildings**, v. 281, p. 112770, 2023.

[84] PAL, S. K.; MITRA, S. Multilayer perceptron, fuzzy sets, and classification. **IEEE Transactions on Neural Networks**, v. 3, n. 5, p. 683–697, 1992.

[85] BERGSTRA, J.; YAMINS, D.; COX, D. D. **Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures**. Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28. **Anais**...: ICML'13.JMLR.org, 2013.

[86] PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, n. 85, p. 2825–2830, 2011.

[87] MASSEY JR., F. J. The Kolmogorov-Smirnov Test for Goodness of Fit. **Journal of the American Statistical Association**, v. 46, n. 253, p. 68–78, 1 mar. 1951.

[88] ASHRAE. **Guideline14-2023: Measurement of Energy, Demand, and Water Savings**. Atlanta, Georgia: American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), 2023.

Appendix A

In this appendix, the supplementary material for Chapter 2 is presented. The following figures show the results of the maximum and minimum records of DBT, RH, and PPT. They also show the degree days results across the Brazilian territory.

Figure A1. Average annual (a) and trend (b) for the maximum DBT, and average annual (c) and trend for minimum DBT (d). For the Mann–Kendall summary, the gray area indicates no trend, red indicates an increasing trend, and blue indicates a decreasing trend.

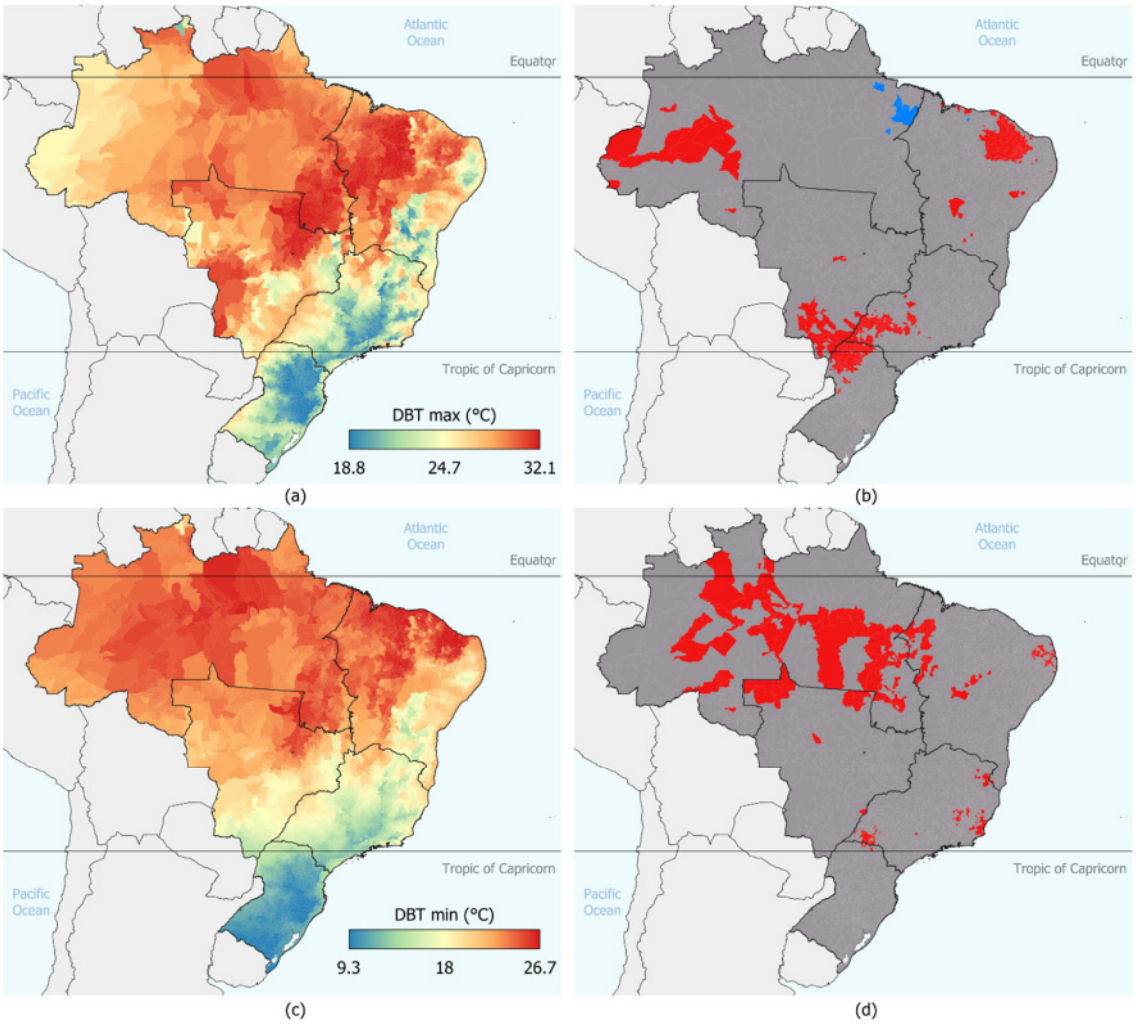


Figure A2. Average annual (a) and trend (b) for the maximum RH, and average annual (c) and trend for the minimum RH (d). For the Mann–Kendall summary, the gray area indicates no trend, red indicates a decreasing trend, and blue indicates an increasing trend.

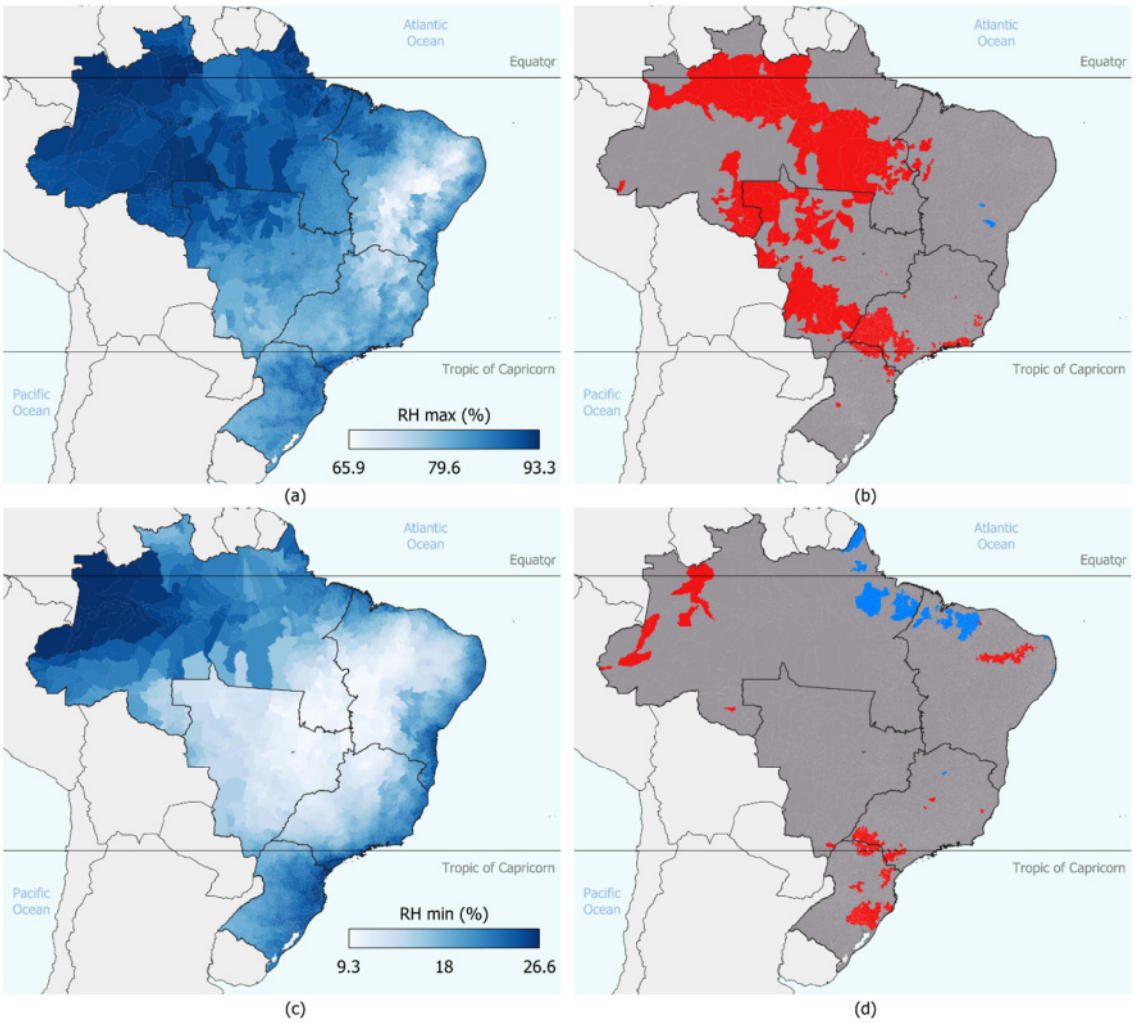


Figure A3. Average annual (a) and trend (b) for the maximum PPT, and average annual (c) and trend for the minimum PPT (d). For the Mann–Kendall summary, the gray area indicates no trend, red indicates a decreasing trend, and blue indicates an increasing trend.

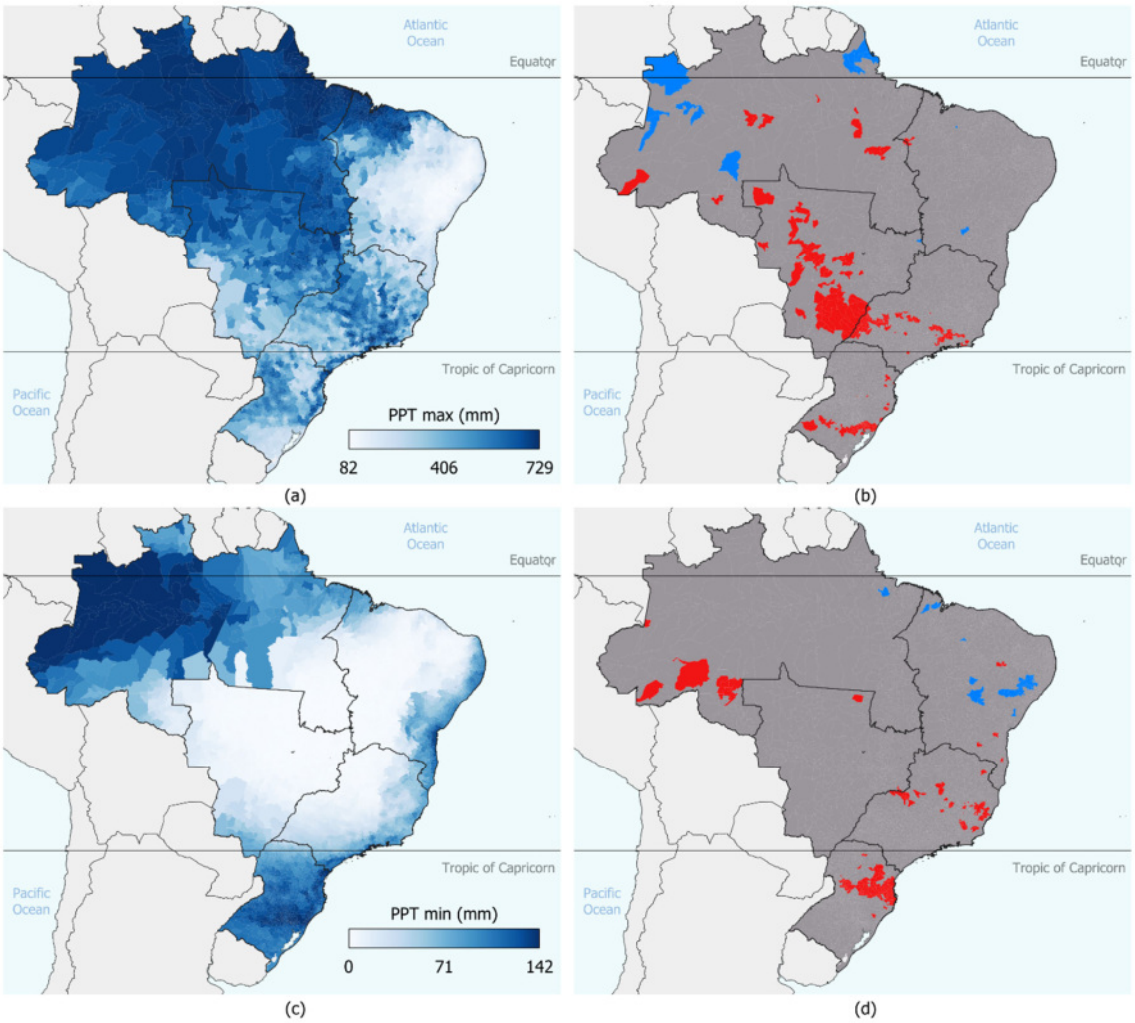
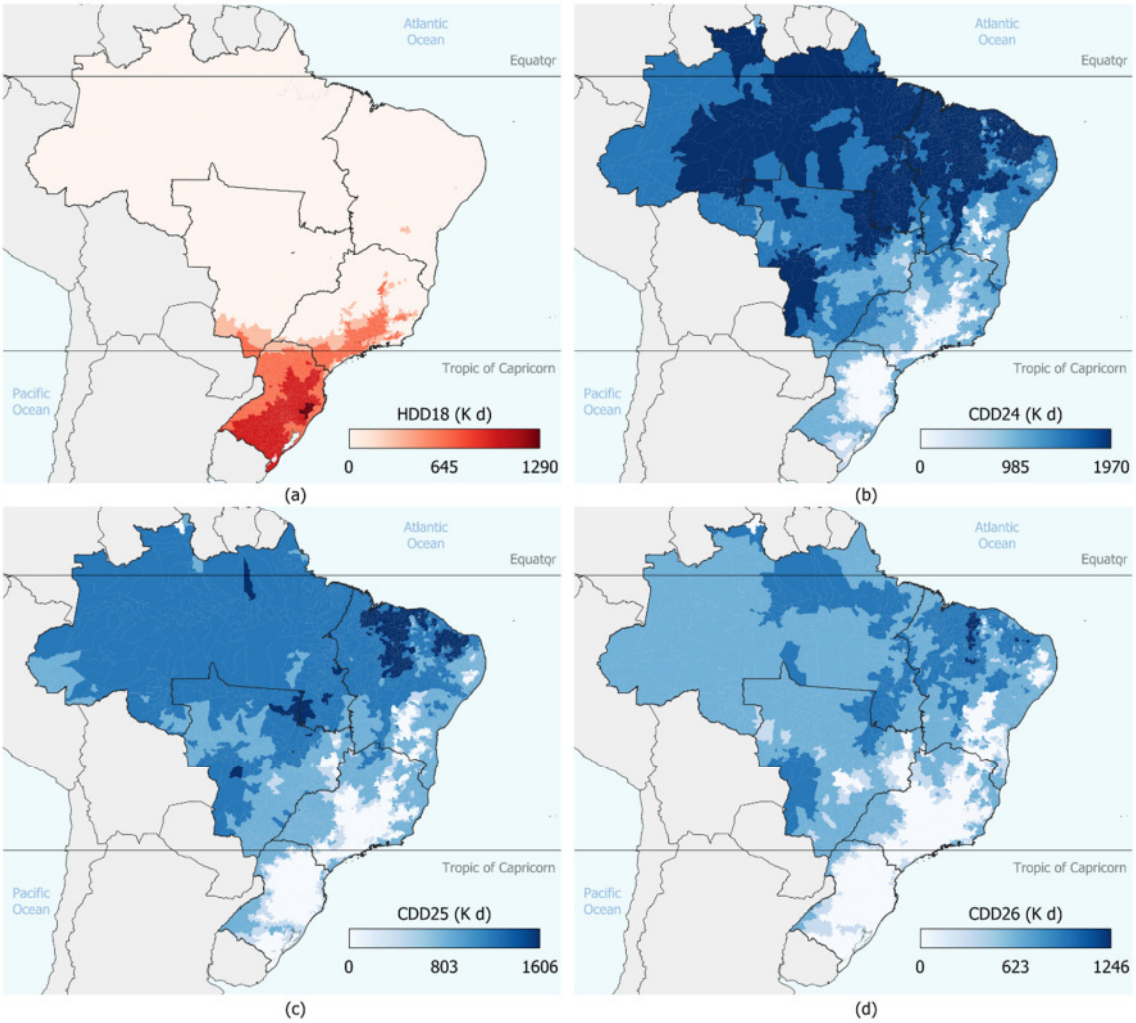


Figure A4. Average annual results for HDD_{18} (a), CDD_{24} (b), CDD_{25} (c), and CDD_{26} (d).



Appendix B

In this appendix, the supplementary material for Chapter 3 is presented. The following figures show the results related to the gaps for different weather variables and linear correlation results between ERA5-Land and INMET weather data.

Figure B1. Number of missing records summarized by year.

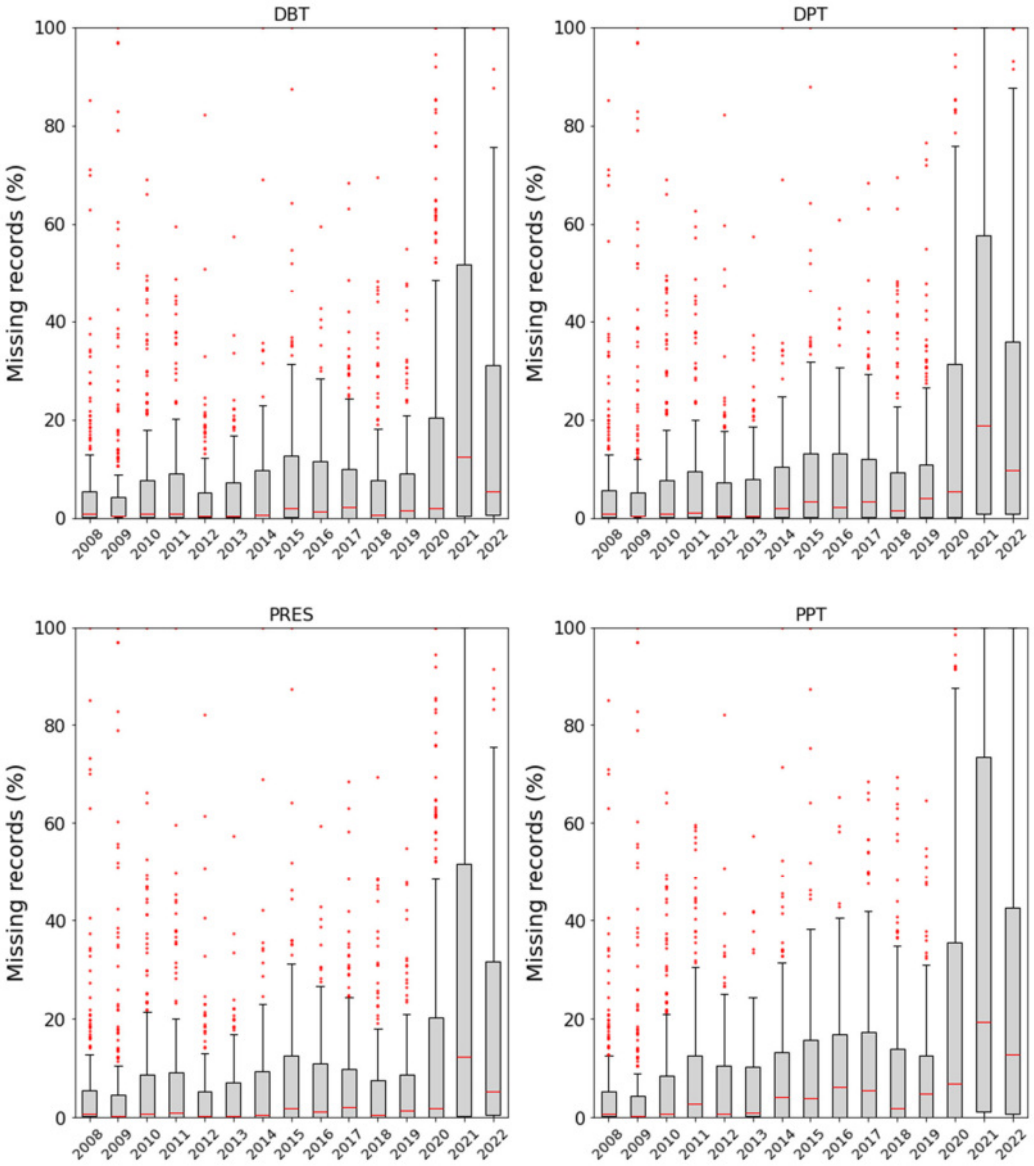


Figure B1. *Cont.*

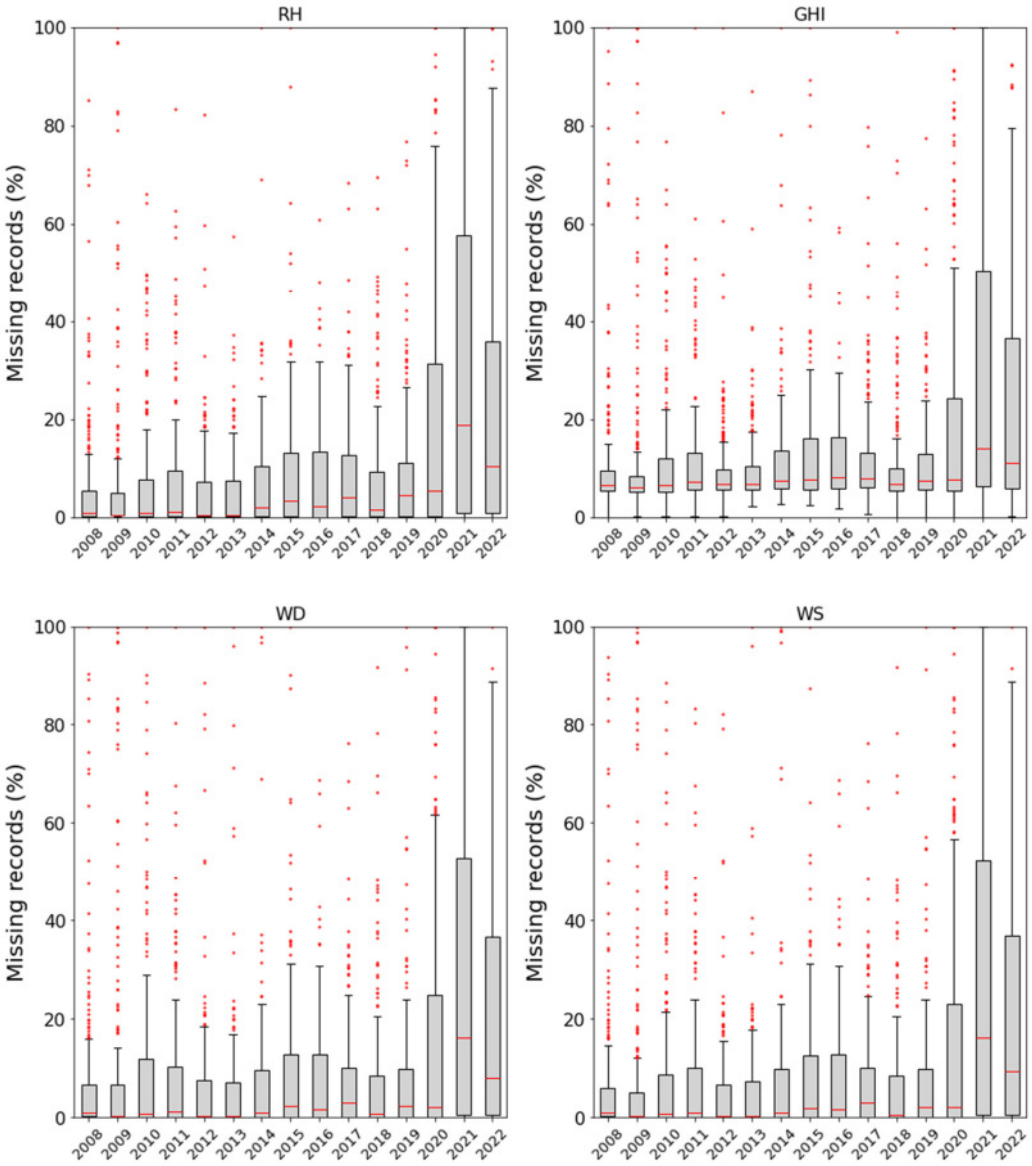
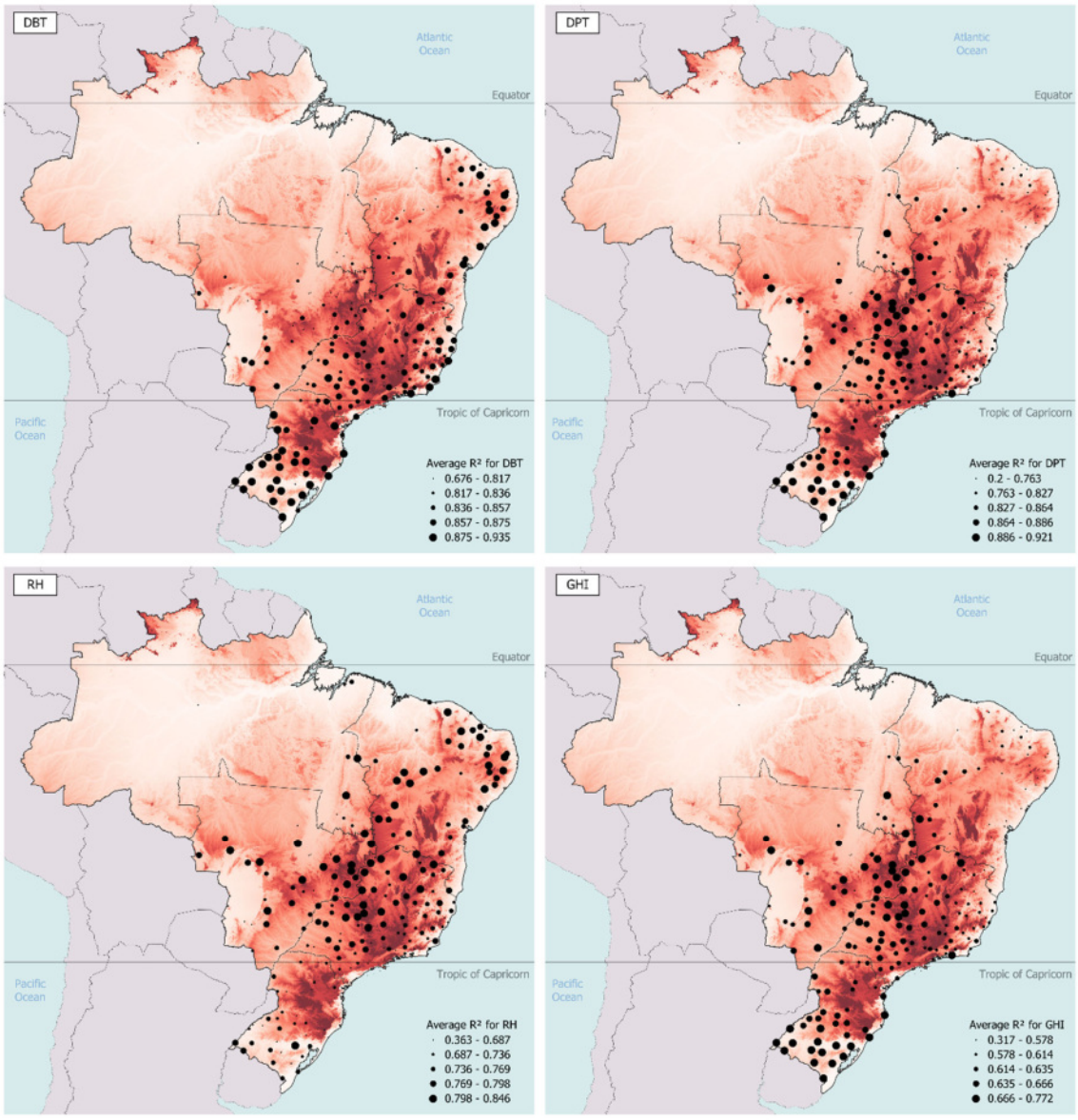
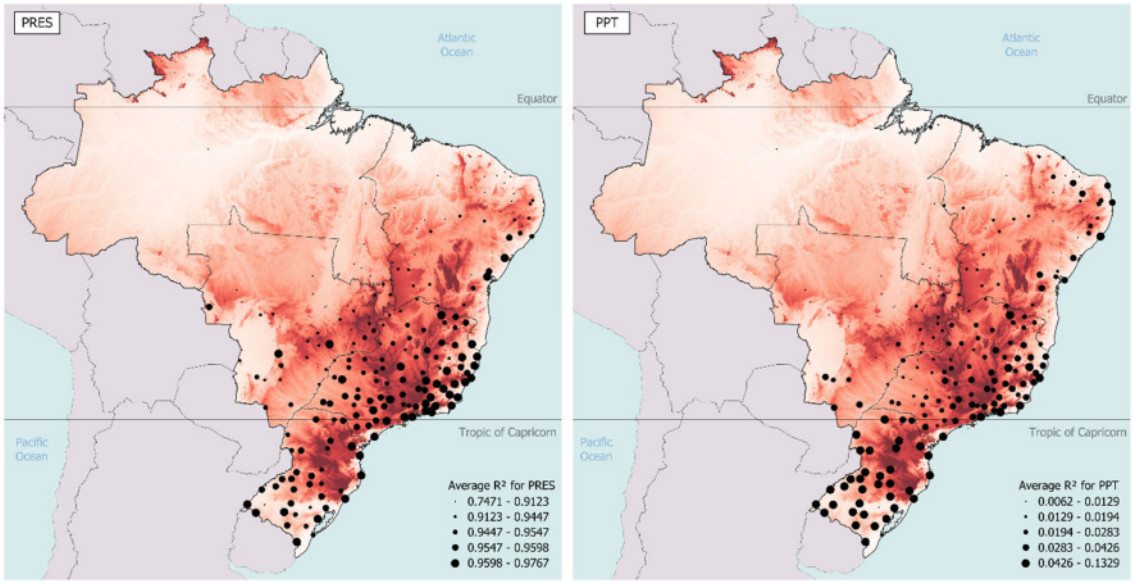


Figure B2. R^2 distribution for DBT, DPT, RH, GHI, PRES, and PPT. The red shading represents a digital elevation model for Brazil, varying from 0 to above 1000 meters.



FigureB2. *Cont.*



Appendix C

In this appendix, the supplementary material for Chapter 4 is presented. The following figures and tables show the results related distribution of the weather variables difference between the 15-year records and the TMY methods, the Spearman correlation, the distribution of the feature importance, and the occurrence of representative years.

Figure C1. Distribution of the difference between the average from AMY and TMY methods for HDH, CDH, DNI, DHI, RHmax, RHmin, and RHmean.

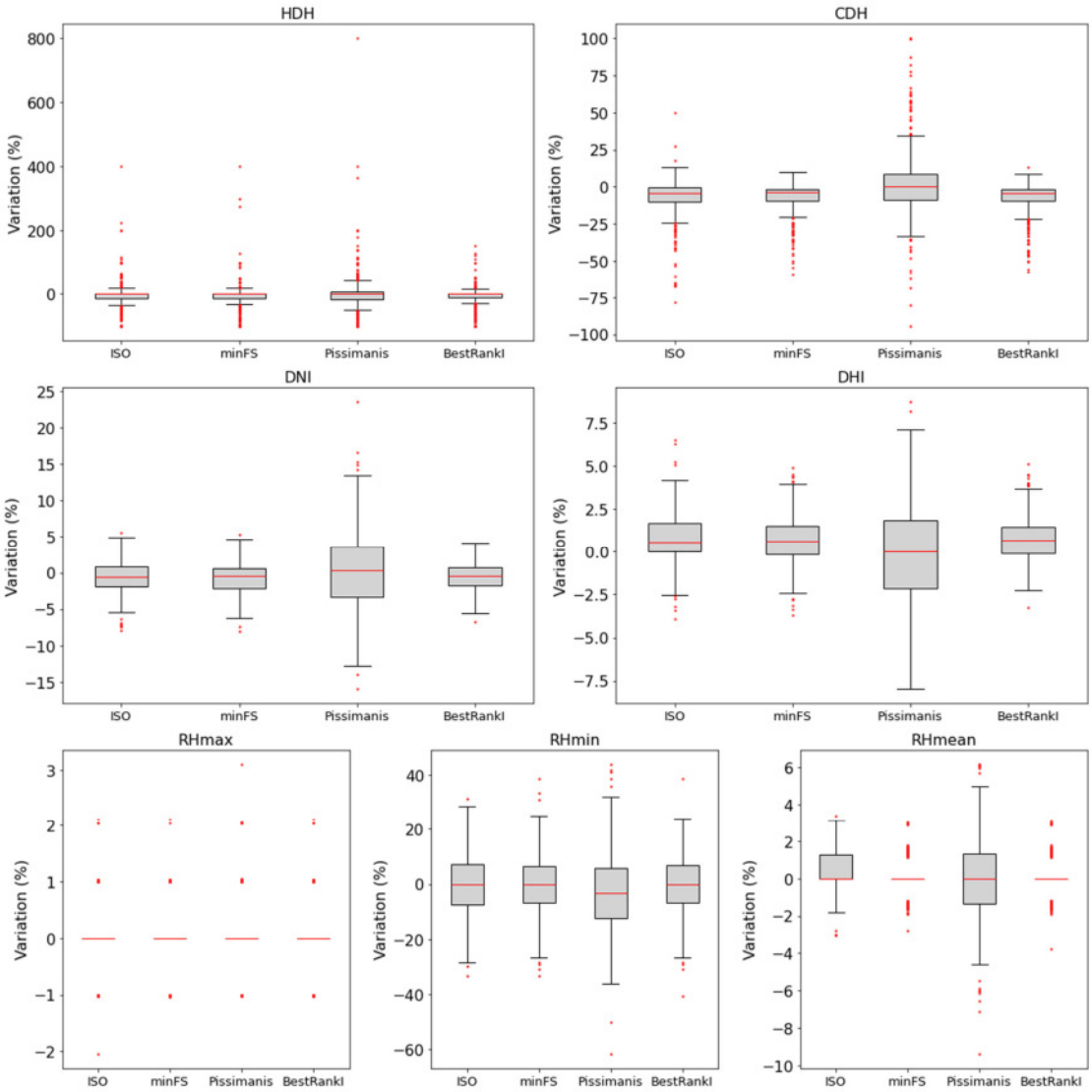


Figure C2. Average Spearman correlation of the weather variables.

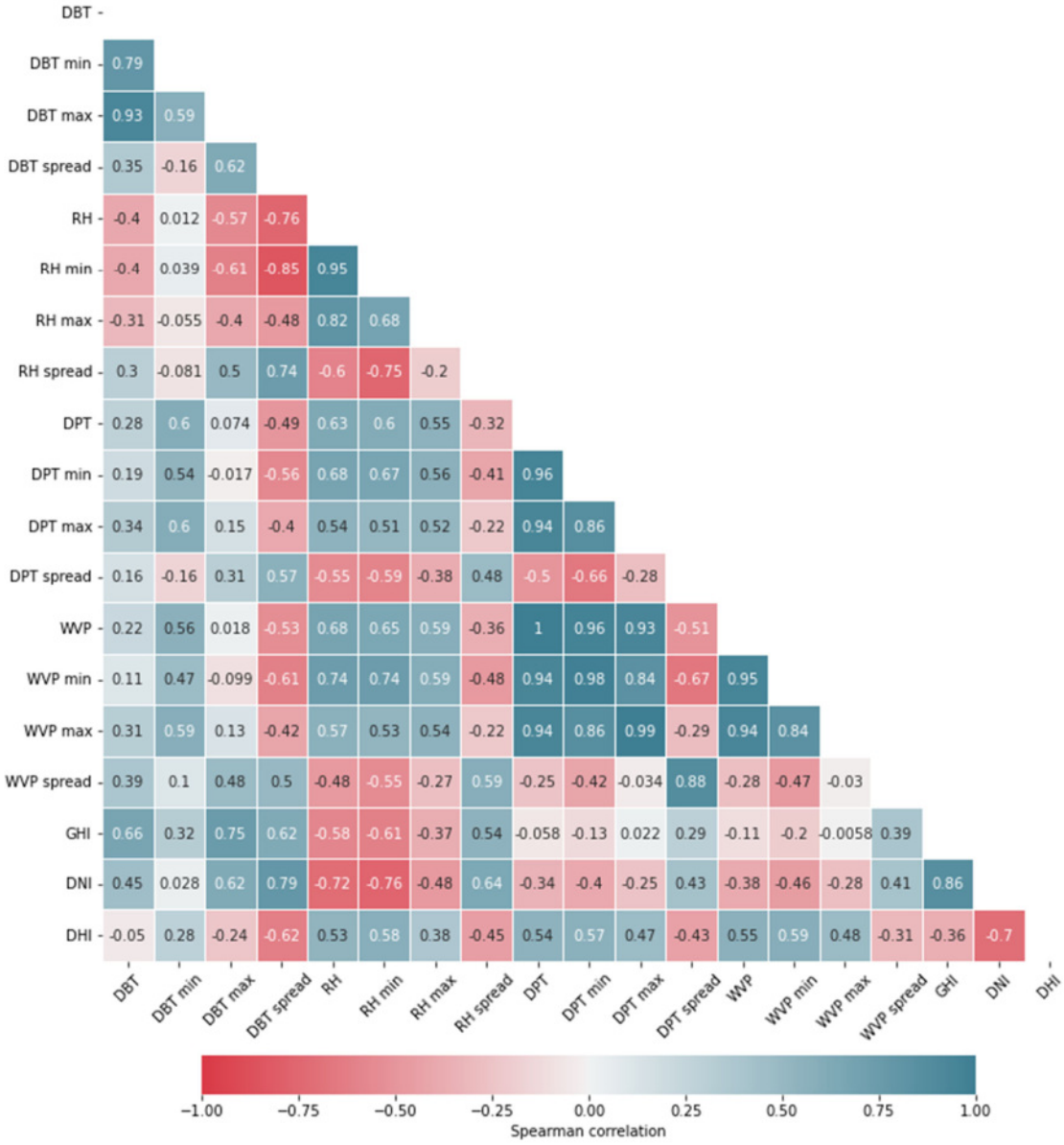


Figure C3. Feature importance distribution for the operative temperature (top) and thermal loads (bottom) machine learning model.

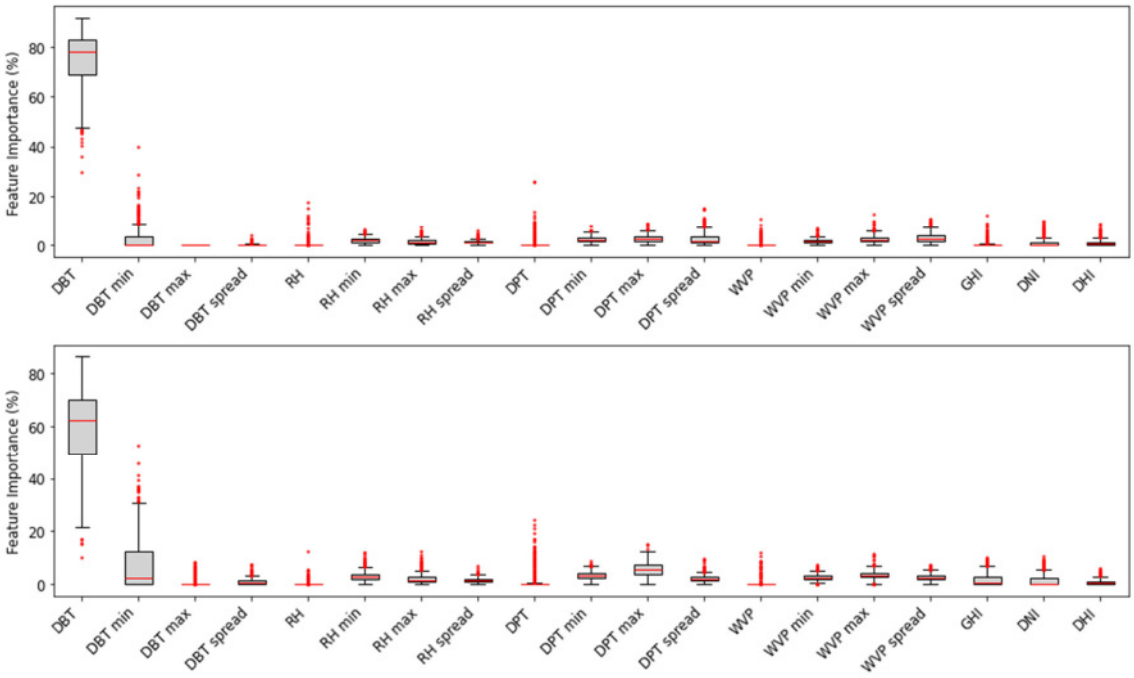


Table C1. Operative temperature based BTMY: Occurrence of reference year for each calendar month.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2008	14	12	6	10	4	9	4	18	27	91	26	19
2009	65	48	60	16	15	1	10	19	16	9	8	35
2010	30	1	20	17	24	27	45	7	58	48	51	44
2011	22	2	19	10	25	37	35	20	15	9	9	8
2012	10	24	25	35	24	23	50	1	83	19	45	4
2013	50	65	31	34	107	8	20	53	44	18	85	39
2014	20	5	38	53	89	57	35	71	69	20	48	28
2015	8	137	41	43	30	41	9	36	35	21	5	20
2016	20	1	7	2	22	14	59	57	12	77	84	42
2017	31	47	61	87	20	40	1	24	5	45	33	120
2018	100	30	39	15	33	38	30	1	28	41	32	44
2019	10	17	70	4	9	18	129	67	14	15	7	27
2020	37	35	18	44	21	24	18	24	35	11	19	33
2021	37	25	22	54	21	106	25	39	2	20	24	7
2022	26	31	23	56	36	37	10	43	37	36	4	10

Table C2. Thermal loads based BTMY: Occurrence of reference year for each calendar month.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2008	14	10	8	5	4	11	1	17	31	71	33	14
2009	72	61	57	11	14	1	3	21	13	7	4	33
2010	29	0	28	21	28	25	39	6	73	42	44	40
2011	23	4	21	7	16	34	36	19	10	15	5	7
2012	8	14	5	32	23	14	52	3	92	16	37	5
2013	40	78	33	39	115	4	25	42	58	20	84	42
2014	13	3	31	69	104	56	28	84	57	12	54	29
2015	4	140	57	40	27	44	14	48	23	15	1	24
2016	17	4	8	3	24	15	55	56	11	86	91	46
2017	30	43	54	99	26	53	1	22	5	59	48	133
2018	127	34	40	11	41	41	21	1	25	28	25	34
2019	6	14	71	1	7	20	144	61	11	28	9	29
2020	36	22	16	41	13	20	23	19	28	12	16	28
2021	33	33	22	50	7	114	26	35	3	17	29	3
2022	28	20	29	51	31	28	12	46	40	52	0	13