



A Novel Building Performance Based Climate Zoning for Ethiopia

Belay Zeleke¹, Manoj Kumar¹ and E. Rajasekar^{1,2*}

¹ Department of Architecture and Planning, Indian Institute of Technology Roorkee, Roorkee, India, ² Net-Zed Lab, Indian Institute of Technology Roorkee, Roorkee, India

OPEN ACCESS

Edited by:

Anir Upadhyay,
University of New South
Wales, Australia

Reviewed by:

Sayanti Mukherjee,
University at Buffalo, United States
Sisay E. Debele,
University of Surrey, United Kingdom

*Correspondence:

E. Rajasekar
raj@ar.iitr.ac.in

Specialty section:

This article was submitted to
Sustainable Infrastructure,
a section of the journal
Frontiers in Sustainable Cities

Received: 22 March 2021

Accepted: 05 January 2022

Published: 08 February 2022

Citation:

Zeleke B, Kumar M and Rajasekar E
(2022) A Novel Building Performance
Based Climate Zoning for Ethiopia.
Front. Sustain. Cities 4:684148.
doi: 10.3389/frsc.2022.684148

Climate zoning plays a vital role in the development and implementation of building energy regulations. This paper presents a novel building performance-based approach for climate zoning. By using a high resolution spatial climate dataset, a climate severity mapping of Ethiopia is presented. Ethiopia represents 13 of the Köppen Geiger global climate zones. Real-time thermal performance measurement of representative residential buildings in three climatic locations is presented. Thermodynamic models of these buildings are developed and validated using EnergyPlus software tool. Hourly building simulations of these buildings are performed for 1,490 locations (15 min spatial resolution) across Ethiopia. Cooling and heating discomfort hours, as well as energy performance index, are computed for each of these geolocations. Spatially interpolated building performance metrics and adaptive thermal comfort limits are presented. The relevance of Köppen Geiger's classification in the context of building performance is statistically tested. We observed that the existing climate zones do not considerably represent the building thermal performance and energy footprint. Effect of thermal severity on building performance is described. Further, climate zones are deduced based on the building performance variables using multivariate statistical clustering. In order to evaluate potential climate responsive strategies, bioclimatic zoning of Ethiopia is presented. The bioclimatic zoning is created using an improved Mahoney's method that incorporates solar radiation. This analysis resulted in 21 bioclimatic strategy zones. A comparative assessment of the new climate zoning with Mahoney's bioclimatic classification is presented. The proposed framework will be relevant for regulating building performance and energy conservation measures.

Keywords: climate clustering, k-means algorithm, PCA, degree discomfort hours, energy performance index, Ethiopia, bioclimatic zoning

INTRODUCTION

Climate of a region impacts the thermal comfort and energy demand which are the two prime movers of building performance (Brager and de Dear, 1998). There are several global and local climate-zoning schemes proposed over the years. These are intended for specific themes such as agro-ecology (Hashemi et al., 1981), and building energy performance (Lee and Kung, 2011; Xiong et al., 2019). These themes demand a careful selection of climate variables, zoning methods, and class intervals in the case of supervised classifications.

Climate classification is a necessary precursor tool for energy efficiency programs. This is demonstrated in studies such as Erell et al. (2003), Wan et al. (2010), Bodach (2014), Walsh et al. (2017a,b), Naveen Kishore and Rekha (2018), Attia et al. (2019), Verichev et al. (2019),

and Xiong et al. (2019). It facilitates the design of climate responsive, comfortable indoor environments and reduced energy demand (Markus, 1982; Xiong et al., 2019). Many countries adopt climate classification as a first step to their thermal comfort and building energy efficiency policies, regulations and guidelines. In their recent review, Walsh et al. (2017b) identified as many as 19 climate variables used across 54 countries. Some of these classifications such as the DOE climate zoning (ASHRAE Standard, 1999) are used in the national and international context, while others such as the National Building Code climate zoning (Bureau of Indian Standards, 2016) are tailored for a specific country. Though climate classification is the common theme, climate variables used for classification vary significantly. In addition to climate variables, methods used for zoning also differ.

Conventional methods such as the Köppen Geiger climate classification (Geiger and Pohl, 1954; Kottek et al., 2006) and Thornthwaite's classification (Thornthwaite, 1948) are based on predefined thresholds of climate variables. These thresholds are derived from observed ecological variabilities and can be influenced by subjective decisions of the researcher (Jacobeit, 2010). In recent years the use of multivariate statistical techniques for climate classification is frequented. Multivariate statistical techniques are useful in identifying classes where observation of the phenomena being studied (for instance, building energy performance) is difficult to categorize. These statistical techniques eliminate the subjective nature of threshold based classifications.

One such climate zoning method is cluster analysis (Lee and Kung, 2011; Walsh et al., 2018). Cluster analysis is a statistical method of pattern recognition to categorize data (Bai et al., 2017). It uses various algorithms to detect within-group similarity and between-group differences among data points. Unsupervised data clustering is exploratory (Jain, 2010) and is essential to pick up structures in data where classes are not predetermined (Johnson and Wichern, 2007). Clustering methods can be categorized into hierarchical, partitional, model-based, density-based and grid-based (Saxena et al., 2017). In climate clustering, hierarchical (Anyadike, 1987; Fovell and Fovell, 1993; DeGaetano, 1996; Unal et al., 2003; Iyigun et al., 2013; Xiong et al., 2019) and partitional (Zscheischler et al., 2012; Gao and Malkawi, 2014; Carvalho et al., 2016; Parracho et al., 2016) methods are widely used. Climate data required

for clustering can be either weather station based time-series data recorded over a long period or spatial grid data with average monthly values. In the case of station based time-series data, a variable reduction is performed to simplify the number of inputs and reduce redundancy that might occur through the years. Statistical or empirical methods can be employed for the reduction. For instance, Badraddin (1997) used factor analysis for the climate clustering of Saudi (Arabia and Wan et al., 2010) simplified daily temperature and humidity variables to seasonal discomfort in terms of heat and cold stress values. Though measured weather data is preferable, lack of reliable data with consistent spatial and temporal resolutions necessitate output interpolation to obtain climate zoning. To avoid this problem some studies delineate the climate zones based on administrative boundaries (Walsh et al., 2017a). This works well when geographic extent is small and the spatial climate diversities are minimal (Daly, 2006). Some studies adopt supervised learning algorithms such as discriminant analysis (DA) to interpolate the station data based clustering to a higher resolution spatial setting (DeGaetano, 1996; Chang et al., 2018). However, one of the major drawbacks in DA is that, new classes which are not visible in the original station data based clustering cannot be introduced while increasing the spatial resolution. Therefore, to represent all unique climate zones in the interpolated result the input weather stations should be inclusive (DeGaetano, 1996). On the other hand, the use of high-resolution spatial monthly climate data has shown a better potential for climate zoning applications. For instance, the Köppen Geiger classification was re-created on gridded datasets in various studies (Kottek et al., 2006; Kriticos et al., 2012; Chen and Chen, 2013; Beck et al., 2018). Other applications include thermal comfort design zoning (Pawar et al., 2015), climate zoning (Attia et al., 2019), and bioclimatic classification (Rivas-Martinez et al., 2011; Pesaresi et al., 2014). Gridded spatial data are also used in unsupervised classifications (clustering). Metzger et al. (2013) presented a global bioclimatic zoning using cluster analysis of 30" resolution climate data. Gridded datasets are a better replacement for station data because they offer reduced time step data that cover large geographic areas. However, depending on the number of input layers a variable reduction might still be required.

In this context, this paper presents a multivariate clustering based climate classification using interpolated high-resolution monthly climate datasets. The proposed climate zoning is intended to serve as a basis for future thermal comfort standards and building energy performance guidelines for residential buildings of Ethiopia. The objectives of this study are (1) to create multivariate clustering-based geospatial climate classification of Ethiopia and (2) to verify the robustness of the climate zones through geospatial profiling of comfort and energy performance of residential buildings. Ethiopia is an East African country with climate diversities originating primarily from altitude differences (Fazzini et al., 2015). Existing climate classifications categorizes the country into six agro-ecological zones (Tadesse et al., 2006). Apart from this, the country does not have climate classifications purpose built for building energy performance assessment or building energy standards in general (Iwaro and Mwashu, 2010).

Abbreviations: T_{max} , mean monthly maximum temperature [°C]; T_{min} , mean monthly minimum temperature [°C]; T_{μ} , mean monthly minimum temperature [°C]; ΔT_{out} , monthly mean temperature range [°C]; $T_{maxA\mu}$, annual maximum temperature [°C]; $T_{minA\mu}$, annual minimum temperature [°C]; $T_{A\mu}$, annual mean temperature [°C]; $\Delta T_{A\mu}$, annual mean temperature range [°C]; $RH_{A\mu}$, annual mean relative humidity [%]; RH_{μ} , monthly mean relative humidity [%]; IA_{μ} , annual mean solar radiation [kWh/m²]; I_{μ} , monthly mean solar radiation [kWh/m²]; T_n , neutral temperature [°C]; WWR, window-to-wall ratio [%]; U-Value, thermal transmittance [W/m²K]; Kappa, thermal mass [kJ/m²K]; DDH_{heating}, heating degree discomfort hours; DDH_{cooling}, cooling degree discomfort hours; EPI, energy performance index [kWh/m²/y]; RMSE, root mean square error; CH index, Calinski-Harabasz pseudo-F-statistic; PC, principal component; PCA, principal components analysis; SST, total sum of squared differences; SSE, sum of the squared errors; CZ, climate zone.

METHODS AND MATERIALS

The study comprises of four major parts: (a) climate severity of Ethiopia is analyzed and mapped. (b) Thermal performance evaluation of representative houses is performed through field measurements. (c) Cooling and heating discomfort hours (DDH), as well as energy performance index (EPI), are computed for these representative residential buildings. This is realized through hourly energy simulation of these buildings across 1,490 sample locations gridded approximately at 27 km × 27 km interval. (d) The climate of Ethiopia is clustered using the k-means clustering method that includes k-means++ for initial cluster centroid selection. Clustering involves variable selection, standardization and variable reduction using principal components analysis (PCA). Clustering is performed on extracted component scores (PCs). The accuracy of clustering is tested through Calinski-Harabasz (CH) pseudo-f-statistic index. Similarities and separations of DDH and EPI in the climate clusters are statistically tested. (e) In order to evaluate potential climate responsive strategies, bioclimatic zoning of Ethiopia is presented. The bioclimatic zoning is created using an improved Mahoney's method that incorporates solar radiation. A comparative assessment of the new climate zoning with Mahoney's bioclimatic classification is presented. A summary of passive design strategy recommendations applicable to each climate zone is created. Additionally, the robustness of the proposed zones to climate change is discussed considering representative locations. A Step-by-step methodology of the study is presented as follows.

Data Selection and Processing

Variable Selection

Current climate classifications in Ethiopia are either based on altitude (the traditional agro-ecological climate classification), or a combination of temperature and rainfall (Köppen Geiger climate classification). In this study, climate variables that influence building thermal performance are given precedence. A raster data of 30" resolution is sourced from WorldClim repository (Fick and Hijmans, 2017) for temperature and solar radiation variables (Fick and Hijmans, 2017). RH_μ data in 10' resolution is obtained from CRU CL v. 2.0: (A high-resolution data set of surface climate over global land areas) (New et al., 2002). The confidence level of climate variables (T_{max}, T_{min}, T_μ, RH_μ, and I_μ) are iteratively tested. Pilot clusters are developed using Iso-cluster unsupervised classification in combination with the maximum likelihood classification. Various climate variable combinations are considered and the confidence level of variables evaluated. The climate clusters developed using T_{max}, T_{min}, RH_μ, and I_μ yielded the best confidence levels. The clusters developed with T_μ fail to differentiate moderate locations with low diurnal temperature range and hotter locations with high diurnal temperature ranges.

Data Pre-processing

Relative humidity datasets are interpolated into a high-resolution raster dataset from a 10-min resolution ASCII file. Since the units of each variable group vary, the input data is standardized to (z)

data ranging from a minimum value of (0) and a maximum value of (1) using the following equation.

$$z = \frac{(x - \text{old min}) \times (\text{new max} - \text{new min})}{(\text{old max} - \text{old min})} + \text{new min} \quad (1)$$

The standardization is based on the assumption that all variables contribute equally to the variance inherent in the clusters (Fovell and Fovell, 1993). The resulting standardized variables are combined in a composite band raster dataset containing 48 bands. Bands 1–12 represent standardized scores of T_{max} from January to December. Bands 13–24 contain standardized T_{min}, bands 25–36 represent standardized monthly RH_μ and bands 37–48 represent standardized monthly I_μ values. These variables are exported as a point data for further analysis.

Variable Reduction

Multivariate data exhibits complex relationships among different variables. A similarity check is performed using Pearson's correlation to evaluate redundancy among input variables. A correlation matrix reveals a significant relationship among multiple variables (result not included). This correlation indicates that a variable reduction is necessary before cluster analysis. This study employs PCA as the dimension reduction technique. PCA is a multivariate statistical technique (Wilks, 2011) that creates uncorrelated variables from the linear transformation of original variables and a transposed eigenvector matrix. Given a data matrix X (n locations and p variables), PCA reduces p variables X₁, X₂, ..., X_p into fewer (k) principal components while retaining as much information as the original p variables in reproducing the total system variability (Johnson and Wichern, 2007). The estimation of principal components or non-correlated linear transformations Y = (Y₁, Y₂... Y_p) for a random vector X' = (X₁, X₂... X_p) having covariance matrix Σ or correlation matrix ρ, eigenvalues λ_i, (i=1, 2, ... p) and eigenvectors e_i is given as (Johnson and Wichern, 2007):

$$Y_p = e_{p1}X_1 + e_{p2}X_2 + \dots + e_{pp}X_p \quad (2)$$

The first few PCs with the highest variance are termed as principal components. PC1 is constrained to the condition e₁^T e₁ = 1, while PC2 is constrained to the condition covariance (PC1, PC2) = 0 (Ramos et al., 2017) and e^T denotes transpose of the eigenvector matrix.

Climate Severity Assessment

climate severities in the country are investigated through annual summaries of T_μ, RH_μ, and I_μ. Additionally, derived variables such as monthly average diurnal temperature range (ΔT) is also used.

Real-Time Thermal Performance Measurement of Representative Residential Buildings

Indoor climate data consisting of air temperature and relative humidity are collected from sample residential buildings in three cities (Addis Ababa located at 8.9806° N, 38.7578° E, Bahir Dar

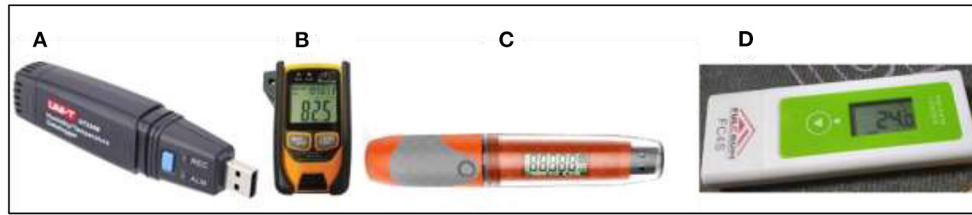


FIGURE 1 | Temperature and humidity data loggers used for indoor measurements. **(A)** UNI_T temperature and humidity data logger, **(B)** HTC Easy Log humidity/temperature data logger, **(C)** Elitech RC-51 High Accuracy USB Temperature Data Logger, and **(D)** Fulcrum temperature and humidity pdf data logger.

at 11.5742° N, 37.3614° E, and Nekemte at 9.0893° N, 36.5554° E). These cities are located at an elevation of 2,355, 1,800, and 2,088 m above sea level.

The measurements are taken at a sub-hourly interval of 10 min. In addition to indoor measurements, one data logger is deployed outdoors to make direct comparisons of indoor and outdoor conditions. Thermal performance is analyzed by investigating temperature (T), relative humidity (RH), temperature difference (T_{diff}), diurnal temperature range (ΔT), and thermal damping (T_d). **Figure 1** shows the devices used for the assessment. The third logger (1c) only measures temperature while the others measure both temperature and humidity.

All software used to handle modeling, simulation, and post-processing of climate data are either free or are licensed to IIT Roorkee. ArcGIS 10.5 software licensed to IIT Roorkee is used to process spatial data.

Representative Houses

In Addis Ababa and Bahir Dar, the selected typology is multi-family low-rise condominiums. The height of these buildings vary from two to seven stories. Each flat's functional designs share similarities in the size of rooms, openings, and circulation areas. **Figure 2A** presents the plan layout of a typical condominium block in Addis Ababa. This building contains three of the houses (AA_H1, AA_H5, and AA_H6) in the field study.

Figure 2B presents a photograph of the building discussed above (**Figure 2A**). This photograph is included here for visualization purposes and does not represent all houses surveyed.

Statistical Multivariate Clustering

Unsupervised clustering algorithms are broadly divided into hierarchical, and partitional (Jain, 2010). Hierarchical methods create nested clusters either by using agglomerative (AHC) or divisive (DHC) techniques. The k-means algorithm is one of the widely used partitional clustering techniques. Compared to other methods the k-means method performs well for large datasets (Kidson, 2000). The k-means clustering algorithm divides n points in p dimensions into K clusters through minimizing the within-cluster sum of squares (Hartigan and Wong, 1979). It works by combining two steps: first the initial cluster centroids are estimated, and then the distance between cluster centroids and data points are calculated (Malinen et al., 2014). Cluster centroids are moved each time a new iteration is run until

the centers become stationary. The first seed location is taken randomly while the rest are taken by using an optimized initialization technique also known as k-means++ (David and Vassilvitskii, 2006) which ensures separation between seeds in the attribute space. Multiple iterations are tested before final decision is made. The algorithm uses Euclidean distance to group data points (Arora et al., 2016). Euclidean distance estimates the distance between object (i) and group centroid (C) as:

$$d = \sqrt{\sum_{i=1}^v (x_i - x_C)^2} \quad (3)$$

Multivariate clustering results in varying clusters even while using the same data due to the randomness of initialization. A measure of effectiveness is required to ensure the accuracy of the clustering.

Cluster Evaluation

There is no unified technique to ascertain the accuracy of unsupervised classifications. The difficulty arises from the variations in the clustering algorithms, methods, and input variables (Kumar et al., 2006). Another reason is that in unsupervised techniques there is no prior knowledge about the type and number of classes. In this work, accuracy of clustering is evaluated using two methods: pseudo-f-statistic is used to solve the number of groups problem, and building performance assessment is used to evaluate the clustering robustness.

Pseudo-F-Statistic

CH pseudo-F-statistic is an indicator to measure the grouping effectiveness and suggest optimum number of clusters (Caliński and Harabasz, 1974). It is a ratio reflecting within-group similarity and between-group difference. Higher pseudo-f-statistic values indicate ideal number of classes. The ratio is expressed as:

$$CH = \frac{(R^2/n_C - 1)}{(1 - R^2)/(n - n_C)} \quad (4)$$

Where:

$$R^2 = \frac{SST - SSE}{SST} \quad (5)$$



FIGURE 2 | A typical floor layout of a condominium block in Addis Ababa. This building contains studio type apartments (light blue), one-bed type apartments (light green), and two-bed type apartments (light yellow). **(A)** Floor plan, **(B)** Photograph of the building (Photo by Mekides Worku).

SST measures the total sum of squared differences, and SSE represents sum of the squared errors in i^{th} group, reflects within-group similarity.

$$SSE = \sum_{i=1}^{nc} \sum_{j=1}^{ni} \sum_{k=1}^{nv} (v_{ij}^k - \overline{v_i^k})^2 \tag{6}$$

where, n is the number of features. The number of features in group i is given as n_i , while n_c stands for the number of classes (groups), and n_v is the number of variables used to group features. v_{ij}^k represents the value of the k th variable of the j th feature in the i th group, and $\overline{v_i^k}$ is the mean value of the k th variable in group i . SST is also calculated using a similar equation but

substitutes the mean value of the k th variable, (\bar{v}^k) instead of v_i^k .

Building Energy Performance Assessment

In recent years, building energy performance is being used as an indicator to evaluate the relevance of climate zones. One such attempt uses the mean percentage of misclassified areas (MPMA) to identify overlap of climate zones based on discomfort hours (Walsh et al., 2018). This method demonstrated good results when the number of classes is few (2-4 zones) and the geographic area is small. In a similar note, Xiong et al. (2019) used building heating and cooling load to verify a hierarchical climate zoning. In this study, validation of climate zoning is done through a statistical comparison of heating and cooling DDH, and EPI. Two residential units from the Integrated Housing Development Program of Ethiopia are modeled using energyplus software tool. The first (case 1) is a one-bed room condominium house. The second (case 2), is a single room studio. Two alternatives a low mass (case 1a and 2a) and a high mass (case 1b and 2b) are generated. The low mass buildings are modeled according to the existing specifications of the integrated housing development program of Ethiopia (Table 1). High mass building is generated by modifying kappa value (κ) of walls without altering the U-value and reducing the window-to-wall ratio (WWR) as shown in Table 1. Hourly simulations are performed at 1,490 locations across Ethiopia with an approximate grid interval of 27 x 27 km. The hourly weather data for simulations are synthesized using Meteororm software tool verified with available ground station data. For each location, simulations are performed for four cases (cases 1a, 1b, 2a, and 2b) operated in free running (natural ventilation) and mixed mode scenarios. This resulted in 11,920 simulation runs. DDH and EPI are extracted from the simulation output using a python post processing script developed by the

authors. The results are interpolated using the geospatial analyst tool in ArcMap 10.5.

The discomfort degree hours are estimated as the difference between the adaptive comfort neutral temperature and the zone air temperature. Neutral temperature is calculated from the running mean outdoor temperature through Nicol and Humphrey's equation (Nicol and Humphreys, 2010). Upper and lower comfort temperature limits are determined by adding $\pm 3^\circ\text{C}$ for moderate expectation, as demonstrated by Rajasekar et al. (2015). The whole process was automated using a python post-processing script to reduce errors. The cooling DDH is the cumulative fraction of the inside air temperature that is more than adaptive comfort neutral temperature. Heating DDH is the cumulative fraction of indoor temperature that is less than the adaptive comfort neutral temperature. EPI is the annual energy consumption of a building per unit area in kWh/m²/year (BEE, 2017).

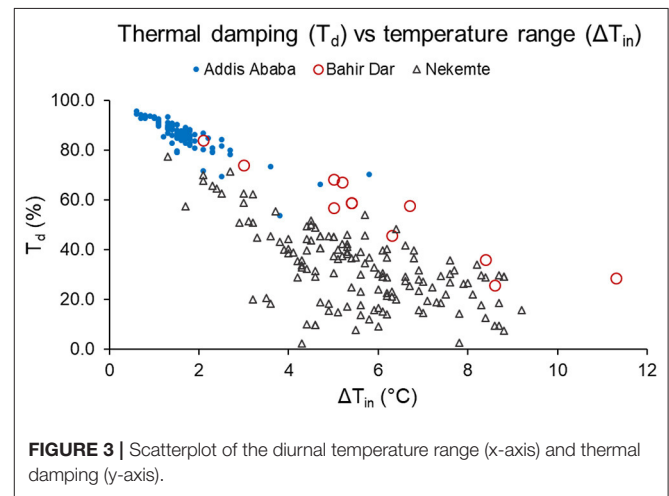


TABLE 1 | Details of building typology selected for analysis.

Building details	Case 1a	Case 1b	Case 2a	Case 2b
U –Value (W/m ² K)	1.78	1.75	1.78	1.75
Wall Thickness (mm)	225 mm	325 mm	225 mm	325 mm
WWR (%)	25%	10%	25%	10%
Kappa (kJ/m ² K)	74	217	74	217
Shade	no	0.5 m	No	0.5 m
Natural Ventilation conditions				
Ventilation- Indoor set point	24°C	24°C	24°C	24°C
Ventilation-Minimum Outdoor Temperature	18°C	18°C	18°C	18°C
Ventilation-Maximum Outdoor Temperature	35°C	35°C	35°C	35°C
Delta T	0°C	0°C	0°C	0°C
Natural ventilation mode	Temperature Controlled			
Air conditioning operation schedule				
Bedroom	22:00 to 06:00 h	22:00 to 06:00 h	-	-
Living	06:00 to 22:00 h	06:00 to 22:00 h	-	-
Studio	-	-	24 h	24 h

Floor and roofs are modeled as adiabatic, while occupancy is kept similar for all cases.

Bioclimatic Zoning

Bioclimatic zoning is performed using Mahoney’s method. This method takes monthly temperature, humidity, and precipitation data to analyze thermal severity, identify remedial actions (indicators), and recommend passive building design strategies. This method is suitable for applications on high-resolution monthly data.

Uncertainties in Multivariate Clustering

Uncertainties in cluster analysis can occur due to a combination of factors. Among these are uncertainties due to methodological, latent, and information biases (Fovell and Fovell, 1993). The

methodological biases are discussed in section Data Selection and Processing above. Measurement related errors are also dealt with through standardization of the variables in the data pre-processing stage.

Data Related Errors

Two types of climate data are used in the study. The first type of data is a monthly spatial raster and vector data. Root mean square error (RMSE) of T_{max} is 1.29°C, while T_{min} has an RMSE of 1.39°C, and I_{μ} 1.45 MJ/m². Errors associated with RH is reported as the square root of generalized cross validation (RTGCV). Over the domain of North Africa including Ethiopia, monthly RTGCV of RH is reported between 4.6 and 5.6% (New et al., 2002). Further interpolation of RH from 10’ to 30” spatial resolution results in RMSE below 0.6%. The second type of data is TMY datasets synthesized from Meteonorm for hourly simulation of building performance. Meteonorm uses a combination of interpolation and stochastic models (Meteonorm, 2018). Errors of interpolation associated with the yearly means of these datasets, as given in the Meteonorm handbook II are discussed here. Cross-correlation of measured and interpolated values is used to calculate the RMSE for individual variables. Accordingly, relative RMSE of I_{μ} over Africa is reported as 7.4% while RMSE of temperature is estimated at 1.8°C.

RESULTS

Climate Severity

In Ethiopia, $T_{A\mu}$ varies from 12 to 30°C. $RH_{A\mu}$ varies from 45% to 76%. $I_{A\mu}$ varies between 4.8 KWh/m²/day and 6.4 KWh/m²/day. The central parts of the country are characterized

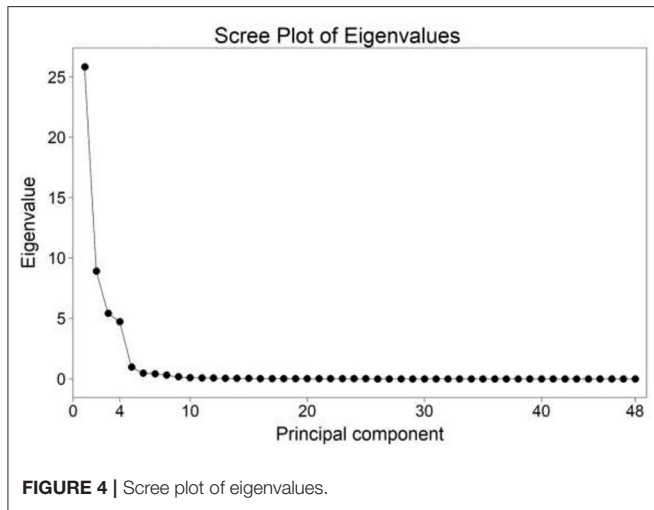


FIGURE 4 | Scree plot of eigenvalues.

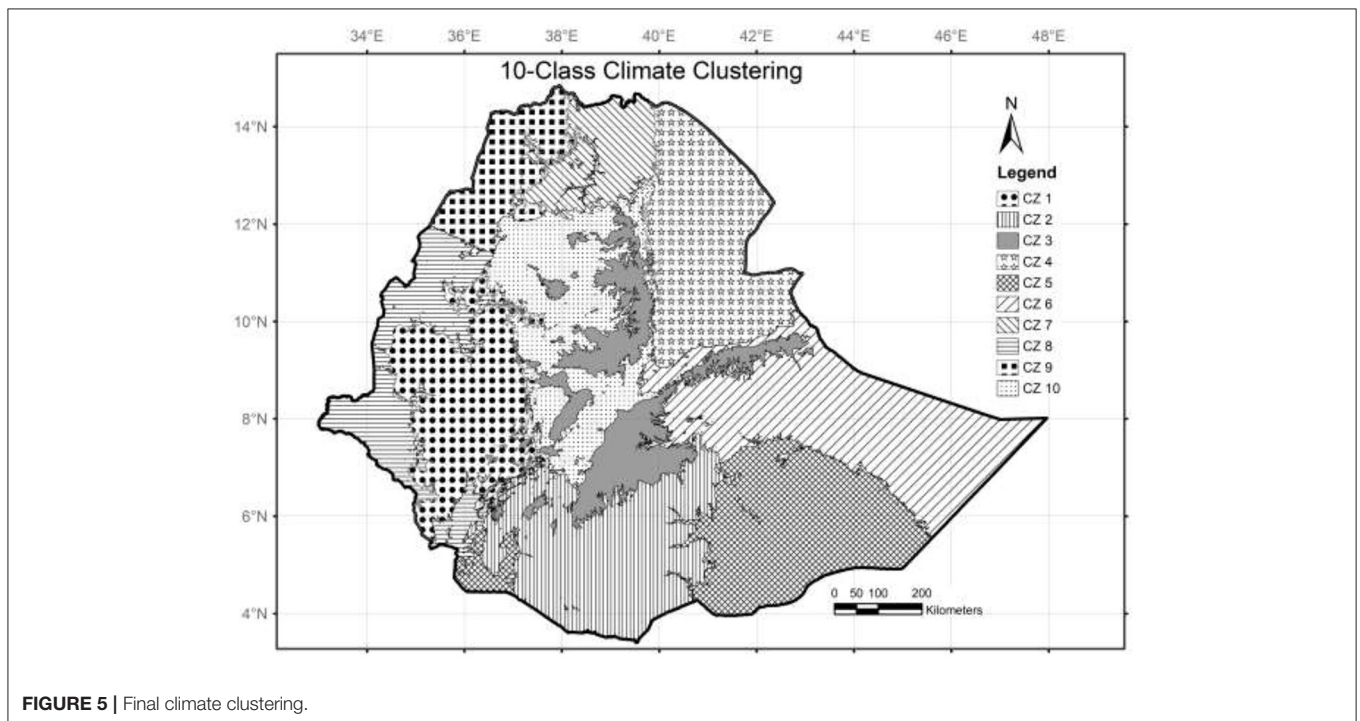


FIGURE 5 | Final climate clustering.

by highlands and plateaus resulting in lower temperatures. An exception to this is the central rift valley that intersects the country in a northeast to southwest axis. Regions in the rift valley exhibit warm to hot temperatures with varying humidity. The highest values of $I_{A\mu}$ occur in the eastern and northern parts of the country.

The variations of temperature range (ΔT) indicate both spatial and temporal differences (map not included). Some parts of Tigray and Amhara regional states exhibit high levels of ΔT ($>15^\circ\text{C}$) in winter as well as summer. Some climate severity indicators like the Mahoney’s method consider a ΔT that is $>10^\circ\text{C}$ as high. Accordingly, only small parts of the country in the Northeast indicate low levels of temperature swings in January and western and southwestern regions in June.

Thermal Performance Measurement of Representative Residential Buildings

The types of houses studied in Addis Ababa, and Bahir Dar share common characteristics. For one, these houses are public condominiums constructed under the IHDP scheme. These buildings have similarities in design, construction materials, and overall layout of the buildings themselves. The site settings of these buildings also share a similarity. Both sites are located in outbounds of the respective city and have an open plan with ample spacing between buildings allowing for unobstructed access to sunlight and air movement. Keeping these factors in

mind, a direct comparison of thermal characteristics between T_{diff} , ΔT , and T_d of all houses in these locations is made.

A general comparison shows the variation in thermal performance across cities and similarity within a given city. A thermal damping comparison between the three locations indicate that there is a visible difference between T_d in all three locations. In Addis Ababa, the mean T_d is 85%, while T_d in Bahir Dar is near 60%. Nekemte houses have the lowest T_d at $< 40\%$.

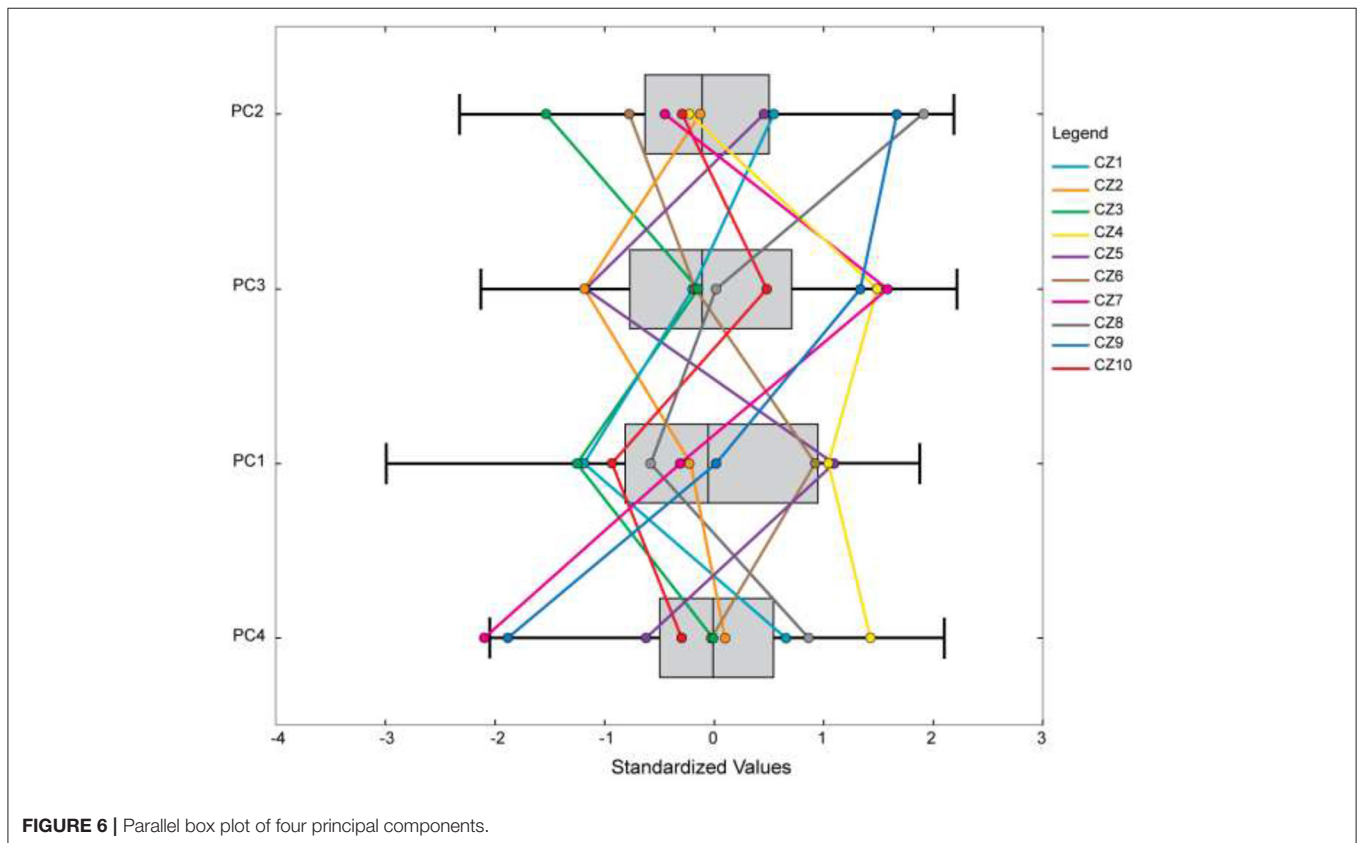
Figure 3 presents ΔT_{in} against T_d of all locations. Houses in Addis Ababa have a short range of ΔT_{in} as well as T_d . On the other hand, Bahir Dar houses show the broadest range of ΔT_{in} , while Nekemte houses have a larger spread indicating a weak relationship between ΔT_{in} and T_d .

Since the houses in Addis Ababa demonstrated a better thermal performance, a representative house is selected for further analysis through simulations.

Climate Zoning Through Multivariate Clustering

Principal Components Analysis

A strong correlation among temperature variables is reflected in the resulting principal components. Optimum number of PCs is estimated using the scree plot. As indicated in Figure 4, the first four PCs are well-separated from the rest. Among 48 layers, four dimensions in the component space account for 93.7% of the variance. Individually, PC1, PC2, PC3, and PC4 account for 53.9, 18.6, 11.3, and 9.9% of the variance respectively. The component



loadings reveal the variable groups represented in each extracted PC (Appendix a).

The component loading on the first PC explains T_{max} and T_{min} of all months. PC two explains RH_{μ} and August I_{μ} at $|R| > 0.60$. PC 3 represents I_{μ} of the months March to June and October to November. The fourth PC shows correlation with late summer RH_{μ} and winter I_{μ} . Additionally, the communality shows all input variables are well-represented ($> 84\%$) in the extracted PCs. The variable with the lowest variance explained by all PCs is I_{μ} for November at 72%, followed by RH_{μ} of September at 76%.

The latent variables on the extracted PCs are shown in Appendix b (PC1 to PC4) while the last principal component before extraction (PC48) is displayed for comparison.

Climate Clustering

Grouping is performed iteratively to determine the optimum number of groups inherent in the data. At each stage (number of groups $k = 2, \dots, 15$), 10 iterations are run each yielding a different pseudo-f-statistic value (CH index). The highest mean pseudo-f value is used to determine the optimum number of groups. Based on the pseudo-f-statistic, a 10 class-clustering yields the best results for the current data. However, the differences between 6 and 10 group classification is minimal. In order to investigate this

the clustering is repeated by varying the number of groups from 6 to 13. The 6-group classification results in a much-generalized climate zoning and fails to highlight sub zones. On the other hand, any classification that has more than 10-groups is left out due to smaller mean pseudo-f-statistic.

Figure 5 presents the final climate clustering of Ethiopia obtained by this method. The country is grouped into 10 zones based on the highest mean CH index. Some of the zones such as CZ3 and CZ4 correspond to the altitude. However, some of the larger geographical swathes with similar altitude are sub categorized. For instance, climate zones CZ8 and CZ9 in the western lowlands have a similar elevation. This is in contrast with the existing agro ecological climate zoning, which groups geographies based on altitude similarity.

A parallel box plot of principal components used for the clustering is presented in Figure 6. The graph indicates the overlap and separation between each class in terms of the four PCs. It is evident that each group is separated from others at least in one input (PC). The use of optimized seed initialization proves significant by producing less variability of pseudo-f-statistic values as compared to random seed selection. The difference between the maximum and minimum CH index values is 68,187 for random seed selection, and 40,656 for optimized seed selection.

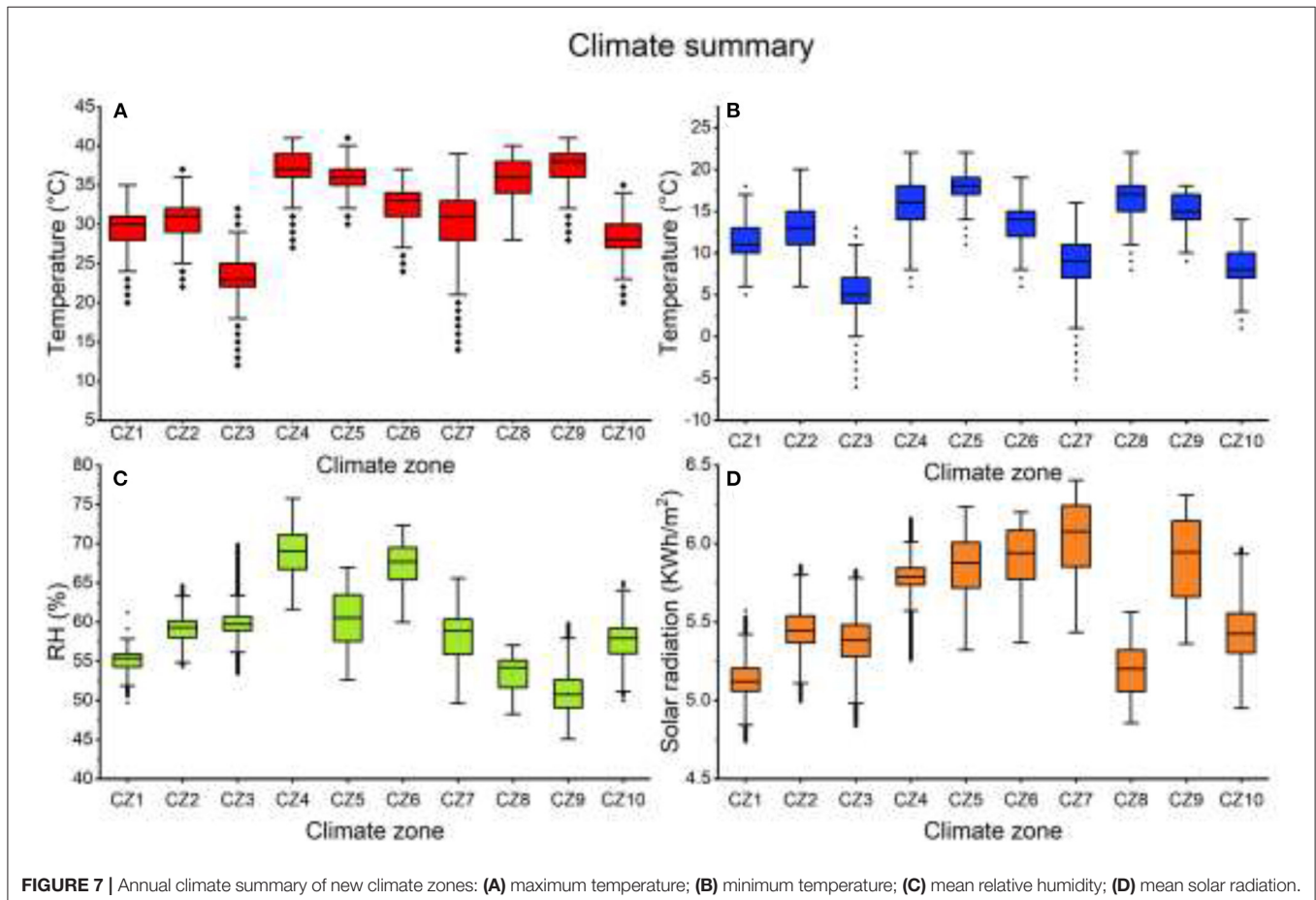


FIGURE 7 | Annual climate summary of new climate zones: (A) maximum temperature; (B) minimum temperature; (C) mean relative humidity; (D) mean solar radiation.

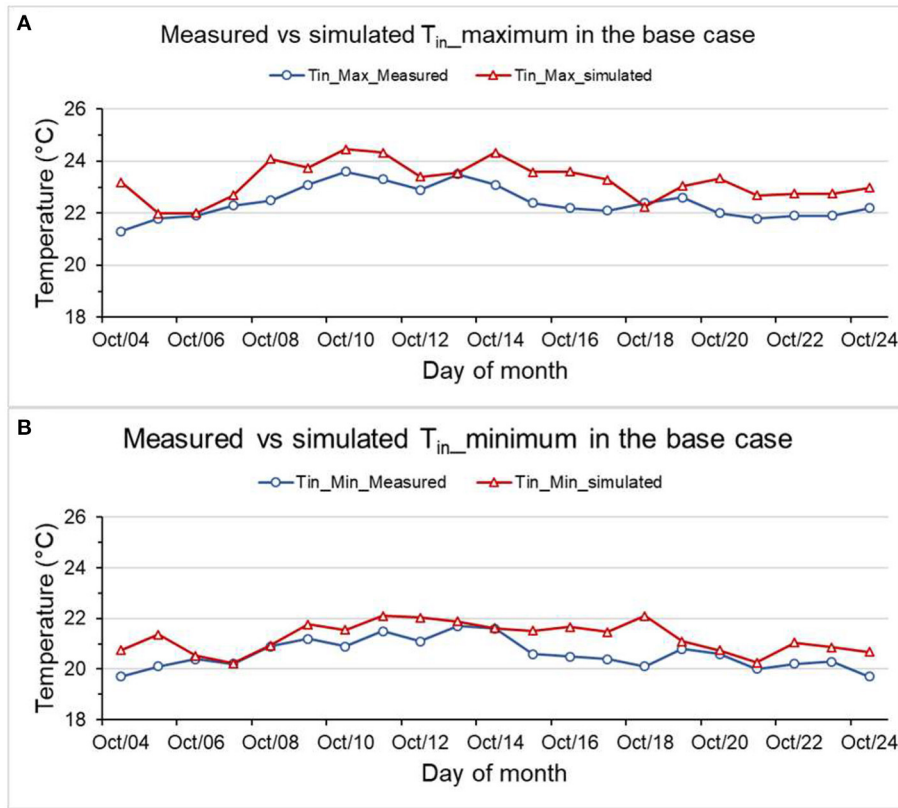


FIGURE 8 | Comparison between measured vs. simulated (A) T_{in} maximum of Case 1a, and (B) T_{in} minimum of Case 1a.

Cluster Evaluation Through Pseudo-F-Statistic

A statistical summary of pseudo-f-statistic values for groups between 2 and 15 is presented in Appendix c. The f-statistic values remain significantly lower for $K \leq 4$. On the other hand, for $K > 10$ these values show a decreasing trend. Since the maximum and minimum values represent extreme conditions, the highest mean value is considered to be better in the current analysis.

Climate Profile of the New Zones

A statistical summary of $T_{maxA_{\mu}}$, $T_{minA_{\mu}}$, $RH_{A_{\mu}}$ and $I_{A_{\mu}}$ of the proposed climate zones is presented in **Figures 7A–D**. As indicated, a difference between medians and distributions are observed.

Assessment of Clusters Through Building Simulations

The robustness of the climate zones is verified by comparing the spatial profiles of DDH and EPI. The purpose of this is to test if there is a significant difference in cooling DDH, heating DDH, and EPI between each of the proposed climate zones. It works under the assumption that different climate conditions lead to varying levels of discomfort and thermal performance requirements for a similar building.

Validation of the Base Case

The correlation between the measured outdoor temperature and relative humidity and the locations outdoor climate data is tested. The result shows a significant relationship. For instance, in Nekemte, a simple regression analysis of T_{out} between measured and climate station data shows a positive correlation ($R^2 = 0.94$) and RH_{out} shows a positive correlation ($R^2 = 0.92$). A paired sample t-test shows, at the 95% CI, the differences between measured T_{out} and station data T_{out} are $< 0.2^{\circ}C$.

Validation of the model is performed on the low-thermal mass 1-bed type apartment (Case 1a). The inside air temperature taken from the field measurement in Addis Ababa (Haile Garment condominium site) is compared with the same room’s simulated T_{in} . Climate data for the simulation is selected from a nearby location to the actual location of the building. Since the simulation model is designed according to the existing building, the validation doesn’t include different scenarios (for instance, optimizing various aspects of the model to enhance performance).

Figure 8A presents a line graph comparing the base case’s maximum inside air temperature (Case 1a). The result shows that the model fits the measured values significantly. At 95%CI, the difference between measured and simulated T_{in_Max} is between 0.4 and $1.2^{\circ}C$. A 2-sample standard

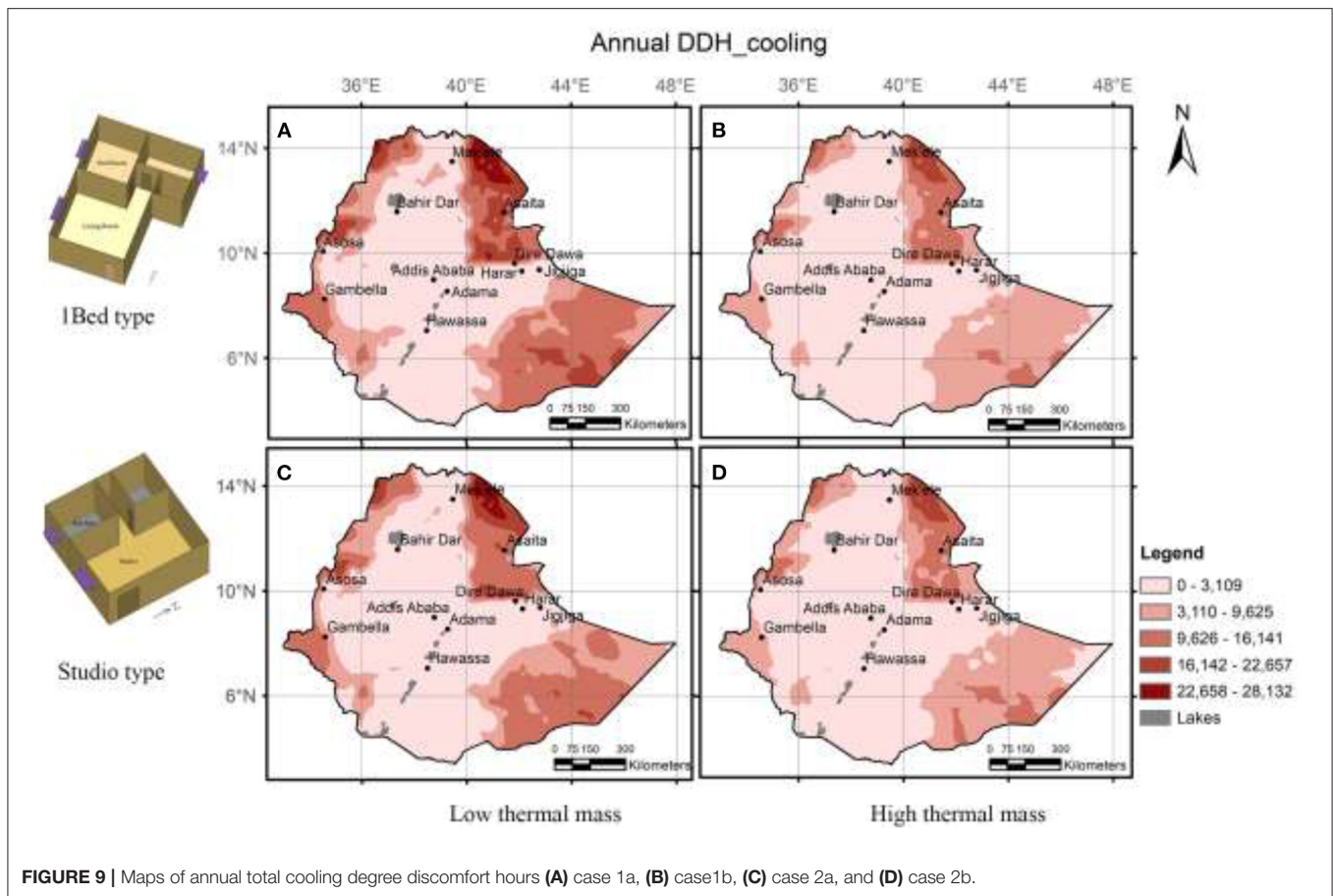


FIGURE 9 | Maps of annual total cooling degree discomfort hours (A) case 1a, (B) case1b, (C) case 2a, and (D) case 2b.

deviation test for the measured T_{in_Max} and simulated T_{in_Max} is performed. The result shows at the 0.05 level; there is no significant difference between standard deviations of measured T_{in_Max} and simulated T_{in_Max} ($P = 0.431$). A simple regression analysis between T_{in_max} measured and T_{in_max} simulated indicated a positive correlation ($r = 0.69$).

A similar analysis on measured and simulated T_{in_Min} is performed. At the 95%CI, the difference between T_{in_Min} measured and T_{in_Min} simulated is below $1.01^{\circ}C$ at the highest. Once again, a 2-sample standard deviation test for the measured T_{in_Min} and simulated T_{in_Min} is performed. The result shows at the 0.05 level; there is no significant difference ($P = 0.43$). **Figure 8B** compares simulated and measured minimum inside air temperatures in Case 1a.

The differences observed above do not solely result from the error in modeling. The raw climate data itself has some inherent differences from the actual measured outdoor climate. The difference stems from (a) climate data for simulation represents a typical meteorological year (TMY) while measured data only represents instant data. (b) the meteorological stations are not located at the building site, and slight variations are expected. Considering these, the building model is validated for further batch processing and analysis.

Limitations

Though the houses used in this analysis are considered representative of emerging residential trends in the country, they are not enough to account for the vast traditional building practices throughout Ethiopia.

Cooling Degree Discomfort Hours

The cooling DDH in this context indicates hot thermal stress that requires cooling. **Figure 9** presents the geo-spatial profile of cooling DDH for case 1a, 1b, 2a, and 2b.

A case-by-case analysis of cooling degree discomfort hours partially reveals the variation between some groups. For instance, in (case1a) distribution of cooling DDH for climate zone 3 and 4 show a considerable difference in terms of both medians as well as distribution. However, these variations are not pronounced between other zones, as is indicated between zone 2 and 7.

Heating Degree Discomfort Hours

The heating DDH in this context indicates cold thermal stress which requires heating. Heating DDH is found to significantly vary among the proposed climate zones (figure not included). Testing of similarities and differences in building performance necessitated the building layout and specifications to be kept consistent across all the 1,490 locations. Though it poses certain practical limitations, this is essential for assessing the climate

TABLE 2 | Statistics of mean daily cooling and heating degree discomfort hours for each house type.

House type	Thermal mass	Climate zone	Cooling degree discomfort hours			Heating degree discomfort hours		
			Mean	P5*	P95*	Mean	P5	P95
House type 1 (1bed)	Low thermal mass	CZ1	5.4	1.7	15.9	70.7	35.3	87.1
		CZ2	4.7	1.9	12.8	61.3	37.1	75.6
		CZ3	2.0	1.3	3.5	90.2	72.2	118.3
		CZ4	43.3	8.1	70.0	23.2	8.1	54.9
		CZ5	33.1	8.0	48.9	15.4	5.4	44.9
		CZ6	20.1	1.9	39.6	41.6	12.9	74.7
		CZ7	5.6	1.6	15.5	68.7	36.2	108.8
		CZ8	25.7	4.4	48.9	34.3	14.0	74.4
		CZ9	25.7	3.2	57.1	35.5	15.3	72.4
		CZ10	2.5	1.7	4.1	80.4	67.0	98.1
	High thermal mass	CZ1	2.6	1.6	6.1	104.1	47.9	131.6
		CZ2	2.2	1.7	4.4	85.3	51.2	111.4
		CZ3	1.7	1.4	2.0	136.9	106.5	177.5
		CZ4	28.1	2.9	49.9	35.8	15.4	77.5
		CZ5	16.0	2.0	28.3	24.8	11.5	62.6
		CZ6	9.3	1.6	21.3	60.9	21.7	109.9
		CZ7	2.4	1.5	6.6	102.7	51.6	163.8
		CZ8	12.9	1.8	31.5	49.6	21.9	109.8
		CZ9	12.9	1.7	38.0	52.5	24.1	107.1
		CZ10	1.8	1.6	2.0	119.0	96.9	150.1
House type 2 (Studio)	Low thermal mass	CZ1	2.7	1.5	7.0	81.0	37.4	99.8
		CZ2	2.3	1.5	5.4	68.2	39.2	88.4
		CZ3	1.6	1.3	2.0	105.4	82.7	136.3
		CZ4	27.8	2.6	52.0	22.8	6.7	59.3
		CZ5	17.9	2.2	30.7	17.1	6.5	48.1
		CZ6	8.8	1.5	20.2	46.9	14.5	89.1
		CZ7	2.4	1.4	7.1	74.9	37.5	124.1
		CZ8	13.8	2.0	30.2	37.7	14.2	88.9
		CZ9	13.4	1.6	37.2	38.0	13.5	86.7
		CZ10	1.7	1.5	2.1	94.4	75.6	116.3
	High thermal mass	CZ1	3.2	0.5	11.8	93.2	41.2	116.1
		CZ2	2.9	0.7	10.8	76.7	43.5	100.2
		CZ3	0.7	0.4	1.3	123.8	96.3	163.1
		CZ4	37.3	2.6	66.2	24.9	6.0	69.1
		CZ5	29.1	6.5	43.4	16.0	3.5	55.1
		CZ6	15.0	0.7	32.4	51.1	12.4	100.7
		CZ7	3.3	0.5	10.3	88.7	41.3	149.2
		CZ8	22.4	4.4	41.1	41.8	12.3	98.6
		CZ9	23.2	3.0	51.4	42.0	14.2	98.3
		CZ10	1.3	0.4	2.8	108.1	87.4	135.1

*P5, P95 represent the 5th and 95th percentile respectively.

impact. Further discussions and interpretations related to the performance of high mass and low mass buildings at different locations are limited in this paper considering the intent of the simulations.

A statistical comparison of climate clusters based on cooling DDH and heating DDH confirms a significant difference among different climate clusters (**Table 2**).

Energy Performance Index of Climate Zones

The geospatial profile of EPI for the four cases is presented in Appendix d. CZ4 is associated with higher EPI followed by parts of CZ5, CZ6, CZ8, and CZ9. CZ3 is associated with lowest EPI. As indicated in these maps, EPI varies from case to case. To understand the relationship between the climate zones and

combined EPI, mean EPI of all cases is calculated (maps not included). The results show similar trends for individual cases.

Figure 10 presents the case wise statistical summary of annual EPI for the proposed climate zones. These results highlight the statistical difference in EPI between climate zones and reinforce the findings presented in **Figure 10**. CZ4 and CZ5 have higher EPI values across all cases. However, CZ4 shows wider distribution from the mean compared to CZ5. Compared to all zones CZ6 has the widest distribution of EPI. A further investigation of cooling DDH and heating DDH of this particular climate zone indicates high seasonal variability. Climate clusters CZ4 and CZ5 show higher mean EPI while CZ2 and CZ10 show lower mean EPI values.

A one way between-clusters ANOVA of EPI shows significant differences among the means at the $p = 0.05$ level. Additionally, pairwise comparisons aimed to analyse the overlap or separation of clusters based on EPI show significant differences ($P < 0.001$) between each cluster. **Table 3** presents the climate zone wise EPI summary for case 2 as an example. At the mean and 95% confidence interval EPI of every cluster is significantly different. Similar results are observed for the other cases (case 1, 3 and 4).

Mood's median test is performed to check the separation between clusters. In this test, median values of EPI for case 4 are compared across climate clusters. At $p = 0.05$ level the medians of EPI are significantly different across climate zones

TABLE 3 | Grouping information using the Tukey method and 95% confidence.

Climate zone	Mean EPI_HM	Grouping
CZ5	100.0	A
CZ4	93.8	B
CZ8	77.1	C
CZ9	72.8	D
CZ6	60.2	E
CZ7	30.6	F
CZ1	26.6	G
CZ2	22.4	H
CZ3	19.3	I
CZ10	17.5	J

Means that do not share a letter are significantly different.

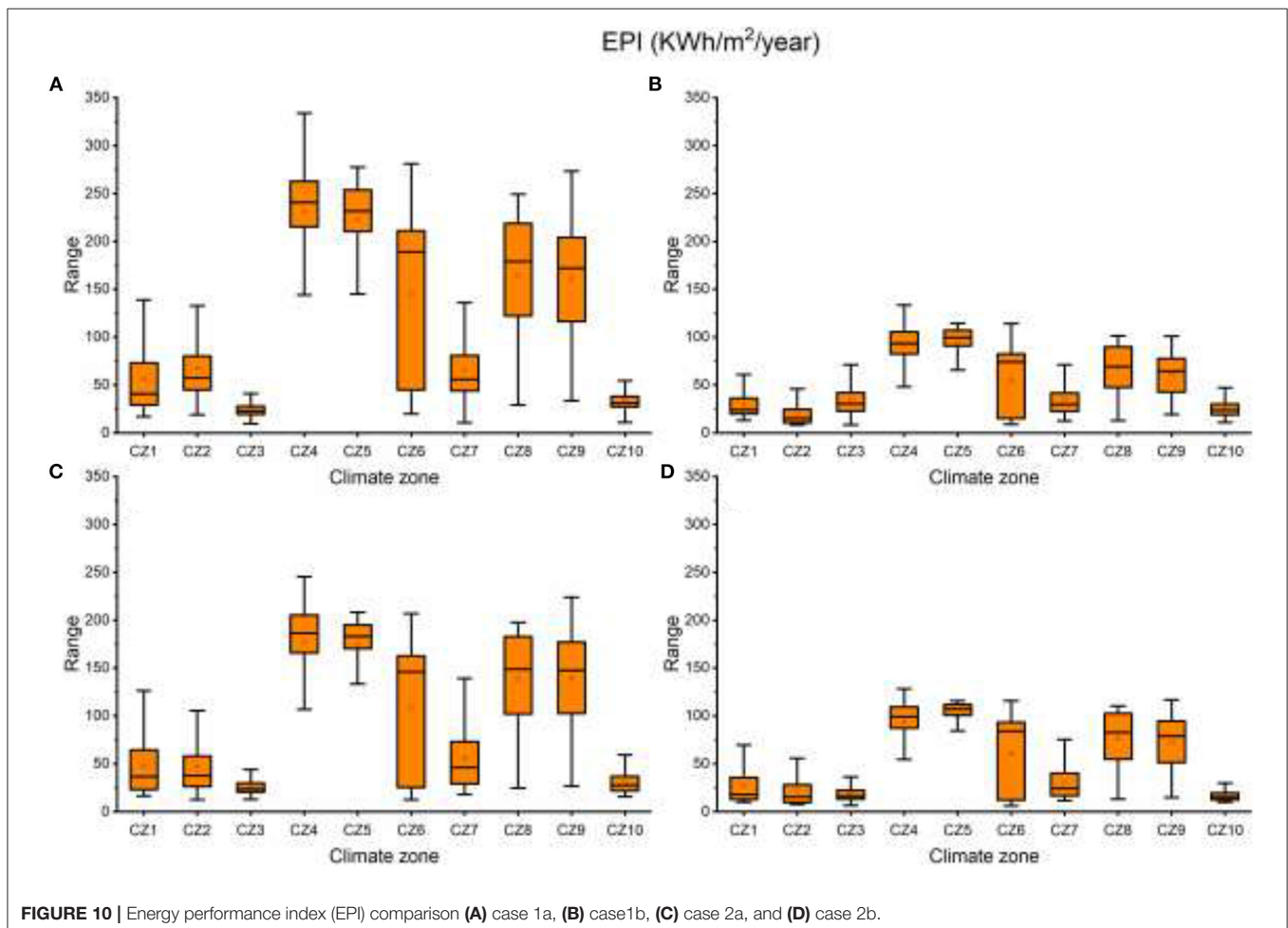


FIGURE 10 | Energy performance index (EPI) comparison (A) case 1a, (B) case 1b, (C) case 2a, and (D) case 2b.

with ($P < 0.001$). Statistical summary of medians is presented in **Table 4**. The Mood's median test does not assume normality.

Bioclimatic Zoning Through Mahoney's Method

The bioclimatic zoning of the country represents the overall passive design strategies recommended individually. Each zone represents places that share similar sets of passive design strategies and in turn, similar climatic preconditions. The eight recommended specifications (a-h) are overlaid with solar

radiation groups to create the final bioclimatic strategy group (**Figure 11**). The methods and resulting bioclimatic zones are documented in a previous work (Zekele and Rajasekar, 2020). Here the bioclimatic zoning is presented to provide a comparison.

Comparative Analysis Between the New Climate Zones and Bioclimatic Zones

This section presents a comparison between bioclimatic zones and the new climate zones in Ethiopia. Each climate zone includes a minimum of five bioclimates zones. Since the number of bioclimate zones is higher than climate zones, this result is not unexpected. This comparison is performed after discretizing both datasets to the district level administrative boundary. Six bioclimatic zones (16a, 7b, 10b, 19b, 13c, and 14c) are found in single climate zones. However, some bioclimatic zones such as 17b are found in as many as seven climate zones.

Energy Efficient Design Strategy Recommendations for Climate Zones

Based on the results of this research, the following general strategies are recommended. In colder climate zones, including CZ3 and CZ10, the use of passive solar heating is encouraged. This is especially important in CZ7, where the cold climate is complimented with relatively high solar radiation. In hotter zones such as CZ4, design strategies that maximize natural ventilation and solar shading are recommended. In extremely cold zones (CZ3) where passive solar heating is unattainable and extremely hot humid zones, HVAC can be used coupled with

TABLE 4 | Mood's median test: EPI of case 4 vs. climate zone.

Climate zone	Median	N <= Overall Median	N > Overall Median	Q3-Q1	95% Median CI
CZ 1	17.8	103952	28,826	22.7	(17.7, 17.9)
CZ 2	15.2	146632	26,843	18.5	(15.2, 15.4)
CZ 3	16.3	104444	5,935	9.2	(16.3, 16.4)
CZ 4	99.0	9,767	143410	21.9	(98.9, 99.1)
CZ 5	107.2	7019	169686	11.0	(107.1, 107.2)
CZ 6	84.1	81,305	128664	81.1	(84.0, 84.2)
CZ 7	24.4	46,100	16,410	23.5	(24.2, 24.5)
CZ 8	82.6	13,124	89,477	47.8	(82.4, 82.8)
CZ 9	79.1	10,063	54,435	43.4	(78.6, 79.4)
CZ 10	14.9	144590	3,310	6.9	(14.9, 15.0)
Overall	38.9				

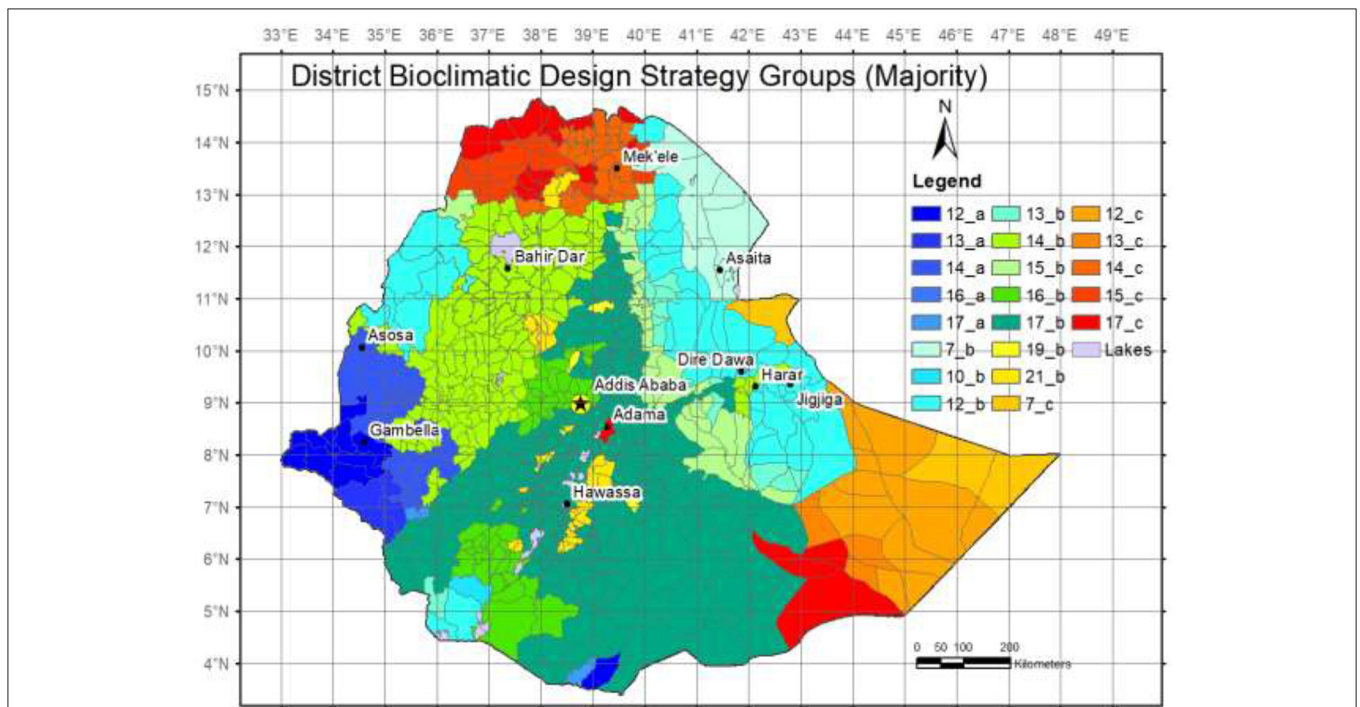


FIGURE 11 | District-level discretization of modified bioclimatic strategy groups (An expanded legend is presented in Appendix e).

high thermal mass design. A list of useful bioclimatic strategies for each climate zone is summarized based on the statistical majority (Appendix f). This analysis is meant to provide a simplified representation of design strategy combinations in each climate zone. However, since most of the climate zones cover large geographic areas, a more comprehensive summarization of strategies as well as climate zones to the smallest administrative boundary is performed (table not included).

CONCLUSIONS

This study presented a multivariate climate clustering of Ethiopia using high-resolution climate datasets for building energy performance applications. The clustering was performed using the k-means algorithm coupled with the k-means++ initial cluster centroid selection method. Reduction of redundant variables was performed using PCA and four PCs were extracted (Figure 4). Calinski-Harabasz pseudo-f-statistic identified the optimum number of groups as 10 (Appendix c). The use of optimized seed initialization proves significant by producing less variability of pseudo-f-statistic values as compared to random seed selection. The optimized algorithm results in robust climate clusters with resemblances to known climate zones with extreme features while introducing new zones. Robustness of the proposed climate clusters is evaluated by analyzing the cooling DDH, heating DDH and EPI of representative residential cases. These values are obtained through 11,920 runs of building energy simulations performed by considering alternate building layouts, thermal characteristics and operational scenarios for 1,490 locations across Ethiopia. A significant difference in DDH and EPI between climate zones was demonstrated. This study highlights the need to include monthly maximum and minimum temperatures, mean relative humidity and mean solar radiation in the climate zoning for building energy efficiency purposes. The use of multivariate clustering on a high-resolution climate data can be a better alternative to station based clustering used in combination with other supervised interpolation techniques, which might increase uncertainties. The results of this research can be used as a step toward building performance regulations related to building performance, energy efficiency, and usage and

will have a potential effect on architectural and urban design for Ethiopia. The results of this study can be used to inform design at the front end and drive the adaptation of climate responsive strategies. However, further tests incorporating other building types and operational characteristics are essential for evolving thermal performance criteria for the proposed climate zones.

Limitations of the Study

This study is limited to exploring a data-driven climate classification of Ethiopia for building energy performance applications. The dimension reduction technique adopted in the study (PCA) assumes a linear relationship between climate variables. The study did not take into account the effects of vegetation cover, availability of open spaces, and differences of rural/urban microclimate due to urban heat island. The effects of the number of residents on heating/cooling is also not investigated.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

All authors contributed in the concept, design, analysis, and writing of the manuscript.

ACKNOWLEDGMENTS

The authors are thankful to the Indian Institute of Technology Roorkee, India, for providing necessary facilities. The authors would like to thank the Ministry of Education of Ethiopia for providing a study abroad fellowship for one of the authors.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frsc.2022.684148/full#supplementary-material>

REFERENCES

- Anyadike, N. C. (1987). A multivariate classification and regionalization of West African climates. *J. Climatol.* 7, 157–164. doi: 10.1002/joc.3370070206
- Arora, P., Deepali, and Varshney, S. (2016). Analysis of K-Means and K-Medoids algorithm for big data. *Phys. Procedia* 78, 507–512. doi: 10.1016/j.procs.2016.02.095
- Ashrae Standard (1999). Energy standard for buildings except low-rise residential buildings. *ASHRAE/IESNA Stand.* 2010, 404–636.
- Attia, S., Lacombe, T., Rakotondramiarana, H. T., Garde, F., and Roshan, G. R. (2019). Analysis tool for bioclimatic design strategies in hot humid climates. *Sustain. Cities Soc.* 45, 8–24. doi: 10.1016/j.scs.2018.11.025
- Badraddin, A. (1997). Climatic classification of Saudi Arabia: an application of factor - cluster analysis. *GeoJournal* 41, 69–84. doi: 10.1023/A:1006827322880
- Bai, L., Cheng, X., Liang, J., Shen, H., and Guo, Y. (2017). Fast density clustering strategies based on the k-means algorithm. *Pattern Recognit.* 71, 375–386. doi: 10.1016/j.patcog.2017.06.023
- Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Sci. Data* 5, 1–12. doi: 10.1038/sdata.2018.214
- BEE (2017). *Energy Conservation Building Code, 2017*. Bureau of Energy Efficiency, Ministry Of Power, New Delhi, India.
- Bodach, S. (2014). “Developing Bioclimatic zones and passive solar design strategies for Nepal,” in *30th International PLEA Conference* (Ahmedabad), 1–8.
- Brager, G. S., and de Dear, R. J. (1998). Thermal adaptation in the built environment: a literature review. *Energy Build.* 27, 83–96. doi: 10.1016/S0378-7788(97)00053-4
- Bureau of Indian Standards (2016). *National Building Code of India 2016: Volume 2*. New Delhi.
- Calinski, T., and Harabasz, J. (1974). Communications in statistics - theory and methods. *Commun. Stat.* 3, 1–27. doi: 10.1080/03610927408827101
- Carvalho, M. J., Melo-Gonçalves, P., Teixeira, J. C., and Rocha, A. (2016). Regionalization of Europe based on a K-Means cluster analysis of the climate

- change of temperatures and precipitation. *Phys. Chem. Earth* 94, 22–28. doi: 10.1016/j.pce.2016.05.001
- Chang, C., Zhu, N., Yang, K., and Yang, F. (2018). Data and analytics for heating energy consumption of residential buildings: the case of a severe cold climate region of China. *Energy Build.* 172, 104–115. doi: 10.1016/j.enbuild.2018.04.037
- Chen, D., and Chen, H. W. (2013). Using the Köppen classification to quantify climate variation and change: an example for 1901–2010. *Environ. Dev.* 6, 69–79. doi: 10.1016/j.envdev.2013.03.007
- Daly, C. (2006). Guidelines for assessing the suitability of spatial climate data sets. *Int. J. Climatol.* 26, 707–721. doi: 10.1002/joc.1322
- David, A., and Vassilvitskii, S. (2006). *K-means++: The Advantages of Careful Seeding*. Available online at: <http://ilpubs.stanford.edu:8090/778/> (accessed December 18, 2008).
- DeGaetano, A. T. (1996). Delineation of mesoscale climate zones in the northeastern United States using a novel approach to cluster analysis. *J. Clim.* 9, 1765–1782. doi: 10.1175/1520-0442(1996)09<1765:DOMCZI>2.0.CO;2
- Erell, E., Portnov, B., and Etzion, Y. (2003). Mapping the potential for climate-conscious design of buildings. *Build. Environ.* 38, 271–281. doi: 10.1016/S0360-1323(02)00119-1
- Fazzini, M., Bisci, C., and Billi, P. (2015). “The climate of Ethiopia,” in *Landscapes and Landforms of Ethiopia*, ed P. Billi (Berlin: Springer Nature Switzerland AG), 65–87. doi: 10.1007/978-94-017-8026-1_3
- Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. doi: 10.1002/joc.5086
- Fovell, R. G., and Fovell, M. Y. C. (1993). Climate zones of the conterminous United States defined using cluster analysis. *J. Clim.* 6, 2103–2135. doi: 10.1175/1520-0442(1993)06<2103:CZOTCU>2.0.CO;2
- Gao, X., and Malkawi, A. (2014). A new methodology for building energy performance benchmarking : an approach based on intelligent clustering algorithm. *Energy Build.* 84, 607–616. doi: 10.1016/j.enbuild.2014.08.030
- Geiger, R., and Pohl, W. (1954). Eine neue Wandkarte der Klimagebiete der Erde nach W. Köppens Klassifikation (A new wall map of the climatic regions of the world according to W. Köppen's classification). *Erdkunde* 8, 58–61. doi: 10.3112/erdkunde.1954.01.04
- Hartigan, J. A., and Wong, M. A. (1979). Algorithm AS 136 : A K-means clustering algorithm. *J. R. Stat. Soc.* 28, 100–108. doi: 10.2307/2346830
- Hashemi, F., Smith, G. W., and Habibian, M. T. (1981). Inadequacy of climatological classification systems in agroclimatic analogue evaluations - suggested alternatives. *Agric. Meteorol.* 24, 157–173. doi: 10.1016/0002-1571(81)90041-8
- Iwano, J., and Mwashia, A. (2010). A review of building energy regulation and policy for energy conservation in developing countries. *Energy Policy* 38, 7744–7755. doi: 10.1016/j.enpol.2010.08.027
- Iyigun, C., Türkeş, M., Batmaz, I., Yozgatligil, C., Puruçuoğlu, V., Koç, E. K., et al. (2013). Clustering current climate regions of Turkey by using a multivariate statistical method. *Theor. Appl. Climatol.* 114, 95–106. doi: 10.1007/s00704-012-0823-7
- Jacobson, J. (2010). Classifications in climate research. *Phys. Chem. Earth* 35, 411–421. doi: 10.1016/j.pce.2009.11.010
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* 31, 651–666. doi: 10.1016/j.patrec.2009.09.011
- Johnson, R. A., and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Hoboken, NJ: Pearson Education, Inc.
- Kidson, J. W. (2000). An analysis of New Zealand synoptic types and their use in defining weather regimes. *Int. J. Climatol.* 20, 299–316. doi: 10.1002/(SICI)1097-0088(20000315)20:3<299::AID-JOC474>3.0.CO;2-B
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. *Meteorol. Zeitschrift* 15, 259–263. doi: 10.1127/0941-2948/2006/0130
- Kriticos, D. J., Webber, B. L., Leriche, A., Ota, N., Macadam, I., Bathols, J., et al. (2012). CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods Ecol. Evol.* 3, 53–64. doi: 10.1111/j.2041-210X.2011.00134.x
- Kumar, V., Tan, P.-T., and Steinbach, M. (2006). *Cluster Analysis: Basic Concepts and Algorithms. Introduction to data Mining* (Pearson India), 488–568.
- Lee, W. S., and Kung, C. K. (2011). Using climate classification to evaluate building energy performance. *Energy* 36, 1797–1801. doi: 10.1016/j.energy.2010.12.034
- Malinen, M. I., Mariescu-Istodor, R., and Fränti, P. (2014). K-means*: clustering by gradual data transformation. *Pattern Recognit.* 47, 3376–3386. doi: 10.1016/j.patcog.2014.03.034
- Markus, T. A. (1982). Development of a cold climate severity index. *Energy Build.* 4, 277–283. doi: 10.1016/0378-7788(82)90057-3
- Meteonorm (2018). *Meteonorm Handbook, Part II: Theory*. Available online at: https://meteonorm.com/assets/downloads/mn80_theory.pdf
- Metzger, M. J., Bunce, R. G. H., Jongman, R. H. G., Sayre, R., Trabucco, A., and Zomer, R. (2013). A high-resolution bioclimate map of the world: a unifying framework for global biodiversity research and monitoring. *Glob. Ecol. Biogeogr.* 22, 630–638. doi: 10.1111/geb.12022
- Naveen Kishore, K., and Rekha, J. (2018). A bioclimatic approach to develop spatial zoning maps for comfort, passive heating and cooling strategies within a composite zone of India. *Build. Environ.* 128, 190–215. doi: 10.1016/j.buildenv.2017.11.029
- New, M., Lister, D., Hulme, M., and Makin, I. (2002). A high-resolution data set of surface climate over global land areas. *Clim. Res.* 21, 1–25. doi: 10.3354/cr021001
- Nicol, F., and Humphreys, M. (2010). Derivation of the adaptive equations for thermal comfort in free-running buildings in European standard EN15251. *Build. Environ.* 45, 11–17. doi: 10.1016/j.buildenv.2008.12.013
- Parracho, A. C., Melo-Gonçalves, P., and Rocha, A. (2016). Regionalisation of precipitation for the Iberian Peninsula and climate change. *Phys. Chem. Earth* 94, 146–154. doi: 10.1016/j.pce.2015.07.004
- Pawar, A. S., Mukherjee, M., and Shankar, R. (2015). Thermal comfort design zone delineation for India using GIS. *Build. Environ.* 87, 193–206. doi: 10.1016/j.buildenv.2015.01.009
- Pesaresi, S., Galdenzi, D., Biondi, E., and Casavecchia, S. (2014). Bioclimate of Italy: application of the worldwide bioclimatic classification system. *J. Maps* 10, 538–553. doi: 10.1080/17445647.2014.891472
- Rajasekar, E., Udaykumar, A., Soumya, R., and Venkateswaran, R. (2015). Towards dynamic thermal performance benchmarks for naturally ventilated buildings in a hot-dry climate. *Build. Environ.* 88, 129–141. doi: 10.1016/j.buildenv.2014.08.011
- Ramos, N. M. M., Almeida, R. M. S. F., Simões, M. L., and Pereira, P. F. (2017). Knowledge discovery of indoor environment patterns in mild climate countries based on data mining applied to in-situ measurements. *Sustain. Cities Soc.* 30, 37–48. doi: 10.1016/j.scs.2017.01.007
- Rivas-Martinez, S., Rivas-Saenz, S., and Penas-Merino, A. (2011). Worldwide bioclimatic classification system. *Glob. Geobot.* 1, 1–638. doi: 10.5616/gg110001
- Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., et al. (2017). A review of clustering techniques and developments. *Neurocomputing* 267, 664–681. doi: 10.1016/j.neucom.2017.06.053
- Tadesse, M., Alemu, B., Bekele, G., Tebikew, T., Chamberlin, J., and Benson, T. (2006). *Atlas of the Ethiopian Rural Economy*. Washington, DC: International Food Policy Research Institute (ifpri) Central Statistical Agency Ethiopian Development Research Institute.
- Thorntwaite, C. W. (1948). An approach toward a rational classification of climate. *Geogr. Rev.* 38, 55. doi: 10.2307/210739
- Unal, Y., Kindap, T., and Karaca, M. (2003). Redefining the climate zones of Turkey using cluster analysis. *Int. J. Climatol.* 23, 1045–1055. doi: 10.1002/joc.910
- Verichev, K., Zamorano, M., and Carpio, M. (2019). Assessing the applicability of various climatic zoning methods for building construction: case study from the extreme southern part of Chile. *Build. Environ.* 160, 106165. doi: 10.1016/j.buildenv.2019.106165
- Walsh, A., Cóstola, D., and Labaki, L. C. (2017a). Comparison of three climatic zoning methodologies for building energy efficiency applications. *Energy Build.* 146, 111–121. doi: 10.1016/j.enbuild.2017.04.044
- Walsh, A., Cóstola, D., and Labaki, L. C. (2017b). Review of methods for climatic zoning for building energy efficiency programs. *Build. Environ.* 112, 337–350. doi: 10.1016/j.buildenv.2016.11.046

- Walsh, A., Cóstola, D., and Labaki, L. C. (2018). Performance-based validation of climatic zoning for building energy efficiency applications. *Appl. Energy* 212, 416–427. doi: 10.1016/j.apenergy.2017.12.044
- Wan, K. K. W., Li, D. H. W., Yang, L., and Lama, J. C. (2010). Climate classifications and building energy use implications in China. *Energy Build.* 42, 1463–1471. doi: 10.1016/j.enbuild.2010.03.016
- Wilks, D. S. (2011). “Principal component (EOF) analysis,” in *International Geophysics* 519–562. doi: 10.1016/B978-0-12-385022-5.00012-9
- Xiong, J., Yao, R., Grimmond, S., Zhang, Q., and Li, B. (2019). A hierarchical climatic zoning method for energy efficient building design applied in the region with diverse climate characteristics. *Energy Build.* 186, 355–367. doi: 10.1016/j.enbuild.2019.01.005
- Zelege, B., and Rajasekar, E. (2020). “Geospatial thermal stress assessment and a new bioclimatic classification for Ethiopia,” in *11th Windsor Conference: Resilient Comfort (London)* 647–660.
- Zscheischler, J., Mahecha, M. D., and Harmeling, S. (2012). Climate classifications: the value of unsupervised clustering. *Procedia Comput. Sci.* 9, 897–906. doi: 10.1016/j.procs.2012.04.096

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zelege, Kumar and Rajasekar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.