Climatic Zoning Methodology
Based On Data-Driven Approach

Leonardo Mazzaferro, Rayner Mauricio e Silva Machado, Ana Paula Melo, Roberto Lamberts
Federal University of Santa Catarina, Florianópolis, Brazil

Abstract
Climatic zoning is largely used by building energy efficiency applications. However, in some cases, climatic zoning is performed by oversimplified methodologies. Further study in climatic variables and building simulation could increase the robustness of climatic zoning methodologies. The aim of this study was to develop a data-driven climatic zoning methodology. The methodology is composed of four main steps. The initial steps were focused on weather data and building simulation, followed by a combined data analysis and a clustering process. Climatic zones validation was performed by cluster quality measures and building simulation results. The proposed methodology was applied in a simplified case study to demonstrate each step. Sensible and latent thermal loads as performance indicators demonstrated the differences in energy needs for each climatic zone.

Introduction
There is a worldwide growing concern about building energy use and its impact on the environment. According to the International Energy Agency (IEA (2018)), global demand for cooling in buildings tends to increase in the next decades. In order to achieve comfortable and energy efficient buildings, it is essential to design buildings according to the climate (Givoni (1992)). Climatic zoning is often connected to performance strategies in building regulations, by generic recommendations and requirements for each climatic zone (Walsh et al. (2018)). The problem with this approach is that building typologies respond differently to climate. To address this problem, building performance analysis could be included in climatic zoning methodologies. Despite its importance for energy efficiency applications, there is no consensus about which climatic zoning methodology should be used among the many available (Walsh et al. (2017)). The most common approach for climatic zoning is the degree-days method, as it is highly correlated with building energy use in cold and mild climates (De Rosa et al. (2015)). The degree-days method does not consider other climatic variables that influence building energy needs in warm or hot climates. The climatic variable relevance strongly depends on the building typology and vary for cold and hot climates. Variables as wind speed, solar radiation and relative humidity can be particularly relevant to building performance in hot climates. Buildings located in hot climates are typically more sensitive to climate effect, because these buildings usually have less envelope insulation than buildings in cold climates (Walsh et al. (2018)). Therefore, the application of oversimplified climatic zoning methodologies in hot climates is even more problematic.

Countries of continental dimensions as Australia, China and India are divided into eight, five and six climatic zones, respectively. Dry and wet bulb temperature, global and diffuse radiation, wind speed and direction were the main climatic variables used to establish those countries climatic zoning (Lam et al. (2006)). Brazilian territory is currently divided into eight climatic zones (ABNT (2003)). Roriz (2012) defined four climatic indicators to propose a new Brazilian climatic zoning with 24 climatic zones. The climatic indicators were: average annual temperature, average annual amplitude, standard deviation of average annual temperature and standard deviation of average annual amplitude (Roriz (2014)).

The optimal number of climatic zones is another critical issue. Higher numbers of climatic zones usually demand a wide range of recommendations, if compared to fewer climatic zones. Too many specific recommendations may create difficulties in regulation acceptance by the construction industry. Furthermore, some typologies are less affected by the climate and do not require a high number of climatic zones. Regarding energy needs, large buildings are intrinsically dependent on internal gains, while small buildings with low insulated envelopes are much more sensitive to external air temperatures and solar radiation. Even more specifically, buildings with highly glazed facades have their performance strongly affected by solar radiation and surroundings. Among the many capabilities of building performance simulation, support for policies such as climate zoning has been gaining relevance in the last decades (Crawley...
Therefore, climatic zoning methodologies may be improved by considering building typologies and by estimating an optimal number of climatic zones. Thus, building energy performance is related to climatic variables in quite complex relations, depending on the boundary conditions. This fact opens possibilities for testing new methods to solve the problem. Deductive (theory-driven) research approaches can be inefficient in solving complex problems (McAbee et al. (2017)). Direct observation of data greatly helps the formulation of a new hypothesis, reinforcing an inductive (data-driven) approach. Data analysis helps to provide a comprehensive climatic understanding for building energy efficiency applications.

This study developed a climatic zoning methodology that includes weather data, building simulation and cluster quality validations to match the climatic zoning purpose. The methodology proposed in this paper focused on the definition and validation of climatic zones. No building design recommendations or passive strategies were addressed in this study.

**Method**

In the proposed methodology, weather data and building energy simulations were analyzed to provide valuable information for the climatic zoning definition and validation. The methodology flowchart is divided into four main steps, as shown in Figure 1.

![Methodology flowchart](image)

**Application**

The methodology first step is the definition of the climatic zoning purpose. A Brazilian climatic zoning for certain conditioned office buildings was obtained, as a methodology demonstration.

**Weather data**

The methodology first step was the data collection from all available EnergyPlus weather files. Weather files are data sets of 8760 hourly values for each climatic variable, that compose a typical climatic year for a given location. An exploratory data analysis was performed on the collected weather files, aiming to extract relevant relationships among the climatic variables or even to discover possible data errors. The adoption of exploratory data analysis tend to facilitate the solution of scientific and engineering problems, by maximizing the value of the data (Jebb et al. (2017)).

In the field of statistics, the exploratory data analysis involves a set of procedures for analyzing data sets (Behrens (1997)). The procedures can be divided into graphical techniques and quantitative techniques (as statistical measures). Statistical measures are descriptive coefficients that summarize a given data set. Therefore, measures of central tendency, dispersion, modality, skewness and kurtosis were obtained for each climatic variable.

Skewness is a measure of the variable distribution asymmetry and kurtosis describes the shape of distribution tails in relation to the distribution overall shape. The bimodality coefficient has a range of zero to one, where a value greater than 0.555 implies that the variable distribution is not normal. A multimodal distribution is a continuous distribution with at least two different modes (Knapp (2007)). Currently, Brazil has 411 locations with EnergyPlus weather data files. Frequency histograms allowed to analyze the cited statistical measures from 411 climates at once.

**Building simulation**

The climate influence on the building performance may vary according to building typology, internal loads, envelope and surroundings. Therefore, building energy simulations must be executed in a reliable energy and load simulation tool to quantify the building performance. The climatic zoning purpose must be defined to establish which building typologies and variations are included in the analysis.

Since artificially conditioned, naturally ventilated and hybrid buildings do not share the exact same relevant climatic variables, the proposed methodology will point to distinct climatic zoning proposals depending on the energy efficiency application purpose. Building thermal loads or energy consumption can be used for conditioned buildings. For naturally ventilated buildings, the percentage of occupied comfort hours can be a proper quantitative performance indicator. Therefore, a proper quantitative indicator must be established to measure the influence of different factors on the building performance.
Sensible and latent cooling thermal loads were the performance indicators of case study models. The temperature setpoint was set on 24°C for cooling. Office building thermal zones were modeled as core and perimeter zones. The office building model had a squared geometry, thus length and width were equal. To reduce outputs variance in the case study, only three office building sizes were defined: small, medium and large office. From small to large office, the building lengths were 15 m, 30 m and 45 m. Those variations were defined to exemplify how a single parameter variation can be relevant to the climatic influence over the building performance. Other office building parameters were fixed, as shown in Table 1.

**Table 1: Building parameters and fixed values.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building Ratio</td>
<td>1.00</td>
</tr>
<tr>
<td>Number of floors</td>
<td>5.00</td>
</tr>
<tr>
<td>Ceiling height (m)</td>
<td>3.00</td>
</tr>
<tr>
<td>People (W/m²)</td>
<td>13.33</td>
</tr>
<tr>
<td>Equipment (W/m²)</td>
<td>10.00</td>
</tr>
<tr>
<td>Lights (W/m²)</td>
<td>10.00</td>
</tr>
<tr>
<td>Infiltration (air changes/hour)</td>
<td>0.30</td>
</tr>
<tr>
<td>Occupancy schedule (hours/day)</td>
<td>8.00</td>
</tr>
<tr>
<td>Wall absorptance</td>
<td>0.50</td>
</tr>
<tr>
<td>Roof absorptance</td>
<td>0.50</td>
</tr>
<tr>
<td>Wall thermal transmittance (W/m²K)</td>
<td>2.34</td>
</tr>
<tr>
<td>Roof thermal transmittance (W/m²K)</td>
<td>2.20</td>
</tr>
<tr>
<td>Wall thermal capacity (kJ/m²K)</td>
<td>180.00</td>
</tr>
<tr>
<td>Roof thermal capacity (kJ/m²K)</td>
<td>200.00</td>
</tr>
<tr>
<td>Window-to-Wall Ratio</td>
<td>0.50</td>
</tr>
<tr>
<td>Glass SHGC</td>
<td>0.87</td>
</tr>
</tbody>
</table>

The surroundings influence was considered by a grid of buildings with the same size than the analyzed building. The office building model was in the grid center, as shown in Figure 2. The surroundings surfaces absorptance and reflectance values were fixed to be equal as the building model and allowed to partially consider the urban climate influence, in the form of shadings and reflections. This type of surroundings modeling did not consider long-wave radiation between building external surfaces, since surface temperatures of surroundings were equal to the air temperature, due to an EnergyPlus algorithm simplification.

**Figure 2: Surroundings grid.**

Building energy simulations were executed in the widely used EnergyPlus software, version 8.9. For wider experiments, more parameter variations are recommended for the building envelope, internal gains and surroundings modeling.

**Combined analysis**

The combined analysis of weather data and building simulation enabled the selection of relevant climatic variables for the energy efficiency application purpose. Correlations between climatic variables and performance indicators must be considered. Correlations tests only between climatic variables were performed to identify redundant information in the weather data and to reduce the number of variables that describe each climate. The general rule for variable selection was: climatic variables highly correlated to the performance indicator and possibly uncorrelated with other climatic variables.

**Clustering process**

Once the climatic variables were selected, it was necessary to scale the data in order to run the clustering algorithm. Clustering is an unsupervised data analysis technique used to discover inherent cluster structures in data, dividing data into groups of similar individuals (Adolfsson et al. (2019)). Clusters can be hierarchical, centroid-based, density-based, distribution-based. The clustering algorithm must be compatible with the data type and application purpose, otherwise the clustering results will not be satisfactory. The k-means clustering algorithm was selected for this methodology, as it generally suits low dimensional numerical data and euclidean distances. Furthermore, the definition of clusters number is usually constrained by a cost versus benefit relation. Normally, more climatic zones guarantee less variation within a single zone, but a high number of zones complicates the climatic zoning application in energy standards and regulations. A higher number of clusters also tend to lower the variance within the same climatic zone. To guarantee a certain cluster quality, a clustering process should include a validation phase. There are many statistical indexes to test cluster quality and determine an optimum number of clusters. Most indexes try to balance within-cluster homogeneity against separation between clusters.

Therefore, a climatic zoning validation was performed using cluster quality measures to guarantee a fair trade-off between the variation in a single climatic zone and the number of zones. The ease of application of climatic zoning in regulations is inversely related to the number of climatic zones. Cluster quality measures were defined according to the climatic zoning previous assumptions. Since the assumption was to group climates with similar variable values, a measure such as the sum of squares (SS) was adopted to estimate the optimal number of clusters. As a final step and assessment, the simulation results were associated with the obtained climatic zoning, aiming to extract the performance indicator typical ranges per climatic zone.
Results

Results section was divided and described in four modules: weather data analysis, building simulation results, combined analysis and clustering results.

Weather data

Following the proposed methodology, a total of 411 Brazilian weather data locations were analyzed. To start the exploratory data analysis, frequency distributions of climatic variables from the 411 Brazilian locations were obtained, as shown in Figure 3.

From the histograms in Figure 3, the locations latitude is uniformly spread, although the same is not true for the locations longitude. Most of the locations are below 500 meters. A predominance of hot climates was noticed in the annual mean dry bulb temperature histogram. Most of Brazilian climates have high values of annual mean relative humidity. The radiation distributions followed normal tendencies. The annual mean wind speed varied between 1 and 3 meters per second for most locations.

Descriptive statistics analysis were applied to the weather data to summarize its main characteristics. Skewness, kurtosis and bimodality coefficients were calculated for climatic variables distributions in all climates. The bimodality coefficient analysis pointed that only 10% of the climates presented bimodality in dry bulb temperature. Climates that presented bimodal frequency distributions were located in the north portion of Brazil, where temperatures do not vary considerably during the year. These results pointed to a predominance of normal distributions in the rest of weather files dry temperature values. On the other side, 75% of the climates presented bimodality in the global horizontal radiation and 73% in relative humidity. The exploratory data analysis provided knowledge about weather data. Also, due to the exploratory data analysis proceedings, an error in data values of a single weather file was identified and properly corrected.

Building simulation

Annual building cooling thermal loads were chosen as indicators to evaluate the climate influence on the building performance. Heating thermal load was not included in the study due to the insignificant heating loads for office buildings in the majority of Brazilian climates. Cooling thermal loads were divided into sensible and latent fractions, to better describe the climatic variables influence in simulation outputs. A boxplot of simulation results per office building size in 411 Brazilian locations was presented in Figure 4.

The building size was quite relevant to the range of annual cooling thermal loads, as shown in Figure 5.
Due to relatively low internal loads in case study buildings, the smallest office was more affected by external temperatures and presented higher cooling values. Cooling loads of the smallest office building also presented more variance, within higher annual mean dry bulb temperature climates. Correlations between annual mean dry bulb temperatures and annual cooling thermal loads were high for the case study office buildings. A simple raise in people density or occupation schedule hours would impact negatively this correlation. Again, the relevance of climatic variables depended on the building typology.

**Combined analysis**

Correlations between climatic variables and building energy simulation indicators were analyzed (in order): latitude, longitude, altitude, dry bulb temperature, dew point temperature, relative humidity, atmospheric pressure, global horizontal radiation, direct normal radiation, diffuse horizontal radiation, wind speed, annual total cooling, annual sensible cooling and annual latent cooling. Figure 6 contains small office building results, where the tendencies were more pronounced. Smaller buildings thermal loads presented higher correlation with climatic variables, as the small office building performance was more sensitive to the climatic influence. As shown in the correlation matrix, high correlations between variables were identified by stretched out ellipses filled with darker colors.

![Figure 6: Climatic variables and simulation results correlation matrix.](image)

Positive or negative correlations are visually distinguishable by the colors blue and red, respectively, or by the ellipse inclination. Considering only climatic variables, the most noticeable positive correlation was between global horizontal radiation and direct normal radiation, followed by the dry bulb temperature and dew point temperature correlation. As expected, altitude and atmospheric pressure negative correlation reached the value of 0.99. The combined analysis focus was on the performance indicator of the case study buildings: the annual thermal cooling loads. The annual total cooling was extremely positively correlated to dry bulb temperature, to dew point temperature and to latitude. Altitude, relative humidity and wind speed were the only three climatic variables inversely correlated to annual cooling thermal load. The relative humidity correlations presented an inverse behavior when compared to annual sensible cooling and annual latent cooling. Global direct radiation was highly correlated to direct normal radiation and partially correlated to diffuse horizontal radiation. Therefore, three climatic variables were selected by its correlation with sensible and latent cooling loads: dry bulb temperature, relative humidity and global horizontal radiation. After the correlation analysis was performed, the tendencies between climatic variables and simulation indicators were mapped in a scatter plot matrix, as shown in Figure 7.

![Figure 7: Climatic variables and simulation results scatter plot matrix.](image)

From the scatterplot matrix, dry bulb temperature was impressively correlated with annual sensible cooling values, but not enough with annual latent cooling values (especially in very hot climates). Climates with lower annual mean dry bulb temperatures presented high values of annual mean relative humidity. As a result of the combined analysis, redundant climatic variables were removed and a few relevant climatic variables were selected for the clustering process: dry bulb temperature, relative humidity and global horizontal radiation. These variables showed high variance for all climates and had a significant impact on case study buildings performance.

Although the usage of mean annual values (of climatic variables) to describe a climate can result in losing details and differences between climates, the considera-
tion of smaller time intervals, as months, complicates the following clustering process. The mean annual values were used for the clustering process.

**Clustering process**

Several numbers of clusters were tested using k-means clustering algorithm in the selected climatic variables. The ideal number of clusters was defined by cluster quality measures. The “NbClust” package for R provided different cluster quality measures to determine the number of clusters by varying all combinations of clusters number, distance measures, and clustering methods (Charrad et al. (2014)). Within-cluster sum of squares ($withinSS$) and Hubert index were the main quality measures analyzed. The $withinSS$ was obtained by summing the squared distance between each data point and the respective cluster centroid. The sum of all within-cluster sum of squares ($withinSS$) resulted in the $tot.withinSS$, also known as total within-cluster sum of squares, i.e. $\text{sum}(withinss)$. The analysis of $tot.withinSS$, proved that a higher number of clusters tend to diminish squared distances between data points and centroids, as shown in Figure 8.

In Figure 9, a knee on the index values (left plot) is equivalent to a peak in the index second differences (right plot) and corresponds to a recommended number of clusters. Considering both cluster quality measures, the optimal number of clusters was defined in 5 climatic zones.

Two-dimensional graphs were elaborated to illustrate the defined clusters in the selected climatic variables mean annual values, as shown in Figure 10.

![Figure 8: Total within-cluster sum of squares.](image)

![Figure 9: Hubert Index.](image)

![Figure 10: Cluster distributions over selected climatic variables mean annual values.](image)
Since each dot represents a single climate in a three-dimensional space, many dots were overlapped. It was observed that climates with high dry bulb temperature values were not always climates with relatively high radiation values. Climatic zones 01 and 03 presented similar global horizontal radiation and relative humidity values, but differed in dry bulb temperatures. The first two climatic zones (01 and 02) presented lower values of dry bulb temperature, compared to the other zones. Analyzing the last three climatic zones (03, 04, 05), notable differences were verified in global horizontal radiation and relative humidity values. All these differences had an impact in the annual building cooling loads, as shown in Figure 11.

Figure 11: Cooling thermal loads distributions per climatic zone.

As described previously, annual cooling thermal loads results of three different sized office buildings were obtained for each climate. An assessment was performed by comparing building thermal loads results within each climatic zone. The cooling thermal load distributions were plotted per zone. The black dot in each distribution represented the mean value. The analysis of building annual sensible and latent cooling loads exposed differences among the climatic zones. The zone number index was designated according to an increasing sequence (from 01 to 05) in annual sensible cooling loads. Consequently, there was a clear crescent tendency in the sensible cooling thermal load as the climatic zone index number increases. Although dry bulb temperatures values were very similar in the three last climatic zones, the highest sensible cooling loads were found in Climatic Zone 05, due to considerably higher solar radiation values. Inversely, latent cooling thermal loads of Climatic Zone 05 are lower than those of Climatic Zone 03 and 04, as the consequence of differences in relative humidity values of each climatic zone.

Oversimplified methods as degree-days are not able to consider this kind of differences. The chosen performance indicators reinforced that conditioned office buildings energy needs were not related to only dry bulb temperature in hot climates. Neither the solar radiation or the relative humidity are considered in degree-days method for climatic zoning. These climatic variables have more impact in hot climates. Conditioned buildings with higher external air infiltration rates and low insulation envelope are some scenarios where the limitations of degree-days method are even more serious. The selection of climatic variables by a combination of weather data and building simulation allowed to cluster climates in a way that reflected the building performance in each climatic zone. Ideally, the more building variables and typologies are considered in the simulations, the broader and more generic the climatic zoning validation will be. However, simulations of one building per typology may be not enough to define a climatic zoning. More simulation variations result in a wider range of possible outputs per climatic zone. Therefore, the wider the application purpose is, the more building typologies and performance indicators must be included in the analysis.

Unfortunately, to develop a climatic zoning for different applications and/or building typologies implies in a trade-off between accuracy and ease of application. Merging different building typologies in the same analysis might result in an underwhelming climatic zoning, but will probably have a much easier implementation in energy regulations, than different climatic zoning proposals (one per building typology). The procedures presented in this study may also inspire studies that involve climate description in other building energy study research areas, such as building energy surrogate models, by providing a method to extract relevant climatic features for the prediction of building performance indicators.

Conclusion

Climatic zoning is frequently present in building energy efficiency applications, but there is great variance on the methods, climatic variables and performance indicators to consider. The choice of those
parameters shall vary according to the weather data and the climatic zoning purpose. Despite the intrinsic complexity, classifications as climatic zoning allow to generalize knowledge and help to ensure a simple communication between performance standards or regulations and the building construction industry. A robust climate zoning substantially reduces building performance discrepancies within climatic zones and is still applicable in energy codes. The use of simulation results, data analysis and clustering techniques greatly supported the development of a modular and versatile climatic zoning methodology. A case study application was entirely performed, considering Brazilian climates and office conditioned buildings. Dry bulb temperature, global horizontal radiation and relative humidity were the selected climatic variables to perform a climate zoning for case study conditions. Sensible and latent thermal loads as performance indicators demonstrated the differences in energy needs for each climatic zone. The proposed methodology (or derivations of it) can be applied in different countries to support building energy efficiency programs.

Acknowledgment

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and by the National Council for Scientific and Technological Development (CNPq).

References


