Purdue University
Purdue e-Pubs

International High Performance Buildings Conference | School of Mechanical Engineering

2021

Development Of Climate Classification Through Hierarchical Clustering For Building Energy Simulation

Giovanni Pernigotto
Free University of Bozen - Bolzano, Italy

Andrea Gasparella
Free University of Bozen - Bolzano, Italy, andrea.gasparella@unibz.it

Jan Hensen
TU Eindhoven - Eindhoven

Development of Climate Classification through Hierarchical Clustering for Building Energy Simulation

Giovanni PERNIGOTTO[1]*, Andrea GASPARELLA[2] and Jan L. M. HENSEN[3]

[1]Free University of Bozen-Bolzano, Faculty of Science and Technology,
Bolzano, Italy
Tel: +39 0471 017632, Fax: +39 0471 017009, email: giovanni.pernigotto@unibz.it

[2]Free University of Bozen-Bolzano, Faculty of Science and Technology,
Bolzano, Italy
Tel: +39 0471 017200, Fax: +39 0471 017009, email: andrea.gasparella@unibz.it

[3]Eindhoven University of Technology, Department of the Built Environment,
Eindhoven, The Netherlands
Tel: +31 (0)40 2472988, Fax: +31 (0) 40 2438595, email: j.hensen@tue.nl

* Corresponding Author

ABSTRACT

Climate classification plays an important role for the identification of homogeneous groups of climates, from which representative locations can be extracted and used for building energy simulation analyses. Nevertheless, according to the current state-of-the-art, the main reference systems consider just a fraction of those weather quantities which are relevant in the building energy balance, i.e., ambient temperature and humidity and solar radiation. To overcome this issue, in previous researches a new methodology was defined, based on monthly series of weather quantities, statistical analyses and data-mining techniques for climate clustering. In this work, with the aim of further developing such approach, a shorter time-discretization of weather quantities, i.e., a weekly discretization, was tested, alongside additional variables describing the daily range of ambient temperature and humidity. In order to investigate the potential of those modifications, a dataset with more than 300 European reference climates was analyzed and subdivided into climate classes according to the proposed clustering procedure.

1. INTRODUCTION

An accurate climate zoning is particular important for policy makers, when it comes to define effective strategies for building energy retrofitting or minimum energy performance requirements for new buildings. Indeed, a poor climate classification can undermine the efficacy of energy savings strategies or bring to requirements difficult to comply with by professionals, companies and public administrations operating in the building sector. Typical classifications adopted in several countries worldwide are based on heating – and sometimes cooling – degree-days or temperature-based quantities (Walsh *et al.*, 2018 and 2019). For example, popular climate classifications are the Köppen-Geiger system (Peel *et al.*, 2007) and the one based on the ANSI/ASHRAE 169 (ASHRAE, 2013), which account also for precipitation in order to group the different climates. Other weather variables, like humidity and solar irradiance, are rarely included in climate zoning, even though they play a significant role on building energy balance and performance.

In order to discuss the suitability of the climate zoning methods according to the state of the art and overcome their limitations, a new weather-based classification approach was proposed in previous works (Pernigotto and Gasparella, 2018; Pernigotto *et al.*, 2019). Specifically, all weather variables relevant for building energy balance were accounted for, i.e., dry bulb temperature, water vapour partial pressure, solar global horizontal irradiation, and processed by

means of Hierarchical Clustering. Involved quantities were analyzed on a monthly basis and characterized through annual averages and spreads of monthly series, in a fashion similar to the Köppen-Geiger system. Although achieved results were found promising, the adopted time discretization brought to loss of information which may be significant for the design of high-performance buildings, such as the variability of weather quantities within each day.
With the aim of further assessing the proposed climate classification and possibly adding improvements, on one hand a weekly time discretization of variables was tested while, on the other hand, statistics able to describe also the daily ranges of the weather quantities were considered as inputs for the Hierarchical Clustering.

2. METHODOLOGY

2.1 Dataset of climates

The analysis performed in this work exploited the same dataset of climates considered in (Pernigotto *et al.*, 2019). Specifically, all European weather files available in EnergyPlus online weather database (https://energyplus.net/weather) were considered except for the Italian climates. Indeed most of Italian weather files included in the EnergyPlus online database refer to the IGDG series, i.e., they were developed from multi-year series collected in the period 1951-1970 and are now outdated. Consequently, the most recent weather files, published by the Italian CTI (Comitato Termotecnico Italiano, https://try.cti2000.it/) and in agreement with the Italian technical standard UNI 10349-1:2016 on weather data for energy calculations, were preferred. For the other European countries, when more than a weather source was found for the same locality, the International Weather for Energy Calculations IWEC files were selected.
As a whole, 318 typical or reference years were included in the analysis. Table 1 and Figure 1 show, respectively, the distribution of the considered locations according to 11 different Köppen-Geiger climatic classes and their position on a map of Europe. *Dfb* (cold climate without dry season and with warm summer, in central and eastern Europe), *Csa* (temperate climate with dry and hot summer, in the Mediterranean areas), and *Cfb* (temperate climate without dry season and with warm summer, in western Europe) are the three most populated climate classes, with respectively 29.2 %, 23.6 %, and 17.9 % of the dataset. For further details regarding homogeneity and overlapping of the Köppen-Geiger climatic classification applied to this sample, see (Pernigotto *et al.*, 2019).

Table 1: Distribution of the dataset of climates according to the Köppen-Geiger classes

Köppen-Geiger class	Description	Number of locations	Fraction of the sample
BSk	Arid cold steppe climate	14	4.4 %
BWk	Arid desert cold climate	1	0.3 %
Csa	Temperate climate with dry and hot summer	75	23.6 %
Csb	Temperate climate with dry and warm summer	12	3.8 %
Cfa	Temperate climate without dry season and with hot summer	49	15.4 %
Cfb	Temperate climate without dry season and with warm summer	57	17.9 %
Dsb	Cold climate with dry and warm summer	2	0.6 %
Dfa	Cold climate without dry season and with hot summer	3	0.9 %
Dfb	Cold climate without dry season and with warm summer	93	29.2 %
Dfc	Cold climate without dry season and with cold summer	11	3.5 %
ET	Polar and tundra climate	1	0.3 %
Total		318	100 %

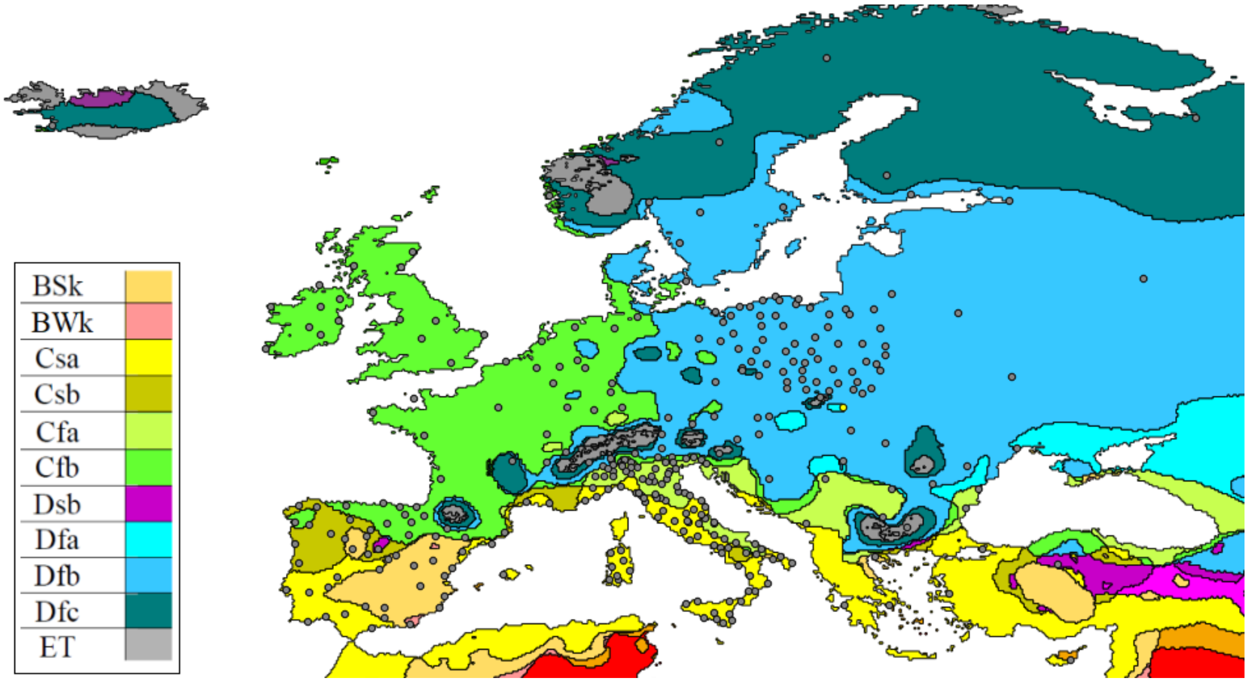

Figure 1: European map showing the position of the considered locations in the Köppen-Geiger climate classes (Pernigotto *et al.*, 2019; prepared with QGIS v. 3.4.2 based on the Köppen-Geiger GIS climate map by NASA ORNL DAAC)

Adopting the same approach as in (Pernigotto and Gasparella, 2018; Pernigotto *et al.*, 2019), the focus was put on those weather variables considered the most significant in the building energy balance, i.e., ambient temperature and humidity, and solar radiation. Respectively, ambient temperature was expressed as dry bulb temperature *DBT*, air humidity as water vapour partial pressure *WVP*, and solar radiation as global horizontal solar radiation *GHI*. While *DBT* and *WVP* were defined as average values, *GHI* was integrated over the considered time period, i.e., with monthly discretization. As explained in previous researches, wind speed was neglected due to representativeness issues incompatible with the goals of this analysis. Furthermore, considering the different number of weather files available for the different European countries in the dataset (see Figure 1), in particular with the majority located in Italy (110 locations equal to 34.6 % of the sample), in Poland (61 equal to 19.2 %) and in Spain (46 equal to 14.5 %), geographical distances and altitudes were not considered as variables for clustering. In such a way, uneven distributions of localities did not affect the classification.

2.2 Climate clustering

In previous researches (Pernigotto and Gasparella, 2018; Pernigotto *et al.*, 2019), preliminary analyses required for climate clustering were performed for each location, considering a series of 12 monthly values. These series were used to determine *annual average* and *spread* of monthly values for each weather quantity, as in Figure 2 (left). Then, these statistics were normalized and used as inputs for a Hierarchical Clustering, chosen for its ability to allow for a climate classification without the need of a predefined number of groups, as it is, for instance, in the *k*-means approach. Classes were identified with the goal of enhancing group homogeneity and avoiding overlapping, assessed respectively by means of *standard deviations* and *mean values* of the annual averages and spreads of *DBT*, *WVP* and *GHI*.
Even if the clustering can be performed for each weather quantity at a time (e.g., for dry bulb temperature), a more meaningful classification was found including all of them in the procedure. Moreover, each quantity was given the same relative importance, without any predetermined hierarchy. Although clustering can be repeated in multiple steps to run sub-classification, the study of the dendrogram characterizing the Hierarchical Clustering was recommended to target potential subdivisions when useful to increase group homogeneity and reduce the risk of overlapping.

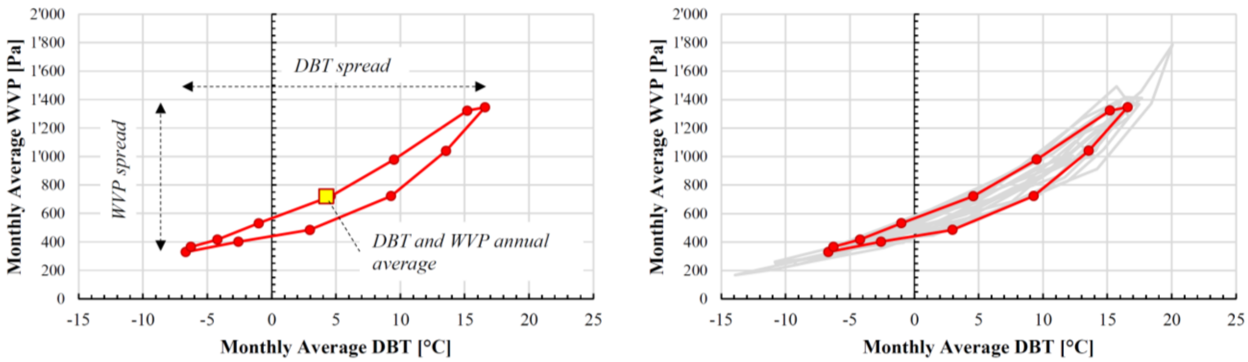

Figure 2: Annual average and spread for *DBT* and *WVP* in a given location (**left**); example of a class of climates to identify its homogeneity (**right**)

2.3 Modifications to the proposed methodology

As explained in the introduction, this work wants to further improve the proposed procedure for climate classification to get distinct groups with homogenous weather conditions, suitable to identify representative locations and to run building energy simulation studies to ease the definition of regional or national energy policies. With that goal in mind, it was decided to discuss if (1) a different time discretization of the input data series can improve clusters' uniformity, and if (2) an alternative definition of weather quantities, based on metrics assessing daily ranges, can lead to a different classification.

In order to investigate the first goal, instead of a series of 12 monthly values, 52 weekly values were used as inputs. It was assumed that each typical year begins conventionally with the first day of the week, Monday, and the last day of the year was included in the calculation of the previous week average value. The procedure was either repeated as in Section 2.2, i.e., with (A) annual *averages* and *spreads* of weekly weather quantities, and considering also different statistics. Specifically, in order to account for the fact that weekly series are more sensitive to extreme or anomalous input data, which can affect the calculation of overall statistics like spreads as well, alternatives were tested. In details, the clustering procedure was applied also working with (B) annual *averages* and *standard deviations* of weekly weather quantities, and with (C) annual *averages* coupled with *minimum* and *maximum* values in the weekly series weather quantities. All three approaches share a statistical variable, i.e., the *average*, which is clearly the same as in the case of monthly *DBT* and *WVP* series due to the properties of the arithmetic mean. To determine if a shorter time-discretization of the input data could be beneficial for the procedure, homogeneity of climate classes was expressed in terms of standard deviations of weather statistics and compared with those found in (Pernigotto *et al.*, 2019). Both partial clusterings, with all considered statistics for a single weather quantity at a time, and global clusterings were analysed for a more comprehensive understanding of the findings.

As regards the second goal, the original methodology with monthly series of weather quantities was applied. Nevertheless, along with monthly averages of *DBT* and *WVP*, also those of daily ranges of the same quantities, $\widehat{DBT}$ and $\widehat{WVP}$, were included. As regards the solar radiation, the monthly *GHI* integrals were analysed without any additional metrics. For all considered weather quantities annual *averages* and *spreads* were calculated, normalized and used as inputs for the Hierarchical Clustering. Besides the qualitative comparison of the obtained climate classes with those in (Pernigotto *et al.*, 2019), standard deviations were discussed as well. Again, for sake of completeness, both partial and global clusterings were performed.

3. RESULTS AND DISCUSSION

3.1 Impact of time-discretization of input weather data series

To discuss the impact of time-discretization of the weather inputs on the results of the climate classification, both partial and global clustering were performed. As indicated in Section 2.3, three alternative sets of statistics were considered (i.e., A – *averages* and *spreads*, B – *averages* and *standard deviations*, C – *averages*, *maximum* and *minimum* values). The number of groups was fixed to 7, as in (Pernigotto *et al.*, 2019), for both partial and global clusterings. Very different dendrograms were obtained in partial clusterings, highlighting the importance of the chosen weather statistics, as it can be seen for instance in Figure 3 for the global clusterings. Nevertheless, for both (A) and (B) 7 groups were identified at a height in the dendrogram around 0.5 for partial clustering and at a height of 1 for the global one, similarly to what found in the previous studies. Regarding (C), instead, slightly higher heights were

observed, respectively around 0.7 and 1.2, for partial and global clustering. This can be associated to the higher number of variables included in the Hierarchical Clustering, i.e., 3 and 9 instead of 2 and 6.
The selection of weather statistics influenced also the distribution of the locations in the different classes. For instance, partial clustering (A) led to groups ranging from 4 to 91 elements for *DBT*, from 8 to 98 for *WVP*, and from 15 to 71 for *GHI*. The size of groups from partial clustering (B), instead, is between 3 or 4 and 130 elements for *DBT*, 6 or 7 and 138 for *WVP*, and 11 and 80 for *GHI*. Finally, partial clustering (C) generated groups with from 3 to 101 locations for *DBT*, from 19 to 104 for *WVP*, and from 7 to 58 for *GHI*. In case of global clustering, the minimum group size is 12, 24, and 15, and the maximum one 90, 94, and 68, respectively for (A), (B), and (C).

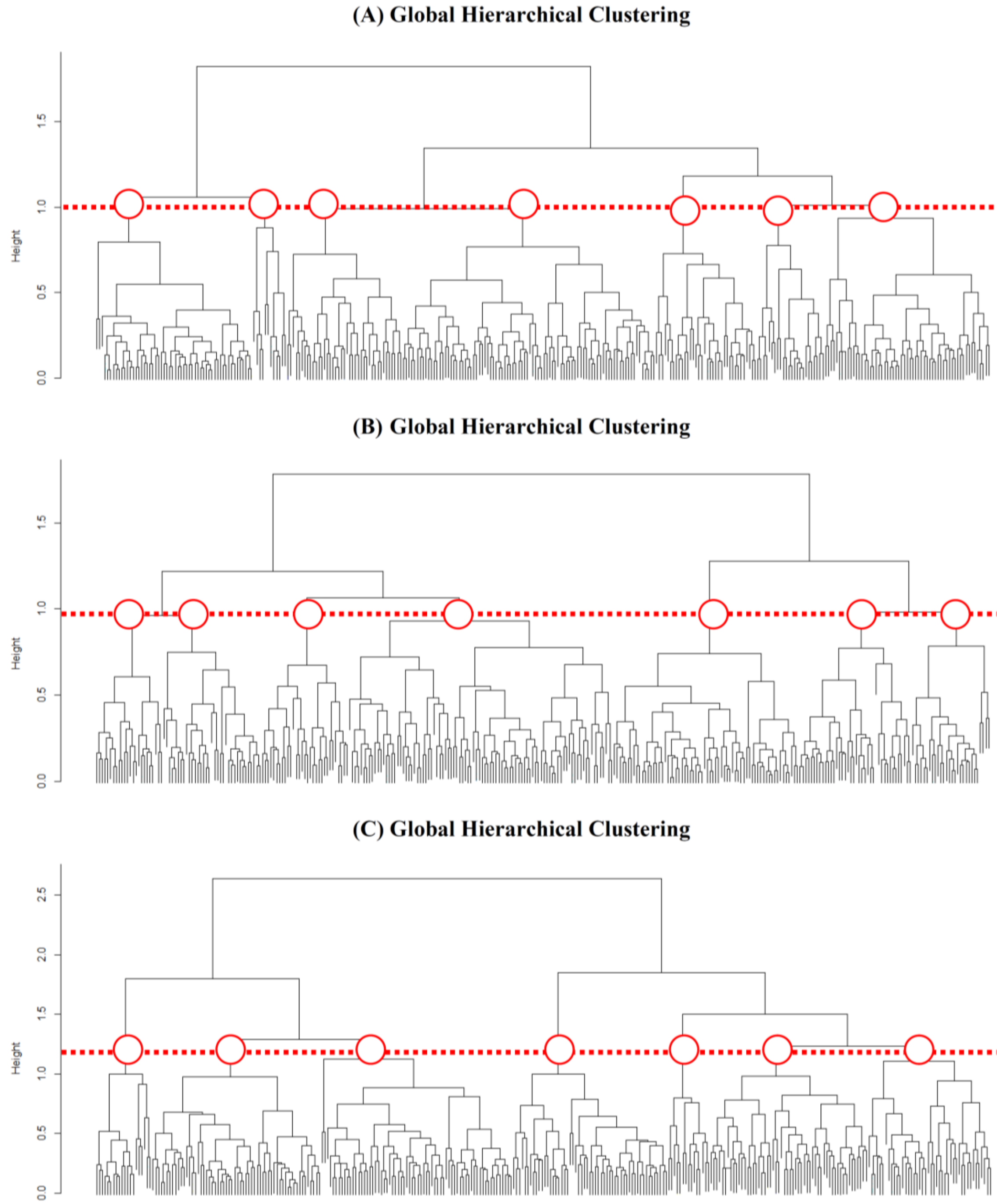

Figure 3: Global hierarchical clusterings according to different groups of weather statistics (A, B, C). The red line represents the height chosen for determining the number of clusters, identified with the circles.

As it can be seen in Tables 2-4, in some cases partial clustering gave groups with very close annual averages, as for instance classes $\mathbf{4^{DBT}}$ and $\mathbf{5^{DBT}}$, and classes $\mathbf{6^{DBT}}$ and $\mathbf{7^{DBT}}$ for *DBT* in clustering (C) in Table 2, classes $\mathbf{1^{WVP}}$ and $\mathbf{2^{WVP}}$, $\mathbf{3^{WVP}}$ and $\mathbf{4^{WVP}}$, $\mathbf{5^{WVP}}$ and $\mathbf{6^{WVP}}$ for *WVP* in clustering (B) in Table 3, classes $\mathbf{1^{GHI}}$, $\mathbf{2^{GHI}}$ and $\mathbf{3^{GHI}}$ for *GHI* in clustering (C) in Table 4. As a whole, the adoption of annual *spreads* as in clustering (A) seems to reduce the risk of overlapping with respect to the other tested alternatives. Indeed, as confirmed also in the comparison of global clusterings in Table 5, while just two classes show close annual averages in clustering (A) (specifically, classes **3** and **4** for *WVP*), this occurs more frequently for clustering (B) (see, for instance, groups **2** and **3** and groups **5** and **6** for *WVP*) and (C) (e.g., groups **1** and **2** and groups **4** and **5** for *GHI*).

Table 2: *DBT* Partial Hierarchical Clustering with weekly series of weather quantities: mean values and standard deviations for each class and annual averages and spreads

***DBT* Hierarchical Clustering**									
			Classes and fraction of sample						
A	**Annual**	**Quantities**	**1 (1.3%)**	**2 (9.4 %)**	**3 (28.6%)**	**4 (6.9%)**	**5 (18.6%)**	**6 (23.3%)**	**7 (11.9%)**
	Average	*DBT* [°C]	2.7±3.0	6.8±2.3	9.1±1.5	9.7±1.4	12.6±1.1	15.3±1.4	16.5±1.6
		WVP [Pa]	694±21%	860±13%	940±8%	1005±10%	1131±10%	1215±13%	1357±9%
		GHI [kWh m^{-2} w^{-1}]	17.6±18%	18.7±12%	20.5±20%	18.9±13%	25.8±7%	29.2±8%	29.8±12%
	Spread	*DBT* [°C]	37.3±2.0	28.9±2.3	24.3±2.7	15.4±2.1	27.1±2.0	22.4±1.7	17.2±2.0
		WVP [Pa]	1554±23%	1434±14%	1326±19%	1052±10%	1709±15%	1443±24%	1441±23%
		GHI [kWh m^{-2} w^{-1}]	42±5%	41.7±9%	41.9±11%	40.2±9%	46.8±8%	45.9±8%	43.3±8%
B	**Annual**	**Quantities**	**1 (1.2%)**	**2 (0.9%)**	**3 (18.5%)**	**4 (6.2%)**	**5 (28.6%)**	**6 (40.8%)**	**7 (3.4%)**
	Average	*DBT* [°C]	0.8±1.9	3.9±2.0	8.6±1.7	9.5±1.3	10.0±2.1	15.0±1.8	15.7±1.6
		WVP [Pa]	572±12%	759±10%	933±9%	1004±10%	995±11%	1230±13%	1300±9%
		GHI [kWh m^{-2} w^{-1}]	16.0±8%	18.6±15%	20.8±17%	18.3±9%	21.5±20%	28.6±10%	29.5±10%
	Spread	*DBT* [°C]	27.1±6.2	38.2±0.5	27.0±3.3	15.2±2.2	24.6±3.4	22.1±3.4	20.1±5.1
		WVP [Pa]	1040±8%	1728±4%	1447±15%	1064±9%	1403±20%	1499±24%	1535±20%
		GHI [kWh m^{-2} w^{-1}]	39.3±13%	42.2±5%	42.3±9%	39.5±8%	43.1±10%	45.3±8%	47.1±7%
C	**Annual**	**Quantities**	**1 (0.9%)**	**2 (4.7%)**	**3 (11.3%)**	**4 (9.4%)**	**5 (24.5%)**	**6 (31.7%)**	**7 (17.2%)**
	Average	*DBT* [°C]	1.3±2.1	6.4±1.2	7.8±1.9	9.81±1.8	10.0±2.0	14.7±2.1	14.8±1.8
		WVP [Pa]	595±14%	825±6%	923±9%	1012±13%	1007±10%	1186±16%	1244±10%
		GHI [kWh m^{-2} w^{-1}]	15.7±8%	18.8±7%	18.1±8%	19.1±15%	21.8±17%	29.0±10%	27.8±8%
	Spread	*DBT* [°C]	34.4±4.1	29.3±4.3	26.3±2.7	17.7±4.2	25.9±3.1	21.2±4.0	23.1±2.8
		WVP [Pa]	1262±25%	1418±15%	1431±10%	1196±24%	1484±15%	1387±30%	1584±16%
		GHI [kWh m^{-2} w^{-1}]	41.9±6%	43.9±5%	39.6±9%	40.7±9%	42.7±10%	46.1±7%	45.3±8%

Table 3: *WVP* Partial Hierarchical Clustering with weekly series of weather quantities: mean values and standard deviations for each class and annual averages and spreads

***WVP* Hierarchical Clustering**									
				Classes and fraction of sample					
A	**Annual**	**Quantities**	**1 (3.7%)**	**2 (2.5%)**	**3 (30.8%)**	**4 (20.7%)**	**5 (11.9%)**	**6 (17.2%)**	**7 (12.8%)**
	Average	*DBT* [°C]	6.3±4.7	12.5±2.1	9.7±2.3	9.6±2.6	14.1±2.1	14.0±1.6	16.8±1.2
		WVP [Pa]	725.±17%	873±10%	951±7%	982±9%	1172±5%	1254±6%	1396±7%
		GHI [kWh m^{-2} w^{-1}]	21.4±31%	30.7±6%	20.7±21%	21.7±17%	27.0±16%	26.9±7%	30.2±9%
	Spread	*DBT* [°C]	22.9±5.2	20.6±0.8	23.5±4.5	27.7±3.3	19.2±4.1	24.3±3.2	20.0±2.6
		WVP [Pa]	972.±9%	651±13%	1254±13%	1591±6%	1182±14%	1850±8%	1559±14%
		GHI [kWh m^{-2} w^{-1}]	43.9±12%	48.1±4%	42.1±11%	43.2±9%	44.1±8%	46.4±7%	43.7±9%
B	**Annual**	**Quantities**	**1 (1.8%)**	**2 (2.2%)**	**3 (43.3%)**	**4 (19.8%)**	**5 (6.6%)**	**6 (22.3%)**	**7 (3.7%)**
	Average	*DBT* [°C]	4.9±3.7	5.8±5.0	10.3±2.5	10.1±2.6	14.7±1.2	15.5±1.7	17.0±1.4
		WVP [Pa]	706±18%	712±15%	989±9%	991±10%	1293±8%	1299±7%	1414±9%
		GHI [kWh m^{-2} w^{-1}]	18.0±30%	22.3±31%	22.0±21%	22.6±21%	27.5±9%	28.8±9%	30.9±8%
	Spread	*DBT* [°C]	25.1±6.8	25.9±8.7	23.8±4.7	25.4±4.6	22.5±3.9	21.6±3.4	19.4±2.7
		WVP [Pa]	1003±8%	1106±40%	1338±18%	1411±24%	1721±21%	1575±17%	1678±18%
		GHI [kWh m^{-2} w^{-1}]	44.0±9%	42.8±14%	43.3±10%	42.6±11%	46.5±08%	44.9±8%	43.6±6%
C	**Annual**	**Quantities**	**1 (6.6%)**	**2 (12.2%)**	**3 (24.8%)**	**4 (32.7%)**	**5 (9.1%)**	**6 (8.4%)**	**7 (5.9%)**
	Average	*DBT* [°C]	5.9±2.4	8.94±2.9	11.5±3.0	13.1±3.0	13.5±2.3	13.0±4.4	14.0±1.5
		WVP [Pa]	786±13%	899±11%	1069±14%	1143±15%	1174±8%	1178±20%	1248±11%
		GHI [kWh m^{-2} w^{-1}]	18.4±17%	21.8±25%	22.9±23%	25.9±16%	26.4±14%	25.9±25%	26.8±7%
	Spread	*DBT* [°C]	28.3±6.0	24.0±3.6	21.7±5.0	23.8±3.7	22.3±5.1	23.4±4.7	24.0±3.1
		WVP [Pa]	1302±19%	1223±26%	1339±21%	1504±18%	1507±23%	1511±19%	1770±15%
		GHI [kWh m^{-2} w^{-1}]	43.2±7%	42.0±12%	43.5±11%	44.2±9%	43.5±9%	43.9±10%	46.5±8%

Table 4: *GHI* Partial Hierarchical Clustering with weekly series of weather quantities: mean values and standard deviations for each class and annual averages and spreads

***GHI* Hierarchical Clustering**

			Classes and fraction of sample						
A	**Annual**	**Quantities**	**1 (14.4%)**	**2 (12.2%)**	**3 (15.0%)**	**4 (21.3%)**	**5 (22.3%)**	**6 (9.7%)**	**7 (4.7%)**
	Average	*DBT* [°C]	7.9±2.2	7.6±2.2	10.2±2.4	12.6±2.2	13.9±1.9	15.9±1.7	17.4±1.2
		WVP [Pa]	922±11%	894±13%	1037±13%	1124±13%	1136±16%	1312±11%	1310±12%
		GHI [kWh m^{-2} w^{-1}]	17.2±6%	18.0±6%	21.4±9%	25.2±8%	29.1±6%	29.2±6%	33.8±2%
	Spread	*DBT* [°C]	23.0±5.6	25.6±5.5	24.2±5.5	25.1±3.4	22.6±3.5	20.6±2.8	19.3±2.3
		WVP [Pa]	1289±15%	1329±15%	1473±15%	1607±18%	1385±31%	1524±16%	1246±24%
		GHI [kWh m^{-2} w^{-1}]	36.8±4%	44.1±5%	40.2±5%	46.8±3%	48.4±4%	42.2±4%	42.4±5%
B	**Annual**	**Quantities**	**1 (12.2%)**	**2 (24.2%)**	**3 (25.1%)**	**4 (11.3%)**	**5 (10.0%)**	**6 (13.5%)**	**7 (3.4%)**
	Average	*DBT* [°C]	7.1±2.0	8.6±1.9	12.5±2.2	13.6±1.9	14.1±1.7	15.7±2.1	17.4±1.1
		WVP [Pa]	883±12%	947±10%	1109±15%	1182±11%	1219±11%	1212±18%	1335±8%
		GHI [kWh m^{-2} w^{-1}]	17.8±7%	18.4±8%	25.9±9%	26.2±8%	27.0±4%	31.7±5%	32.1±3%
	Spread	*DBT* [°C]	26.8±3.1	23.2±5.8	23.7±5.3	24.6±2.7	23.5±2.5	20.3±2.6	20.1±2.3
		WVP [Pa]	1400±11%	1312±15%	1496±24%	1621±17%	1642±16%	1242±31%	1435±17%
		GHI [kWh m^{-2} w^{-1}]	40.2±9%	40.8±9%	45.5±7%	46.2±9%	45.6±9%	45.8±6%	42.1±6%
C	**Annual**	**Quantities**	**1 (6.2%)**	**2 (17.9%)**	**3 (14.7%)**	**4 (22.6%)**	**5 (18.2%)**	**6 (17.9%)**	**7 (2.2%)**
	Average	*DBT* [°C]	8.2±2.0	7.8±1.7	8.6±2.4	14.0±2.3	13.2±2.2	15.0±2.0	15.5±2.4
		WVP [Pa]	935.±9%	914±10%	941±13%	1208±12%	1098±17%	1223±13%	1289±17%
		GHI [kWh m^{-2} w^{-1}]	18.2±9%	18.3±9%	18.8±9%	27.0±10%	27.6±9%	29.1±11%	30.8±12%
	Spread	*DBT* [°C]	26.7±3.6	25.8±3.8	22.4±6.7	23.6±3.4	22.2±4.5	21.9±4.1	22.7±4.9
		WVP [Pa]	1424±10%	1392±13%	1290±17%	1598±15%	1334±31%	1460±27%	1620±12%
		GHI [kWh m^{-2} w^{-1}]	40.1±9%	41.5±10%	39.8±6%	44.8±9%	46.1±7%	46.3±6%	47.3±8%

Table 5: Global Hierarchical Clustering with weekly series of weather quantities: mean values and standard deviations for each class and annual averages and spreads

Global Hierarchical Clustering

			Classes and fraction of sample						
A	**Annual**	**Quantities**	**1 (3.7%)**	**2 (17.6%)**	**3 (18.5%)**	**4 (7.5%)**	**5 (11.3%)**	**6 (28.3%)**	**7 (12.8%)**
	Average	*DBT* [°C]	4.0±2.7	7.9±1.3	10.0±2.2	10.6±1.8	13.2±2.0	13.9±1.7	16.7±1.4
		WVP [Pa]	722±16%	918±6%	1016±12%	1064±10%	982±15%	1204±10%	1345±9%
		GHI [kWh m^{-2} w^{-1}]	17.4±11%	18.0±9%	21.2±13%	19.4±12%	29.4±08%	27.3±6%	30.8±8%
	Spread	*DBT* [°C]	30.8±5.9	26.1±3.0	25.4±2.9	14.9±2.2	20.3±2.8	24.7±3.1	19.7±2.3
		WVP [Pa]	1310±24%	1420±10%	1502±13%	1106±11%	911±20%	1707±13%	1428±15%
		GHI [kWh m^{-2} w^{-1}]	43.6±6%	38.2±6%	43.2±7%	39.2±8%	47.5±3%	47.9±4%	42.3±5%
B	**Annual**	**Quantities**	**1 (8.8%)**	**2 (12.5%)**	**3 (20.7%)**	**4 (29.5%)**	**5 (7.5%)**	**6 (11.0%)**	**7 (9.7%)**
	Average	*DBT* [°C]	6.2±2.8	8.7±1.7	8.9±1.2	13.1±1.6	14.6±1.5	15.2±2.0	16.9±1.2
		WVP [Pa]	833±16%	950±8%	963±7%	1099±13%	1269±11%	1253±12%	1354±10%
		GHI [kWh m^{-2} w^{-1}]	17.3±7%	20.4±13%	18.6±9%	27.0±10%	28.2±8%	28.3±8%	31.4±7%
	Spread	*DBT* [°C]	27.3±3.6	26.9±4.3	22.6±5.4	23.4±4.2	22.0±4.0	22.3±2.6	19.7±2.8
		WVP [Pa]	1350±13%	1486±12%	1304±15%	1408±29%	1615±22%	1563±19%	1495±20%
		GHI [kWh m^{-2} w^{-1}]	40.8±10%	41.5±10%	40.8±8%	46.3±8%	46.6±6%	45.1±8%	44.0±6%
C	**Annual**	**Quantities**	**1 (6.2%)**	**2 (17.2%)**	**3 (18.8%)**	**4 (21.3%)**	**5 (13.8%)**	**6 (17.6%)**	**7 (4.7%)**
	Average	*DBT* [°C]	5.6±2.3	7.7±1.5	9.9±1.2	14.1±2.3	14.2±1.7	14.2±2.3	16.3±1.6
		WVP [Pa]	784±13%	919±8%	1005±7%	1211±12%	1197±12%	1124±18%	1332±14%
		GHI [kWh m^{-2} w^{-1}]	17.9±11%	18.1±8%	19.8±11%	27.3±8%	27.3±9%	29.4±10%	30.8±9%
	Spread	*DBT* [°C]	28.9±5.8	26.3±2.5	21.7±5.1	24.0±3.5	23.8±4.0	20.2±3.7	21.0±3.9
		WVP [Pa]	1344±19%	1428±9%	1313±16%	1601±16%	1616±21%	1216±33%	1541±22%
		GHI [kWh m^{-2} w^{-1}]	43.4±6%	39.6±9%	40.9±8%	45.3±8%	47.1±6%	45.4±7%	47.0±7%

In order to discuss the groups' homogeneity in the different clusterings, the maximum standard deviations found for annual *averages* and *spreads* were compared to those determined in (Pernigotto *et al.*, 2019, Table 2 and Figure 6). Clusters with less than 10 climates were discarded in this analysis and the largest standard deviations observed in the current research compared to those detected in previous analyses. Regarding clustering (A), (1) in *DBT* partial clustering, slightly larger standard deviations were found for *DBT*, both *average* and *spreads*, and for *WVP averages*, while slightly lower ones were observed for *WVP spreads* and *GHI*; (2) in *WVP* partial clustering, improvements were recorded just for *WVP* and *GHI spreads*; finally, in (3) *GHI* partial clustering limited reductions of the maximum

standard deviations were detected for *DBT averages* and *GHI*. As far as clustering (B) is concerned, (1) in *DBT* partial clustering significant worsening of standard deviations was seen for *DBT spreads*, alongside negligible variations for the other statistics; (2) in *WVP* partial clustering larger standard deviations were found for *DBT* and *WVP spreads*, while reductions were noticed for *DBT* and *WVP averages* and for *GHI*; in (3) *GHI* partial clustering, standard deviations of all the weather statistics but *DBT averages* increased. Clustering (C) gave the worst performance in terms of groups' homogeneity: limited improvements were registered for *GHI* (in *DBT* partial clustering and, just for *GHI spreads*, in *WVP* partial clustering) and for *DBT averages* (in *GHI* partial clustering); worsening in homogeneity was observed in all partial clustering for the remaining weather statistics, particularly for *DBT spreads*. Global clusterings do not differ too much each other in terms of homogeneity: a significant worsening is observed for *DBT spread*, a general worsening for all variables except for *GHI*, which was found slightly improved.
In conclusion of this section, the adoption of weekly series of weather quantities brought mixed changes with respect to the original method based on monthly series, with improvements in climate clusters' homogeneity limited to the quantity characterizing solar radiation but with significant worsening in case of the ambient temperature. As a consequence, it should be observed that the calculation of shorter data series has not practical benefits in the framework of the proposed methodology.

3.2 Daily range weather quantities

Again, in this second analysis 7 climate classes were identified. This time, due to higher number of variables involved in clustering, i.e., 4 in each partial clustering and 10 in the global one, the 7 groups were determined at heights in the dendrograms equal to 0.8 and 1.3, respectively. Groups' sizes varied from 3 to 94 (*DBT* partial clustering), from 1 to 172 (*WVP* partial clustering), and from 4 to 130 (global clustering). As a whole, with respect to the classification in Section 3.1 and to previous researches (Pernigotto *et al.*, 2019), dominant larger groups emerged.
Although partial clusterings with *DBT* and *WVP* (Table 6) revealed some overlapping between classes considering the annual *averages* of weather statistics, e.g., *DBT* or *WVP*, those groups were found differentiated by $\widehat{DBT}$ or $\widehat{WVP}$ (see for instance in *DBT* partial clustering, classes $\mathbf{4^{DBT}}$, $\mathbf{5^{DBT}}$ and $\mathbf{6^{DBT}}$ as regards *DBT* and $\widehat{DBT}$ values). The same is confirmed in global clustering (Table 7): for example, while groups **2** and **3** have similar *DBT averages*, they have very different $\widehat{DBT}$ *averages*.
As expected, due to the presence of one or more larger classes, homogeneity is lower compared to the monthly-based clustering performed in (Pernigotto *et al.*, 2019). For both partial and global clusterings, a general worsening of homogeneity was detected, in particular for *DBT*; limited or negligible improvements were observed just for *GHI spreads*.

Table 6: Partial Hierarchical Clustering with additional weather quantities: mean values and standard deviations for each class and annual averages and spreads

***DBT* Hierarchical Clustering**								
		Classes and fraction of sample						
Annual	**Quantities**	**1 (2.5%)**	**2 (0.9%)**	**3 (29.5%)**	**4 (19.4%)**	**5 (11.0%)**	**6 (26.4%)**	**7 (10.0%)**
Average	*DBT* [°C]	4.0±2.4	10.3±5.5	11.0±3.3	11.9±3.1	12.2±3.2	12.5±3.4	13.6±3.9
	$\widehat{DBT}$ [°C]	6.4±1.0	1.7±0.1	7.2±1.2	5.3±1.1	2.9±0.9	6.3±1.0	2.9±0.9
	WVP [Pa]	734±14%	1039±34%	1040±17%	1077±14%	1096±16%	1122±16%	1198±18%
	$\widehat{WVP}$ [Pa]	320±22%	156±53%	385±51%	411±40%	310±70%	301±18%	363±74%
	GHI [kWh m^{-2} m^{-1}]	17.7±12%	21.7±26%	23.7±21%	24.2±24%	23.4±25%	25.1±16%	26.5±22%
Spread	*DBT* [°C]	33.9±3.9	20.2±5.2	24.9±3.9	22.1±4.3	18.5±4.7	24.3±3.0	22.8±4.3
	WVP [Pa]	1459±19%	1518±35%	1461±22%	1309±24%	1247±26%	1529±16%	1517±21%
	GHI [kWh m^{-2} m^{-1}]	42.9±4%	40.5±12%	43.5±10%	43.4±10%	42.6±11%	44.5±8%	44.7±11%

***WVP* Hierarchical Clustering**								
		Classes and fraction of sample						
Annual	**Quantities**	**1 (0.3%)**	**2 (8.4%)**	**3 (54.0%)**	**4 (12.2%)**	**5 (12.5%)**	**6 (4.7%)**	**7 (7.5%)**
Average	*DBT* [°C]	16.2	11.0±3.8	11.6±3.4	11.6±4.7	11.5±2.8	11.8±5.3	14.4±1.7
	$\widehat{DBT}$ [°C]	2.9	5.9±2.1	6.1±1.4	4.7±2.2	5.2±2.0	5.9±2.5	3.98±2.3
	WVP [Pa]	1022	1041±19%	1069±16%	1078±23%	1096±14%	1131±27%	1192±12%
	$\widehat{WVP}$ [Pa]	1112	619±26%	348±26%	545±17%	260±41%	131±59%	49.1±47%
	GHI [kWh m^{-2} m^{-1}]	32.0	23.4±29%	24.2±20%	24.7±24%	22.4±25%	24.4±20%	27.4±8%
Spread	*DBT* [°C]	21.9	23.8±3.9	24.5±4.1	22.5±4.7	18.9±5.0	24.0±5.1	24.6±3.8
	WVP [Pa]	582	1345±21%	1449±19%	1389±26%	1268±24%	1615±25%	1677±15%
	GHI [kWh m^{-2} m^{-1}]	49.4	42.4±10%	43.8±9%	42.4±10%	42.4±10%	44.5±9%	48.6±5%

Table 7: Global Hierarchical Clustering with additional weather quantities: mean values and standard deviations for each class and annual averages and spreads

Global Hierarchical Clustering								
		Classes and fraction of sample						
Annual	**Quantities**	**1 (1.2%)**	**2 (40.8%)**	**3 (6.2%)**	**4 (29.8%)**	**5 (8.4%)**	**6 (10.0%)**	**7 (3.1%)**
Average	DBT [°C]	9.5±4.7	10.9±3.5	11.2±2.9	11.9±4.0	13.3±3.9	13.7±1.5	14.7±1.6
	$\widehat{DBT}$ [°C]	2.2±1.0	6.3±1.6	2.9±0.9	6.1±1.1	3.1±1.1	6.9±1.0	2.1±0.6
	WVP [Pa]	988±31%	1027±16%	1081±13%	1096±19%	1153±19%	1173±14%	1259±10%
	$\widehat{WVP}$ [Pa]	172±43%	459±27%	151±47%	290±18%	594±30%	156±74%	52.4±45%
	GHI [kWh m^{-2} m^{-1}]	19.8±30%	23.6±24%	21.0±24%	24.2±19%	26.4±23%	26.8±8%	27.5±9%
Spread	DBT [°C]	20.4±4.2	24.0±4.4	16.6±4.5	24.3±4.3	22.5±4.5	24.5±3.2	22.6±4.9
	WVP [Pa]	1373±38%	1362±22%	1211±22%	1491±18%	1423±26%	1621±19%	1670±18%
	GHI [kWh m^{-2} m^{-1}]	42.1±12%	42.8±10%	41.1±11%	44.2±9%	43.1±10%	47.0±6%	48.9±5%

Considering the geographical distribution of the obtained groups (Figure 4):

- Class **1** (grey) includes two alpine climates (Kasprowy Wierch and Sniezka), at the border between Poland and Czech Republic, Reykjavik, Iceland, and Bergen, Norway;
- Class **2** (yellow) is the largest and most varied one, with 130 climates located mainly in Italy, in the Alpin region, and in the Balkans;
- Class **3** (blue) is composed by 20 localities in the British Islands and in the French and Spanish regions surrounding the Bay of Biscay;
- Class **4** (green) includes 95 climates, located mostly in Scandinavia, central and eastern Europe and Russia;
- Class **5** (purple) has 27 coastal localities, mainly in Italy;
- Class **6** (light blue) has 32 locations, distributed mostly in inland Spain;
- Class **7** (red) includes 10 cities along the coast of the Iberian peninsula.

Although different from the classification developed in (Pernigotto *et al.*, 2019), it can be noticed a good degree of land and geographical continuity in the clusters. According to the dendrogram, further subdivisions show the northern Scandinavian and the Russian locations separated from the rest of class **4**, and class **2** divided into 3 sub-groups, distinguishing mountain, inland and coastal climates, in particular for the Italian peninsula.

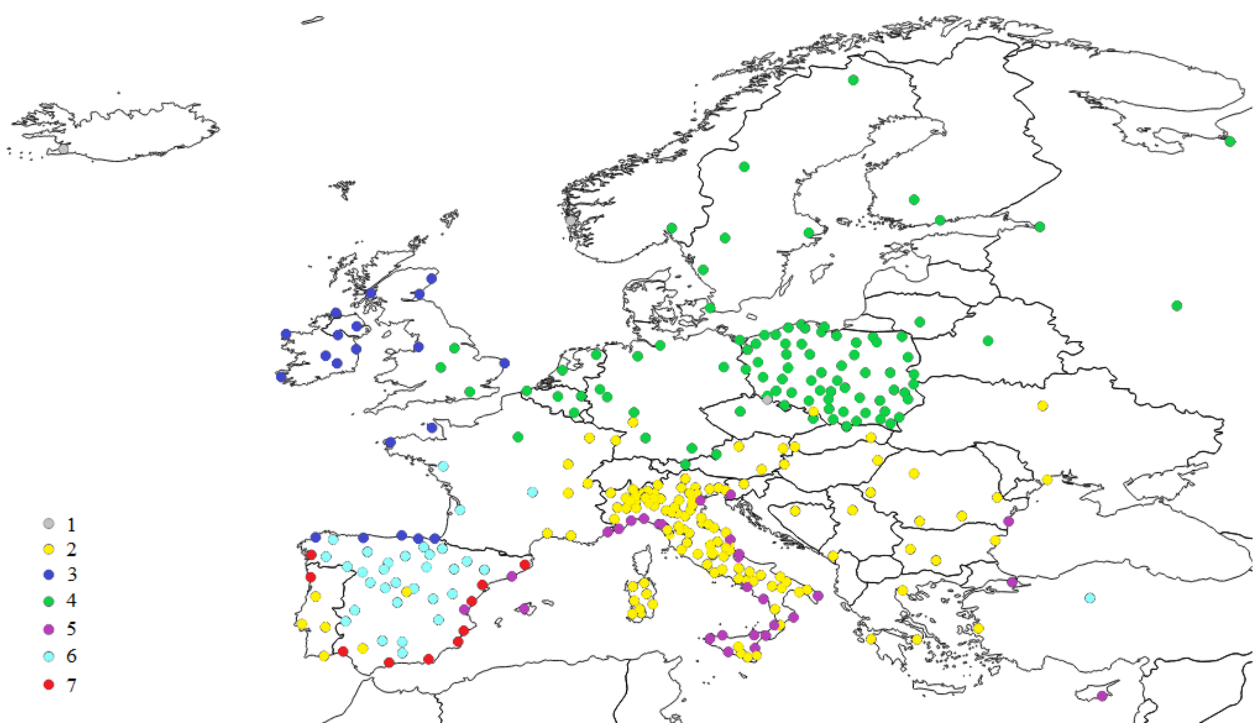

Figure 4: Distribution of the dataset of climates into 7 classes

4. CONCLUSIONS

This work tested two possible improvements to a methodology previously proposed by the Authors to perform climate clustering with the aim of identifying groups of locations sufficiently homogeneous for a robust selection of representative climates to use in building energy simulation analyses. According to the presented approach, typical or reference years are used as input to calculate monthly series of weather quantities which play a primary role on the building energy balance, i.e., the ambient temperature and humidity and the solar radiation. The series are then characterized by means of some weather statistics, e.g., annual average and spread of input series, and processed by means of a Hierarchical Clustering. The first goal of this research was to understand if a monthly time-discretization was adequate or some improvements were achievable using a shorter one, i.e., working with weekly series of weather quantities. The second goal, instead, aimed at discussing the impact of additional quantities describing the daily range of weather data, specifically for ambient temperature and humidity.
In order to answer to those research questions, the same dataset of more than 300 typical European climates already analyzed in previous studies by the Authors, was considered. Features of the generated climate classes were discussed and compared, considering also additional statistics besides average and spreads. In particular, the presence of distinct groups and their level of homogeneity were studied through mean and standard deviation values of weather statistics. We found that:

1. The adoption of weekly series of weather quantities can lead to groups different from those obtained with monthly series. Furthermore, significantly different results can be observed in Hierarchical Clustering, depending on the employed statistics. Among the tested alternatives, the combination of annual averages and spreads as statistical indicators seems the one giving the better results in terms of limiting interclass overlapping and enhancing uniformity.
2. As a whole, the improvements achieved by adopting shorter time-discretization of the inputs are limited and often related just to solar radiation. Furthermore, they are largely counterbalanced by important reductions in homogeneity for the other weather quantities, in particular for the ambient temperature. Consequently, the adoption of weekly series of inputs does not bring advantages in the framework of the proposed methodology.
3. As regards additional weather quantities descriptive of daily range of temperature and humidity, their inclusion in the clustering procedure led to a geographically meaningful classification with good degree of land continuity. Nevertheless, homogeneity was found decreased compared to previous analyses, requiring a further refinement of the largest classes. In general terms, the increase of the number of groups and a slightly loss of homogeneity look reasonable, especially for complex building systems for which daily variations of weather solicitation can have large effects on the energy performance.

ACKNOWLEDGEMENT

Funded by the project “Klimahouse and Energy Production”, in the framework of the programmatic-financial agreement with the Autonomous Province of Bozen-Bolzano of Research Capacity Building.

REFERENCES

ASHRAE. (2013). ANSI / ASHRAE Standard 169-2013, Climatic Data for Building Design.

Peel, M.C., Finlayson, B.L., & McMahon, T.A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11, 1633–1644.

Pernigotto, G., & Gasparella, A. (2018). Classification of European Climates for Building Energy Simulation Analyses. *Proceedings of the V International High Performance Buildings Conference at Purdue*. West Lafayette (Indiana, U.S.), 9-12 July 2018.

Pernigotto, G., Walsh, A., Gasparella, A., & Hensen, J.L.M. (2019). Clustering of European Climates and Representative Climate Identification for Building Energy Simulation Analyses. *Proceedings of Building Simulation 2019*. Rome (Italy), 2-4 September 2019.
Available at: http://www.ibpsa.org/proceedings/BS2019/BS2019_210938.pdf

Walsh, A., Cóstola, D., & Labaki, L.C. (2017). Review of methods for climatic zoning for building energy efficiency programs. *Building and Environment*, 112, 337-350.

Walsh, A., Cóstola, D., & Labaki, L.C. (2018). Performance-based validation of climatic zoning for building energy efficiency applications, *Applied Energy*, 212, 416-427.