Mapping the Urban Heat Island at the territory scale: an unsupervised learning approach for urban planning applied to the Canton of Geneva

Alessia Boccalatte, Marco Fossa, Martin Thebault, Julien Ramousse, Christophe Ménézo

Mapping the Urban Heat Island at the territory scale: an unsupervised learning approach for urban planning applied to the Canton of Geneva

Abstract. This study presents a fully reproducible clustering-based methodology for the assessment of the urban heat island intensity (UHII) at the territory scale, using parametric microclimate models and limited computational resources. In large-scale climate modelling, a common preliminary operation is to utilize the well-established Local Climate Zone classification to characterize the thermal response of urban areas based on morphology. With the increasing availability of urban datasets, data-driven approaches can be implemented to quantitatively derive meaningful urban features without relying on a standardized classification. The proposed methodology employs a Gaussian Mixture Model clustering algorithm to partition the urban territory into a suitable number of homogeneous microclimate zones, enabling the calculation and mapping of the UHII for each zone through the Urban Weather Generator (UWG) tool. The developed approach is applied to the Canton of Geneva, Switzerland, identifying ten microclimatic areas and analyzing the spatiotemporal variation of UHII. Results show yearly average values of UHII ranging from 1.7°C to 2.2°C, depending on urban morphology. The simulated values are partially validated by comparison with on-site measurements from two urban weather stations, yielding a satisfactory agreement. The methodology can support urban planning with the goal of avoid overheating through a large-scale mapping.

Keywords: *Urban Heat Island; Local Climate Zones; Urban Clustering; GIS-data; Urban Microclimate*

1. Introduction

The urban heat island (UHI) phenomenon, i.e. the local increase of the urban air temperature compared to the rural surrounding areas, is a major issue for global climate disruption (Palme & Salvati, 2021). In

addition to the global temperature rise of about 1.5°C and the ever more frequent climate anomalies such as heatwaves (IEA, 2021; IPCC, 2022; Pielke et al., 2022; Pyrgou et al., 2017), urbanization is responsible for an air temperature increase that may reach up to 12°C at peak in cities (Oke, 1982). This condition strongly increases the vulnerability of modern cities (Grimmond et al., 2010; Rajagopal et al., 2023) especially in Europe, which is particularly affected by global warming. Urban overheating negatively impact building energy consumption (He, 2019; Hwang et al., 2020), public health (Tong et al., 2021), air pollution (Y. Wang et al., 2021), thermal comfort (Alvarez et al., 2021), ecosystems (Dissanayake et al., 2020), economics and productivity (Memme & Fossa, 2022; Raalte et al., 2012). In the literature, numerous field studies have been presented for more than 450 worldwide major cities including London, U.K. (Kolokotroni et al., 2012), Barcelona, Spain (Salvati, Coch Roura, et al., 2017), Basel, Switzerland (Parlow et al., 2014), Sydney, Australia (Santamouris et al., 2018), Singapore (Bueno et al., 2015a), Wuhan, China (Huang et al., 2018), Los Angeles, United States (Vahmani & Ban-Weiss, 2016), Toronto, Canada (Y. Wang et al., 2015).

Despite the evidence of this phenomenon, most research related to building energy performance evaluation still fails to integrate UHI into energy demand and thermal comfort analyses (Lauzet et al., 2019; Mirzaei & Haghighat, 2010; Santamouris, 2014). The majority of building energy-related studies exploit non-local weather data from reference weather station measurements located outside the cities (e.g. airports). As demonstrated in previous research by the Authors (Boccalatte et al., 2020) and other studies (X. Li et al., 2019; Lima et al., 2019; Palme et al., 2017; Salvati, Coch, et al., 2017), this affects to a great extent the building energy-use predictions. The UHI strongly modifies the energy demand related to building HVAC systems and buildings, in turn, negatively impact urban air temperature and thermal comfort through heat losses. In this sense, microclimate models are becoming essential for both building design and urban planning to adequately consider local climate conditions and plan mitigation strategies.

1.1 Experimental Measurements of Urban Heat Island (UHI)

Given the relevance of the Urban Heat Island (UHI) phenomenon, substantial research has been conducted to investigate its magnitude and the characteristics through experimental observations. Remote Sensing (RS) data acquired through satellites (e.g. Landsat, MODIS, ASTER), drones, aircrafts have been extensively used to map the urban heat at the city scale (Venter et al., 2020). By utilizing thermal images from RS, land use and land surface temperature (LST) data can be derived to investigate the spatiotemporal variation of urban heat (AlDousari et al., 2022; Chen et al., 2023; Unal Cilek & Cilek, 2021). Most of the large-scale Urban Heat Island (UHI) estimates derived from Land Surface Temperature (LST) data refers to the Surface Urban Heat Island (SUHI), which measures the temperature of the surface of the built environment. Nevertheless, for thermal comfort studies and Building Energy Modeling (BEM), it is crucial to consider the air temperature instead of the surface temperature.

In recent years, the use of sensors to measure air temperature in urban areas has become increasingly popular (de Almeida et al., 2021). Besides the traditional urban meteorological networks, various techniques exist for crowdsourcing, citizen science weather stations (CWS), and mobile data (Chàfer et al., 2022; Muller et al., 2015; Romero Rodríguez et al., 2020). For instance, Netatmo urban weather stations provide small, flexible, and affordable sensors that can be autonomously installed by citizens at multiple locations throughout a city, offering a good level of spatial coverage for experimental data measurements (Benjamin et al., 2021; Brousse et al., 2022; Meier et al., 2015). Despite their value in UHI studies, sensors are subject to limitations related to device accuracy, placement, and maintenance, which may hinder capturing all the relevant information for comprehending the UHI effect. The data generated by these sensors provide uncertified observations, which may be misrepresentative. Several studies have reported significant daytime biases mainly resulting from improper shading of outdoor sensors (Varentsov et al., 2020). Additionally, while sensors may provide valuable insights, their usage is limited to providing point data, limiting the ability to evaluate the impact of mitigation strategies, changes in urban texture, or even projections with future weather scenarios. In this context, numerical simulation and modeling is still indispensable for decision-making procedures. To obtain a

comprehensive understanding of the UHI effect, it is thus essential to combine sensors with modeling approaches, such as street scale, local scale, and city scale microclimate models.

1.2 Modelling the Urban Heat Island (UHI) at different spatial scales

In recent decades, several modeling approaches have been proposed to investigate the Urban Heat Island (UHI) phenomenon at different spatial scales, ranging from the street scale to the city scale. However, a primary research gap still exists, which relates to the differences between microscale (street and local scale) and macroscale (city scale) models (Lauzet et al., 2019; Masson et al., 2020). Microscale models, while providing higher resolution and accuracy, are more computationally expensive and often limited to small areas (a street or a district), while macroscale models, although more computationally efficient, lack the necessary spatial resolution to capture the fine-grained features of urban environments. To overcome this gap, the main objective of this study is to implement a data-driven approach based on local scale parametric models that can bridge the gap between spatial resolution and computational efficiency. Reviews related to urban climate simulation and modeling tools illustrate the main differences across the urban scales and can be used as a reference to identify the most appropriate modeling approach for a given research question (Johari et al., 2020; Lauzet et al., 2019; Lobaccaro et al., 2021; Lun et al., 2009; Mutani & Todeschi, 2020; Sola et al., 2020; Tyagi et al., 2021).

1.2.1 City scale models

Regarding city scale models, several limitations have been identified and discussed in the literature (Mirzaei, 2015). The simulation domain is often up to several kilometers, encompassing an entire city and its surroundings. Urban morphological features can be estimated through approximated values of roughness length or parametrized using Urban Canopy Models (UCM), such as the popular Town Energy Balance (TEB) model (Afshari & Ramirez, 2021; Lemonsu et al., 2012; Masson, 2000). However, mesoscale models, such as MESO-NH (Lac et al., 2018) and the Weather Research and Forecast model (WRF) (Grimmond, 2017), have inherent limitations in terms of resolution making it challenging to observe local phenomena and capture differences among various urban morphologies.

1.2.2 Street and local scale models

Street scale models only cover the volume of air within the urban canyon, including local phenomena and detailed modeling of 3D geometry, heat transfers, and airflow regimes (Jänicke et al., 2021). Local scale models can be categorized into two types: detailed models and parametric models. Detailed models, such as Envi-met (ENVI-met, 2021), SOLENE-microclimat (Morille et al., 2015), take into consideration both fluid mechanics equations and 3D radiation equations, while SOLWEIG/UMEP (Lindberg et al., 2018) solves the detailed 3D radiations equations in real geometries. These models provide the most accurate representation of the urban environment, as they rely on a detailed representation of the area. However, due to the complex calculations involved, particularly those related to fluid mechanics (as in the case of ENVI-met or SOLENE-microclimat), performing year-long simulations over large spatial areas can be challenging.

In contrast, parametric models, including the Canyon Air Temperature (CAT) model (Erell & Williamson, 2006) and the Urban Weather Generator (UWG) (Bueno et al., 2012, 2014, 2015a; Bueno, Hidalgo, et al., 2013; Bueno, Norford, et al., 2013) define urban morphological features through a set of urban parameters that characterize the thermal properties of the district, making them computationally efficient even for year-long simulations. Despite their computational efficiency, they are typically limited to simulating a single district and cannot be directly applied at the city scale comprising multiple districts. This can be overcome if the whole studied area can be represented by a limited set of representative districts on which these parametric models can be applied.

1.3 Definition of representative districts for microclimate studies

The identification of representative urban morphologies within a city can be achieved through different approaches. Some studies rely on expert knowledge or on the administrative boundaries to identify homogeneous district morphologies within a city and calculate the related urban parameters for simulation purposes (Litardo et al., 2020; Salvati et al., 2020). Another “top-down” approach is the Local Climate Zone (LCZ) classification, originally introduced by Stewart and Oke (Oke, 1982; Stewart & Oke, 2012). The LCZ is used to categorize the landscape into 17 representative local climate zone

typologies that are assumed to have a unique air temperature regime under similar atmospheric conditions (Stewart & Oke, 2012). A typical range of urban parameter values that describe the urban shape, the characteristics of the vegetation, the human activity levels, the land cover is associated with each LCZ. If some data are lacking, the scheme also allows deriving the values of unknown parameters from look-up tables for the parameters of the other categories (e.g., for mean building height and density, aspect ratio, sky view factor, anthropogenic heat emissions, etc.). LCZs have been extensively applied to city-scale UHI studies based on both numerical simulations and field measurements (Brousse et al., 2022; Dimitrov et al., 2021; Fenner et al., 2014; Hashemi, 2020; Houet & Pigeon, 2011; Huang et al., 2018; Leconte et al., 2015; O'Malley & Kikumoto, 2022; Richard et al., 2018). Despite the advantages of this qualitative approach, as recently highlighted by Lipson et al. (Lipson et al., 2022), the growing availability of high resolution urban datasets (Biljecki et al., 2021; Milojevic-Dupont et al., 2023) and unsupervised machine learning classification techniques enables a transition to quantitative "bottom-up" approaches (Boccalatte et al., 2022).

This research aims to develop a data-driven approach that utilizes clustering techniques to quantitatively identify representative urban morphologies based on microclimate-related parameters. The proposed approach aims to bypass the subjectivity of expert knowledge as well as the abstraction of LCZs.

Unsupervised learning methods have been applied to a few urban studies (D'Acci & Batty, 2019; J. Wang & Biljecki, 2022), for example to identify representative building groups and predict the energy use at the city scale (Tardioli et al., 2018), to derive a detailed morphological classification of the urban form (Fleischmann et al., 2021, 2022), to identify typo-morphologies and perform thermal comfort simulations with Envi-met (Maiullari et al., 2021).

However, the use of unsupervised learning for investigating Urban Heat Island (UHI) at the city scale remains relatively unexplored. While previous studies have utilized clustering techniques to investigate the Surface Urban Heat Island (SUHI) at the city scale (Kwak et al., 2020), the coupling of such approaches with parametric models for large scale UHI mapping is still unexplored, making it the original contribution of this study. Specifically, this research aims at identifying representative microclimate zones within the Canton of Geneva (whose area is approximately 300 km^2) through GIS

data and clustering. Subsequently, the microclimate-related parameters of each zone are fed into the Urban Weather Generator (UWG) tool (Bueno et al., 2014, 2015a; Bueno, Hidalgo, et al., 2013; Bueno, Norford, et al., 2013), a parametric microclimate model that allows predicting the urban air temperatures at the district level based on urban parameters. The resulting workflow enables the simulation of the spatiotemporal variation of the UHI at the city scale with comparable accuracy to a local scale model but with much lower computational time. The simulated results have been partially validated against experimental measurements from two urban weather stations located in the city of Geneva yielding a satisfactory agreement.

2. Data sources, models and methods

As briefly introduced in Section 1, this study aims to evaluate and map the Urban Heat Island intensity (UHII) at the city scale. Given the vast extent, using a parametric model is the most effective way to achieve the objective due to computational limitations that hinder the use of detailed 3D modeling. The starting point of the presented methodology is the determination of the urban parameters required for numerical simulations. These parameters, primarily related to urban shape, vegetation, human activity levels, and building characteristics, serve as inputs for the simulations. In this study, the Urban Weather Generator tool is selected because of its validated accuracy and fast computational time (Section 2.1 and Section 2.2). Once the model is chosen, the following step is GIS data collection and pre-processing. At this stage, the urban parameters required for the selected model (in this case, the UWG tool) are extracted from GIS urban datasets (Section 2.3). Since GIS data can be either at the level of individual buildings or at a bigger scale (e.g. district), a certain number of pre-processing operations are performed to uniform data and obtain the averaged values of the urban parameters within a spatial area that is coherent with the selected simulation tool. This step is fundamental since a single building do not determine the urban microclimate, but a group of them does. To this end, the morphological tessellation technique is applied (Section 2.4) avoiding rough data averaging by means of a fixed grid. Then, starting from the averaged parameter values, an advanced clustering algorithm based on Gaussian Mixture Models (GMM) is employed to classify urban areas with similar characteristics, under the assumption of a similar impact

on the microclimate (Section 2.5). Finally, the UWG tool is used to numerically simulate the intensity of the UHI phenomenon in each urban zone, resulting in a map of the entire city that accounts for the local urban characteristics.

The developed workflow is schematically represented in Figure 1. The total area under consideration in this study corresponds to the administrative boundaries of the Canton of Geneva, which are depicted in the figure (boxes 4 and 5) by the black line.

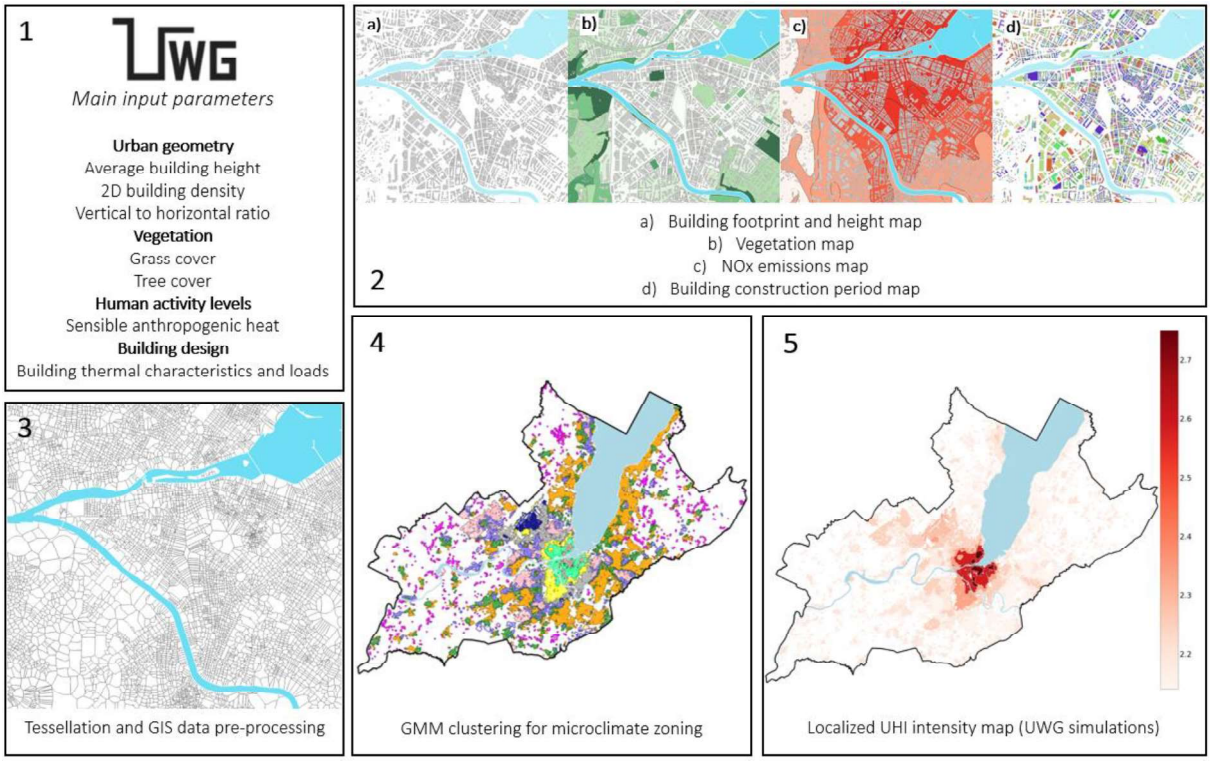

Figure 1: Schematic representation of the developed methodological steps applied to the Canton of Geneva.

2.1 The choice of the urban microclimate model: The Urban Weather Generator

The UHI intensity (UHII) is calculated through the Urban Weather Generator (UWG) model (Bueno et al., 2012, 2014, 2015b; Bueno, Norford, et al., 2013). The UWG is a parametric microclimate model that uses the principles of energy transfer and heat balance equations within the urban canopy to account for various physical processes such as solar radiation, thermal radiation, conduction, and convection. It is built upon the well-established Town Energy Balance (TEB) scheme (Masson, 2000), which is a two-dimensional representation of an urban canyon consisting of three surfaces: a wall, a road, and a

roof. The TEB scheme calculates the climate conditions, drag force, and heat fluxes of a district composed of identical urban canyons. To improve the representation of the interactions between buildings and the urban climate, the UWG integrates the original TEB scheme with a detailed Building Energy Model (BEM) derived from EnergyPlus algorithms (*EnergyPlus*, 2023). The UWG comprises four calculation components, including the Rural Station Model (RSM), Vertical Diffusion Model (VDM), Urban Boundary-layer Model (UBL), and Urban Canopy and Building Energy Model (UC-BEM), which are fully described in (Bueno et al., 2015b). The UWG structure allows to predict urban canopy air temperature based on rural weather data and a parametric description of the urban area. The UWG tool is freely available to download at (*GitHub - Ladybug Tools/Uwg*, 2022). Originally written in MATLAB in 2013 (*GitHub-Jiachen-Mao/UWG_Matlab*, 2021), the tool has been continuously updated and is now available into a Python library and as a grasshopper plugin. This study utilizes the latest Python version of the tool (V5).

To run the simulations the UWG takes two inputs: a meteorological file in a .epw format from the nearest rural weather station and an .xml format file (Extensible Markup Language) describing the urban parameters. The output is a modified .epw format weather file in which air temperatures have been adjusted based on the local urban characteristics. The main input urban parameters can be subdivided into four main categories: urban geometry, vegetation, human activity levels, and building characteristics (including thermal properties of construction elements and building loads). Table 1 summarizes the most relevant parameters required for this study. These parameters have a significant impact on local microclimate changes and can vary widely across different areas of the city. It's worth noting that a certain number of other parameters related to general simulation settings (such as latitude and longitude of the city, day and night boundary layer height, road materials albedo and conductivity, etc.), or to specific building-related variables (such as wall and roof albedo, glazing ratio, solar heat gains from windows, etc.), are also needed, as fully described in (Bueno et al., 2014), but for sake of brevity they are reported in Appendix A.

Table 1: UWG main input parameters for UHI simulation

Urban geometry		
$\overline{H}$	[m]	Average height of buildings
ρ_{urb}	[-]	Urban area building plan density: ratio between built and un-built area
VH	[-]	Vertical to horizontal ratio: ratio between façade area and plan area of the site
Vegetation[1]		
ρ_{grass}	[-]	Grass coverage: fraction of the urban ground covered in grass/shrubs
ρ_{trees}	[-]	Tree coverage: fraction of the urban ground covered in trees
Human activity levels		
Q_{sens}	[W/m^2]	Non-building sensible heat at street level: heat from cars, pedestrians, street cooking, etc.
Building characteristics[2]		
R_{wall}	[m^{2}K/W]	Thermal resistance of walls
R_{roof}	[m^{2}K/W]	Thermal resistance of roof
U_{window}	[W/m^{2}K]	Thermal transmittance of windows
i	[ach]	Building infiltration rate

Notes:

1. Since in the reference vegetation map the vegetation coverage is not differentiated between grass and trees, the vegetation coverage value is split in half between ρ_{grass} and ρ_{trees} resulting in ρ_{grass}=ρ_{trees}
2. Building characteristics are derived from the building construction period

Several studies have conducted sensitivity analyses on the input parameters of the UWG model to determine their significance (Alchapar et al., 2019; Litardo et al., 2020; Mao et al., 2017; Salvati, Palme, et al., 2017). The results indicated that the most influential parameters are those related to urban geometry and sensible heat from traffic. In contrast, the impact of vegetation parameters in some cases

is lower due to the simplified modeling of vegetation in the UWG model (Salvati, Palme, et al., 2017). The model assumes that a fraction of absorbed solar radiation is transformed into latent heat and does not contribute to the temperature increase in the urban canyon and it neglects the impact of tree shading on building walls and roofs, considering only its effect on the road. Regarding building characteristics, the sensitivity analysis performed by (İsmet Berke, 2010) showed that the thermal resistance of walls and infiltration rate are among the most influential building characteristics. However, the model shows a low sensitivity to building albedo values. Overall the parameters considered for this study (Table 1) are deemed relevant for conducting the simulations.

2.2 Validation and Limitations of the Urban Weather Generator Model for Urban Microclimate Analysis

The accuracy of the UWG has been extensively validated using both field measurements and simulations in different urban environments including Basel (Switzerland), Toulouse (France) (Bueno, Norford, et al., 2013), Singapore (Bueno et al., 2015a), Boston (USA) (Street et al., 2013), Abu Dhabi (UAE) (Mao et al., 2017), Rome (Italy), Barcelona (Spain) (Salvati et al., 2016). The validation results show an average RMSE error of about 1-2K with respect to hourly temperature predictions.

However, it is crucial to recognize that the UWG model is based on a simplified computational model that may not fully capture site-specific microclimate effects beyond spatially averaged results. Nonetheless, these simplifications are necessary for computational efficiency, enabling large-scale simulations over extended periods which is the main objective of this study. The limitations of the model in simulating vegetation and advection from rural to urban boundary layers are considered acceptable due to the computational efficiency required for such simulations. Additionally, the UWG model does not account for the effects of large water bodies, which have been excluded from the study to date, leading to the assumption that the air above Lake Geneva has similar characteristics to rural air.

Despite these limitations, the UWG model has been extensively validated and shown to be suitable for a variety of urban environments. The UWG limitations are considered acceptable for this case study bearing in consideration that the UWG model performs best for urban sites characterized by low wind conditions, where the UHI intensity is primarily due to urban features and anthropogenic heat release.

2.3 Urban data sources for urban parameter derivation in the Geneva Canton case study

The case study is the whole Canton of Geneva (46°13'05" N, 6°09'58" E, Switzerland), sizing about 300 km^2 and comprising about 60,000 buildings. The reference urban dataset is the Geneva Territory Information System (SITG) (*SITG | Le Territoire Genevois à La Carte*, 2023) which is an open-source repository of hundred geodata sets related to town planning, mobility, energy, nature, and even climatic analyses. Data are stored into geospatial vector data or into raster data.

Here, a set of SITG vector data in shapefile format are used to derive the urban parameters required for simulations. The shapefile format is used to store the geometric location and one or multiple attributes of a geographic feature, which can be a point, a line, or a polygon (area). The reference geographic feature can be represented by individual buildings (e.g. building height data) or a portion of space delimited by predefined boundaries that fit the stored information (e.g. vegetation data). In this study, three shapefiles are used to derive the urban parameters. *Cad.batiment.hors.sol* stores several data about each of the 60,000 buildings. Among them, the building height data are used to calculate the urban parameters related to building geometry and the building construction year data are used to derive the building element thermal characteristics, as explained in more detail in Section 2.5. *Ecopot.za* includes data related to biodiversity and vegetation for each district of the city, and it is used to derive the vegetation parameters. Finally, *Immissions.no2.moyenne* is a map of the yearly average NO_2 concentration within the city and it is used to derive the anthropogenic heat emissions, as explained in more detail in Section 2.5. Indeed, NO_2 emissions is related to vehicle exhausts, which account for the largest part of the anthropogenic heat emissions.

It is important to note that the reliability and accuracy of data obtained from large open source urban datasets are critical, particularly with respect to data frequency and updates. While sources like OpenStreetMap provides a comprehensive and accessible source of urban data, the accuracy and reliability of the data may not always be sufficient. In the present study, building height data were obtained from a high-precision LIDAR survey conducted by swisstopo in 2019, while the other data (building construction period, vegetation, and NO_2 emissions) were updated less than one year ago.

2.4 Data pre-processing: calculating the average urban parameters with tessellation and spatial weights

The key point for the calculation of the urban parameters is to define the reference area that fits the microclimate model used for numerical simulation. The Urban Weather Generator is a district scale model able to simulate the urban temperatures within a 200/300 m radius reference area. This value aligns with the definition of a district in the Local Climate Zone (LCZ) framework. This framework considers a district to be a 300 m radius urban area with similar land use, ground cover, and building density. The most straightforward approach to quantify the average urban parameters within a coherent reference area is through a squared grid defining a certain number of cells measuring about 500 m each. Although the use of a fixed regular grid is a fast technique for data averaging, it represents an arbitrary segmentation that is not consistent with the complexity of the urban texture. Furthermore, since the UWG has to be fed with 'average' urban parameters of the site, using a square grid may lead to a loss of complexity and local specificities. In what follows an alternative approach is proposed to partition the city into homogeneous areas with respect to the input parameters needed for simulations. To this aim, a polygon-based adaptation of Voronoi tessellation is applied to this study. In particular, a recently developed Python toolkit named *momepy* (Fleischmann, 2017; Fleischmann et al., 2020, 2021) is employed to define morphological cells based on building footprints. The morphological tessellation function aims to derive the "surface of influence" of a building, i.e. the smallest spatial unit that delineates the portion of space around each building (tessellation cell). In a first phase, the tessellation cell is used to calculate the building area plan density (ρ_{bld}) and the vertical–to–horizontal ratio (VH_{bld}) of each building through Eq. 1 and 2:

$$\rho_{bld} = \frac{A_{bld}}{A_{tess}} \quad (1)$$

$$VH_{bld} = \frac{P_{bld} \cdot H_{bld}}{A_{tess}} \quad (2)$$

where A_{bld}, P_{bld}, and H_{bld} are respectively the footprint area, the perimeter, and the height of each building and A_{tess} is the area of the related tessellation cell.

As briefly introduced in Section 2.3, urban datasets collect GIS data that can be related to different geometrical entities based on the type of information described. In particular, building height (as well as the derived ρ_{bld} and VH_{bld}) and construction period data are related to individual buildings, whereas vegetation and NO_2 data relate to coherent portions of land. In the first case, where data are related to each building, the value assigned to each tessellation cell is the one of the building itself, whereas in the second case, the value assigned is obtained by superimposing the vegetation and NO_2 maps.

At this stage, the raw GIS data has undergone processing to compute the required urban parameters pertaining to each building/tessellation cell. As a result, a series of values relating to urban geometry, vegetation, human activity levels, and building characteristics are associated to each individual element (building/tessellation cell). However, measuring individual characters is insufficient to capture the impact on microclimate, as it is necessary to calculate the average values of these parameters within a defined area that corresponds to the extent of the UWG simulations (200/300 m radius). This is because the urban microclimate is altered by a group of buildings rather than by a single isolated building. To this aim *spatial weights* are used to calculate the contextual tendency of each urban parameter. Spatial weights are mathematical structures that identify the neighbouring buildings of a given building, i.e. the buildings whose tessellation cells are adjacent to the building in question. As depicted in Figure 2, it is possible to define n orders of spatial weights depending the number of times adjacency has been verified. For instance, if a generic reference building (shown in red in Figure 2) is considered, the 1st order spatial weights are the buildings (shown in yellow) whose tessellation cells touch the reference. This can be seen as the first 'belt' of the buildings (in yellow) around the reference one. The 2nd order spatial weights refer to the buildings (shown in green), whose tessellation cells are adjacent to the first order cells. This process is repeated for all the buildings and till an nth order level corresponding to the radius (from reference building) required by the UWG model.

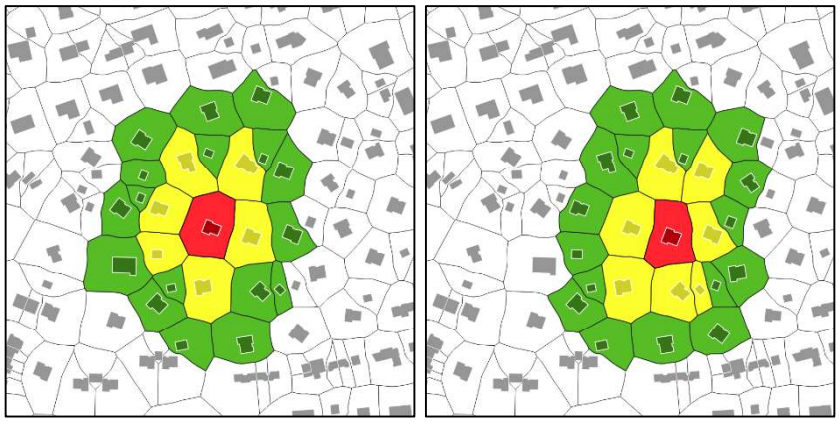

Figure 2: 1st and 2nd order cells (yellow and green respectively) determined through spatial weights with respect to a reference building (red). The operation is repeated for all the buildings.

In the present study, 3rd order spatial weights are chosen to calculate the averaged urban parameters so that a group of approximately 50 adjacent buildings are included, coherently with the extent of the UWG simulation (200-300 m radius). The averaged urban parameters, noted $U_{dist,j}$, replace the individual urban parameters for the jth building, $U_{bld,j}$, and they are representative of the district immediately surrounding the considered building. In more detail, $U_{dist,j}$ are calculated as in Eq 3:

$$U_{dist,j} = \frac{U_{bld,j} + \sum_{i=0}^{N_{neigh,3}} U_{bld,i}}{1 + N_{neigh,3}} \quad (3)$$

where $N_{neigh,3}$ is the number of 3rd order neighbouring buildings (i) around the reference jth building, and $U_{bld,i}$ are the parameter values associated with each of them. In this way, for each building it is possible to calculate averaged urban parameters, that are representative of the district composed of this building and the surroundings 3rd order neighbours. This approach allows for the input to the UWG model to be representative of a district rather than just a single building. Therefore, it results in values that are not significantly different for neighbouring buildings, which are likely to be part of the same

microclimate. This approach also allows for the consideration of morphological transition zones between the high-density city centre and the sparsely built rural areas.

2.5 Inference of human activity levels and building characteristics

Unlike urban geometry and vegetation parameters, which can be directly derived from GIS pre-processing operations on shapefiles (respectively *Cad.batiment.hors.sol* and *Ecopot.za*), the non-building sensible heat at street level (Q_{sens}) and building characteristics (R_{wall}, R_{roof}, U_{window}, i) need to be inferred from other data.

In order to determine the value of non-building sensible heat (Q_{sens}) at ground level, a maximum value of 20 W/m^2 is assumed based on literature studies (Mao et al., 2017). This value represents the highest heat emissions in the densest and busiest areas of the Canton according to similar studies. To determine the values for the other areas, a linear relationship is established between Q_{sens} and the yearly average NO_2 concentration (whose maximum value is 38 µg/m^3 for the Canton of Geneva), according to the *Immissions.no2.moyenne* shapefile. The reasoning behind this is that NO_2 emissions are related to vehicle exhausts, which can be considered proportional to the anthropogenic heat emissions. By scaling the maximum value with the maximum NO_2 concentration, a range of Q_{sens} values is obtained, representing the non-building sensible heat for the different areas of the Canton. However, it should be noted that in future scenarios where a significant increase in electric vehicle usage is expected, this assumption may need to be re-evaluated, taking into account the heat emission rates associated with electric vehicles (Mussetti et al., 2022).

The building characteristics are inferred based on their construction period contained in *Cad.batiment.hors.sol* shapefile. The reference values are obtained from a study by Tardioli et al. (Tardioli et al., 2020), which derived the main building characteristics based on building energy modelling standards in Switzerland (SIA 380/4, SIA 382/1, SIA 385/2) and on the expertise of the Department of Planning, Housing and Energy (DALE). The UWG tool allows for consideration of three different construction periods, i.e. before 1980, between 1980 and 2000, and after 2000. In this study, the building characteristics from Tardioli at al. is averaged to align with the required construction

periods, as reported in Table 2. It should be noted that it would have been technically feasible to include building refurbishment in the analysis but this aspect is not considered due to the lack of available data and the relatively low refurbishment rate in the Canton of Geneva (Flourentzou, 2019). The other building-related parameters which are assumed to be equal across the three construction periods, are detailed in Appendix A.

Table 2: Reference values of building characteristics based on different construction periods

Name	Units	Before 1980	Between 1980-2000	After 2000
R_{wall}	[m^2K/W]	1.28	0.47	0.21
R_{roof}	[m^2K/W]	1.01	0.38	0.21
U_{window}	[W/m^2K]	3.22	1.95	1.40
i	[ach]	1.10	0.60	0.45

2.6 Gaussian Mixture Model (GMM) Clustering

Clustering is a common unsupervised learning method that groups data based on their similarities. In this study, clustering is used to group portions of land that are expected to have similar microclimatic conditions. This is achieved by normalizing and using as inputs for the clustering algorithm the average urban parameters that are considered representative of the thermal response of a group of buildings.

The choice of the clustering method depends on the type of data, the purpose of the study, and the assumptions made about the data distribution. The selected clustering algorithm is the Gaussian Mixture Model (GMM) clustering due to its suitability to handle complex and varied data distributions such as similar urban applications (Fleischmann et al., 2022; Ma et al., 2021; Quan, 2020; J. Wang & Biljecki, 2022). Other popular methods for clustering include k-means, hierarchical clustering, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN), but they may not always be optimal for urban studies with similar purposes, as they have certain limitations. For example, k-means assumes that clusters are spherical and equally sized, which may not hold for this type of urban data. DBSCAN is a density-based algorithm and could not achieve satisfactory results with multi-density data

distributions (Cesario et al., 2020). Hierarchical clustering is computationally expensive for large datasets as in this case. GMM, unlike other unsupervised algorithms, assumes that each cluster corresponds to a multi-dimensional Gaussian probability distribution, that is often used in statistical modeling as it provides a suitable way to represent complex data distributions. GMM uses the Expectation-Maximization (EM) approach (Reynolds, 2015) to fit data points to a mixture of K multi-dimensional Gaussian distributions (one for each cluster) which are randomly generated starting from a set of means and a covariance matrices. The EM algorithm iterates over the Expectation (E-step) and the Maximization (M-step) until it converges. Once the total number of clusters (K) is determined, each k^{th} component can be defined through a mean of μ_k and a covariance matrix $\sum_k$. The mixing coefficients for the k^{th} cluster are defined as π_k, with the constraint that $\sum_{k=1}^{K} \pi_k = 1$, so that the total probability distribution is 1. The probability density function of point x is calculated as in Eq. 4:

$$f(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \sum_k) \quad (4)$$

Where x is the data point being evaluated, $N(x|\mu_k, \sum_k)$ is the Gaussian distribution with mean μ_k and covariance $\sum_k$, and π_k is the mixing coefficient for the k^{th} cluster. Then, the M-step is used to determine the maximum value of the log-likelihood function. In this study, the GMM clustering algorithm is implemented through a Python script by means of the *scikit-learn* library (Pedregosa et al., 2011). To identify the optimal number of clusters, the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) scores (Schwarz, 2007) are calculated and plotted for a range of potential cluster numbers, from 2 to 15, as shown in Figure 3. The selection of 15 clusters as the maximum limit is made to ensure that each cluster represented a sufficient number of buildings. Choosing more clusters would have added unnecessary complexity without significantly identifying representative clusters. Lower BIC and AIC scores indicate a higher quality of clustering performance. In this study, two distinct plateaux are identified in the results, at 8 and 10 clusters. Both options are carefully evaluated and it is

determined that 10 clusters provide a more fine-grained classification of the urban textures in the Geneva Canton, effectively capturing the required diversity. On the other hand, 8 clusters offered a simpler, more generalized representation, but failed to distinguish between the old medieval area and the high-density city center.

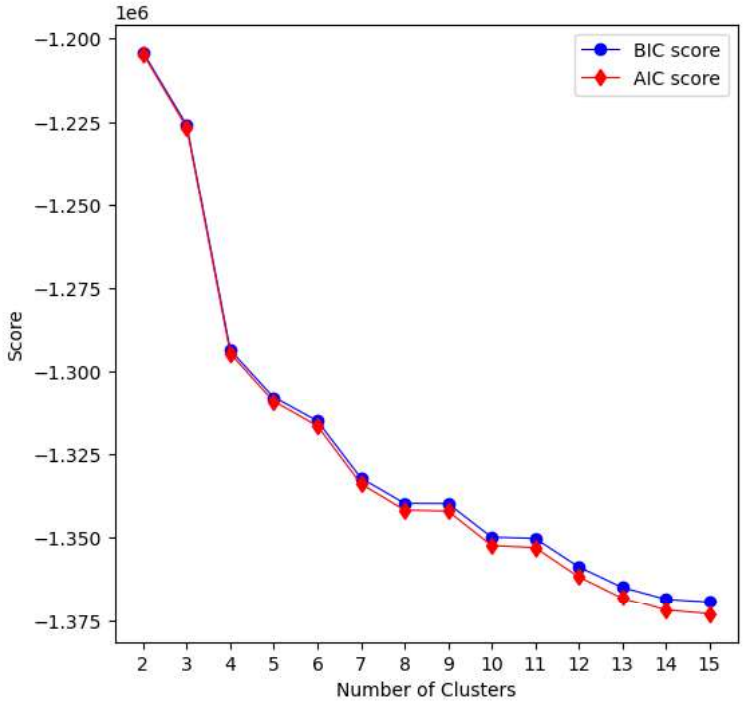

Figure 3: BIC and AIC scores per number of clusters

Once the cluster are formed, the average urban parameters of each cluster ($U_{cluster}$) are calculated and used as inputs to the UWG simulations.

2.7 UHII calculation and mapping through the Urban Weather Generator and validation towards real data

The systematic approach developed in this study allow the classification of urban textures in the Geneva Canton. In brief, the first step involves the determination of the building urban parameters (U_{bld}) for each of the 60,000 buildings in the study area using GIS data (Section 2.3). Next, the district urban parameters (U_{distr}) associated with each building are calculated through tessellation and spatial weights (Section 2.4). Finally, a Gaussian Mixture Model (GMM) is applied to classify the buildings into representative microclimatic clusters, reducing the number of alternatives from 60,000 to 10 (Section

2.6). From these clusters, the average cluster urban parameters ($U_{cluster}$) are calculated. This three-step approach is critical in gaining a comprehensive understanding of the representative microclimate zones within the Canton and classifying them effectively.

The $U_{cluster}$ parameters are then used as inputs to perform hourly microclimate simulations (10 simulations, one for each cluster) over the course of a typical meteorological year. The Urban Weather Generator (UWG) modifies the rural weather station data file (.epw format) to create a new urban weather file with adjusted air temperatures based on the predetermined urban parameters. The source rural weather station data is obtained from the reference EnergyPlus weather file for the Canton of Geneva (Geneva 067000 IWEC) available at (*EnergyPlus*, 2023). The latter arises from the TMY (Typical Meteorological Year) weather file which was built based on more than ten years of real measurements recorded at the Geneva International Airport weather station (MeteoSwiss). The station is located on the city boundaries and it is officially recognised by the World Meteorological Organization (WMO). As a result of the UWG simulations, the spatiotemporal variation of the Urban Heat Island Intensity (UHII) is analysed and a UHII map of the whole Canton is created. The UHII is here defined as the positive difference between the urban and the rural air temperature as in Eq. 4:

$$\mathrm{UHII} = (\mathrm{T_{urb}} - \mathrm{T_{rur}}) \qquad \text{if } \mathrm{T_{urb}} > \mathrm{T_{rur}} \tag{4}$$

In order to verify the accuracy of the spatiotemporal variation of the Urban Heat Island intensity (UHII), a partial validation is carried out. Since the Typical Meteorological Year (TMY) file used in the simulation is a standard weather file and does not represent actual measurements, real data was necessary for validation. As a result, measurements from a rural weather station located in Brenex (used as the input rural weather file for UWG simulations) and two urban weather stations in the Canton of Geneva (situated in the Battelle and Prairie areas) were used for validation. The urban weather stations were monitored by the University of Geneva throughout 2019, and the measured data were compared with simulated data obtained from the UWG. It is important to note that this validation using real data is only partial, as only two out of the ten total clusters (i.e., Battelle and Prairie areas)

could be compared to actual data due to the lack of supplementary reliable urban weather stations. The Results section presents the outcomes of this partial validation.

3. Results

The following sections present the results of the previously outlined methodology. Section 3.1 presents the classification of the Canton of Geneva into representative microclimate zones using the Gaussian Mixture Model (GMM) clustering. In Section 3.2 the temperature distribution of the rural and urban areas is analysed as a result of the UWG simulations. Section 3.3 focuses on the spatiotemporal variation of the Urban Heat Island Intensity (UHII) among the identified urban clusters. This section evaluates the microclimatic impact of the different urban clusters in terms of average and maximum monthly values ($UHII_{ave,month}$ and $UHII_{max,month}$ respectively). Section 3.4, explores the temporal variation of the UHII through the analysis of average hourly intensity values ($UHII_{ave,hour}$), highlighting the differences between nighttime and daytime. Finally, Section 3.5 is dedicated to the comparison between the UWG simulated values and on-site measurements.

3.1 Clustering results

As a result of the GMM clustering algorithm the 10 different homogeneous microclimatic zones have been identified and a map of them is shown in Figure 4. Each cluster C_j (with j ranging from 1 to 10) is represented by a different colour and the map shows that the clusters are well-defined and compact, effectively showcasing the morphological variations within the urban environment. White areas in the map represent unbuilt areas where the local conditions are thus expected to pertain the rural ones.

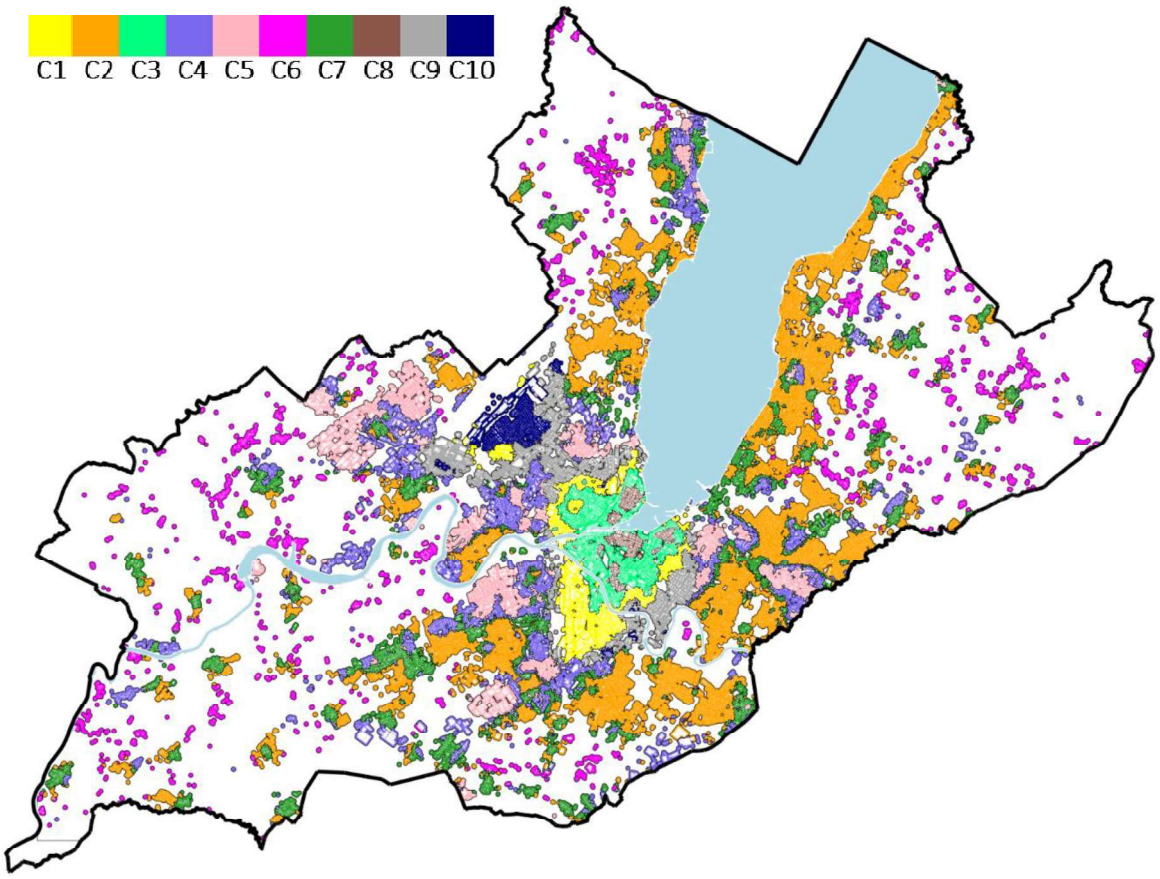

Figure 4: Ten homogeneous microclimatic clusters of the Geneva Canton identified through GMM clustering. Each cluster is represented by a different colour and white parts represent the surrounding rural areas.

Table 3 lists the number of buildings and the average urban parameters for each cluster ($U_{cluster}$), which served as inputs for the Urban Weather Generator. From a morphological perspective, clusters C2, C4, C6, and C7 can be categorized as low-density urban areas, with the highest number of buildings and low impact of sensible anthropogenic heat. These areas generally feature well-spaced, low-rise buildings. On the other side, C1, C3, C8, and C9 are high-density and poorly vegetated urban areas. Here the impact of sensible anthropogenic heat is higher and the building construction period is typically before the year 1980. Clusters C5 and C10 can be considered intermediate urban typologies. Cluster C5 consists of medium/high-rise buildings in a compact urban texture, with a significant share of vegetation, low

impact of human activities, and more recent building constructions. On the other hand, cluster C10 represents recent low-rise building areas that are significantly impacted by nearby human activities.

Table 3: $U_{cluster}$ related to the ten clusters identified through GMM algorithm

Cluster ID	N° of buildings	$\overline{H}$ [m]	ρ_{urb} [-]	VH [-]	$\rho_{grass/trees}$ [-]	Q_{sens} [W/m^2]	Construction period [%][1]		
							Pre1980	1980-2000	Post2000
C1	2856	17.8	0.31	1.5	0.13	8.3	70	15	15
C2	21676	7.5	0.11	0.4	0.33	<1.5	47	32	21
C3	3916	21.6	0.39	2.1	0.06	16.5	77	13	10
C4	6826	11.3	0.15	0.5	0.27	<1.5	51	23	27
C5	4195	16.81	0.22	1.0	0.21	<1.5	60	23	17
C6	3135	7.7	0.08	0.3	0.47	<1.5	58	27	16
C7	10402	9.5	0.18	0.7	0.29	<1.5	52	29	19
C8	1827	20.6	0.55	3.2	0.03	20.0	75	10	15
C9	2767	16.3	0.17	0.8	0.22	4.0	64	20	16
C10	1234	9.5	0.14	0.5	0.28	11.7	46	30	24

Notes:

1. The building construction period is expressed as percentage with respect to the total number of buildings within each cluster

3.2 Comparison of rural and urban temperatures using Probability Density Functions

In order to compare the simulated urban temperatures from the UWG with rural conditions, the probability density functions (PDFs) are calculated for each month of the year. The PDFs provide a smoothed representation of data that is useful to estimate and visualize the distribution of temperatures evidencing their density (probability of observing a specific temperature value) and distribution among rural areas and the ten identified urban clusters. To provide a more concise representation of the results,

two representative months are selected to show the temperature distribution in the coldest and hottest periods of the year. February is chosen to represent the coldest months (September to April) while August represents the hottest months (April to September). Figure 5 shows the PDFs for February (left) and August (right). The black line represents the rural conditions, whereas the different colored lines (coherent with the map shown in Figure 4) are related to the ten urban clusters. In February, the distribution of urban temperatures has the typical bell shape, with values uniformly distributed around the maximum value which is between 0°C and 5°C. The distribution of the urban temperatures is similar in the shape but the curves are shifted upward and slightly to the right, indicating a higher frequency of higher temperatures. It can be also noticed that cluster C8 shows an inflection point between 5°C and 10°C, indicating higher temperatures compared to the other urban clusters. In August, the temperature distribution has a different and asymmetrical shape, with both rural and urban distributions having two peaks (of which one has an intensity equal to about half of the highest value). The rural temperature has its maximum peak reached between 15°C and 20°C and the lower one between 25°C and 30°C. The pattern for all urban temperature distributions is exactly the same, with the two peaks of the rural temperatures being inverted in comparison to the urban temperatures.

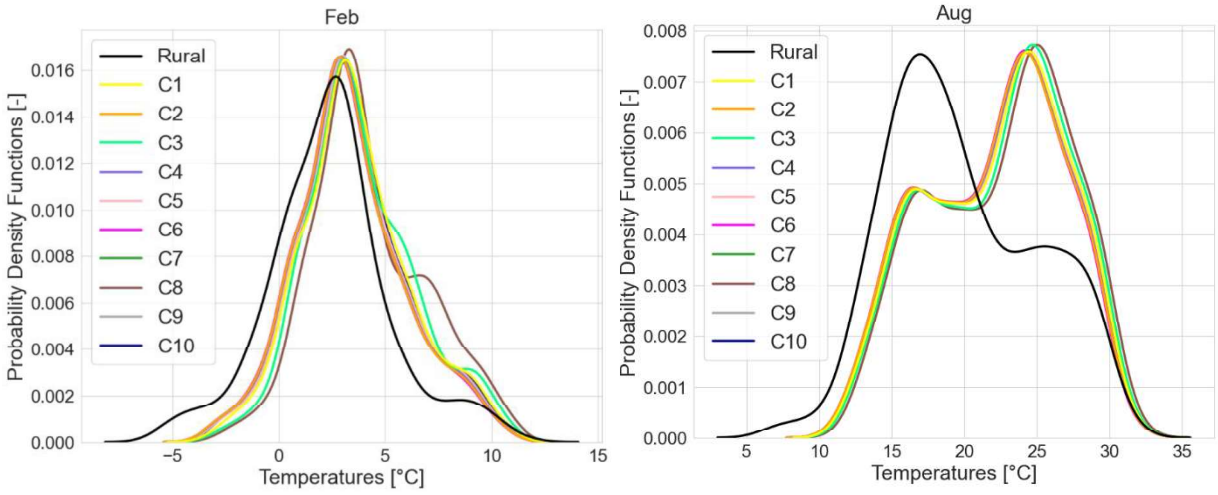

Figure 5: Probability density functions of rural and urban temperatures. The representative months are February (left) and August (right)

3.3 Average and maximum monthly UHI intensity comparison among the identified clusters

In this section, the Urban Heat Island Intensity (UHII) for each urban cluster is investigated. As a yearly average the UHII ranges from a maximum value of 2.2°C for cluster C8 and a minimum value of 1.7°C for the low-density areas, namely clusters C2, C4 and C6, evidencing the impact of urban morphological features on overheating. To investigate the annual variability of the UHI the average monthly intensity ($UHII_{ave,month}$) is calculated for each cluster and shown in Figure 6. In general, it can be observed that the thermal behaviour of the identified urban clusters is clearly distinguishable. The ten clusters may be separated into four distinct groups according to their impact on microclimate. The groups, starting from the most impactful to the least, are clusters C3 and C8 (group 1), cluster C1 (group 2), clusters C5, C9, C10 (group 3), and clusters C2, C4, C6, C7 (group 4). The results align with the expectations based on the average urban characteristics of each group: group 1 represents the high-density historical city center, group 2 is a transition area near the city center, group 3 includes both medium-rise suburban areas (C5 and C9) and low-rise suburban areas with high anthropogenic impact (C10), and group 4 encompasses the peripheral open low-rise suburbs. The $UHII_{ave,month}$ difference between group 1 and 4 is considerable, ranging from about 0.4°C between April and September to 0.6°C between October and March. Considering all clusters, the minimum $UHII_{ave,month}$ value is of 0.73°C (cluster C6 in December), whereas the maximum value is of 3.13°C (cluster C8 in May). The months with the highest $UHII_{ave,month}$ are May, July, and August, with an overall overheating of 2.8°C, compared to the 0.9°C observed during November, December, and January. The results indicate that the UHII is not necessarily highest in the hottest months and that the combination of solar radiation and morphological features such as vertical-to-horizontal ratios plays a role. In high-density urban areas (such as C8 and C3), the sun's lower angle during shoulder months leads to more heat accumulation on urban vertical surfaces (façades) and longer periods of elevated temperatures.

Finally, it is also important to note that although the UWG has been validated effectively for urban districts, its ability to predict the UHII in low-density suburban areas seems here limited. This is evident in cluster C6. This cluster is representative of very sparse building, with a very low density, and sometimes isolated building. Therefore, it would be expected, for this cluster to have temperatures closer

to the rural ones. However, the UWG predicts a relatively significant UHII in these areas (Figure 5). Confrontation with experimental measurements would be necessary to fix this issue. But in case it would be likely linked to the assumption of the UWG model that the district being analyzed is surrounded by other urban districts rather than rural areas, potentially leading to an overestimation of overheating in boundary and suburban areas.

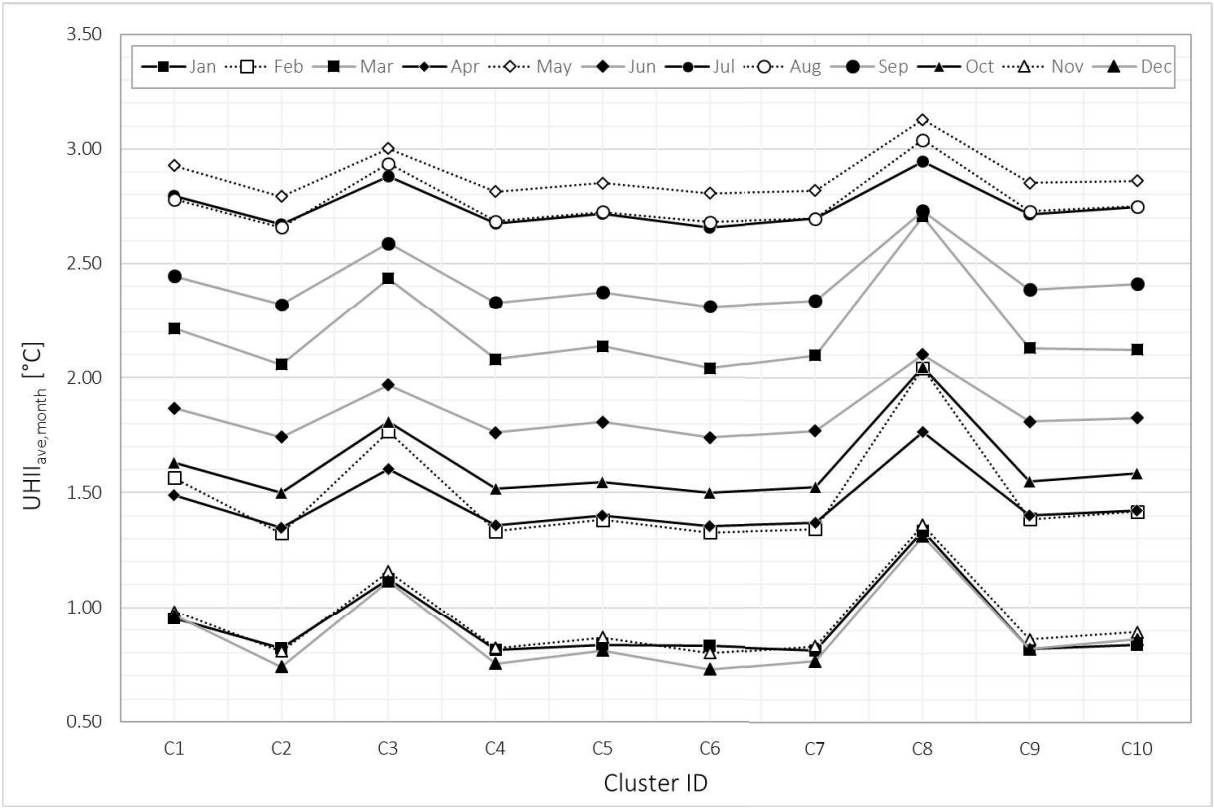

Figure 6: Average monthly urban heat island intensity among the urban clusters

The spatial variability of the $UHII_{ave,month}$ can be further elaborated through GIS tools to create a UHI map of the whole area visualizing the most impacted clusters. As an example, Figure 7 shows the $UHII_{ave,month}$ in August within the Geneva Canton. Consistent with previous numerical findings, the high-density urban areas (C8 and C3) closest to the lake experience the greatest UHII. While it is acknowledged that the absence of lake modeling represents a limitation of the UWG tool, it should be noted that the UHI phenomenon is most relevant in conditions of high irradiance and low wind speed.

To address this, GEO-NET, a bureau specializing in climatic analyses, conducted a detailed analysis of the current climate in the Canton of Geneva, identifying areas most affected by urban overheating (Gmbh, 2020). GEO-NET employed a mesoscale model, FITNAH 3D (Gross, 1992), to simulate the impact of wind and temperature on the entire Canton, generating UHII estimates for a single representative summer day under conditions of high irradiance and low wind speed. Although it is not possible to compare results directly due to the different temporal scales of the two studies (a single day compared to one year), detailed modeling results including are comparable to those obtained in the present study. In particular, they found that, despite the presence of the lake, the same high-density urban areas are the most impacted by overheating and that the air exchange is very limited due to the obstruction effect of buildings.

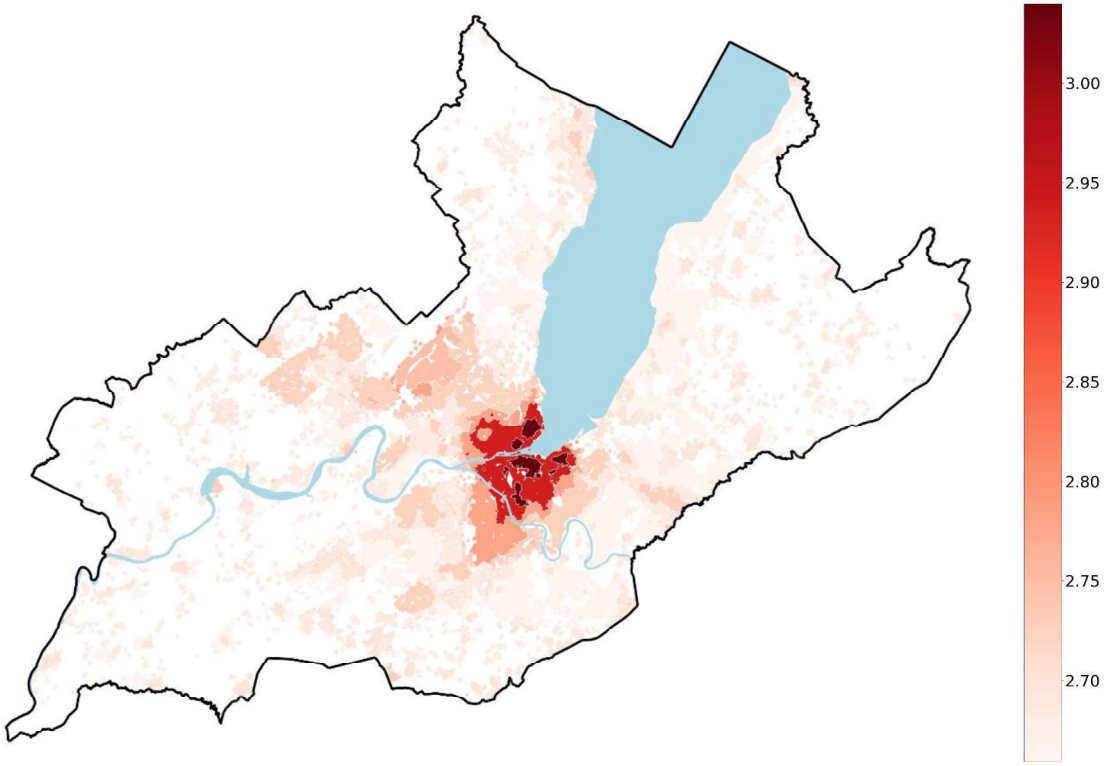

Figure 7: Spatial variability of the monthly average UHII in August within the Canton of Geneva

The peak hourly values of urban heat island intensity ($UHII_{max,month}$) for each month among the urban clusters are shown in Figure 8. During the colder months (October to February), the hourly peaks range

from 5.1°C in clusters C2 and C6 to 8.6°C in cluster C8. However, during the warmer months (March to September), the calculated values are much higher, with a minimum of 7.9°C in cluster C6 and a maximum of 11.5°C in cluster C8. In general, the differences among the clusters are less evident compared to the average values, except for the two high density clusters C3 and C8, whose maximum values in some months are considerably higher. This is an effect of the non-simultaneous between nocturnal and diurnal conditions in rural and urban areas. At sunrise, when incoming solar radiation starts warming the urban surfaces, high-density clusters experience a faster warming process due to their significantly higher thermal mass, resulting in a larger $UHII_{max,month}$ compared to low-density areas. This phenomenon is explained in more detail in Section 3.4.

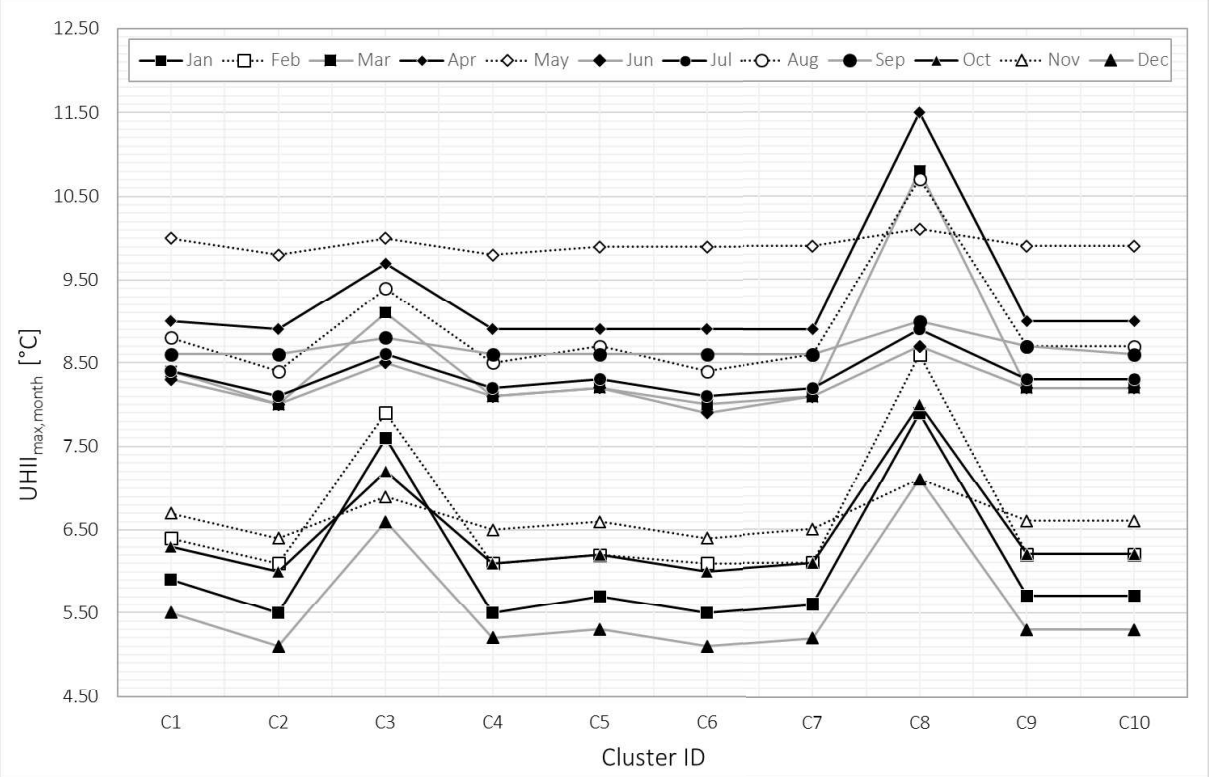

Figure 8: Maximum monthly urban heat island intensity among the urban clusters

3.4 Average daily variation of the UHI intensity among the identified urban clusters

This section is dedicated to the analysis of the diurnal cycle of the UHII. As shown in Figure 9 the average hourly urban heat island intensity ($UHII_{ave,hour}$) among the urban clusters is higher during

nighttime, reaching a minimum between 11:00 h and 14:00 h, with values lower than 0.5°C. During the late afternoon, both the $UHII_{ave,hour}$ and the differences among the urban clusters increase, reaching values ranging from 2.5°C to 3.5°C, depending on the urban morphology. As observed in the previous Section 3.3, the high-density urban clusters (C3 and C8) exhibit the highest UHII peak values which are generally observed at sunrise (around 6:00 h). As previously anticipated, at this time, the atmosphere is transitioning between nocturnal and diurnal conditions, and the incoming solar radiation is beginning to warm the urban surfaces (Oke, 1982). Observing the UWG simulation process (which is based on an hourly timestep) it can be noticed that generally when sun rises the rural temperature is still at its minimum value (nighttime conditions), whereas the high-density urban clusters have already started the warming process. For the urban clusters with lower density this warming process is generally delayed by 1 hour, when the rural temperatures have also started to rise, resulting in a considerably lower $UHII_{max,month}$ compared to high density clusters. These results evidence that the UWG predictions are strongly influenced by the input geometric parameters (in particular the vertical-to-horizontal ratio), which has been also highlighted in previous literature studies where the same behaviour has been observed (Alchapar et al., 2019; Salvati, Coch Roura, et al., 2017; Salvati, Coch, et al., 2017).

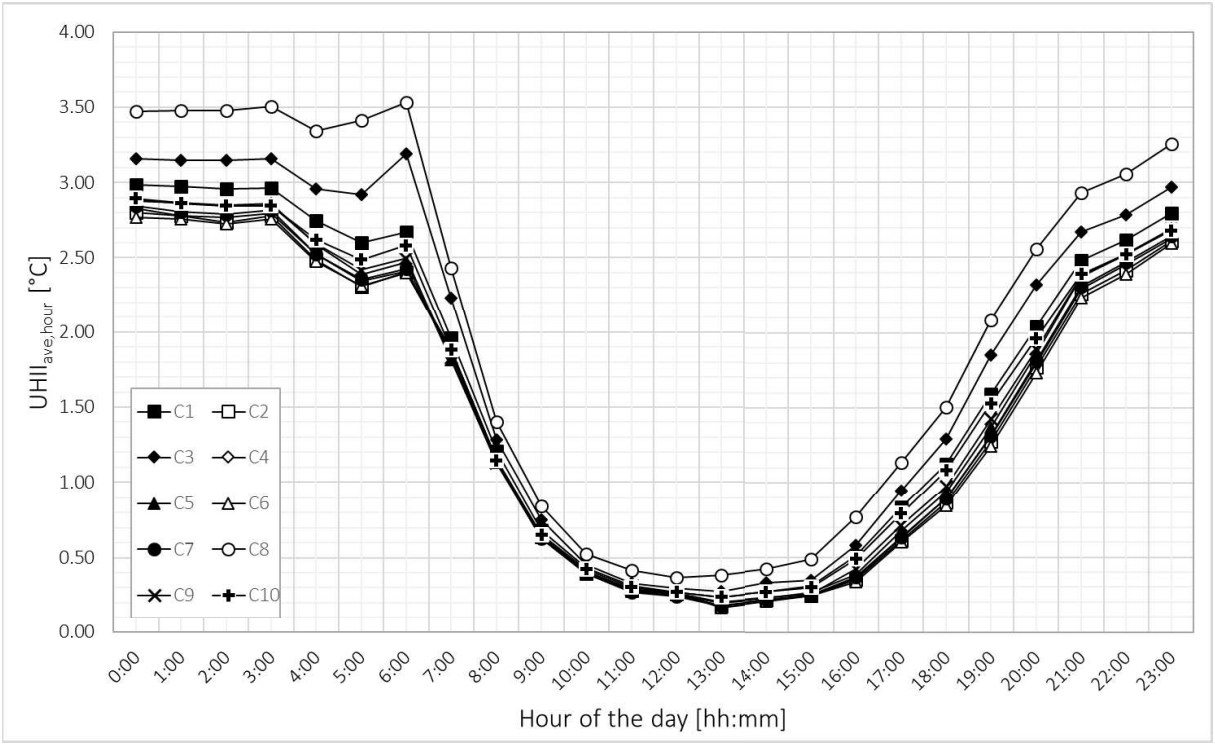

Figure 9: Average hourly urban heat island intensity among the urban clusters

The boxplots of the UHII during daytime and nighttime are plotted for each cluster and shown in Figure 10, left for daytime and right for nighttime. Each whisker represents the interquartile range (the 25th percentile for the lower quarter of the values, and the 75th for the upper quarter) of the yearly UHII data, with the median value represented by the line inside the box. Outliers are represented as individual points outside the whiskers. Daytime and nighttime periods have been determined based on solar radiation. It can be observed, as expected, that the median values of UHII during daytime is about 0.5°C for all the clusters, whereas the nighttime values range between 1.5°C and 2.2°C depending on urban morphology. It can be also observed that during nighttime the UHII experience a higher variability and that the differences between the clusters are more evident compared to daytime. Peak values reaching up to 10°C are also more likely to be observed during the night than during the day.

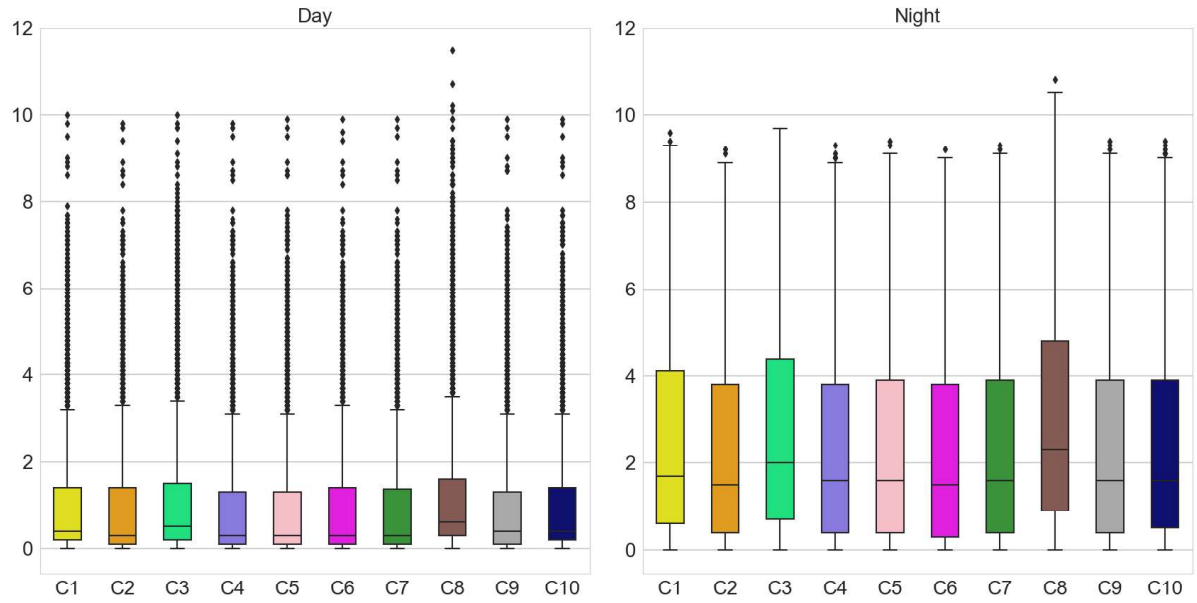

Figure 10: Boxplots of UHII during daytime (left) and nighttime (right)

3.5 Comparison between predicted and measured average UHII monthly values

This section includes the results of two additional simulations using the Urban Weather Generator (UWG) to validate the developed workflow and the simulated urban heat island intensity (UHII) values with metered data. Real temperature measurements from a rural site (Bernex, used as input weather file for UWG simulations) and two urban weather stations (Battelle and Prairie, used for comparison). The meteorological data from Bernex were recorded and provided by AgroMétéo (Confederation suisse, 2022), while the data from Battelle and Prairie were monitored by the University of Geneva during the same year (University of Geneva, 2022). The recorded UHII of 2019 ($UHII_{real}$) is calculated for both Battelle and Prairie sites as the positive temperature difference with the rural site of Brenex. $UHII_{real}$ is then compared with the UWG simulated values ($UHII_{UWG}$) from two additional UWG simulations performed using Bernex data as the source rural weather file and the urban parameters of Battelle and Prairie areas derived from clustering. Based on the clustering results (Section 3.1), the Prairie weather station is located within the high-density cluster C3, whereas Battelle falls into the suburban cluster C4. Thus, the urban parameters used as an input to the UWG simulations are the ones of the related clusters (C3 and C4). The results are shown in Figure 11 in terms of average monthly values (markers) and related standard deviation (bars). Two different colors are used for Prairie and Battelle, and the different

markers identify the simulated and the metered data. It can be observed that, overall, the simulated values follow the same yearly trend as the measured values while overestimating them in some cases. It should be noted that the standard deviation of the simulated data is generally larger than that of the measured data. With the exception of the months of February and June, where the simulated values are about 0.6°C higher than the measured ones, the average absolute difference in all other cases is around 0.15°C. The substantial discrepancy in the results of February and June could be due to the presence of various sources of error and uncertainty which may have occurred in real-world conditions, in contrast to simulated values which are based on model assumptions and parameterizations.

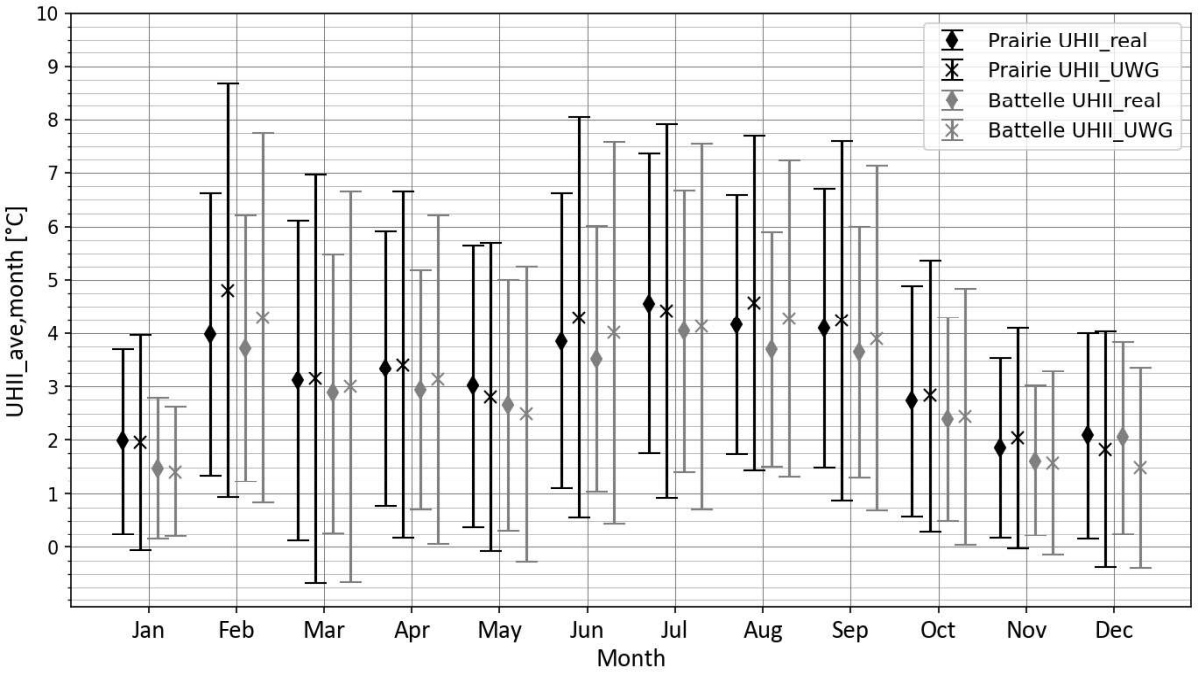

Figure 11: Average monthly UHII (markers) and standard deviation (bars) arising from UWG simulations (UHII_UWG) and from real measurements (UHII_real) for two selected urban weather stations

A comparison of the average hourly values and their related standard deviations is also conducted and shown in Figure 12. The results indicate that the UWG tool tends to overestimate the UHII during nighttime and underestimate it during daytime, which may be attributed to various factors such as model assumptions and input parameter quality. While improving the model is not within the scope of this

article, these findings provide valuable insight for future improvements and refinement of the model accuracy, and highlight its potential limitations.

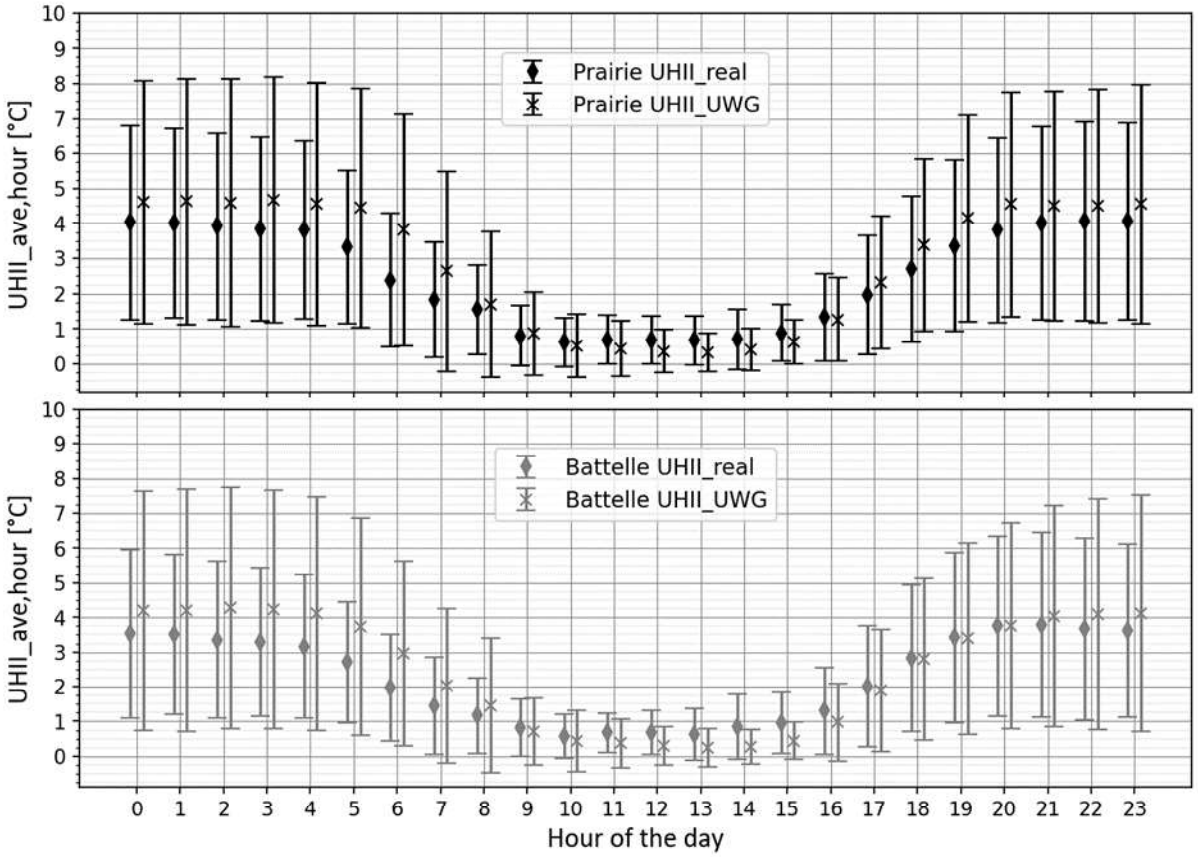

Figure 12: Average hourly UHII (markers) and standard deviation (bars) arising from UWG simulations (UHII_UWG) and from real measurements (UHII_real) for two selected urban weather stations

4. Discussion

This study proposes a data-driven and machine learning-based approach to perform computationally efficient year-long Urban Heat Island (UHI) analyses at the city scale. The developed workflow aims to derive meaningful urban parameters and to identify representative microclimatic clusters within a city. The urban parameters are quantitatively obtained from detailed urban datasets, GIS pre-processing operations, and Gaussian Mixture Models (GMM) for clustering. The results are then used as inputs to the Urban Weather Generator (UWG), a well-established microclimate parametric model, to simulate urban weather conditions from rural weather station data. The methodology is applied to the Canton of Geneva and results in the identification of ten representative microclimatic clusters.

The simulated urban heat island intensity (UHII) has a considerable variability over the year and among the urban clusters, with a range of values from 0.7°C to 3.1°C and differences up to 0.6°C depending on the urban morphology. The results also show that the UHII monthly peak values range from 5.1°C up to 11.5°C. The daily variation of the UHII is also found to be much higher during nighttime (about 2.5°C) than during daytime (about 0.5°C), consistent with similar literature studies. Finally, the accuracy of the simulated values is compared towards real measured data of two urban weather stations during 2019. In general, the accuracy of the simulated values is effective in capturing the average monthly UHI intensity and the differences between the two reference urban weather stations. However, for two months (February and June), the simulated values are significantly higher (about 0.6°C) than the measured ones, compared to an average difference of only 0.15°C for the other months. The model also slightly overestimates nighttime values while underestimating them during daytime.

5. Limitations and future perspectives

The proposed approach has some limitations that should be noted. Firstly, the approach is data-driven, and reproducing it in other cities is possible provided that a detailed urban dataset like the one used in this study is available. Secondly, as discussed in Section 2.2, the UWG model presents some simplifications mainly related to the modelling of advection, vegetation, and the presence of large water bodies. These simplifications may prevent from capturing site-specific phenomena and seems to overestimate the UHII in peri-urban areas. Nevertheless, the proposed workflow enables the simulation of one year on a large scale, making it suitable for city-wide analyses. Precise temperature time series in specific areas of the city would require more accurate simulations, not a parametric model.

However, despite these limitations, the proposed approach has several strengths. It offers a computationally efficient way to perform city-scale UHI analyses, providing valuable insights for urban planners to accurately plan UHI mitigation strategies, identify areas at the greatest risk of overheating, and plan energy-related interventions.

Future improvements to the Urban Weather Generator (UWG) model offer promising prospects for achieving even higher accuracy in simulating urban microclimates. Recent studies, such as the Vertical

Urban Weather Generator (VUWG) (Moradi et al., 2021) and the improvements introduced by Xu et al. (Xu et al., 2022), are refining the capabilities of the UWG. The VUWG resolves vertical profiles of climate variables, including temperature, wind, specific humidity, and turbulence kinetic energy, in relation to urban design parameters. The improvements by Xu et al. enhance the radiation, vegetation, and convective heat transfer calculation processes, better matching the physical representation of urban districts.

In the future, the integration of Internet of Things (IoT) sensors for weather data, crowdsourcing, satellite data, and other data sources could potentially enhance the accuracy of the proposed approach. These future prospects hold the potential for more reliable data and more precise modeling of urban microclimates.

6. Acknowledgements

This research has been supported by the program MITI (Mission for Transversal and Interdisciplinary Initiatives) of CNRS and CSTB through the CITYBIOM project, and by the French National Research Agency, through the Investments for Future Program (ref. ANR-18-EURE-0016 - Solar Academy). The research unit LOCIE is a member of the INES Solar Academy Research Center.

References

Afshari, A., & Ramirez, N. (2021). Improving the accuracy of simplified urban canopy models for arid regions using site-specific prior information. *Urban Climate*, *35*(October 2020), 100722. https://doi.org/10.1016/j.uclim.2020.100722

Alchapar, N. L., Pezzuto, C. C., Correa, E. N., & Salvati, A. (2019). Thermal Performance of the Urban Weather Generator Model as a Tool for Planning Sustainable Urban Development. *Geographica Pannonica*, *23*(4), 374–384. https://doi.org/10.5937/gp23-24254

AlDousari, A. E., Kafy, A. Al, Saha, M., Fattah, M. A., Almulhim, A. I., Faisal, A. Al, Al Rakib, A., Jahir, D. M. A., Rahaman, Z. A., Bakshi, A., Shahrier, M., & Rahman, M. M. (2022). Modelling the impacts of land use/land cover changing pattern on urban thermal characteristics in Kuwait. *Sustainable Cities and Society*, *86*(August), 104107. https://doi.org/10.1016/j.scs.2022.104107

Alvarez, I., Quesada-Ganuza, L., Briz, E., & Garmendia, L. (2021). Urban heat islands and thermal comfort: A case study of zorrotzaurre island in Bilbao. *Sustainability (Switzerland)*, *13*(11), 1–13. https://doi.org/10.3390/su13116106

Benjamin, K., Luo, Z., & Wang, X. (2021). Crowdsourcing urban air temperature data for estimating urban heat island and building heating/cooling load in london. *Energies*, *14*(16). https://doi.org/10.3390/en14165208

Biljecki, F., Chew, L. Z. X., Milojevic-Dupont, N., & Creutzig, F. (2021). *Open government geospatial data on buildings for planning sustainable and resilient cities*. http://arxiv.org/abs/2107.04023

Boccalatte, A., Fossa, M., Gaillard, L., & Menezo, C. (2020). Microclimate and urban morphology effects on building energy demand in different European cities. *Energy and Buildings*, *224*, 110129. https://doi.org/10.1016/j.enbuild.2020.110129

Boccalatte, A., Thebault, M., Ménézo, C., Ramousse, J., & Fossa, M. (2022). Evaluating the impact of urban morphology on rooftop solar radiation: A new city-scale approach based on Geneva GIS data. *Energy and Buildings*, *260*, 111919. https://doi.org/10.1016/j.enbuild.2022.111919

Brousse, O., Simpson, C., Walker, N., Fenner, D., Meier, F., Taylor, J., & Heaviside, C. (2022). Evidence of horizontal urban heat advection in London using six years of data from a citizen

weather station network. *Environmental Research Letters*, *17*(4), 44041. https://doi.org/10.1088/1748-9326/ac5c0f

Bueno, B., Hidalgo, J., Pigeon, G., Norford, L., & Masson, V. (2013). Calculation of air temperatures above the urban canopy layer from measurements at a rural operational weather station. *Journal of Applied Meteorology and Climatology*, *52*(2), 472–483. https://doi.org/10.1175/JAMC-D-12-083.1

Bueno, B., Nakano, A., & Norford, L. (2015a). Urban weather generator: a method to predict neighborhood-specific urban temperatures for use in building energy simulations. *ICUC9 - 9th International Conference on Urban Climate Jointly with 12th Symposium on the Urban Environment*.

Bueno, B., Nakano, A., & Norford, L. (2015b). Urban weather generator: a method to predict neighborhood-specific urban temperatures for use in building energy simulations. *ICUC9 - 9th International Conference on Urban Climate Jointly with 12th Symposium on the Urban Environment*, *i*.

Bueno, B., Norford, L., Hidalgo, J., & Pigeon, G. (2013). The urban weather generator. *Journal of Building Performance Simulation*, *6*(4), 269–281. https://doi.org/10.1080/19401493.2012.718797

Bueno, B., Pigeon, G., Norford, L. K., Zibouche, K., & Marchadier, C. (2012). Development and evaluation of a building energy model integrated in the TEB scheme. *Geoscientific Model Development*, *5*(2), 433–448. https://doi.org/10.5194/gmd-5-433-2012

Bueno, B., Roth, M., Norford, L., & Li, R. (2014). Computationally efficient prediction of canopy level urban air temperature at the neighbourhood scale. *Urban Climate*, *9*, 35–53. https://doi.org/10.1016/j.uclim.2014.05.005

Cesario, E., Uchubilo, P. I., Vinci, A., & Zhu, X. (2020). Discovering Multi-density Urban Hotspots in a Smart City. *Proceedings - 2020 IEEE International Conference on Smart Computing, SMARTCOMP 2020*, 332–337. https://doi.org/10.1109/SMARTCOMP50058.2020.00073

Chàfer, M., Tan, C. L., Hien, W. N., Pisello, A., & Cabeza, L. F. (2022). Mobile Measurements of Microclimatic Variables Through the Central Area of Singapore. *SSRN Electronic Journal*,

83(June). https://doi.org/10.2139/ssrn.4065782

Chen, Y., Shu, B., Zhang, R., & Amani-Beni, M. (2023). LST determination of different urban growth patterns: A modeling procedure to identify the dominant spatial metrics. *Sustainable Cities and Society*, *92*(February), 104459. https://doi.org/10.1016/j.scs.2023.104459

Confederation suisse. (2022). *Agrometeo*. https://www.agrometeo.ch/

D'Acci, L., & Batty, M. (2019). The Mathematics of Urban Morphology Foreword by Michael Batty. In *The Mathematics of Urban Morphology. Modeling and Simulation in Science, Engineering and Technology.* http://www.springer.com/series/4960%0Ahttp://dx.doi.org/10.1007/978-3-030-12381-9_21

de Almeida, C. R., Teodoro, A. C., & Gonçalves, A. (2021). Study of the urban heat island (Uhi) using remote sensing data/techniques: A systematic review. In *Environments - MDPI* (Vol. 8, Issue 10). https://doi.org/10.3390/environments8100105

Dimitrov, S., Popov, A., & Iliev, M. (2021). An application of the LCZ approach in surface urban heat island mapping in Sofia, Bulgaria. *Atmosphere*, *12*(11). https://doi.org/10.3390/atmos12111370

Dissanayake, K., Kurugama, K., & Ruwanthi, C. (2020). Ecological Evaluation of Urban Heat Island Effect in Colombo City, Sri Lanka Based on Landsat 8 Satellite Data. *2020 Moratuwa Engineering Research Conference (MERCon)*, 531–536. https://doi.org/10.1109/MERCon50084.2020.9185277

EnergyPlus. (2023). https://energyplus.net

ENVI-met. (2021). *ENVI-met*. https://www.envi-met.com/

Erell, E., & Williamson, T. (2006). Simulating air temperature in an urban street canyon in all weather conditions using measured data at a reference meteorological station. *International Journal of Climatology*, *26*(12), 1671–1694. https://doi.org/10.1002/joc.1328

Fenner, D., Meier, F., Scherer, D., & Polze, A. (2014). Spatial and temporal air temperature variability in Berlin, Germany, during the years 2001-2010. *Urban Climate*, *10*(P2), 308–331. https://doi.org/10.1016/j.uclim.2014.02.004

Fleischmann, M. (2017). *A Systematisation of Attributes for Quantitative Urban Morphology Measuring*

Urban Form (Issue November).

Fleischmann, M., Feliciotti, A., & Kerr, W. (2021). Evolution of Urban Patterns: Urban Morphology as an Open Reproducible Data Science. *Geographical Analysis*, 1–23. https://doi.org/10.1111/gean.12302

Fleischmann, M., Feliciotti, A., Romice, O., & Porta, S. (2020). Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale. *Computers, Environment and Urban Systems*, *80*(November 2019), 101441. https://doi.org/10.1016/j.compenvurbsys.2019.101441

Fleischmann, M., Feliciotti, A., Romice, O., & Porta, S. (2022). Methodological foundation of a numerical taxonomy of urban form. *Environment and Planning B: Urban Analytics and City Science*, *49*(4), 1283–1299. https://doi.org/10.1177/23998083211059835

Flourentzou, F. (2019). Possible strategies and obstacles in the pathway towards energy transition of residential building stocks in Switzerland. *IOP Conference Series: Earth and Environmental Science*, *323*(1). https://doi.org/10.1088/1755-1315/323/1/012171

GitHub-Jiachen-Mao/UWG_Matlab. (2021). https://github.com/Jiachen-Mao/UWG_Matlab

GitHub - ladybug tools/uwg. (2022). https://github.com/ladybug-tools/uwg.git

Gmbh, G. U. (2020). *Situation climato-écologique du canton de Genève : Analyse climatique sur la base d ′ un modèle Décembre*.

Grimmond, C. S. B. (2017). *The integrated WRF / urban modelling system : Development , evaluation , and applications to urban environmental problems The integrated WRF / urban modelling system : development , evaluation , and applications to urban environmental*. *March*, 1–38.

Grimmond, C. S. B., Roth, M., Oke, T. R., Au, Y. C., Best, M., Betts, R., Carmichael, G., Cleugh, H., Dabberdt, W., Emmanuel, R., Freitas, E., Fortuniak, K., Hanna, S., Klein, P., Kalkstein, L. S., Liu, C. H., Nickson, A., Pearlmutter, D., Sailor, D., & Voogt, J. (2010). Climate and more sustainable cities: Climate information for improved planning and management of cities (Producers/Capabilities Perspective). *Procedia Environmental Sciences*, *1*(1), 247–274. https://doi.org/10.1016/j.proenv.2010.09.016

Gross, G. (1992). Results of supercomputer simulations of meteorological mesoscale phenomena. *Fluid Dynamics Research*, *10*(4–6), 483–498. https://doi.org/10.1016/0169-5983(92)90035-U

Hashemi, F. (2020). *A novel approach for investigating canopy heat island effects on build - ing energy performance : A case study of Center City of Philadelphia , PA*. *November*, 1–6.

He, B. J. (2019). Towards the next generation of green building for urban heat island mitigation: Zero UHI impact building. *Sustainable Cities and Society*, *50*. https://doi.org/10.1016/j.scs.2019.101647

Houet, T., & Pigeon, G. (2011). Mapping urban climate zones and quantifying climate behaviors - An application on Toulouse urban area (France). *Environmental Pollution*, *159*(8–9), 2180–2192. https://doi.org/10.1016/j.envpol.2010.12.027

Huang, Q., Huang, J., Yang, X., Fang, C., & Liang, Y. (2018). Quantifying the seasonal contribution of coupling urban land use types on Urban Heat Island using Land Contribution Index: A case study in Wuhan, China. *Sustainable Cities and Society*, *44*. https://doi.org/10.1016/j.scs.2018.10.016

Hwang, R. L., Lin, T. P., & Lin, F. Y. (2020). Evaluation and mapping of building overheating risk and air conditioning use due to the urban heat island effect. *Journal of Building Engineering*, *32*(November), 101726. https://doi.org/10.1016/j.jobe.2020.101726

IEA. (2021). World Energy Outlook 2021 : Part of the World Energy Outlook. *International Energy Agency*, 386. https://www.iea.org/reports/world-energy-outlook-2021

IPCC. (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.

İsmet Berke, Ç. (2010). *Global Sensitivity Analysis for Urban Heat Island Effect: A Case Study in a Residential Neighbourhood in Ankara, Turkey*. https://open.metu.edu.tr/handle/11511/99521

Jänicke, B., Milošević, D., & Manavvi, S. (2021). Review of user- friendly models to improve the urban micro- climate. *Atmosphere*, *12*(10), 1–22. https://doi.org/10.3390/atmos12101291

Johari, F., Peronato, G., Sadeghian, P., Zhao, X., & Widén, J. (2020). Urban building energy modeling: State of the art and future prospects. *Renewable and Sustainable Energy Reviews*, *128*. https://doi.org/10.1016/j.rser.2020.109902

Kolokotroni, M., Ren, X., Davies, M., & Mavrogianni, A. (2012). London's urban heat island: Impact on current and future energy consumption in office buildings. *Energy and Buildings*, *47*, 302–311. https://doi.org/10.1016/j.enbuild.2011.12.019

Kwak, Y., Park, C., & Deal, B. (2020). Discerning the success of sustainable planning: A comparative analysis of urban heat island dynamics in Korean new towns. *Sustainable Cities and Society*, *61*(January), 102341. https://doi.org/10.1016/j.scs.2020.102341

Lac, C., Chaboureau, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., Leriche, M., Barthe, C., Aouizerats, B., Augros, C., Aumond, P., Auguste, F., Bechtold, P., Berthet, S., Bielli, S., Bosseur, F., Caumont, O., Cohard, J.-M., Colin, J., … Wautelet, P. (2018). Overview of the Meso-NH model version 5.4 and its applications. *Geoscientific Model Development*, *11*(5), 1929–1969. https://doi.org/10.5194/gmd-11-1929-2018

Lauzet, N., Rodler, A., Musy, M., Azam, M. H., Guernouti, S., Mauree, D., & Colinart, T. (2019). How building energy models take the local climate into account in an urban context – A review. *Renewable and Sustainable Energy Reviews*, *116*(August), 109390. https://doi.org/10.1016/j.rser.2019.109390

Leconte, F., Bouyer, J., Claverie, R., & Pétrissans, M. (2015). Using Local Climate Zone scheme for UHI assessment: Evaluation of the method using mobile measurements. *Building and Environment*, *83*, 39–49. https://doi.org/10.1016/j.buildenv.2014.05.005

Lemonsu, A., Masson, V., Shashua-Bar, L., Erell, E., & Pearlmutter, D. (2012). Inclusion of vegetation in the Town Energy Balance model for modelling urban green areas. *Geoscientific Model Development*, *5*(6), 1377–1393. https://doi.org/10.5194/gmd-5-1377-2012

Li, X., Zhou, Y., Yu, S., Jia, G., Li, H., & Li, W. (2019). Urban heat island impacts on building energy consumption: A review of approaches and findings. In *Energy* (Vol. 174, pp. 407–419). https://doi.org/10.1016/j.energy.2019.02.183

Lima, I., Scalco, V., & Lamberts, R. (2019). Estimating the impact of urban densification on high-rise office building cooling loads in a hot and humid climate. *Energy and Buildings*, *182*, 30–44. https://doi.org/10.1016/j.enbuild.2018.10.019

Lindberg, F., Grimmond, C. S. B., Gabey, A., Huang, B., Kent, C. W., Sun, T., Theeuwes, N. E., Järvi, L., Ward, H. C., Capel-Timms, I., Chang, Y., Jonsson, P., Krave, N., Liu, D., Meyer, D., Olofson, K. F. G., Tan, J., Wästberg, D., Xue, L., & Zhang, Z. (2018). Urban Multi-scale Environmental Predictor (UMEP): An integrated tool for city-based climate services. *Environmental Modelling and Software*, *99*, 70–87. https://doi.org/10.1016/j.envsoft.2017.09.020

Lipson, M., Nazarian, N., Hart, M. A., Nice, K. A., & Conroy, B. (2022). A transformation in city-descriptive input data for urban climate models. *Frontiers in Environmental Science*, *10*(July), 1–18. https://doi.org/10.3389/fenvs.2022.866398

Litardo, J., Palme, M., Borbor-Cordova, M., Caiza, R., Macias, J., Hidalgo-Leon, R., & Soriano, G. (2020). Urban Heat Island intensity and buildings' energy needs in Duran, Ecuador: Simulation studies and proposal of mitigation strategies. *Sustainable Cities and Society*, *62*(July), 102387. https://doi.org/10.1016/j.scs.2020.102387

Lobaccaro, G., De Ridder, K., Acero, J. A., Hooyberghs, H., Lauwaet, D., Maiheu, B., Sharma, R., & Govehovitch, B. (2021). Applications of models and tools for mesoscale and microscale thermal analysis in mid-latitude climate regions—A review. *Sustainability (Switzerland)*, *13*(22). https://doi.org/10.3390/su132212385

Lun, I., Mochida, A., & Ooka, R. (2009). Progress in Numerical Modelling for Urban Thermal Environment Studies. *Advances in Building Energy Research*, *3*, 147–188. https://doi.org/10.3763/aber.2009.0306

Ma, R., Li, X., & Chen, J. (2021). An elastic urban morpho-blocks (EUM) modeling method for urban building morphological analysis and feature clustering. *Building and Environment*, *192*, 107646. https://doi.org/10.1016/j.buildenv.2021.107646

Maiullari, D., Pijpers- Van Esch, M., & van Timmeren, A. (2021). A quantitative morphological method for mapping local climate types. *Urban Planning*, *6*(3), 240–257. https://doi.org/10.17645/up.v6i3.4223

Mao, J., Yang, J. H., Afshari, A., & Norford, L. K. (2017). Global sensitivity analysis of an urban microclimate system under uncertainty: Design and case study. *Building and Environment*, *124*,

153–170. https://doi.org/10.1016/j.buildenv.2017.08.011

Masson, V. (2000). A physically-based scheme for the urban energy budget in atmospheric models. *Boundary-Layer Meteorology*, *94*(3), 357–397. https://doi.org/10.1023/A:1002463829265

Masson, V., Heldens, W., Bocher, E., Bonhomme, M., Bucher, B., Burmeister, C., de Munck, C., Esch, T., Hidalgo, J., Kanani-Sühring, F., Kwok, Y. T., Lemonsu, A., Lévy, J. P., Maronga, B., Pavlik, D., Petit, G., See, L., Schoetter, R., Tornay, N., … Zeidler, J. (2020). City-descriptive input data for urban climate models: Model requirements, data sources and challenges. *Urban Climate*, *31*(August 2019), 100536. https://doi.org/10.1016/j.uclim.2019.100536

Meier, F., Fenner, D., Grassmann, T., Jänicke, B., Otto, M., & Scherer, D. (2015). Challenges and benefits from crowdsourced atmospheric data for urban climate research using Berlin, Germany, as testbed. *ICUC9 - 9th International Conference on Urban Climate Jointly with 12th Symposium on the Urban Environment Challenges*, *September*, 6p. https://www.researchgate.net/publication/281647191%0Ahttp://www.meteo.fr/icuc9/LongAbstracts/nomtm6-2-6171335_a.pdf

Memme, S., & Fossa, M. (2022). Maximum energy yield of PV surfaces in France and Italy from climate based equations for optimum tilt at different azimuth angles. *Renewable Energy*, *200*(May), 845–866. https://doi.org/10.1016/j.renene.2022.10.019

Milojevic-Dupont, N., Wagner, F., Nachtigall, F., Hu, J., Brüser, G. B., Zumwald, M., Biljecki, F., Heeren, N., Kaack, L. H., Pichler, P.-P., & Creutzig, F. (2023). EUBUCCO v0.1: European building stock characteristics in a common and open database for 200+ million individual buildings. *Scientific Data*, *10*(1), 147. https://doi.org/10.1038/s41597-023-02040-2

Mirzaei, P. A. (2015). Recent challenges in modeling of urban heat island. In *Sustainable Cities and Society*. https://doi.org/10.1016/j.scs.2015.04.001

Mirzaei, P. A., & Haghighat, F. (2010). Approaches to study Urban Heat Island - Abilities and limitations. *Building and Environment*. https://doi.org/10.1016/j.buildenv.2010.04.001

Moradi, M., Dyer, B., Nazem, A., Nambiar, M. K., Rafsan Nahian, M., Bueno, B., Mackey, C., Vasanthakumar, S., Nazarian, N., Scott Krayenhoff, E., Norford, L. K., & Aliabadi, A. A. (2021).

The vertical city weather generator (vcwg v1.3.2). *Geoscientific Model Development*, *14*(2), 961–984. https://doi.org/10.5194/gmd-14-961-2021

Morille, B., Lauzet, N., & Musy, M. (2015). SOLENE-microclimate: A tool to evaluate envelopes efficiency on energy consumption at district scale. *Energy Procedia*, *78*, 1165–1170. https://doi.org/10.1016/j.egypro.2015.11.088

Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., & Leigh, R. R. (2015). Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology*, *35*(11), 3185–3203.

Mussetti, G., Davin, E. L., Schwaab, J., Acero, J. A., Ivanchev, J., Singh, V. K., Jin, L., & Seneviratne, S. I. (2022). Do Electric Vehicles Mitigate Urban Heat? The Case of a Tropical City. *Frontiers in Environmental Science*, *10*(February), 1–9. https://doi.org/10.3389/fenvs.2022.810342

Mutani, G., & Todeschi, V. (2020). Building energy modeling at neighborhood scale. *Energy Efficiency*, *13*(7), 1353–1386. https://doi.org/10.1007/s12053-020-09882-4

O'Malley, C., & Kikumoto, H. (2022). An investigation into heat storage by adopting local climate zones and nocturnal-diurnal urban heat island differences in the Tokyo Prefecture. *Sustainable Cities and Society*, *83*(March), 103959. https://doi.org/10.1016/j.scs.2022.103959

Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, *108*(455), 1–24. https://doi.org/10.1002/qj.49710845502

Palme, M., Inostroza, L., Villacreses, G., Lobato-Cordero, A., & Carrasco, C. (2017). From urban climate to energy consumption. Enhancing building performance simulation by including the urban heat island effect. *Energy and Buildings*, *145*, 107–120. https://doi.org/10.1016/j.enbuild.2017.03.069

Palme, M., & Salvati, A. (2021). *Urban Microclimate Modelling for Comfort and Energy Studies*. https://doi.org/10.1007/978-3-030-65421-4

Parlow, E., Vogt, R., & Feigenwinter, C. (2014). The urban heat island of Basel - Seen from different perspectives. *Die Erde; Zeitschrift Der Gesellschaft Für Erdkunde Zu Berlin*, *145*, 1–2. https://doi.org/10.12854/erde-145-8

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

Pielke, R., Burgess, M. G., & Ritchie, J. (2022). Plausible 2005-2050 emissions scenarios project between 2 °c and 3 °c of warming by 2100. *Environmental Research Letters*, *17*(2). https://doi.org/10.1088/1748-9326/ac4ebf

Pyrgou, A., Castaldo, V. L., Pisello, A. L., Cotana, F., & Santamouris, M. (2017). On the effect of summer heatwaves and urban overheating on building thermal-energy performance in central Italy. *Sustainable Cities and Society*, *28*, 187–200. https://doi.org/10.1016/j.scs.2016.09.012

Quan, S. J. (2020). Identifying Urban Form Typologies in Seoul with Mixture Model Based Identifying Urban Form Typologies in Seoul with Mixture Model Based Clustering. *ISUF 2020 Cities in the 21st Century*, *October*. https://doi.org/10.13140/RG.2.2.20864.46088

Raalte, L. van, Nolan, M., Thakur, P., Xue, S., Parker, N., & AECOM Australia. (2012). Economic Assessment of the Urban Heat Island Effect. *City of Melbourne*, 1–71. https://www.melbourne.vic.gov.au/sitecollectiondocuments/eco-assessment-of-urban-heat-island-effect.pdf

Rajagopal, P., Shanthi, R., & Senthil, R. (2023). A review of recent developments in the impact of environmental measures on urban heat island. *Sustainable Cities and Society*, *88*(March 2022), 104279. https://doi.org/10.1016/j.scs.2022.104279

Reynolds, D. (2015). Gaussian Mixture Models. In S. Z. Li & A. K. Jain (Eds.), *Encyclopedia of Biometrics* (pp. 827–832). Springer US. https://doi.org/10.1007/978-1-4899-7488-4_196

Richard, Y., Emery, J., Dudek, J., Pergaud, J., Chateau-Smith, C., Zito, S., Rega, M., Vairet, T., Castel, T., Thévenin, T., & Pohl, B. (2018). How relevant are local climate zones and urban climate zones for urban climate research? Dijon (France) as a case study. *Urban Climate*, *26*(October), 258–274. https://doi.org/10.1016/j.uclim.2018.10.002

Romero Rodríguez, L., Sánchez Ramos, J., Sánchez de la Flor, F. J., & Álvarez Domínguez, S. (2020).

Analyzing the urban heat Island: Comprehensive methodology for data gathering and optimal design of mobile transects. *Sustainable Cities and Society*, *55*(January), 102027. https://doi.org/10.1016/j.scs.2020.102027

Salvati, A., Coch, H., & Morganti, M. (2017). Effects of urban compactness on the building energy performance in Mediterranean climate. *Energy Procedia*, *122*, 499–504. https://doi.org/10.1016/j.egypro.2017.07.303

Salvati, A., Coch Roura, H., & Cecere, C. (2016). Urban heat island prediction in the mediterranean context: An evaluation of the urban weather generator model. *Architecture, City and Environment*, *11*(32), 135–156. https://doi.org/10.5821/ace.11.32.4836

Salvati, A., Coch Roura, H., & Cecere, C. (2017). Assessing the urban heat island and its energy impact on residential buildings in Mediterranean climate: Barcelona case study. *Energy and Buildings*, *146*, 38–54. https://doi.org/10.1016/j.enbuild.2017.04.025

Salvati, A., Palme, M., Chiesa, G., & Kolokotroni, M. (2020). Built form, urban climate and building energy modelling: case-studies in Rome and Antofagasta. *Journal of Building Performance Simulation*, *13*(2), 209–225. https://doi.org/10.1080/19401493.2019.1707876

Salvati, A., Palme, M., & Inostroza, L. (2017). Key Parameters for Urban Heat Island Assessment in A Mediterranean Context: A Sensitivity Analysis Using the Urban Weather Generator Model. *IOP Conference Series: Materials Science and Engineering*, *245*(8). https://doi.org/10.1088/1757-899X/245/8/082055

Santamouris, M. (2014). On the energy impact of urban heat island and global warming on buildings. *Energy and Buildings*, *82*, 100–113. https://doi.org/https://doi.org/10.1016/j.enbuild.2014.07.022

Santamouris, M., Haddad, S., Saliari, M., Vasilakopoulou, K., Synnefa, A., Paolini, R., Ulpiani, G., Garshasbi, S., & Fiorito, F. (2018). On the energy impact of urban heat island in Sydney: Climate and energy potential of mitigation technologies. *Energy and Buildings*, *166*, 154–164. https://doi.org/https://doi.org/10.1016/j.enbuild.2018.02.007

Schwarz, G. (2007). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

SITG | Le territoire genevois à la carte. (2023). https://ge.ch/sitg/

Sola, A., Corchero, C., Salom, J., & Sanmarti, M. (2020). Multi-domain urban-scale energy modelling tools: A review. *Sustainable Cities and Society*, *54*(February 2019), 101872. https://doi.org/10.1016/j.scs.2019.101872

Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, *93*(12), 1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1

Street, M., Reinhart, C., Norford, L., & Ochsendorf, J. (2013). Urban heat island in boston - An evaluation of urban airtemperature models for predicting building energy use. *Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association*, 1022–1029.

Tardioli, G., Kerrigan, R., Oates, M., O'Donnell, J., & Finn, D. P. (2018). Identification of representative buildings and building groups in urban datasets using a novel pre-processing, classification, clustering and predictive modelling approach. *Building and Environment*, *140*(May), 90–106. https://doi.org/10.1016/j.buildenv.2018.05.035

Tardioli, G., Narayan, A., Kerrigan, R., Oates, M., O'Donnell, J., & Finn, D. P. (2020). A methodology for calibration of building energy models at district scale using clustering and surrogate techniques. *Energy and Buildings*, *226*, 110309. https://doi.org/10.1016/j.enbuild.2020.110309

Tong, S., Prior, J., McGregor, G., Shi, X., & Kinney, P. (2021). Urban heat: An increasing threat to global health. *The BMJ*, *375*, 1–5. https://doi.org/10.1136/bmj.n2467

Tyagi, A., Kumar, M., Bhan, S. C., Magotra, R., & Sharma, Y. (2021). *Review of Urban Heat Islands : Monitoring , Forecast and Impacts*. *47*(2), 1–28.

Unal Cilek, M., & Cilek, A. (2021). Analyses of land surface temperature (LST) variability among local climate zones (LCZs) comparing Landsat-8 and ENVI-met model data. *Sustainable Cities and Society*, *69*(March), 102877. https://doi.org/10.1016/j.scs.2021.102877

University of Geneva. (2022). *Archives numériques*. https://www.cuepe.ch/html/meteo/archives-numeriques.htm

Vahmani, P., & Ban-Weiss, G. A. (2016). Impact of remotely sensed albedo and vegetation fraction on simulation of urban climate in WRF-urban canopy model: A case study of the urban heat island in Los Angeles. *Journal of Geophysical Research*, *121*(4), 1511–1531. https://doi.org/10.1002/2015JD023718

Varentsov, M. I., Samsonov, T. E., Kargashin, P. E., Korosteleva, P. A., Varentsov, A. I., Perkhurova, A. A., & Konstantinov, P. I. (2020). Citizen weather stations data for monitoring applications and urban climate research: An example of Moscow megacity. *IOP Conference Series: Earth and Environmental Science*, *611*(1), 0–11. https://doi.org/10.1088/1755-1315/611/1/012055

Venter, Z. S., Brousse, O., Esau, I., & Meier, F. (2020). Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data. *Remote Sensing of Environment*, *242*(January), 111791. https://doi.org/10.1016/j.rse.2020.111791

Wang, J., & Biljecki, F. (2022). Unsupervised machine learning in urban studies : A systematic review of applications. *Cities*, *129*(December 2021), 103925. https://doi.org/10.1016/j.cities.2022.103925

Wang, Y., Berardi, U., & Akbari, H. (2015). Comparing the effects of Urban Heat Island Mitigation Strategies for Toronto, Canada. *Energy and Buildings*, *114*. https://doi.org/10.1016/j.enbuild.2015.06.046

Wang, Y., Guo, Z., & Han, J. (2021). The relationship between urban heat island and air pollutants and them with influencing factors in the Yangtze River Delta, China. *Ecological Indicators*, *129*, 107976. https://doi.org/10.1016/j.ecolind.2021.107976

Xu, G., Li, J., Shi, Y., Feng, X., & Zhang, Y. (2022). Improvements, extensions, and validation of the Urban Weather Generator (UWG) for performance-oriented neighborhood planning. *Urban Climate*, *45*(March), 101247. https://doi.org/10.1016/j.uclim.2022.101247

Appendix A

In this Appendix the general simulation settings for the Urban Weather Generator tool simulations are reported.

Meteorological and site parameters		
Location	[-]	Geneva (Switzerland)
Latitude	[°]	46.20
Longitude	[°]	6.14
Daytime boundary layer height[1]	[m]	1000
Nighttime boundary layer height[1]	[m]	50
Inversion height[1]	[m]	150
Temperature measurement at reference site[1]	[m]	2
Air velocity measurements height[1]	[m]	10
Circulation velocity coefficient[1]	[-]	1.2
Exchange velocity coefficient[1]	[-]	1.0
Heat flux threshold for daytime conditions[2]	[W/m^2]	150
Heat flux threshold for nighttime conditions[2]	[W/m^2]	20
Latent fraction of trees[2]	[-]	0.6
Latent fraction of grass[2]	[-]	0.4
Albedo of vegetation	[-]	0.25
Begin month for vegetation participation	[-]	April
End month for vegetation participation	[-]	October
Urban parameters		
City diameter[3]	[m]	7500
Fraction of HVAC waste heat released to urban canyon	[-]	1
Road pavement conductivity	[W/mK]	0.75

Road pavement volumetric heat capacity	[J/m^{3}K]	1600000
Road pavement albedo	[-]	0.05
Roof albedo	[-]	0.2
Wall albedo	[-]	0.2
Glazing ratio of buildings	[-]	0.25
Solar Heat Gain Coefficient from windows	[-]	0.5
Building HVAC system and internal loads		
Occupancy[4]	[m^2/pers]	30
Sensible heat per occupant[1,2]	[W]	100
Latent heat fraction from occupant[1,2]	[-]	0.3
Radiant heat fraction from occupant[1,2]	[-]	0.2
Lighting intensity[4]	[W/m^2]	3.5
Radiant heat fraction from light[1,2]	[-]	0.7
Electric equipment intensity[4]	[W/m^2]	15
Radiant heat fraction from equipment[1,2]	[-]	0.5
Heating set point[4]	[°C]	20

Notes:

1. (Bueno et al., 2012)
2. (Bueno et al., 2014)
3. Measured with GIS tools
4. (Tardioli et al., 2020)

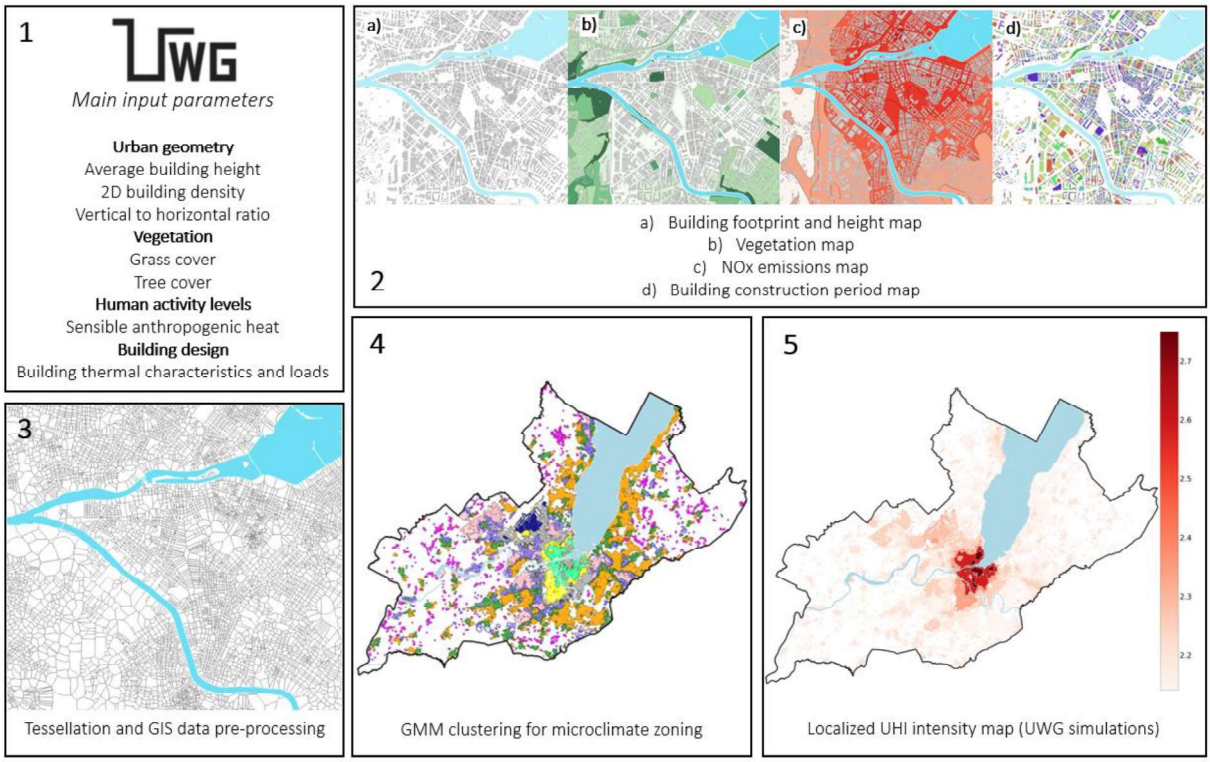

Figure 1: Schematic representation of the developed methodological steps applied to the Canton of Geneva.

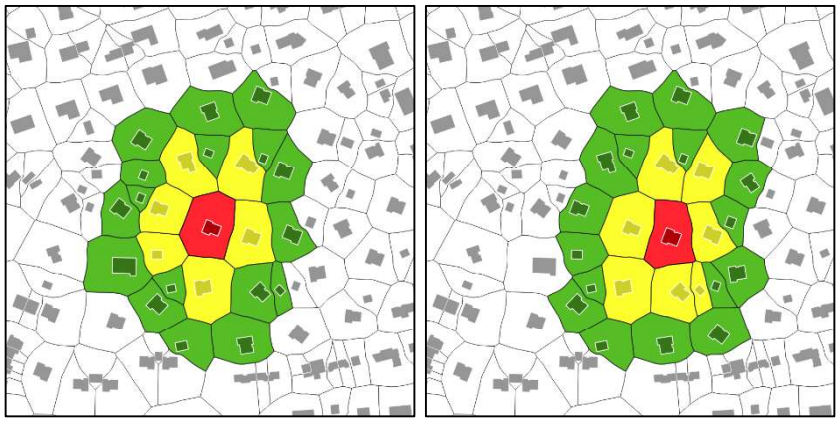

Figure 2: 1st and 2nd order cells (yellow and green respectively) determined through spatial weights with respect to a reference building (red). The operation is repeated for all the buildings.

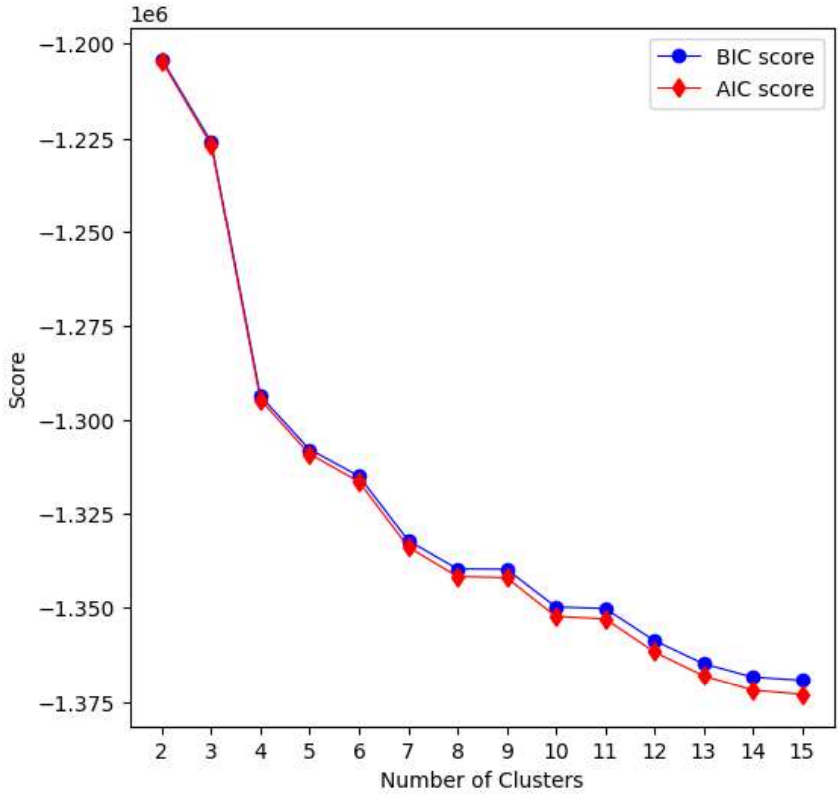

Figure 3: BIC and AIC scores per number of clusters

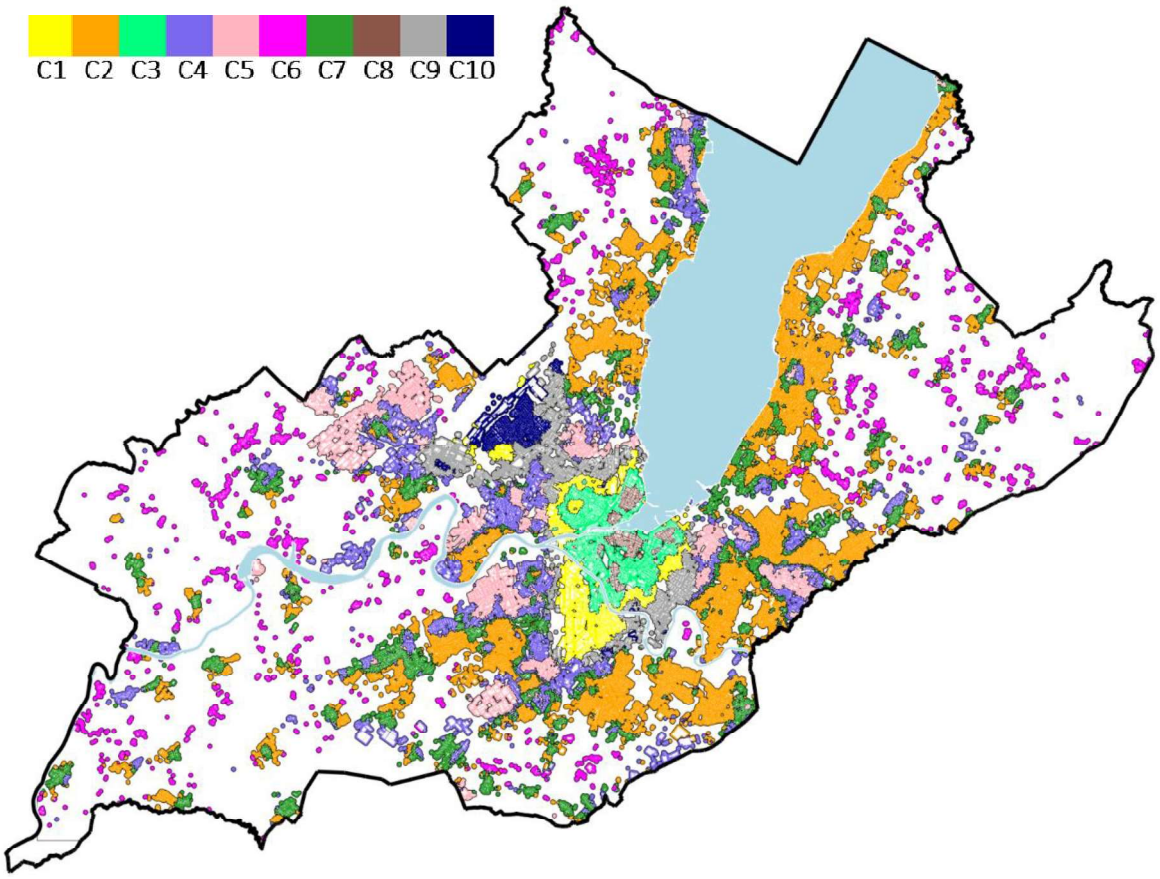

Figure 4: Ten homogeneous microclimatic clusters of the Geneva Canton identified through GMM clustering. Each cluster is represented by a different colour and white parts represent the surrounding rural areas.

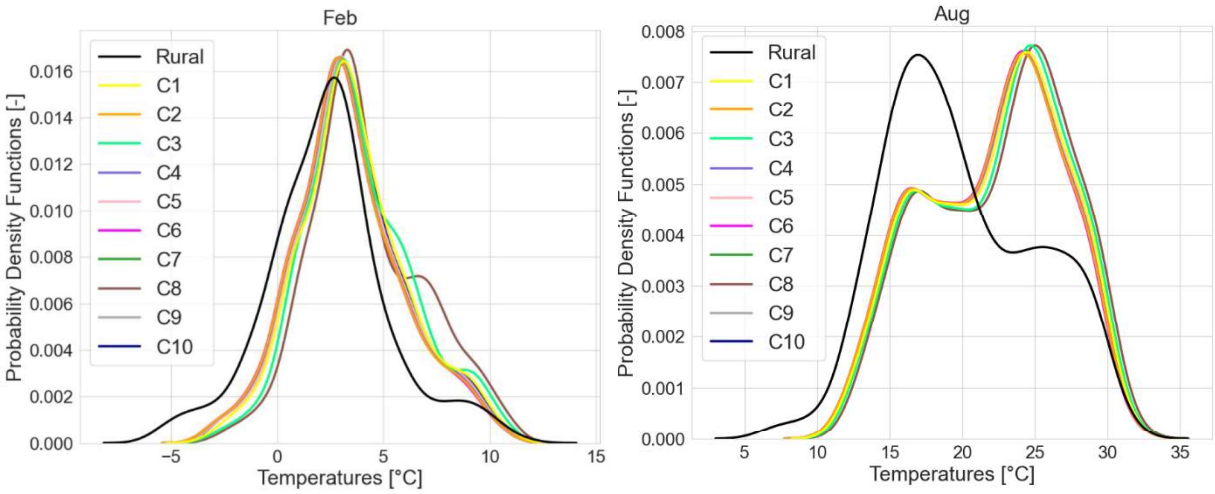

Figure 5: Probability density functions of rural and urban temperatures.

The representative months are February (left) and August (right)

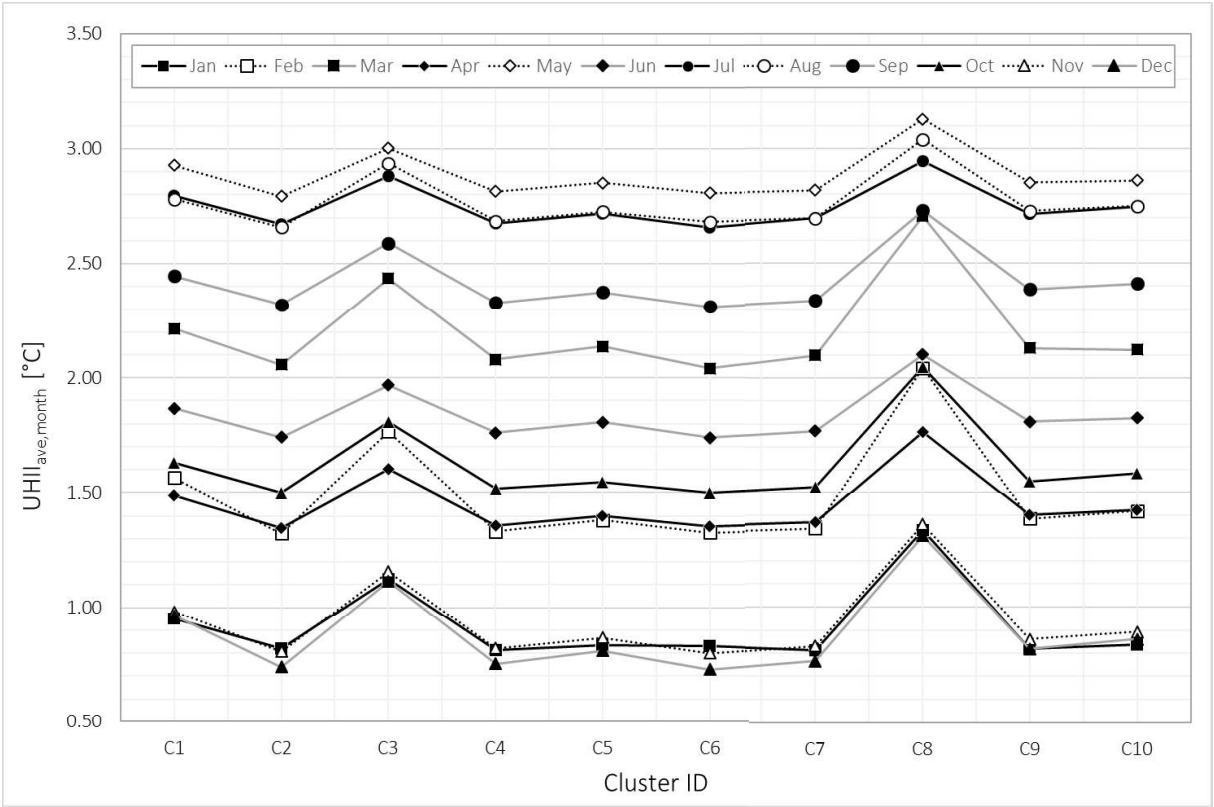

Figure 6: Average monthly urban heat island intensity among the urban clusters

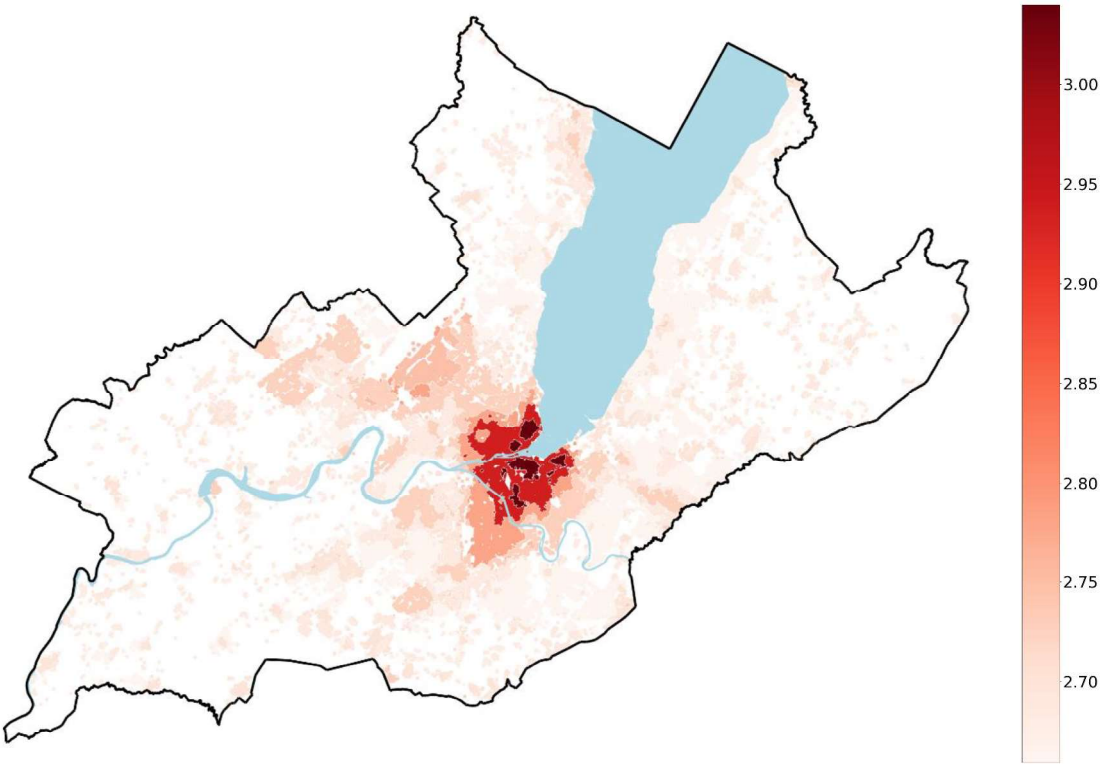

Figure 7: Spatial variability of the monthly average UHII in August within the Canton of Geneva

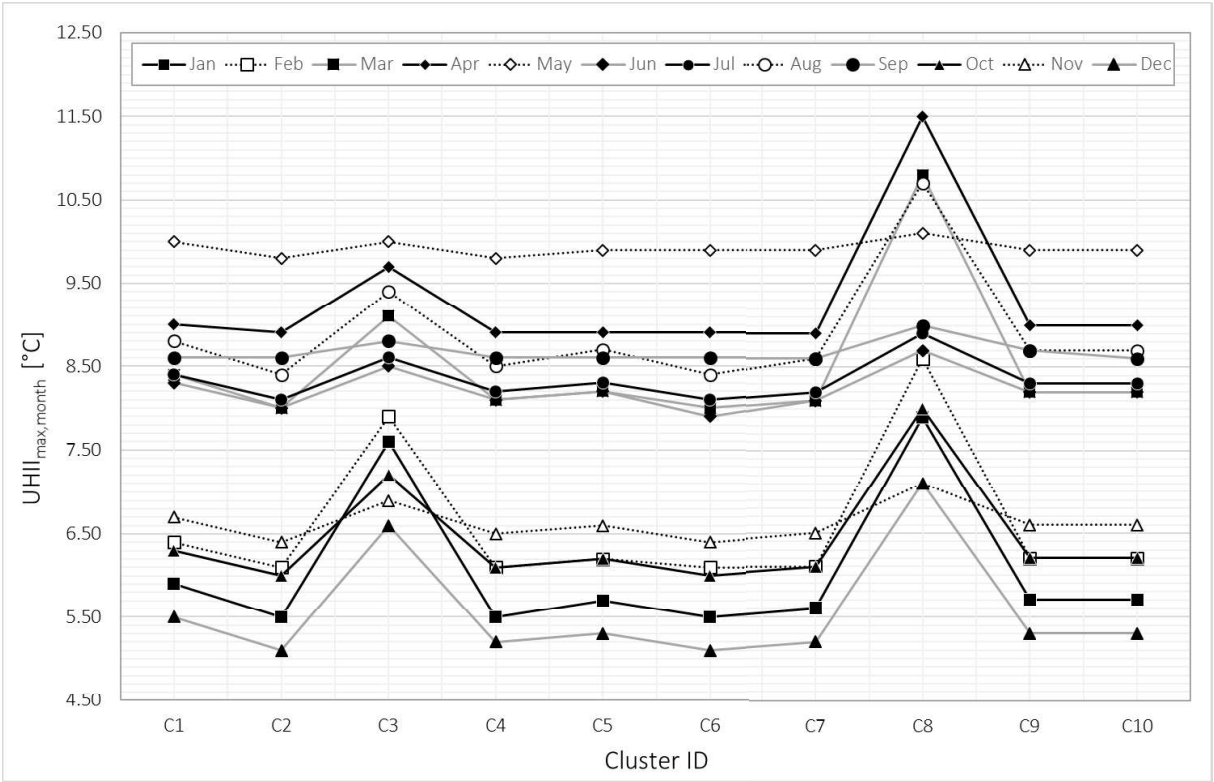

Figure 8: Maximum monthly urban heat island intensity among the urban clusters

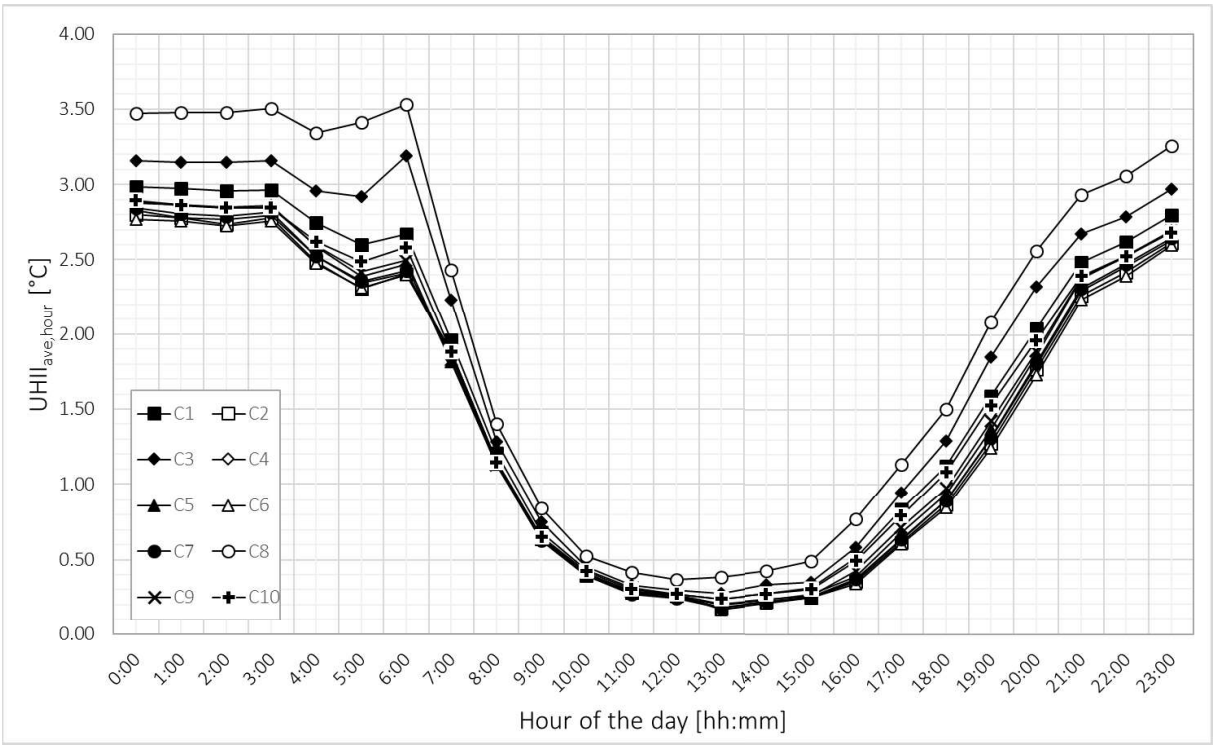

Figure 9: Average hourly urban heat island intensity among the urban clusters

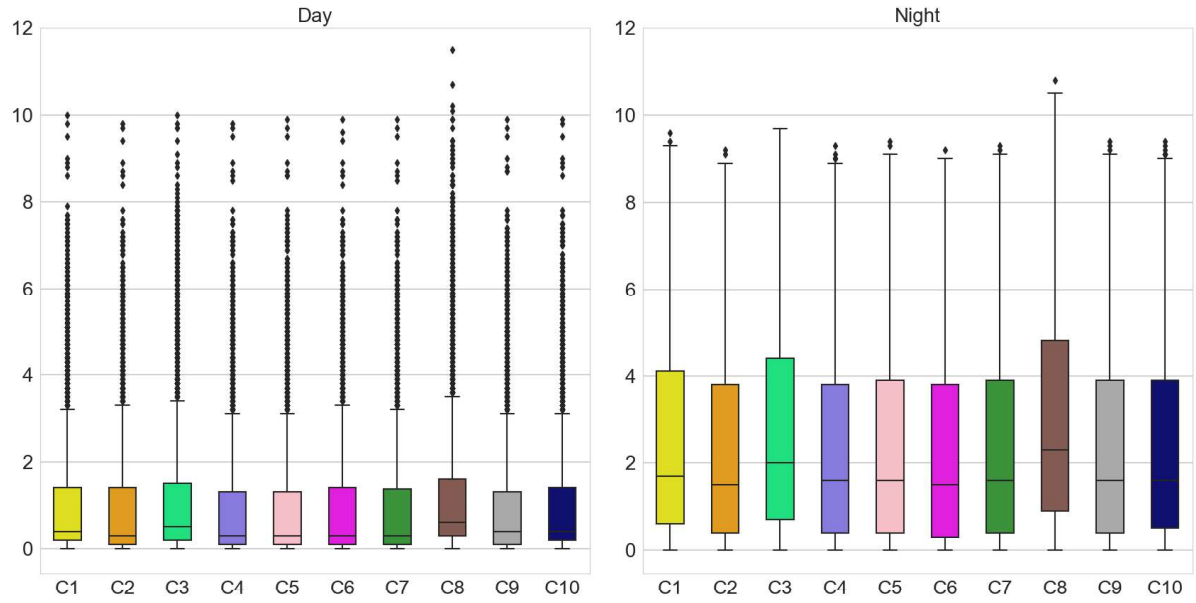

Figure 10: Boxplots of UHII during daytime (left) and nighttime (right)

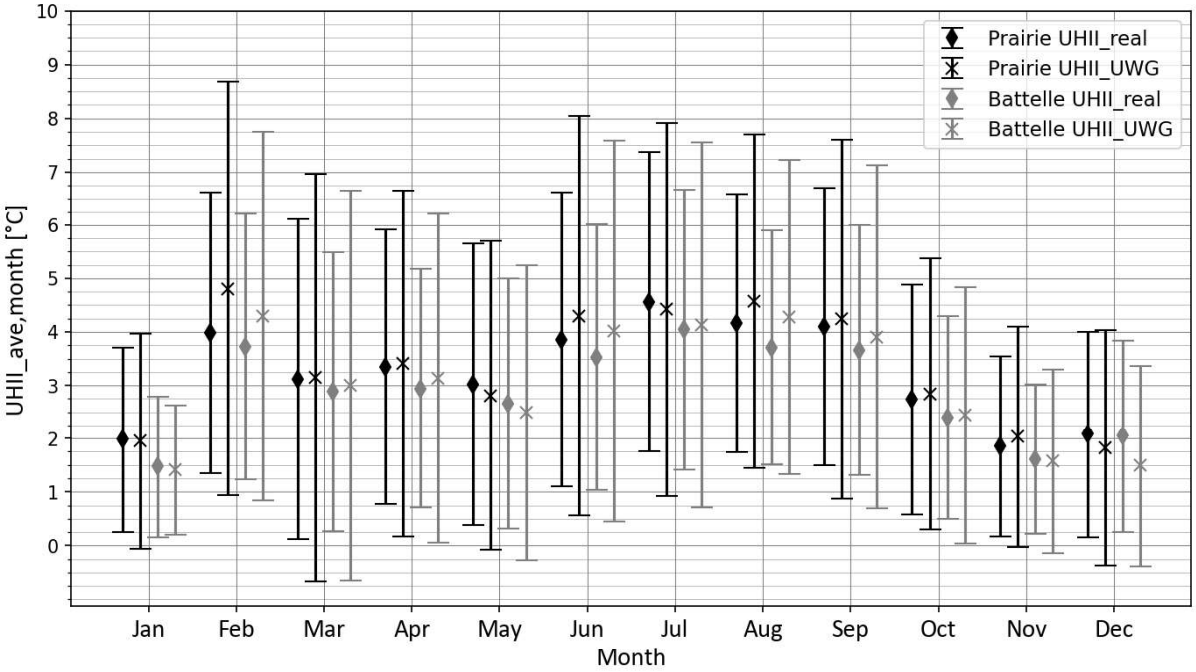

Figure 11: Average monthly UHII (markers) and standard deviation (bars) arising from UWG simulations (UHII_UWG) and from real measurements (UHII_real) for two selected urban weather stations

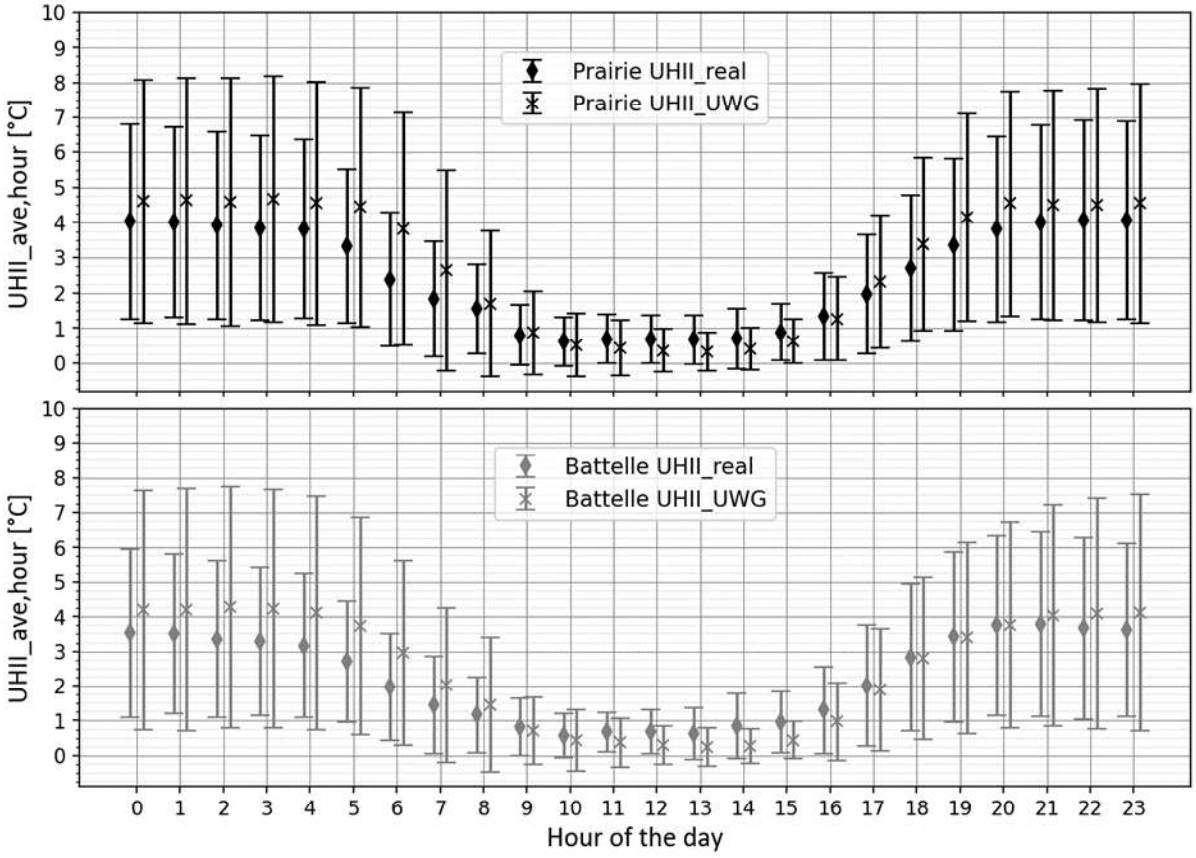

Figure 12: Average hourly UHII (markers) and standard deviation (bars) arising from UWG simulations (UHII_UWG) and from real measurements (UHII_real) for two selected urban weather stations

Table 1: UWG main input parameters for UHI simulation

Urban geometry		
$\overline{H}$	[m]	Average height of buildings
ρ_{urb}	[-]	Urban area building plan density: ratio between built and un-built area
VH	[-]	Vertical to horizontal ratio: ratio between façade area and plan area of the site
Vegetation[1]		
ρ_{grass}	[-]	Grass coverage: fraction of the urban ground covered in grass/shrubs
ρ_{trees}	[-]	Tree coverage: fraction of the urban ground covered in trees
Human activity levels		
Q_{sens}	[W/m^2]	Non-building sensible heat at street level: heat from cars, pedestrians, street cooking, etc.
Building characteristics[2]		
R_{wall}	[m^{2}K/W]	Thermal resistance of walls
R_{roof}	[m^{2}K/W]	Thermal resistance of roof
U_{window}	[W/m^{2}K]	Thermal transmittance of windows
i	[ach]	Building infiltration rate

Notes:

1. Since in the reference vegetation map the vegetation coverage is not differentiated between grass and trees, the vegetation coverage value is split in half between ρ_{grass} and ρ_{trees} resulting in ρ_{grass}=ρ_{trees}
2. Building characteristics are derived from the building construction period

Table 2: Reference values of building characteristics based on different construction periods

Name	Units	Before 1980	Between 1980-2000	After 2000
R_{wall}	[m^2K/W]	1.28	0.47	0.21
R_{roof}	[m^2K/W]	1.01	0.38	0.21
U_{window}	[W/m^2K]	3.22	1.95	1.40
i	[ach]	1.10	0.60	0.45

Table 3: $U_{cluster}$ related to the ten clusters identified through GMM algorithm

Cluster ID	N° of buildings	$\overline{H}$ [m]	ρ_{urb} [-]	VH [-]	$\rho_{grass/trees}$ [-]	Q_{sens} [W/m^2]	Construction period [%][1]		
							Pre1980	1980-2000	Post2000
C1	2856	17.8	0.31	1.5	0.13	8.3	70	15	15
C2	21676	7.5	0.11	0.4	0.33	<1.5	47	32	21
C3	3916	21.6	0.39	2.1	0.06	16.5	77	13	10
C4	6826	11.3	0.15	0.5	0.27	<1.5	51	23	27
C5	4195	16.81	0.22	1.0	0.21	<1.5	60	23	17
C6	3135	7.7	0.08	0.3	0.47	<1.5	58	27	16
C7	10402	9.5	0.18	0.7	0.29	<1.5	52	29	19
C8	1827	20.6	0.55	3.2	0.03	20.0	75	10	15
C9	2767	16.3	0.17	0.8	0.22	4.0	64	20	16
C10	1234	9.5	0.14	0.5	0.28	11.7	46	30	24

Notes:

1. The building construction period is expressed as percentage with respect to the total number of buildings within each cluster

Appendix A

Meteorological and site parameters		
Location	[-]	Geneva (Switzerland)
Latitude	[°]	46.20
Longitude	[°]	6.14
Daytime boundary layer height[1]	[m]	1000
Nighttime boundary layer height[1]	[m]	50
Inversion height[1]	[m]	150
Temperature measurement at reference site[1]	[m]	2
Air velocity measurements height[1]	[m]	10
Circulation velocity coefficient[1]	[-]	1.2
Exchange velocity coefficient[1]	[-]	1.0
Heat flux threshold for daytime conditions[2]	[W/m^2]	150
Heat flux threshold for nighttime conditions[2]	[W/m^2]	20
Latent fraction of trees[2]	[-]	0.6
Latent fraction of grass[2]	[-]	0.4
Albedo of vegetation	[-]	0.25
Begin month for vegetation participation	[-]	April
End month for vegetation participation	[-]	October
Urban parameters		
City diameter[3]	[m]	7500
Fraction of HVAC waste heat released to urban canyon	[-]	1
Road pavement conductivity	[W/mK]	0.75
Road pavement volumetric heat capacity	[J/m^3K]	1600000
Road pavement albedo	[-]	0.05

Roof albedo	[-]	0.2
Wall albedo	[-]	0.2
Glazing ratio of buildings	[-]	0.25
Solar Heat Gain Coefficient from windows	[-]	0.5
Building HVAC system and internal loads		
Occupancy[4]	[m^2/pers]	30
Sensible heat per occupant[1,2]	[W]	100
Latent heat fraction from occupant[1,2]	[-]	0.3
Radiant heat fraction from occupant[1,2]	[-]	0.2
Lighting intensity[4]	[W/m^2]	3.5
Radiant heat fraction from light[1,2]	[-]	0.7
Electric equipment intensity[4]	[W/m^2]	15
Radiant heat fraction from equipment[1,2]	[-]	0.5
Heating set point[4]	[°C]	20

Notes:

1. (Bueno et al., 2012)
2. (Bueno et al., 2014)
3. Measured with GIS tools
4. (Tardioli et al., 2020)