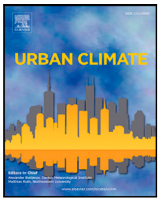
Urban Climate 62 (2025) 102527

Contents lists available at ScienceDirect

Urban Climate

journal homepage: www.elsevier.com/locate/uclim



Predicting the impact of the urban texture on the Urban Heat Island intensity using Machine Learning: The case for the Iberian Peninsula

Josep A. Ferré, Anton Vernet, Alexandre Fabregat *

Universitat Rovira i Virgili, Department of Mechanical Engineering, Av. Països Catalans, 26, 43007, Tarragona, Spain

ARTICLE INFO

Keywords:
Urban Heat Island (UHI)
Machine learning regression models
Urban texture
Urbanization
Meteorological factors
Iberian peninsula

ABSTRACT

Urbanization increases impervious surfaces that store heat, concentrates residual heat sources, and alters atmospheric ventilation. This makes cities warmer than surrounding rural areas, especially at night, a phenomenon known as Urban Heat Island (UHI). UHI intensity depends on meteorology and urban morphology, including building height, population density, green cover, and built-up area. However, estimating each factor's contribution remains challenging due to variability. Furthermore, defining the intensity of the UHI as the temperature difference between urban and rural areas introduces arbitrariness, since factors such as altitude influence the results. To address these issues, we propose a systematic reference temperature definition and a machine learning model to predict daily maximum UHI intensity (UHII). Trained on UHII data from multiple rural–urban pairs, the model uses 100×100 meter hourly raster data sets of meteorological and urban morphology features from 11 Iberian Peninsula cities in 2017. Assuming weak temporal and spatial correlations, it operates independently of past time steps and neighboring cells, ensuring a fully local approach. The performance of the test across cities and a hybrid synthetic region exceeds $R^2 = 0.8$. The main findings indicate that the average rate of change on the daily maximum UHII is 0.09 °C per meter of building height, 0.34 °C per 0.1 point increase in build-up fraction, 0.08 °C per 1000 people/m2 and-0.11 °C per 0.1 point increase in vegetation fraction.

Focused on the Iberian Peninsula, the model accounts for climatic vulnerabilities, helping urban planners in designing strategies to protect public health and improve resilience.

1. Introduction

Urbanization entails the emergence of an urban texture that replaces the previous unaltered land cover with build-up and often impervious surfaces capable of modifying the local radiative energy balance and the ventilation capacity while also accommodating densely concentrated residual heat sources (John Arnfield, 2003; Schrijvers et al., 2015). Consequently, cities frequently exhibit elevated temperatures in comparison to rural regions, particularly noticeable during nocturnal hours when the urban texture radiates back accumulated heat (Lai et al., 2021b; Johnson et al., 1991). This phenomenon, known as Urban Heat Island (UHI) (Oke, 1973, 1982; Nuruzzaman, 2015), amplifies the thermal stress within urban areas, further burdening the health and well-being of city residents (Kabisch et al., 2023).

The threat becomes especially pronounced during heatwave events, which are projected to increase in intensity, duration, and frequency due to ongoing climate change (Grimm et al., 2008; Ruosteenoja and Jylhä, 2023). Southern Europe, and the Iberian Peninsula in particular, is expected to be highly affected (Lorenzo et al., 2021). Recent studies underscore that temperature-related

* Corresponding author.
E-mail addresses: josep.a.ferre@urv.cat (J.A. Ferré), anton.vernet@urv.cat (A. Vernet), alexandre.fabregat@urv.cat (A. Fabregat).

https://doi.org/10.1016/j.uclim.2025.102527
Received 22 April 2025; Received in revised form 9 June 2025; Accepted 24 June 2025
Available online 16 July 2025

mortality is not uniform across populations, but instead varies significantly by age group and region. For instance, projections for Lisbon and Porto metropolitan areas suggest a growing burden of heat-related mortality among elderly populations (65+), particularly under high-emission scenarios, with notable increases anticipated by the end of the century (Rodrigues et al., 2020). Likewise, temperature extremes, both heat and cold, are linked to significant excess mortality, especially among the most vulnerable age groups, such as individuals aged 85 and above, who are affected by both moderate and extreme conditions (Rodrigues et al., 2021). Furthermore, cause-specific analyses indicate that climate change could shift the burden of mortality toward heat-related causes such as respiratory and cardiovascular diseases, even under adaptation scenarios that consider demographic changes (Rodrigues, 2023).

The local Urban Heat Island intensity (UHII) is the (air or surface) temperature difference between a location within the urban area and that which would exist in its absence. This quantity therefore quantifies the impact of emerged urban textures on the local energy balance either by the modification of the land cover and/or the appearance of residual heat sources associated with human activity (Kim et al., 2022). For obvious reasons, the non-urbanized temperature value cannot be accessed and instead the usual approach is to use the temperature of a nearby rural location as a proxy for non-urbanized conditions (Chakraborty and Lee, 2019). To minimize variations due to climate characteristics and elevation, both the rural and urban locations must be close enough.

Along with background climate, urban texture has been shown to impact the UHII (Lai et al., 2021a; Hou et al., 2023). Different artificial land cover exhibit different thermal properties that modify the local radiative heat exchange (Doulos et al., 2004). Building shape characteristics and their spatial distribution and spatial arrangement have also been found to modulate the local temperature by modifying the local urban canopy hydrodynamics and turbulent heat transport that depend on the arrangement of these rugosity elements and the formation of street canyons (Afshari, 2023; Voogt and Oke, 2003). In addition, differences in the amount of greenness modify the local cooling effect associated with evapotranspiration (Yin et al., 2023; Choi et al., 2022; Iungman et al., 2023). All in all, the UHII exhibits significant levels of variability between different cities but also within the same urban area due to distinct local urban texture characteristics (Gago et al., 2013).

Given the central role played by the physics of fluids and heat transport, Computational Fluid Dynamics (CFD) studies can provide highly detailed information on hydrodynamics and thermal fields in both space and time (Huo et al., 2021; Wang and Li, 2016; Aghamolaei et al., 2021; Antoniou et al., 2019). However, this approach has two main limitations. Firstly, solving the coupled system of Partial Differential Equations for mass, momentum, and energy conservation typically requires substantial computational resources, restricting the size of the urban domain that can be studied (Wang and Li, 2016). Secondly, regardless of the specific CFD framework used (typically Reynolds-Averaged Navier Stokes or RANS), the solution depends on various parametrizations, including turbulent transport, which are difficult to estimate accurately. Additionally, this methodology often necessitates a detailed inventory of heat sources, which can be dispersed, transitory, and/or mobile (Kubilay et al., 2020).

Further efforts to simplify the problem have led to the development of various mathematical tools, such as UrbClim (Sobstyl et al., 2018), RayMan (Matzarakis et al., 2007), SOLWEIG (Lindberg et al., 2008), ENVI-met (Bruse and Fleer, 1998), and Ladybug-Grasshopper (Evola et al., 2020), which are capable of predicting the Urban Heat Island Intensity across the entire city extent with reasonable computational resources, provided properly adjusted parametrizations of the underlying physics.

Recent advancements in artificial intelligence and machine learning have significantly enhanced the modeling and forecasting of Urban Heat Islands, enabling more accurate spatial and temporal assessments of urban thermal dynamics. Studies have applied a wide range of Machine Learning (ML) approaches, including deep neural networks (DNNs), recurrent neural networks (e.g., Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU)), gradient boosting (e.g., XGBoost), and support vector regression, to predict UHI intensity or land surface temperature (LST) using multi-source urban, climatic, and remotely sensed data. For instance, Mohammad et al. (2022) combined Artificial Neural Network (ANNs) and XGBoost regression to simulate future land use and LST patterns in Ahmedabad, India, demonstrating the effectiveness of ML in capturing the spatial evolution of urban heat. Yun et al. (2020) employed LSTM networks trained on 18 years of hourly data to predict UHI intensity in Sydney, Australia, achieving high accuracy (R^2 up to 0.932) and revealing temperature as the primary driver of UHI fluctuations in complex coastal-desert environments. Similarly, Tehrani et al. (2024) used deep learning models, including GRU networks, to estimate current and future UHI intensity across 69 European cities, identifying urban morphology features such as building density and green space as critical predictors. Other studies across diverse geographic contexts — including South Korea, Iran, Vietnam, and globally — have consistently shown that machine learning models outperform traditional methods in forecasting UHI behavior, offering high predictive accuracy (often $R^2 > 0.8$) and valuable insights into the environmental and anthropogenic factors driving urban heat. Collectively, these findings underscore the potential of AI-driven modeling to support climate-resilient urban planning by anticipating UHI dynamics under current and future urbanization scenarios.

Data-driven approaches have also been employed to investigate the Urban Heat Island phenomenon using measurements from meteorological station networks or similar sensors that provide time series of temperature records. The UHII is then estimated by computing the temperature difference between two neighboring stations, one regarded as *rural* and the other one as *urban* (Sobstyl et al., 2018). In this case, the sampling size was highly restricted due to the relatively small number of suitable pairs of stations. In addition, the classification between *rural* and *urban* is somewhat arbitrary resulting in significant levels of variability in the urbanization metrics used to classify the stations. In other words, differences in urban texture metrics between the urban and rural stations in one pair could be of the same order as those found between two rural or two urban stations from different pairs (Stewart and Oke, 2009; Martin-Vide et al., 2015; Liu et al., 2023).

A third approach that has gained notable acceptance in recent years is the use of remote sensing data sourced from satellite networks (Naserikia et al., 2022; Dougherty and Jain, 2023; Marando et al., 2019; Liu et al., 2023). While it allows for significantly enhanced spatial resolution for numerous global locations, this methodology encounters a limitation in terms of temporal resolution

due to the intervals between successive images of the same area, a challenge that can be exacerbated by factors like the presence of clouds (Wang et al., 2019).

In previous efforts to develop models that enhance our understanding of the impact of urban texture on Urban Heat Island intensity, several additional limitations have emerged, compounding those already associated with each individual methodology. On the one hand, many studies are limited to a very specific areas of study, usually a single city, (Marando et al., 2019; Oh et al., 2020; Badugu et al., 2023; Wang et al., 2024; Touchaei and Wang, 2015; Huang and Wang, 2019; Güller and Toy, 2024; Zhou and Chen, 2018), hindering the generalization of results. In other cases, studies cover vast spatial domain comprising multiple urban areas but lack fine-grain prediction capabilities below the city-wide stratification (Iungman et al., 2024; Zhou et al., 2013, 2017a; Manoli et al., 2019; Debbage and Marshall Shepherd, 2015; Zhou et al., 2017b).

In this work, we develop a model that accurately captures the relationship between the local UHII and two types of predictors: namely, local atmospheric conditions and urban texture feature, while also alleviating the constraints mentioned before. To achieve this goal, we combine Machine Learning (ML) techniques and a sufficiently extensive working dataset, facilitating the extrapolation of results to any city or urban setting, as long as it predominantly pertains to the same climate region and adheres to comparable urbanization norms and practices. Thus, the current study is limited to the Iberian Peninsula, a region of notable significance due to the projected escalation in heat stress and susceptibility resulting from the Global Warming context (Lorenzo et al., 2021). Importantly, to enhance reproducibility and foster transparency in research, all data is open source, ensuring accessibility and facilitating further analysis by the scientific community.

In this study, the daily Urban Heat Island Intensity is estimated using data from a repository containing hourly raster meteorological variables (European Environment Agency, 2020), derived from high-resolution UrbClim (De Ridder et al., 2015) simulations. Additional predictors include wind speed, precipitation, solar irradiance, urban-rural temperature differences, and rural mask average temperature. UrbClim integrates simplified atmospheric physics, detailed urban surface data, and downscaled ERA5 reanalysis to provide microclimate-scale insights. Its excellent agreement with European meteorological station data underscores the repository's reliability for high-resolution UHII analysis.

These data are combined with urban texture predictors evaluated on the same 100×100 meter pixel grid for the 11 selected cities in the Iberian Peninsula. By segmenting the computational domain into urban and rural sections through the land cover/land use data, it is possible to define multiple virtual urban-rural station pairs for which the UHII is quantified as the temperature difference between the temperature of each hourly urban pixel and the mean temperature of all rural pixels at the same elevation. This strategy minimizes the error introduced when the reference temperature is merely estimated based on some 'rural mask' regardless of the terrain elevation that, for the case of mainland China, has been estimated to be around 1.7 °C (Yao et al., 2018). Specifically, the list of urban texture predictors include normalized building height, build-up surface fraction, population density, vegetation cover fraction, elevation, and land cover/use. Finally, we also used the month of the year and the Köppen-Geiger class.

The model proposed in this work exhibits excellent predictive capabilities when used in both each individual city and a hybrid urban area made up of randomly combining cells from the 11 locations. Specifically, the determination coefficient R^2 in the test data set for the worst performing city is around 0.85. The average R^2 score of the hybrid model when applied to all 11 cities is slightly above 0.80. When used to estimate the mean role of each urban texture predictor, the average rate of change on the daily maximum UHII is 0.09 °C per meter of building height, 0.34 °C per 0.1 point increase in the build-up fraction, 0.08 °C per 1000 people/m^2 and −0.11 °C per 0.1 point increase in vegetation fraction. The findings derived from this study offer significant value to policy makers and officials engaged in decision-making processes aimed at improving the livability and well-being of city dwellers.

The data retrieval, pre-processing, and analysis are presented in Section 2. Modeling methodology is discussed in Section 3, while the results and discussion are presented in Section 4. Finally, Section 5 provides the conclusions.

2. Materials and data

This paper aims to better understand the role of urban texture in the intensity of the urban heat island phenomenon (UHI). Along with local meteorology, the observed increase in temperatures in urban environments with respect to nearby rural areas also depends on the emergence of artificial surfaces made of materials capable of storing radiative energy in the form of heat that is subsequently released during nighttime hours (Oke, 1973). This urban texture also hosts residual heat sources (for example, road traffic) (Husni et al., 2022) and modifies air flow and heat ventilation (Evola et al., 2020). Previous efforts aimed at understanding the role of build-up fraction, distribution and height along with greenness fraction and density of human activities have focused the analysis on city-wide scales (see for instance Debbage and Marshall Shepherd, 2015; Zhou et al., 2017b) and/or relied on measurements of the temperature difference between a pair of nearby points regarded as ''rural'' and ''urban'' to estimate the UHI intensity (see for instance (Sobstyl et al., 2018)). Finding suitable pairs of measurements to draw robust conclusions, either at microclimatic scales within the same city or in inter-city comparisons, is complex due to variations in temperature data associated with imprecise definition of ''rural'' areas and also due to differences in altitude between paired measurements. On the other hand, high spatial resolution studies based on remotely sensed data are often hampered by low temporal resolution and cloudiness effects, which limit the completeness of the data (Diem et al., 2024).

Here we alleviate all these limitations by considering high spatial and temporal resolution UHII data from microclimate simulations expressed as 100×100 m gridded rasters (Hooyberghs et al., 2019; De Ridder et al., 2015). Using Machine Learning (ML) techniques, we have derived numerical models relating this target variable to various meteorological and urban texture predictors that have been projected and interpolated at this native resolution. While urban texture characteristics are assumed constant over time, the present analysis considered an entire year of local atmospheric conditions to feed standard ML workflows. The resulting fine-grained model is used to investigate the UHII dependence on each predictor.

In this study, we only used open source data derived from freely available datasets that facilitate result verification while ensuring transparency and reproducibility.

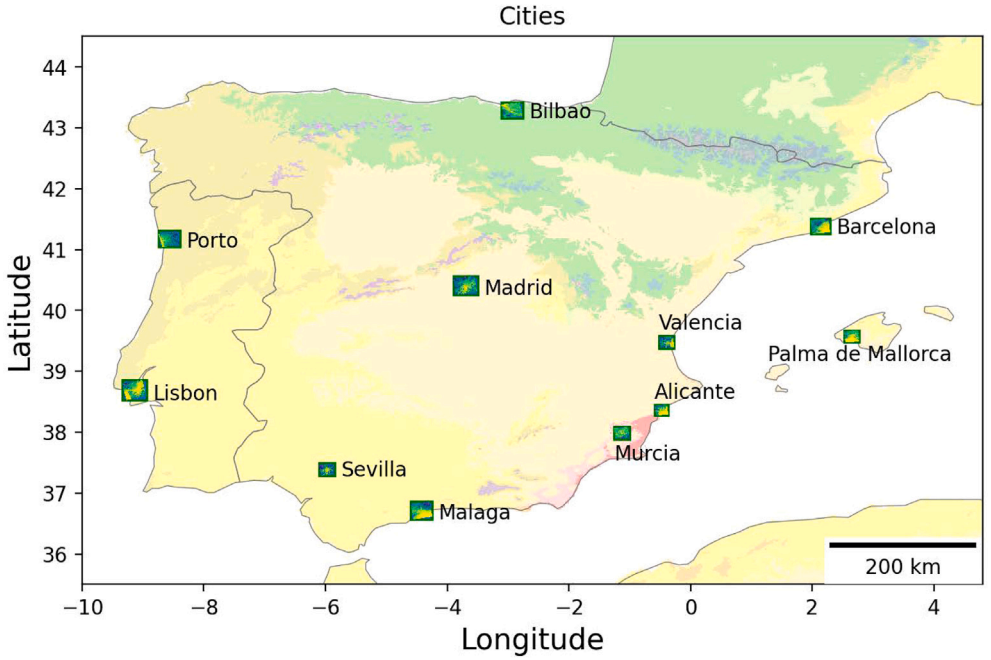


Fig. 1. Location and extent of the 11 urban areas selected from the 100 analyzed in the study *Urban Heat Island (UHI) intensity modeling, Jan. 2020* (European Environment Agency, 2020). Colors represent the Köppen-Geiger climate classifications as in Beck et al. (2023).

2.1. Study area

In this study, we focus on the $N = 11$ cities within the Iberian Peninsula (Spain and Portugal) which form a subset of the 100 urban areas analyzed in the *Urban Heat Island (UHI) intensity modeling, Jan. 2020* (European Environment Agency, 2020), available through the EEA Geospatial Data Catalogue as 100×100 m raster cells. By geographically restricting the cities considered in this study, we ensure that all locations share similar climate characteristics and building construction methods and practices. This approach minimizes the impact of these factors on the variability of the intensity of the urban heat island. Fig. 1 shows the extent of the computational raster domain of the 11 urban areas on top of the Iberian Peninsula map colored according to the Köppen-Geiger climate class at 1 km resolution (Beck et al., 2023).

For each city $k = 1 \dots N$, let $R(i,j)$, $U(i,j)$ and $W(i,j)$ be the *rural*, *urban* and *water* raster masks respectively. The indices $i = 1 \dots n$ and $j = 1 \dots n$ represent the row and column of the raster cell with a total of n^2 cells (square domains). These masks take the value of 1 for cells belonging to the respective category and 0 otherwise. A *land* mask L can be defined by combining U and R, this is, $L(i,j) = U(i,j) + R(i,j)$. Along with these masks, let $E(i,j)$ be a raster containing the elevation of each cell (i,j). The set of cells in each mask is defined as $R_s = \{(i,j)|R(i,j) = 1\}$, $U_s = \{(i,j)|U(i,j) = 1\}$, $L_s = \{(i,j)|L(i,j) = 1\}$ and $W_s = \{(i,j)|W(i,j) = 1\}$ respectively, where $N_R = |R_s|$, $N_U = |U_s|$, $N_L = |L_s|$ and $N_W = |W_s|$ are the number of cells in each mask.

While the raster masks U, R, W and L along with E are all assumed constant in time, time-varying variables including air temperature $T(h,d,i,j)$ depend on the hour of the day ($h = 1 \dots n_h$) and the day of the year ($d = 1 \dots n_d$), where $n_h = 24$ and $n_d = 365$.

The hourly and daily temporal averages of temperature, $\overline{T}_h$ and $\overline{T}_d$ respectively, are defined as:

$$\overline{T}_h(h,i,j) = \frac{1}{n_d} \sum_{d=1}^{n_d} T(h,d,i,j), \tag{1}$$

$$\overline{T}_d(d,i,j) = \frac{1}{n_h} \sum_{h=1}^{n_h} T(h,d,i,j). \tag{2}$$

The spatial averaging operator $\langle \cdot \rangle_M$, applied over any mask $M = \{R, U, L, W\}$, is defined as:

$$\langle T \rangle_M(h,d) = \frac{1}{N_M} \sum_{(i,j) \in M} T(h,d,i,j) \tag{3}$$

Using this notation, the daily average of the temperature over the rural mask $\langle \overline{T}_d \rangle_R$ can be computed as:

$$\langle \overline{T}_d \rangle_R(d) = \frac{1}{N_R} \sum_{(i,j) \in R} \overline{T}_d(d,i,j) = \frac{1}{N_R} \frac{1}{n_h} \sum_{(i,j) \in R} \sum_{h=1}^{n_h} T(h,d,i,j) \tag{4}$$

Table 1
List of the 11 urban areas considered in this work with the name of the city, the country, the location, the raster size, the percentage of cells in the combined rural and urban masks and the percentage in each of these two classifications.

k	City	Country	lon (°)	lat (°)	n	$p = n^2$	F_U	F_R	F_W
1	Alicante	Spain	−0.48	38.36	181	32 761	28	37	35
2	Barcelona	Spain	2.13	41.38	251	63 001	41	30	29
3	Bilbao	Spain	−2.94	43.29	251	63 001	21	74	5
4	Lisbon	Portugal	−9.14	38.7	301	90 601	38	25	38
5	Madrid	Spain	−3.7	40.41	301	90 601	63	37	0
6	Malaga	Spain	−4.42	36.72	281	78 961	15	51	34
7	Murcia	Spain	−1.13	37.99	201	40 401	22	78	0
8	Palma	Spain	2.65	39.57	201	40 401	21	49	30
9	Porto	Portugal	−8.56	41.18	251	63 001	38	52	9
10	Sevilla	Spain	−5.98	37.39	201	40 401	42	55	3
11	Valencia	Spain	−0.39	39.48	201	40 401	38	45	17

For each urban area, Table 1 contains the name of the city, its country, longitude, latitude, and raster size. The three next columns contain the percentage of *urban*, *rural* and *water* cells, $F_U = 100 \times N_U/n^2$, $F_R = 100 \times N_R/n^2$ and $F_W = 100 \times N_W/n^2$ respectively. The cell count for each city raster varies depending on the extent of the city limits. Thus, for example, Alicante features a smaller raster size with 181 × 181 cells covering an area of 18.1 × 18.1 = 327.61 km^2. In contrast, Madrid and Lisbon exhibit the largest city extents with grids spanning 301 × 301 cells (equivalent to 906.01 km^2). Notably, in the case of Lisbon, only 62% of the pixels can be classified as urban or rural, with the remaining 38% comprising water bodies—the Tajo River estuary or the Atlantic Ocean.

2.2. Land cover and elevation data

The *water* and *land* — rural and urban combined — masks are built from the Coordination of Information on the Environment (CORINE) land cover/use data (European Environment Agency, 2018). The urban mask is the result of combining the CORINE 'Artificial surfaces' classes, while the rural mask is the union of 'Agricultural areas' and 'Forest and semi natural areas' classes. Fig. 2 shows the raster plot for the 11 cities colored according to each category along with the main streets layout in the background. Note that the x-scale and the y-scale (not shown for simplicity) are different for each city to keep the same window size.

The terrain elevation raster for each city, $E(i, j)$, was obtained from the 30-meter resolution Digital Terrain Model (Hengl et al., 2020), resampled to the reference 100 × 100 m resolution for each pixel within the *land* mask. Fig. 3 illustrates the elevation distribution within the urban (red) and rural (green) masks for all 11 cities considered in this study. While the degree of variability differs among urban areas, several locations exhibit notable values of skewness — especially within the rural mask — indicating substantial asymmetry in the distribution of elevation values across individual masks.

These results indicate that selecting the *rural* temperature, necessary for calculating the local UHII as the difference between urban and rural temperatures, may not be straightforward. In addition to the ambiguous definition of *rural* in terms of land cover, elevation differences can significantly affect UHII estimation, particularly in cities with notable altitude variations between urban and rural areas (e.g., Bilbao or Málaga).

To address this, for each city k, we computed the percentage of rural cells that have at least one corresponding urban cell within an elevation difference threshold of $\epsilon = 8$ m, denoted as F_E. By definition, F_E must be equal to or smaller than F_R, as it represents a subset of the complete set of rural mask cells.

Table 2 list the value of both fractions F_E and F_R — the latter also shown in Table 1 —for each city. In the case of Málaga, located in the southern coast of Spain, 66% of pixels correspond to land with 51% of those being rural mask and only 40% within the elevation range across the urban mask reflecting the particular orography of this coastal city surrounded by mountain ranges with peaks near 1000 m above the sea level.

As also shown in Fig. 3, except for Madrid — which is situated at 660 m on the central Inner Plateau of the Iberian Peninsula — the collection of urban areas are located at relatively low elevations above the sea level. Elevation values in the urban area of Madrid, Valencia, and Sevilla range approximately over 200 meters. In contrast, Barcelona, Palma de Mallorca, Porto, and Málaga exhibit larger ranges of up to 400 m. It is worth noting that the surrounding rural areas often exhibit higher elevations. This situation is particularly pronounced in cities like Bilbao, Murcia, Palma de Mallorca, and Málaga, where the rural terrain's elevation can easily surpass twice the range of urban altitude values.

2.3. Urban texture data

Data on Population Density (POP), Average Net Building Height (ANBH), Build-up Surface Fraction (BUSURF) and Vegetation Fraction (VEG) have been retrieved from the Global Human Settlement Layer dataset (GHSL) (European Commission and Joint Research Centre, 2023) (all assumed constant in time). After reprojection and interpolation into the base coordinate reference system at 100 × 100 m resolution, Fig. 4 shows ANBH in meters (top left), BUSURF (top right), POP in residents per cell (bottom left), and fraction of vegetation cover (bottom right) for the city of Barcelona.

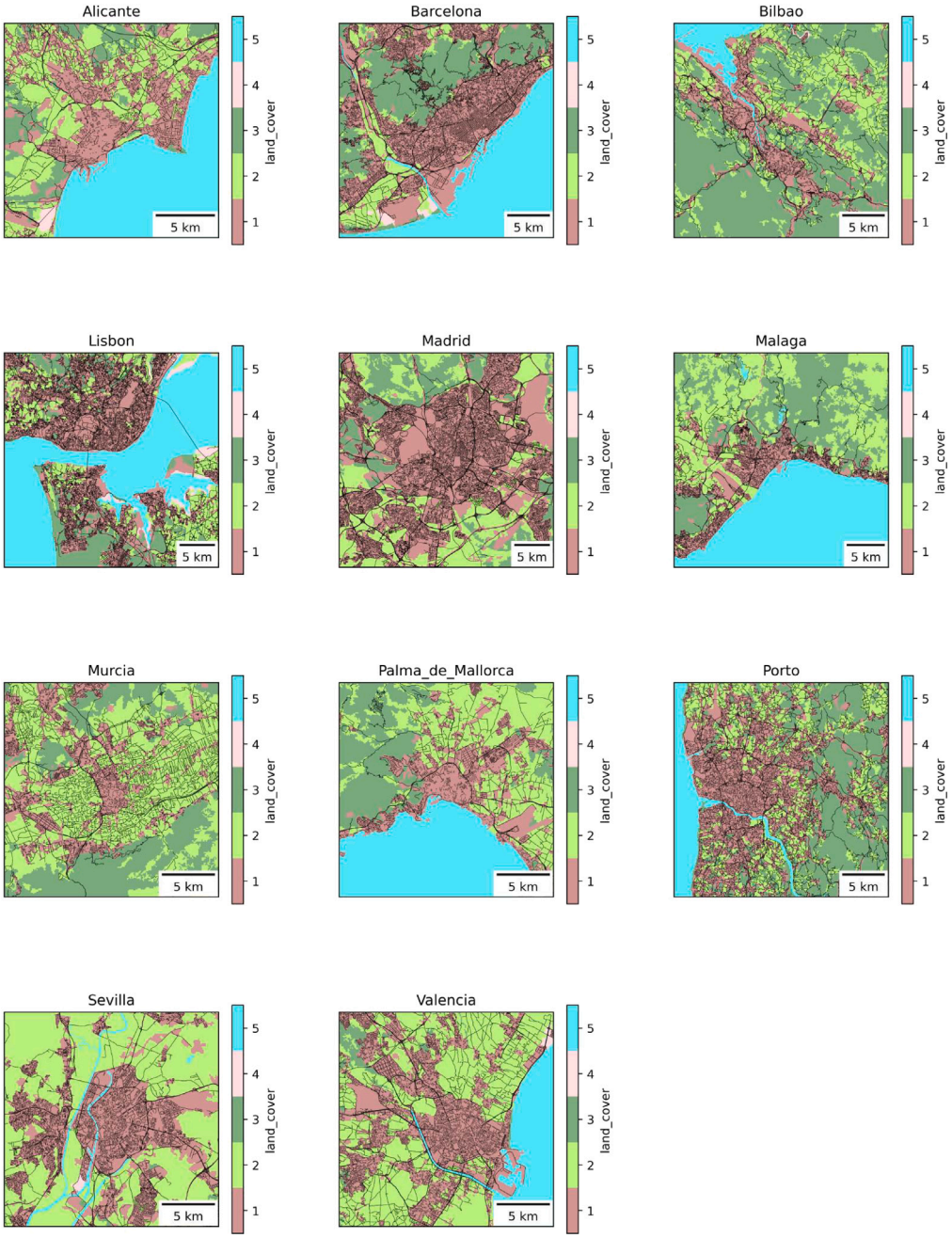


Fig. 2. CORINE land cover/use classification (European Environment Agency, 2018). 1: Artificial surfaces (brown), 2: Agricultural areas (lawngreen), 3: Forest and semi natural areas (darkgreen), 4: Wetlands (pink) and 5: Water bodies (aqua). Major street networks also shown.

2.4. Meteorological data

Hourly data on temperature $T(h,d,i,j)$ and wind speed $w_s(h,d,i,j)$ for 2017, at a 100×100 m resolution, were obtained from the European Environmental Agency repository (European Environment Agency, 2020). These variables were derived from numerical simulations using the UrbClim model (De Ridder et al., 2015) that uses downscaled large-scale weather conditions from ERA5 reanalysis at the agglomeration scale. This tool has demonstrated a high degree of accuracy in predicting microclimate data, with reported correlation coefficients for temperature predictions of $R^2 = 0.95$ for both urban and rural locations in Ghent (Belgium), and $R^2 = 0.90$ for urban locations and 0.94 for rural locations in Bilbao (Spain).

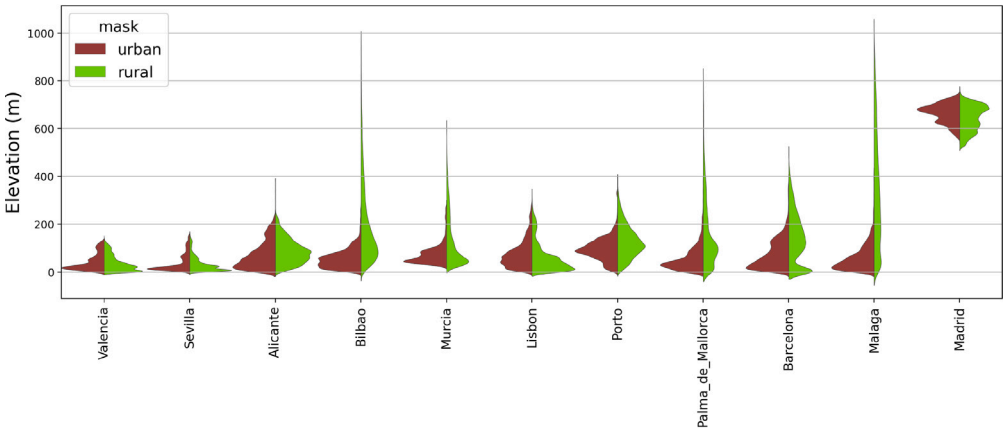


Fig. 3. Elevation range in the rural (green) and urban (red) masked regions for all 11 cities.

Table 2
List of the 11 cities with the percentage of the cells in the rural mask F_R and the percentage of the cells in the rural mask that match at least one urban cell at the same elevation F_E.

City	F_R	F_E
Alicante	37	37
Barcelona	30	30
Bilbao	74	67
Lisbon	25	25
Madrid	37	37
Malaga	51	40
Murcia	78	68
Palma	49	45
Porto	52	52
Sevilla	55	55
Valencia	45	45

In addition, total Precipitation ($\overline{P_T}_d$) and Surface Solar Radiation ($\overline{S}_d$) data were obtained from the CERRA dataset (Verrelle et al., 2022), which provides a regional reanalysis based on ERA5 meteorological forecasts and have been assumed constant across the urban mask extent for each city.

To better characterize the daily thermal dynamics of the rural mask, the set of meteorological predictors is completed with two additional predictors. The first one is the daily average rural temperature at the corresponding urban cell elevation, $\langle \overline{T}_d \rangle_\Delta$. The second feature is the daily rural temperature range also at the corresponding urban cell elevation computed as the difference between the daily extreme values of this quantity, namely, $\langle \overline{\delta T}_d \rangle_\Delta$.

Slight potential climatic differences between cities are accounted using the Köppen-Geiger index (Beck et al., 2023). Finally, yearly seasonality effects are accounted by encoding the month of the year using $\sin(2\pi(j-1)/12)$ and $\cos(2\pi(j-1)/12)$ where $j = 1 \ldots 12$ is the month index.

2.5. Selection of the rural background temperature

To illustrate the importance of selecting an appropriate value of the rural temperature in estimating the Urban Heat Island intensity, the land temperature deviation with respect to its average value ΔT_L for each cell $(i,j) \in L_s$ is computed as:

$$\Delta T_L(h,d,i,j) = T(h,d,i,j) - \langle T \rangle_L(h,d) \tag{5}$$

Using the city of Barcelona (Spain) for illustration purposes, the panels in Fig. 5 show snapshots of ΔT_L at six different hours for an arbitrary day. The arbitrary date used to illustrate Urban Heat Island intensity was randomly selected to avoid periods of extreme average temperatures typically observed over the Iberian Peninsula during summer (July–August) and winter (January–February).

The results reveal distinct warmer and cooler zones within the city, closely corresponding to the urban and rural masks, respectively (see Fig. 2). At 08:00, 12:00, and 16:00, the deviations of land mask temperature from the mean (indicated in red and blue for positive and negative values, respectively) are relatively subtle in both rural and urban areas. However, after sunset, at 20:00, 00:00, and 04:00, these deviations become more pronounced. The urban center exhibits significantly warmer temperatures, with the effect being especially noticeable at 20:00. In contrast, the rural areas show substantial negative deviations, with temperatures dropping up to 6 degrees below the average in the highest elevated locations.

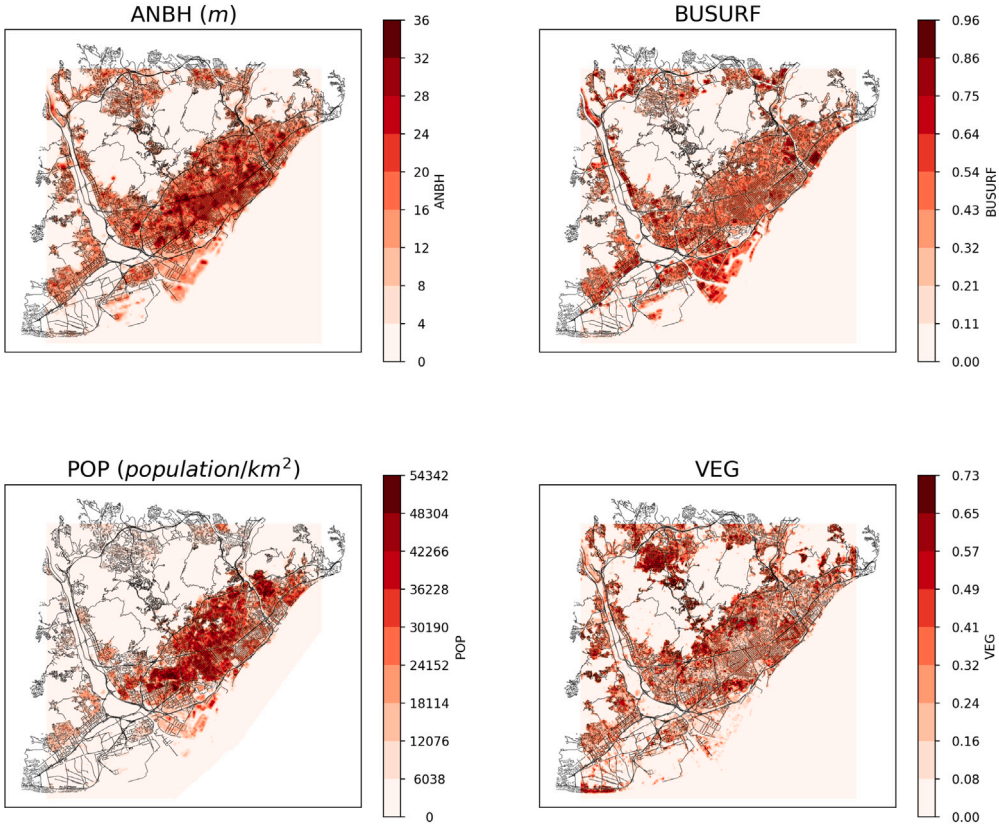


Fig. 4. Raster plots of urban texture predictors for Barcelona at 100 × 100 m resolution. Average Net Building Height in meters (top left), Build-up Fraction (bottom right), Population Density (bottom left), and fraction of vegetation cover (bottom right).

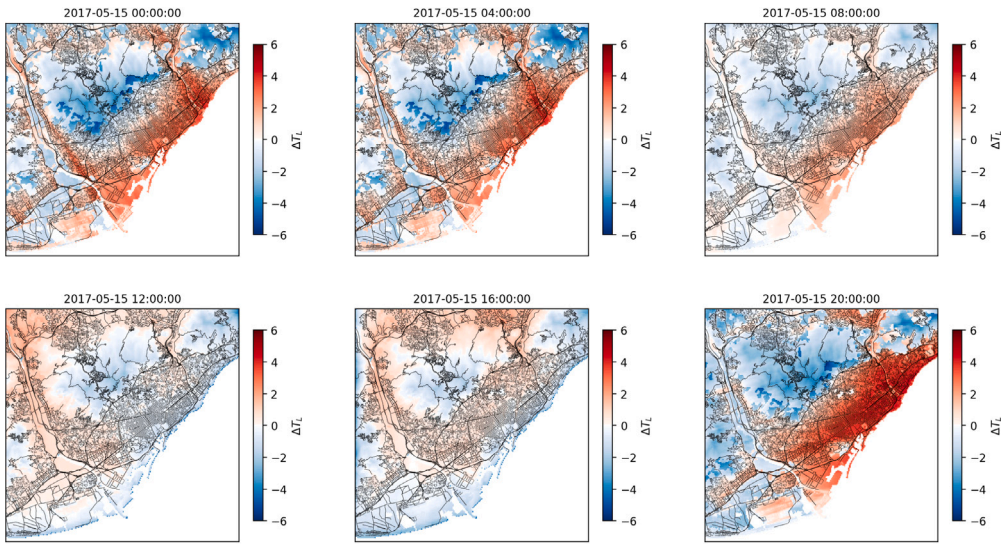


Fig. 5. Land temperature deviation ΔT_L with respect to the average value in Barcelona at 00:00, 04:00, 08:00, 12:00, 16:00 and 20:00 h for an arbitrary day.

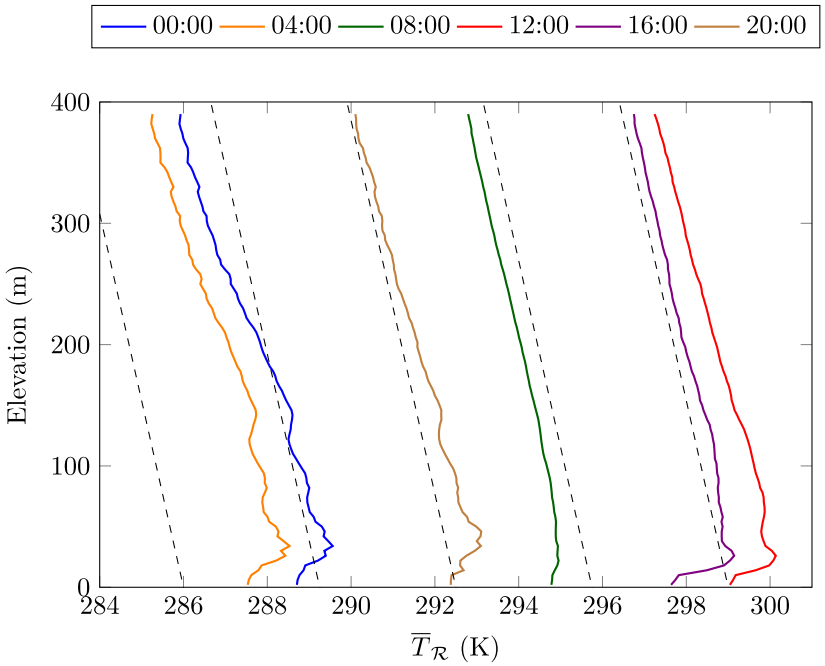


Fig. 6. Vertical profiles of the hourly averaged rural temperature in Barcelona at several times of the day. The black dashed lines correspond to a typical environmental adiabatic lapse rate of $-6.5\ ^{\circ}\mathrm{C\,km^{-1}}$.

Importantly, the disparity of temperature deviations in the rural mask highlights the difficulty to define a *rural* temperature when estimating the UHII, especially when this quantity is based on temperature differences between meteorological stations or, more generally, from point measurements. The temperature in the rural area, that will be subtracted from the urban value to compute the UHII, can strongly depend on the specific chosen location due to differences in the land cover, the elevation above sea level or distinct emissivity values that modify the local surface radiative energy budget (Stathopoulou et al., 2007).

To better illustrate the vertical dependence of the rural temperature and using again Barcelona as an example, Fig. 6 shows the average hourly rural temperature $\overline{T}_h^R(h,z)$ vertical profiles at several times of the day and elevation levels $z = \{E(i,j) | (i,j) \in R_s\}$.

Following a shallow inversion layer, the temperature profiles exhibit an almost linear decrease with increasing land elevation, with a slope closely matching the typical environmental adiabatic lapse rate of $-6.5\ ^{\circ}\mathrm{C\,km^{-1}}$ (International Civil Aviation Organization, 1993) (represented by black dashed lines). Notably, the temperature variation across each of these profiles is approximately 1.5 °C, a magnitude comparable to the Urban Heat Island Intensity itself.

This result underscores the critical importance of accurately defining rural temperature when estimating Urban Heat Island Intensity. To address this issue and standardize the UHII definition, we propose calculating it as the difference between the temperature at each urban location and the average temperature of rural locations at the same elevation. Specifically, for each cell $(i,j) \in U_s$ within the urban mask, we compute the *isohypse rural temperature* $T_{\Delta}(h,d,i,j)$, defined as the average temperature of rural cells $(i',j') \in R_s$ that are at a similar elevation within a threshold ε:

$$T_{\Delta}(h,d,i,j) = \frac{1}{|R_{\Delta}(i,j)|} \sum_{(i',j') \in \Delta(i,j)} T(h,d,i',j') \tag{6}$$

where:

$$\Delta(i,j) = \left\{ (i',j') \in R_s \;\middle|\; |E(i',j') - E(i,j)| \leq \varepsilon \right\} \tag{7}$$

represents the set of rural cells whose elevation differs from the urban cell (i,j) by no more than ε. The term $|\Delta(i,j)| = N_R^{\Delta}$ denotes the number of such rural cells. Note that for each cell in the urban mask $U_s(i,j)$, $\Delta(i,j)$ can be interpreted as the mask that includes the rural cells located at the same elevation.

Now, for each cell in the *urban* mask, $(i,j) \in U_s$, we can define the Urban Heat Island Intensity as:

$$\Delta T(h,d,i,j) = T(h,d,i,j) - T_{\Delta}(h,d,i,j) \tag{8}$$

Using the definition in Eq. (8), the daily maximum Urban Heat Island intensity representing the most extreme value of UHI for any given day, can be computed as:

$$\Delta T_d^*(d,i,j) = \max_{1 \leq h \leq n_h} \Delta T(h,d,i,j) \tag{9}$$

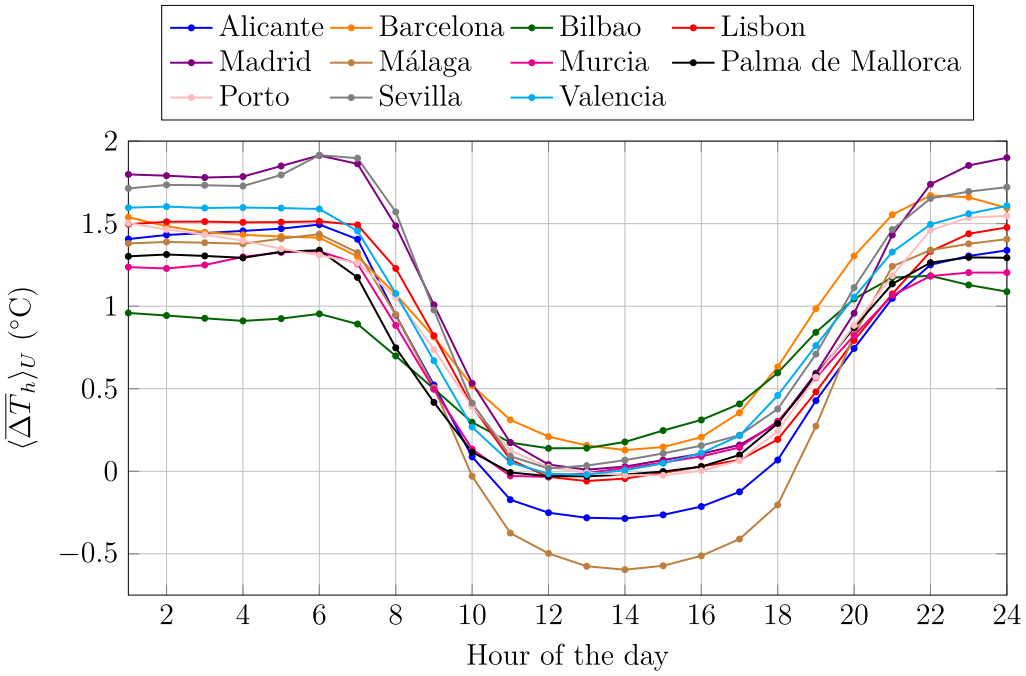


Fig. 7. Hourly evolution of the spatially-averaged hourly maximum UHII over the *urban* mask, $\langle \overline{\Delta T}_h \rangle_U$, for every city.

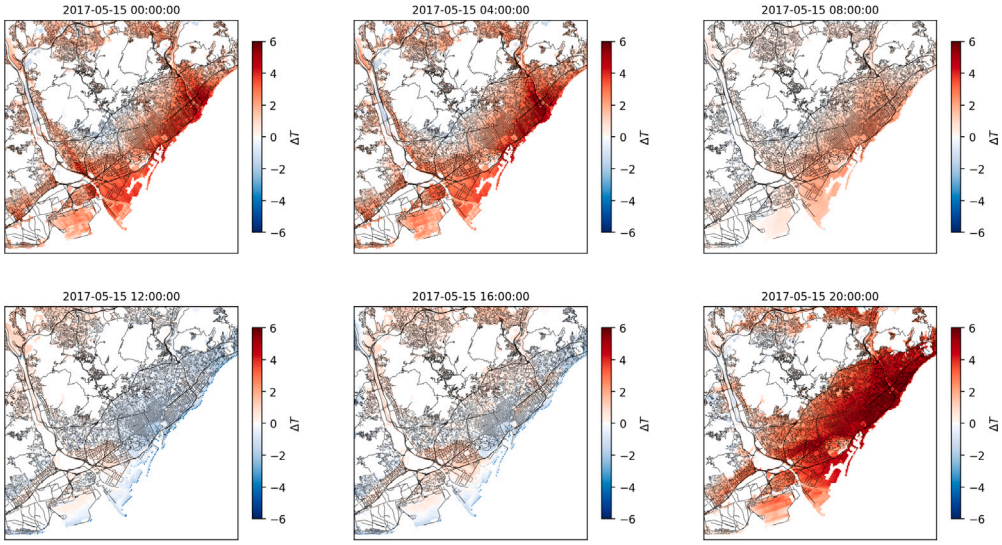


Fig. 8. The Urban Heat Island Intensity $\Delta T(h, d, i, j)$ in the city of Barcelona at six different hours during an arbitrary day.

The quantity ΔT_d^* will serve as the target variable in our model.

Using the definition in Eq. (8), Fig. 7 shows the spatially-averaged hourly maximum UHII over the *urban* mask, $\langle \overline{\Delta T}_h \rangle_U$. During peak sunlight hours, the temperature differences between urban and rural zones are minimal, resulting in UHII values that approach zero across most locations. Notably, Málaga and Alicante exhibit negative UHII values, indicating that, on average, rural temperatures can be slightly higher than those within the urban areas at the same elevation. After sunset, $\langle \overline{\Delta T}_h \rangle_U$ increases, reaching its peak during the late night or early morning hours. A distinctive feature of these UHII profiles is the period from approximately 00:00 to 05:00 h, during which the temperature differences between urban and rural areas remain relatively constant. As solar radiation increases, the rural–urban temperature gradient decreases, leading to a convergence in temperatures between the two zones. On average, the maximum values of $\langle \overline{\Delta T}_h \rangle_U$ remains below 2 °C. However, typical daily variations (not shown here) can range from 0 to 4 °C, with exceptional instances reaching 6 °C or even higher.

Table 3
Nomenclature, description, units, source, spatial scope, temporal dependence, type and class for the target ΔT_d^* and all predictors. Regarding the type, NUM and CAT stand for *Numerical* and *Categorical* respectively. Regarding the predictor class, UT and MET stand for *Urban Texture* and *Meteorology* respectively.

Name	Description	Units	Source	Spatial scope	Constant in time?	Type	Predictor class
ΔT_d^*	Daily maximum UHII	K	UrbClim/CDS	Pixel	N	NUM	Target
ANBH	Normalized building height	m	GHSL	Pixel	Y	NUM	UT
BUSURF	Build-up surface	%	GHSL	Pixel	Y	NUM	UT
POP	Population	inhabitants/km^2	GHSL	Pixel	Y	NUM	UT
VEG	Vegetation cover	%	GHSL	Pixel	Y	NUM	UT
E	Elevation	m	EUDEM	Pixel	Y	NUM	UT
LCLU	Land cover/use	–	CORINE	Pixel	Y	CAT	UT
$\overline{w_{sd}}$	Wind speed	m/s	EEA	Pixel	N	NUM	MET
$\overline{P_{Td}}$	Total precipitation	mm	CERRA	City	N	NUM	MET
$\overline{S_d}$	Surface Solar irradiance	W/m^2	CERRA	City	N	NUM	MET
$\langle\overline{\delta T_d}\rangle_\Delta$	Rural daily temperature span	K	EEA	City	N	NUM	MET
$\langle\overline{T_d}\rangle_\Delta$	Average daily rural temperature	K	EEA	City	N	NUM	MET
month	Month of the year	–	–	–	N	CAT	MET
koppen_index	Köppen-Geiger class	–	Beck et al. (2023)	City	Y	CAT	MET

Using the case of Barcelona, Fig. 8 shows six hourly snapshots of the Urban Heat Island Intensity ΔT during an arbitrary day. Positive values in red denote urban location with higher temperatures than that corresponding to the average rural temperature value and blue otherwise. Rural mask values have been set to 0 (white). These results illustrate the transient nature of the Urban Heat Island phenomenon with rising UHII values after sunset (20:00) with urban temperature reaching values of 4 °C above the rural reference. Positive values persist at 00:00 and 04:00, and continue, albeit at lower levels, beyond sunrise at 08:00. Later, from noon to 16:00, the UHII reverses, showing slightly negative values in a significant portion of the urban area, indicating lower temperatures compared to the rural area.

3. Modeling methodology

3.1. The working dataset: response and predictors

To understand the role of urban texture in the magnitude of the Urban Heat Island phenomenon, here we construct a working dataset containing easily accessible predictors on (i) urban morphology, (ii) meteorology and (iii) supplementary predictors. The first family of features include the Average Building Height, the Build-up Surface Fraction, the Population Density, the fraction of vegetation, and the land cover/use (LCLU).

To summarize, Table 3 shows a list of all the variables in the working dataset and their characteristics including its spatial scope, temporal dependence, data source, format (numerical or categorical) and predictor class.

To ensure reasonable computational cost, only a random selection of 50% of the pixels of each city are included in the analysis. For example, out of the original 56,743 urban pixels in Madrid — the largest urban mask — only 28,367 cells will be used for model training. Therefore, the total number of observations for this city will be 28367 × 365, approximately 10 million observations. The working dataset for Murcia — with the smallest urban mask — will comprise around 1.5 million rows.

The distribution of ΔT_d^* for each city (in red) compared to the overall distribution obtained by combining all urban areas (in blue) is shown in the Supplementary Material Figure S1.

3.2. Learner selection

Predicting the daily maximum value of Urban Heat Intensity using the features listed in Table 3 can be framed as a Supervised Learning problem of the form

$$\Delta T_d^* = f(X_i) \tag{10}$$

where X_i represent the predictors listed in Table 3. For this supervised learning problem, several regression algorithms available in the Python package Scikit-learn were compared, including *support vector machines*, *neural networks*, *random forests*, and *gradient boosting*, along with an alternative *gradient boosting* implementation from the XGBoost library. The results of this benchmark indicate

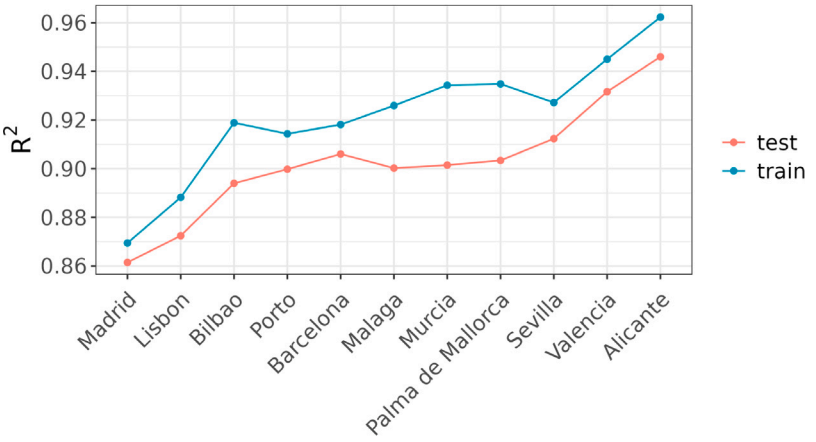


Fig. 9. Train (blue) and test (red) R^2 scores for each city.

that the latter implementation (XGBoost) achieves the best predictive performance, with R^2 scores around 0.9 across all cities considered in this study.

4. Results

4.1. Model performance

A data-driven model for the daily maximum Urban Heat Island Intensity ΔT_d^* has been obtained for each city using a XGBoost learner and a four-fold resampling strategy in which data was randomly divided into four sets with three of them used for training and the fourth for testing. The robustness of each model has been assessed by repeating the procedure four times and collecting average performance scores for the training and testing sets. The coefficient of determination R^2 for both sets, ordered by the test score, is shown in Fig. 9.

The similar R^2 scores observed in both the training set (blue line) and the test set (red line) indicate minimal overfitting of the model, suggesting that the model generalizes well to unseen data. The range of values of R^2 between 0.85 for Madrid and 0.95 for Alicante suggests that the present methodology can accurately predict the maximum daily UHII as defined in Eq. (9).

The model proposed in Eq. (10) assumes that the maximum daily UHII depends on predictors that do not include previous time values of this target variable. By randomly sampling observations from the cell and hourly data pool within the same city, we ignore temporal autocorrelations reducing the implementation complexity and mitigating the risk of temporal overfitting. Furthermore, this method eliminates the need for careful selection of temporal lags, which might be a difficult task given the complexity of the physical phenomena under study. Therefore, the model in Eq. (10) does not require any prior knowledge of the UHII and relies instead solely on basic and readily available information on meteorological conditions and urban texture.

To illustrate the performance of the model, Fig. 10 shows the original and predicted time history of ΔT_d^* at the five selected locations within the urban area of Barcelona characterized by different values of average building height, build-up surface fraction and population density. These results suggest that, despite not incorporating predictors based on previous or neighboring ΔT_d^* values, the proposed model accurately estimates of the daily maximum Urban Heat Island at locations with distinct urban texture characteristics.

The performance of each city-specific model in predicting ΔT_d^* across all other cities is shown in the Supplementary Material Figure S2. The diagonal values represent the test R^2 scores when each model is applied to unseen data from its respective city. For instance, the results for Alicante show that its model can accurately predict the ΔT_d^* within the same urban area, achieving a test R^2 score of 0.95. However, the model's performance significantly declines for other cities, as reflected in the off-diagonal elements. For clarity, only positive test R^2 scores values are displayed.

Comparison of all R^2 values for Alicante suggests that this city model would offer limited performance in predicting the daily maximum UHII in Lisbon, Madrid and Murcia. Notably, the best result for Alicante, a location on the Mediterranean coast, corresponds to a city on the north Atlantic shore of Spain, namely Bilbao. The reciprocal numerical experiment, that is, the prediction of the daily maximum UHII with the Bilbao model (third column in Fig. S2) also results in very poor accuracy with negative values of R^2.

In order to derive an overall model for all 11 combined urban areas, we randomly resampled observations from the entire poll of cities. The resulting working dataset contains as many observations as the largest city (Madrid) with sample sizes from each city contributing equally. This *synthetic city* shares features with all cities, but none of them is dominant.

The test R^2 score for this *synthetic city* city is 0.83. Although, in all cases, the individual city model outperforms this *synthetic city* model, all test R^2 scores are above 0.8 except for Lisbon. Fig. 11 shows the test R^2 score for the overall model used on the *synthetic*

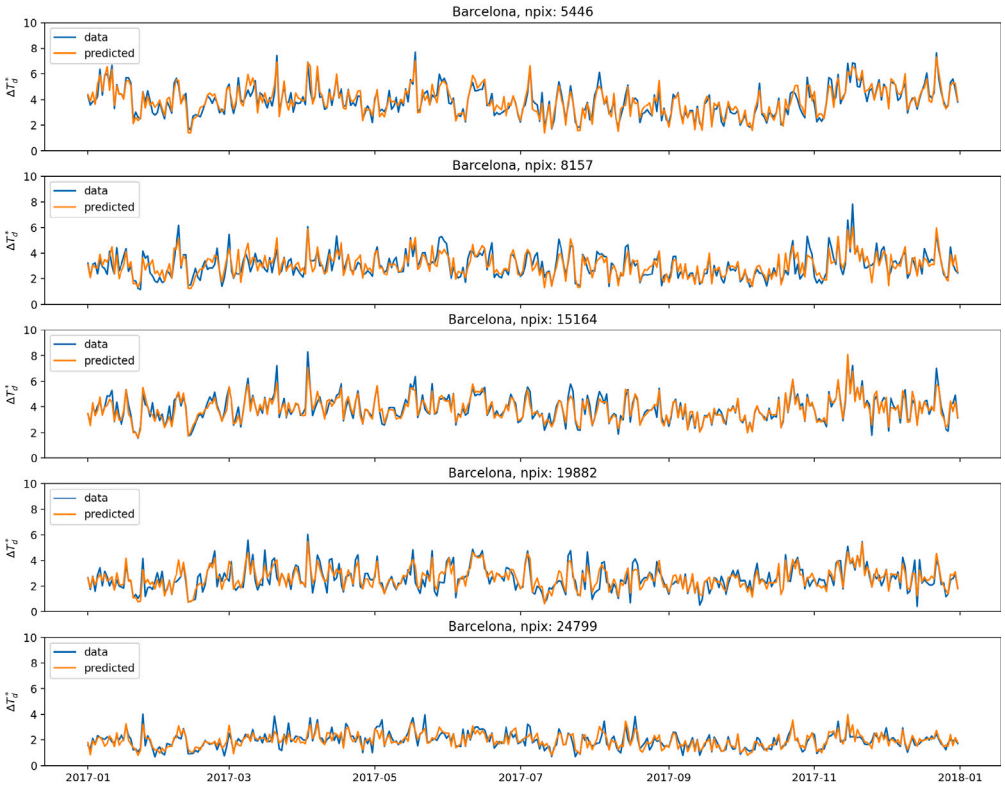


Fig. 10. Original UrbClim simulations (blue) and predicted (orange) time history of ΔT_d^* at the five random location in the city of Barcelona shown in Figure S3.

city (dashed red line), the overall model used in each city (green line), and the individual model for each city (blue line). Cities are ordered according to the test R^2 values obtained using the overall model in each urban area. The resulting ranking almost matches that obtained in Fig. 9, suggesting that cities with better individual model performance also exhibit better predictions when modeled using the *synthetic city* model. The lower predictive capabilities of the *synthetic city* model can be attributed to missing predictors that account for differences between cities, including, for instance, distinct building materials and their thermal radiative properties. Taken together though, these results suggest that the overall model can be used successfully to predict the daily maximum UHII in any city on the Iberian Peninsula.

4.2. Feature importance

The role played by each predictor can be investigated using Permutation Tests (PT) or Partial Dependence Plots (PDP).

In PT, the model is fitted after randomizing, one by one, each individual predictor. The importance of each feature is then quantified as the decay in the R^2 test score with respect to the original non-randomized case. Predictors of limited importance have a modest impact on the R^2 while larger declines in this metric are associated with more important features. The PT results for the overall model are shown in Fig. 12. Note that the train and test R^2 scores for the overall model were 0.83 and 0.82 respectively.

The PT results suggest that the average building height (ANBH) is the most important predictor. Randomization of this feature leads to a reduction of 0.38 points with respect to the original, non-randomized $R^2 = 0.82$ score. The second and third predictors in importance are the daily range of rural temperature ($\langle \overline{\delta T}_d \rangle_\Delta$) and the elevation with reductions in the R^2 scores of 0.32 and 0.31 respectively. The list continues with the population density, the average rural temperature ($\langle \overline{T}_d \rangle_\Delta$), the build-up surface fraction, the solar irradiation, and the wind speed. Finally, randomization of the vegetation fraction and the total precipitation are associated to reduction in R^2 below 0.1.

It should be noted that reductions in R^2 are not additive and combined feature effects can only be assessed by simultaneously randomizing both variables.

In this regards, PDP can be used to gain further insight on the UHII dependence on each predictor. In PDP, the mean impact of a single predictor on the target variables is assessed by letting this vary between its lowest and highest values for the entire set of observations. Thus, to assess the impact of the Average Net Building Height (ANBH), we need to predict the UHII for all observations

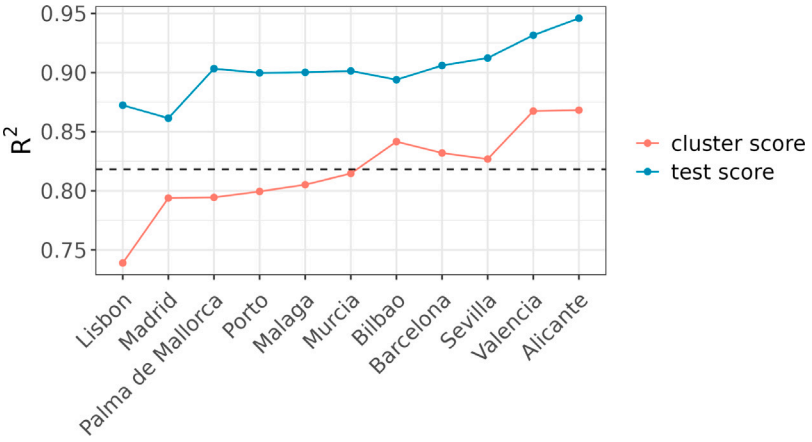


Fig. 11. Test R^2 scores for the overall model used on the *synthetic city* (black dashed line), the overall model used in each city (red line), and the individual model for each city (blue line).

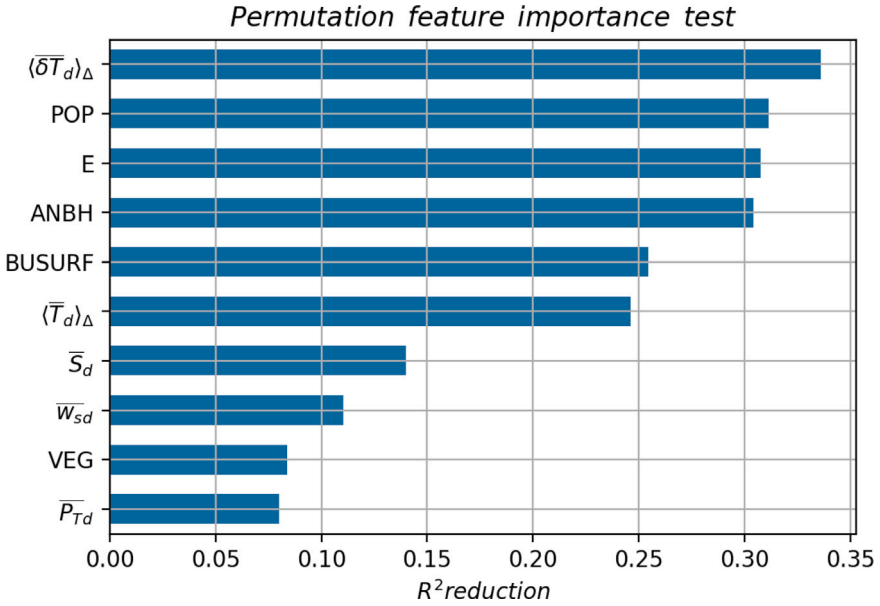


Fig. 12. Permutation Test results showing the R^2 reduction for each predictor.

varying only this feature value across all observations in the training set. Of course, caution must be taken in preventing extrapolation by using modified predictor values outside the range in the training data.

To decide what features should be grouped during the PDP analysis, it is necessary to determine potential correlations between predictors. For example, it seems reasonable to assume that ANBH and POP are to some extent correlated, that is, that part of cities with low average building heights host also lower density populations. Similar trends are also expected between these two features and, for instance, the Build-up fraction.

To elucidate these potential correlations, we conducted a Principal Component Analysis (PCA) for the 10 continuous predictors, that is, excluding the one-hot-encoded features (see Table 3). The PCA results for the standardized data (obtained by subtracting the average and dividing by the standard deviation) are shown in Fig. 13 where the left panel shows the explained variance ratio and the right one its cumulative value.

The results indicate that the first two eigenvectors have almost the same magnitude and both combined account for almost 50% of the total variance with the next eigenvectors weight scoring below 10%. To account for 80% of the total variance it is necessary to use at least the first five eigenvectors.

Fig. 14 shows the PDP importance of each feature in each eigenvector. Interestingly, the first eigenvector is mainly constructed with the meteorological predictors including the solar radiation, the daily rural temperature range and the average daily rural

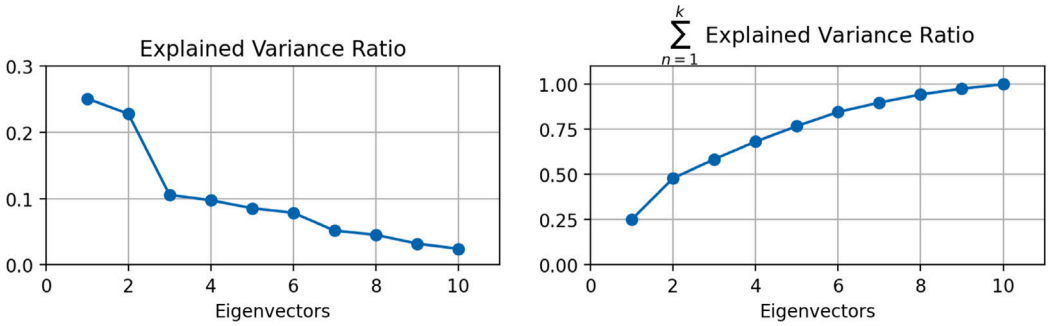


Fig. 13. Principal Component Analysis for the 10 continuous predictors. Left: explained variance ratio. Right: cumulative explained variance ratio.

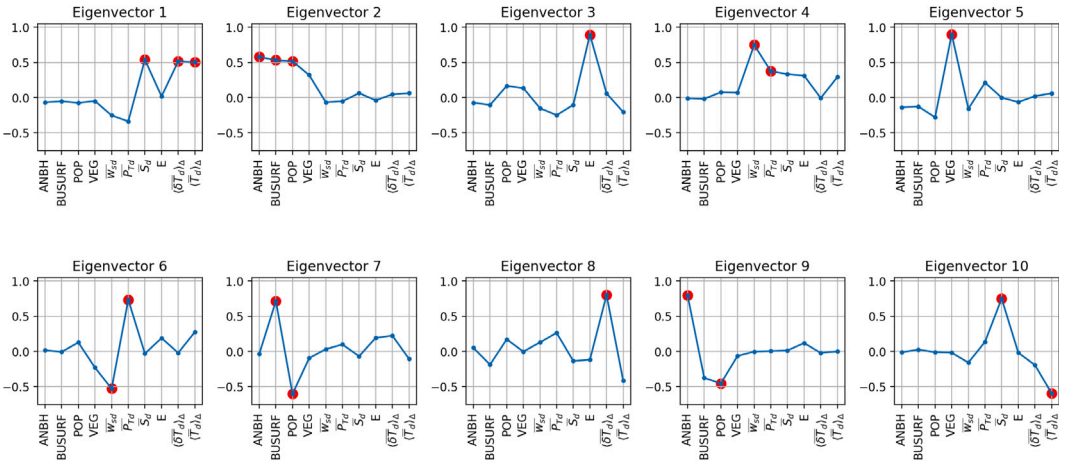


Fig. 14. Partial Dependence Plots results. Red markers indicate the main predictors for each eigenvector.

temperature (red markers), followed by the rest of meteorological variables, namely the wind speed and total precipitation. The rest of the predictors in this eigenvector exhibit near-zero contributions.

The largest contributions in the second PCA eigenvector correspond to urban texture features, namely, the building height, build-up surface fraction, population density and vegetation fraction — with the first three accounting for up to 80% of the total variance. These results suggest that meteorological and urban texture predictors can be independently investigated using PDP when assessing their relevance on the daily maximum Urban Heat Island Intensity. The third and fifth eigenvectors are dominated by a single predictor, the terrain elevation and the vegetation fraction respectively. The main contributors to the fourth eigenvector are the two additional meteorological features: the wind speed and total precipitation. Notably, the features appearing in these first five eigenvectors (red dots) are a partition, this is, they are non-overlapping and mutually exclusive.

The rest of the eigenvectors accounting for less of 20% of the total variance, are mostly paired fine-scale corrections of alternate sign variables. For example, eigenvector 2 points in the direction of data points with same-sign variations of building height, build-up surface fraction and population density while eigenvectors 7 and 9 account for alternate sign variations of build-up fraction and population density and building height and population density respectively.

The PDP results for the correlated predictors based on the PCA analysis are shown in Fig. 15. The top features in the first eigenvector, namely solar radiation, the daily rural temperature range and the average daily rural temperature, are associated with a decrease of 0.68 in the R^2 score with respect to the original value of 0.82.

4.3. Model results: The role of urban texture in urban heat island intensity

The simultaneously randomization of the top predictors in the second eigenvector (the building height, build-up surface fraction, population density and vegetation fraction) lead to a reduction of 0.62 in the R^2 score. These two groups of predictors for the first and second eigenvectors account for the largest impact on the prediction of the daily maximum UHII. PDP analysis for the elevation predictor, the single most important in the third eigenvector, resulted in a modest reduction in R^2 of 0.31. Regarding the fourth and fifth eigenvectors, we found that the wind speed and total precipitation and the vegetation fraction resulted in a decrease of 0.18 and 0.08 in R^2 respectively.

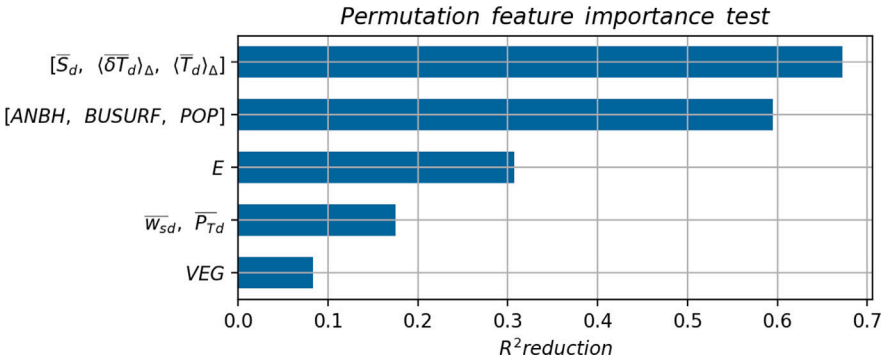


Fig. 15. Partial Dependence Plots showing the R^2 reduction for top PCA eigenvectors.

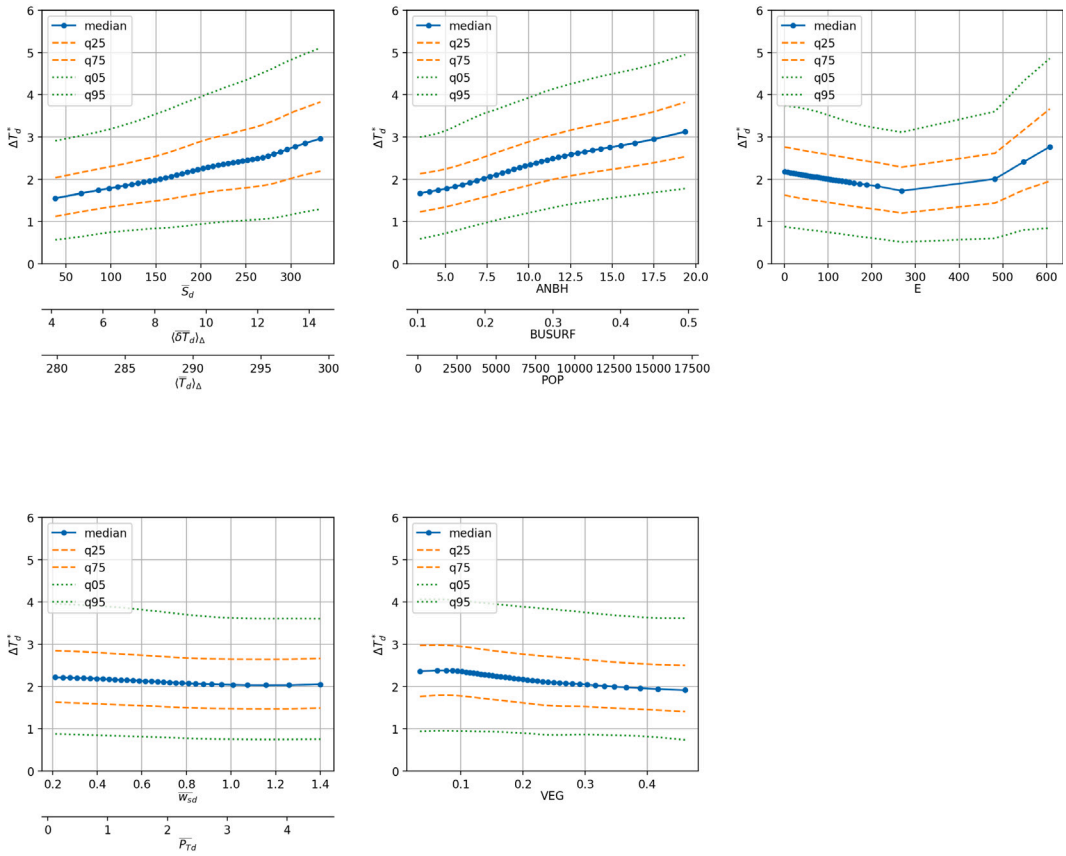


Fig. 16. PDP results for randomization of each eigenvector. Blue, orange range and green range show the median values of the predicted maximum daily UHII, the 25% and 75% quartiles (the interquartile range) and the 5% and 95% quantiles respectively.

The PCA results were also used to randomize each eigenvector predictor data in the PDP analysis. The results of varying simultaneously all the predictors are presented in Fig. 16 that shows the median values of the predicted ΔT_d^*, the 25% and 75% quartiles (the interquartile range) and the 5% and 95% quantiles. The x-axis (one per feature) has been varied over the span of the top predictors in each eigenvector (see Fig. 13). The results for the first eigenvector suggest that an increase in solar radiation, daily rural temperature range and average daily rural temperature leads to an increase in the maximum daily UHII which ranges between approximately 1.5 °C to 3.0 °C across the considered predictor span, an overall increase of 1.5 °C.

In particular, the extreme values in the quartile 95% vary from 3 to 5 °C in the predictor range. The average slope of the median of the maximum daily UHII with respect to the daily rural temperature range is 0.12. Therefore, for each 1 °C increase in the daily range of rural temperature, the maximum daily UHII will increase by 0.12 °C on average. Regarding the 95% percentile curve, the

Table 4
Average rate of change in the daily maximum urban heat island intensity caused by each urban texture predictor.

Predictor	Mean UHII slope in °C	Units
ANBH	0.10	per 1 m of building height
BUSURF	0.41	per 0.1 points in build-up fraction
POP	0.09	per 1000 people/m^2
VEG	−0.13	per 0.1 points in vegetation fraction

average slope is 0.20. This suggests that an increase in the daily range of rural temperature will lead to an intensified urban heat island effect but also to increased likelihood of extreme episodes.

The role of urban texture is shown in the second panel. The PDP results suggest that increased average building height, build-up fraction and population density (the top predictors in the second eigenvector) lead to intensified UHI. Over these features value range, the median maximum daily UHII varies from 1.75 to 3.25 °C while extreme values ranges from 3 to 5 °C. The median value will increase by 0.1 °C for each meter of average building height. In the present study dataset, the lowest values of average building height are around 3 m, a value that corresponds to a single-story houses common in disseminate areas where the build-up fraction is around 0.1. The largest values of average building height, at approximately 20 m, are typical of densely urbanized area where the buildings occupy almost 50% of the surface with population densities around 20000 inhabitants/km^2.

As it occurred for the previous panels, increased values of these predictors lead to larger daily maximum UHII but, notably, to larger extreme values as suggested by the 95% percentile curve. Thus, while the largest values of average building height, build-up fraction and population density are predicted to result in a temperature excess of around 3 °C, the extreme values can reach up to 5 °C. Since we are varying predictor values along the first and second eigenvectors, which are orthogonal by virtue of the PCA, both effects are independent and could therefore lead to combined increases in the UHII.

The highly non-linear behavior in the third panel (and eigenvector), has to be interpreted in the light of Fig. 3 results. While elevations in most cities range from sea level to around 400 m, Madrid elevations range between 500 and 750 m. Thus, while elevation increases in near-sea-level urban areas (0–300 m) generally produce a cooling effect, reducing ΔT_d^*, this trend reverses in Madrid, where higher elevations are associated with an intensified Urban Heat Island effect.

The fourth panel shows the results for the wind speed and total precipitation, the top predictors in the fourth eigenvector. For low to moderate values of these predictors, increases in wind speed and total precipitation exhibit weak reductions in the median of ΔT_d^*. However, at larger values of these predictors, their impact of the daily maximum UHII seems to plateau.

Finally, the vegetation fraction in the fifth panel suggests that increased urban greenness contributes to cool the urban areas. These results suggest that, on average, increasing the vegetation fraction from 0 to 0.5 may reduce the UHII by approximately 0.6 °C, namely, by 0.13 °C per 0.1 points in vegetation fraction or by 1.3 °C between a shaded (vegetation fraction equals to 1) and non-shaded (equal to 0) area. This value of 1.3 °C is in good agreement with that reported in the review on the effectiveness of 'greening' of urban areas in reducing human exposure to the Urban Heat Island effect by Knight et al. (2021) who found that park or garden average cooling effect was 0.8 °C.

To summarize, Table 4 shows the average rate of change in the daily maximum Urban Heat Island intensity caused by each urban texture predictor, expressed in degrees Celsius per unit of the corresponding feature.

Further insight into the role of predictors in the urban heat island effect can be gained from the results in Fig. 17, which shows the predictor distribution for UHII values above and below its 95th percentile. Here we include the variables associated to the first (meteorological) and to the second (urban texture) eigenvalues along with vegetation cover which is the main variable of the fifth eigenvalue.

The results suggest that extreme ΔT_d^* values are linked to high daily rural temperature differences and solar radiation. However, there is no clear association between average rural temperature and ΔT_d^* values exceeding the 95th quantile. Regarding the urban morphology features, extreme UHII values are linked to large values of building height, population density, and build-up fraction. These results are consistent with those presented in Fig. 16 that shows positive values of slope for all three features. The distribution of the vegetation fraction for UHII values above the 95th quantile seems to be only slightly shifted to the left of that for values below this threshold. Again, this would be consistent with the negative slope for this predictor in Fig. 16. Although modest, as previously concluded through analyses of the PT and PDP, these results suggest that low values of vegetation fraction are associated with an increased likelihood of extreme UHII values. For values above 0.3, the mitigation capacity seems to decay to a neutral effect.

5. Conclusions

This study set out to analyze the role of urban texture predictors on the Urban Heat Island intensity (UHII) using data from 11 cities in the Iberian Peninsula. These features include, among others, build-up fraction, vegetation cover percentage and population density. Using high-spatial resolution raster data, the dependence between the UHII and meteorological and urban texture predictor has been elucidated using classical Machine Learning (ML) methodologies.

Firstly, we addressed the need to properly define the background or rural temperature against which the UHII is estimated by computing the temperature difference with respect urban locations.

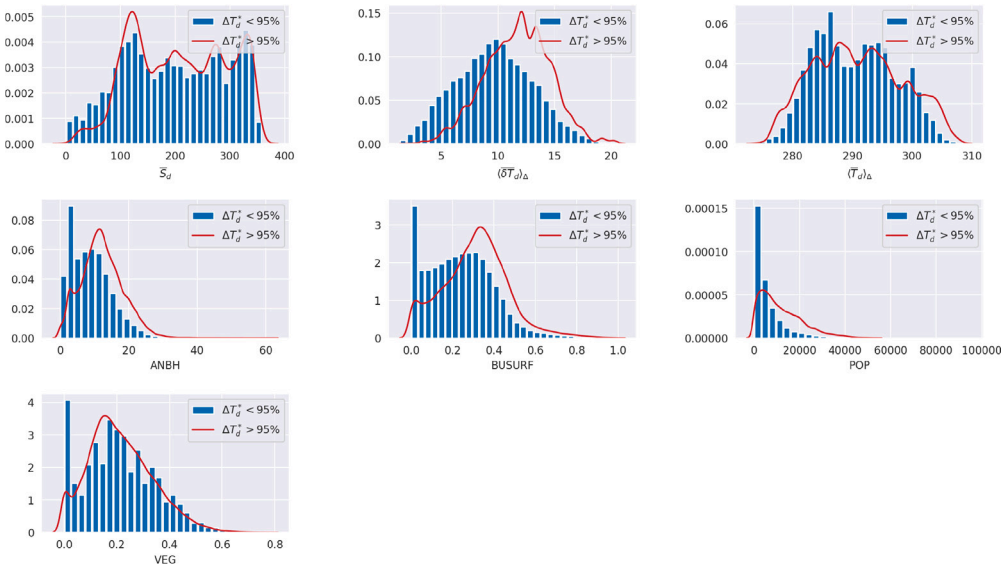


Fig. 17. Predictor distribution for UHII values above and below its 95th percentile.

Equipped with a new definition of the background temperature that takes into account the elevation effect, the city specifics impact on the model performance has been measured by building individual models and comparing the performance results with a hybrid urban area made of randomly combined cell data.

Secondly, we used Principal Components Analysis (PCA) to investigate how predictors group and affect the increased urban temperatures. Results suggest that two groups of predictors impact the UHII the most. On one hand, wind speed and daily rural temperature range (two meteorological features) are responsible for the largest share of the total UHII variability. The second group of predictors in relevance is comprised by the build-fraction, population and building height (urban texture features).

The average rate of change on the daily maximum UHII is 0.09 °C per meter of building height, 0.34 °C per 0.1 point increase in build-up fraction, 0.08 °C per 1000 people/m^2 and −0.11 °C per 0.1 point increase in vegetation fraction.

These urban texture weight coefficients offer a quantitative basis for estimating the contribution of specific urbanization patterns to the intensification of the urban heat island intensity. When combined with projections of increased built-up fraction, construction density, or population densification, these coefficients can be used to model the expected thermal response to urban growth. Similarly, they allow for the spatial assessment of the mitigating effects of urban vegetation on UHII across the urban extent.

It is important to note that the dataset used in this study corresponds to the year 2017. While climatic conditions may have evolved since then — due to factors such as global warming — the urban texture predictors employed (e.g., average building height, built-up area fraction, vegetation cover, and population density) are not expected to have undergone substantial changes over this period. As such, the findings of the present study remain valid. Nonetheless, incorporating more recent data — particularly updated microclimatic variables such as those provided by the UrbClim model — could further enhance the accuracy and relevance of the predictive models.

Future work is aimed at extending the analysis to a larger collection of cities that include all urban areas in Hooyberghs et al. (2019) along with an extended set of predictors that included climatic effects, construction codes and practices and an enriched collection of urban texture features.

CRediT authorship contribution statement

Josep A. Ferré: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Anton Vernet:** Writing – review & editing, Validation, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Alexandre Fabregat:** Writing – review & editing, Writing – original draft, Validation, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by project TED2021-129348B-I00, funded by MCIN/AEI /10.13039/501100011033 and by the European Union NextGenerationEU/PRTR.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.uclim.2025.102527.

Data availability

Data will be made available on request.

References

Afshari, A., 2023. Optimization of urban design/retrofit scenarios using a computationally light standalone urban energy/climate model (SUECM) forced by ERA5 data. Energy Build. 287, 112991.

Aghamolaei, R., Fallahpour, M., Mirzaei, P.A., 2021. Tempo-spatial thermal comfort analysis of urban heat island with coupling of CFD and building energy simulation. Energy Build. 251, 111317.

Antoniou, N., Montazeri, H., Neophytou, M., Blocken, B., 2019. CFD simulation of urban microclimate: Validation using high-resolution field measurements. Sci. Total Environ. 695, 133743.

Badugu, A., Arunab, K.S., Mathew, A., Sarwesh, P., 2023. Spatial and temporal analysis of urban heat island effect over Tiruchirappalli city using geospatial techniques. Geod. Geodyn. 14, 275–291.

Beck, H.E., McVicar, T.R., Vergopolan, N., Berg, A., Lutsko, N.J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A.I.J.M., Miralles, D.G., 2023. High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. Sci. Data 10 (1), 724.

Bruse, M., Fleer, H., 1998. Simulating surface-plant-air interactions inside urban environments with a three dimensional numerical model. Environ. Model. Softw. 13, 373–384.

Chakraborty, T., Lee, X., 2019. A simplified urban-extent algorithm to characterize surface urban heat islands on a global scale and examine vegetation control on their spatiotemporal variability. Int. J. Appl. Earth Obs. Geoinf. 74, 269–280.

Choi, H.M., Lee, W., Roye, D., Heo, S., Urban, A., Entezari, A., Vicedo-Cabrera, A.M., Zanobetti, a., Gasparrini, A., Analitis, A., Tobias, A., Armstrong, b., Forsberg, B., Íñiguez, C., Åström, C., Indermitte, E., Lavigne, E., Mayvaneh, F., Acquaotta, F., Sera, F., Orru, H., Kim, H., Kyselý, J., Madueira, J., Schwartz, J., Jaakkola, J., Katsouyanni, K., Hurtado Diaz, M., Ragettli, M.S., Pascal, M., Ryti, N., Scovronick, N., Osorio, S., Tong, S., Seposo, X., Leon Guo, Y., Guo, Y., Bell, M.L., 2022. Effect modification of greenness on the association between heat and mortality: A multi-city multi-country study. EBioMedicine 87, 104396.

De Ridder, K., Lauwaet, D., Maiheu, B., 2015. UrbClim - A fast urban boundary layer climate model. Urban Clim. 12, 21–48.

Debbage, Neil, Marshall Shepherd, J., 2015. The urban heat island effect and city contiguity. Comput. Environ. Urban Syst. 54, 181–194.

Diem, Phan Kieu, Nguyen, Can Trong, Diem, Nguyen Kieu, Diep, Nguyen Thi Hong, Thao, Pham Thi Bich, Hong, Tran Gia, Phan, Thanh Noi, 2024. Remote sensing for urban heat island research: Progress, current issues, and perspectives. Remote. Sens. Appl.: Soc. Environ. 33, 101081.

Dougherty, T.R., Jain, R.K., 2023. TOM.D: Taking advantage of microclimate data for urban building energy modeling. Adv. Appl. Energy 10, 100138.

Doulos, L., Santamouris, M., Livada, I., 2004. Passive cooling of outdoor urban spaces. The role of materials. Sol. Energy 77 (2), 231–249.

European Commission and Joint Research Centre, 2023. GHSL Data Package 2023. Publications Office of the European Union.

European Environment Agency, 2018. CORINE Land Cover 2018.

European Environment Agency, 2020. Urban heat island (UHI) intensity modelling, Jan. 2020.

Evola, G., Costanzo, V., Magrì, C., Margani, G., Marletta, L., Naboni, E., 2020. A novel comprehensive workflow for modelling outdoor thermal comfort and energy demand in urban canyons: Results and critical issues. Energy Build. 216, 109946.

Gago, E.J., Roldan, J., Pacheco-Torres, R., Ordóñez, R., 2013. The city and urban heat islands: A review of strategies to mitigate adverse effects. Renew. Sustain. Energy Rev. 25, 749–758.

Grimm, N.B., Faeth, S.H., Golubiewski, N.E., Redman, C.L., Wu, J., Bai, X., Briggs, J.M., 2008. Global change and the ecology of cities. Science 319 (5864), 756–760.

Güller, C., Toy, S., 2024. The impacts of urban morphology on urban heat islands in housing areas: The case of Erzurum, Turkey. Sustainability 16 (2), 791.

Hengl, T., Parente, L.L., Krizan, J., Bonannella, C., 2020. Continental Europe Digital Terrain Model at 30 m resolution based on GEDI, ICESat-2, AW3D, GLO-30, EUDEM, MERIT DEM and background layers (v0.3).

Hooyberghs, H., Berckmans, J., Lauwaet, D., Lefebre, F., De Ridder, K., 2019. Climate variables for cities in Europe from 2008 to 2017. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). http://dx.doi.org/10.24381/cds.c6459d3a.

Hou, H., Su, H., Yao, C., Wang, Z., 2023. Spatiotemporal patterns of the impact of surface roughness and morphology on urban heat island. Sustain. Cities Soc. 92, 104513.

Huang, Xin, Wang, Ying, 2019. Investigating the effects of 3D urban morphology on the surface urban heat island effect in urban functional zones by using high-resolution remote sensing data: A case study of Wuhan, Central China. ISPRS J. Photogramm. Remote Sens. 152, 119–131.

Huo, H., Chen, F., Geng, X., Tao, J., Liu, Z., Zhang, W., Leng, P., 2021. Simulation of the urban space thermal environment based on computational fluid dynamics: A comprehensive review. Sensors 21 (20), 6898.

Husni, Emir, Prayoga, Galang Adira, Tamba, Josua Dion, Retnowati, Yulia, Fauzandi, Fachri Imam, Yusuf, Rahadian, Yahya, Bernardo Nugroho, 2022. Microclimate investigation of vehicular traffic on the urban heat island through IoT-Based device. Heliyon 8 (11), e11739.

International Civil Aviation Organization, 1993. Manual of the ICAO Standard Atmosphere: Extended To 80 Kilometres (262, 500 Ft). Doc 7488, third ed. International Civil Aviation Organization, Montreal, ISBN: 9789291940642.

Iungman, T., Cirach, M., Marando, F., Pereira Barboza, E., Khomenko, S., Masselot, P., Quijal-Zamorano, M., Mueller, N., Gasparrini, A., Urquiza, J., Heris, M., Thondoo, M., Nieuwenhuijsen, M., 2023. Cooling cities through urban green infrastructure: a health impact assessment of European cities. Lancet 401, 577–589.

Iungman, T., Khomenko, S., Barboza, E.P., Cirach, M., Gonçalves, K., Petrone, P., Erbertseder, T., Taubenböck, H., Chakraborty, T., Nieuwenhuijsen, M., 2024. The impact of urban configuration types on urban heat islands, air pollution, CO_2 emissions, and mortality in Europe: a data science approach. Lancet Planet. Heal. 8, e489–e505.

John Arnfield, A., 2003. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. Int. J. Climatol. 23 (1), 1–26.

Johnson, G.T., Oke, T.R., Lyons, T.J., Steyn, D.G., Watson, I.D., Voogt, J.A., 1991. Simulation of surface urban heat islands under 'IDEAL' conditions at night part 1: Theory and tests against field data. Bound.-Layer Meteorol. 56 (3), 275–294.
Kabisch, N., Remahne, F., Ilsemann, C., Fricke, L., 2023. The urban heat island under extreme heat conditions: a case study of Hannover, Germany. Sci. Rep. 13, 23017.
Kim, Y., Yeo, H., Kim, Y., 2022. Estimating urban spatial temperatures considering anthropogenic heat release factors focusing on the mobility characteristics. Sustain. Cities Soc. 85, 104073.
Knight, T., Price, S., Bowler, D., Hookway, A., King, S., Konno, K., Richter, R.L., 2021. How effective is 'greening' of urban areas in reducing human exposure to ground-level ozone concentrations, UV exposure and the 'urban heat island effect'? An updated systematic review. Environ. Evid. 10 (1), 12.
Kubilay, A., Allegrini, J., Strebel, D., Zhao, Y., Derome, D., Carmeliet, J., 2020. Advancement in urban climate modelling at local scale: Urban heat island mitigation and building cooling demand. Atmosphere 11 (12).
Lai, Jiameng, Zhan, Wenfeng, Quan, Jinling, Liu, Zihan, Li, Long, Huang, Fan, Hong, Falu, Liao, Weilin, 2021a. Reconciling debates on the controls on surface urban heat island intensity: Effects of scale and sampling. Geophys. Res. Lett. 48 (19), e2021GL094485.
Lai, J., Zhan, W., Voogt, J., Quan, J., Huang, F., Zhou, J., Bechtel, B., Hu, L., Wang, K., Cao, C., Lee, X., 2021b. Meteorological controls on daily variations of nighttime surface urban heat islands. Remote Sens. Environ. 253, 112198.
Lindberg, F., Holmer, B., Thorsson, S., 2008. SOLWEIG 1.0 – Modelling spatial variations of 3D radiant fluxes and mean radiant temperature in complex urban settings. Int. J. Biometeorol. 52 (7), 697–713.
Liu, B., Guo, X., Jiang, J., 2023. How urban morphology relates to the urban heat island effect: A multi-indicator study. Sustainability 15 (14), 10787.
Lorenzo, N., Díaz-Poso, A., Royé, D., 2021. Heatwave intensity on the Iberian Peninsula: Future climate projections. Atmos. Res. 258, 105655.
Manoli, G., Fatichi, S., Schläpfer, M., Yu, K., Crowther, T.W., Meili, N., Burlando, P., Katul, G.G., Bou-Zeid, E., 2019. Magnitude of urban heat islands largely explained by climate and population. Nature 573, 55–60.
Marando, F., Salvatori, E., Sebastiani, A., Fusaro, L., Manes, F., 2019. Regulating ecosystem services and green infrastructure: assessment of urban heat island effect mitigation in the municipality of Rome, Italy. Ecol. Model. 392, 92–102.
Martin-Vide, J., Sarricolea, P., Moreno-García, M.C., 2015. On the definition of urban heat island intensity: The rural reference. Front. Earth Sci. 3.
Matzarakis, A., Rutz, F., Mayer, H., 2007. Modelling radiation fluxes in simple and complex environments - Application of the RayMan model. Int. J. Biometeorol. 51, 323–334.
Mohammad, P., Goswami, A., Chauhan, S., Nayak, S., 2022. Machine learning algorithm based prediction of land use land cover and land surface temperature changes to characterize the surface urban heat island phenomena over Ahmedabad city, India. Urban Clim. 42, 101116.
Naserikia, M., Hart, M.A., Nazarian, N., Bechtel, B., 2022. Background climate modulates the impact of land cover on urban surface temperature. Sci. Rep. 12, 15433.
Nuruzzaman, M., 2015. Urban heat island: Causes, effects and mitigation measures - A review. Int. J. Environ. Monit. Anal. 3, 67–73.
Oh, J.W., Ngarambe, J., Duhirwe, P.N., Yun, G.Y., Santamouris, M., 2020. Using deep-learning to forecast the magnitude and characteristics of urban heat island in Seoul Korea. Sci. Rep. 10 (1), 3559.
Oke, T.R., 1973. City size and the urban heat island. Atmospheric Environ. (1967) 7 (8), 769–779.
Oke, T.R., 1982. The energetic basis of the urban heat island. Q. J. R. Meteorol. Soc. 108 (455), 1–24.
Rodrigues, M., 2023. Projections of cause-specific mortality and demographic changes under climate change in the Lisbon metropolitan area: a modelling framework. Atmosphere 14 (5), 775.
Rodrigues, M., Santana, P., Rocha, A., 2020. Statistical modelling of temperature-attributable deaths in Portuguese metropolitan areas under climate change: who is at risk? Atmosphere 11 (2), 159.
Rodrigues, M., Santana, P., Rocha, A., 2021. Modelling of temperature-attributable mortality among the elderly in Lisbon metropolitan area, Portugal: a contribution to local strategy for effective prevention plans. J. Urban Heal. 98, 516–531.
Ruosteenoja, K., Jylhä, K., 2023. Average and extreme heatwaves in Europe at 0.5-2.0°C global warming levels in CMIP6 model simulations. Clim. Dyn. 61, 4259–4281.
Schrijvers, P.J.C., Jonker, H.J.J., Kenjereš, S., de Roode, S.R., 2015. Breakdown of the night time urban heat island energy budget. Build. Environ. 83, 50–64, Special Issue: Climate adaptation in cities.
Sobstyl, J.M., Emig, T., Abdolhosseini Qomi, M.J., Ulm, F.-J., Pellenq, R.J.-M., 2018. Role of city texture in urban heat islands at nighttime. Phys. Rev. Lett. 120, 108701.
Stathopoulou, M., Cartalis, C., Petrakis, M., 2007. Integrating corine land cover data and landsat TM for surface emissivit y definition: application to the urban area of Athens, Greece. Int. J. Remote Sens. 28 (15), 3291–3304.
Stewart, I., Oke, T., 2009. Classifying urban climate field sites by local climate zones: The case of Nagano, Japan. In: the Seventh International Conference on Urban Climate, 29 June - 3 July 2009, Yokohama, Japan.
Tehrani, A.A., Veisi, O., Kia, K., Delavar, Y., Bahrami, S., Sobhaninia, S., Mehan, A., 2024. Predicting urban Heat Island in European cities: A comparative study of GRU, DNN, and ANN models using urban morphological variables. Urban Clim. 56, 102061.
Touchaei, A.G., Wang, Y., 2015. Characterizing urban heat island in Montreal (Canada)—Effect of urban morphology. Sustain. Cities Soc. 19, 395–402.
Verrelle, A., Glinton, M., Bazile, E., Le Moigne, P., Randriamampianina, R., Ridal, M., Berggren, L., Undén, P., Schimanke, S., Mladek, R., Soci, C., 2022. CERRA-Land sub-daily regional reanalysis data for Europe from 1984 to present.
Voogt, J.A., Oke, T.R., 2003. Thermal remote sensing of urban climates. Remote Sens. Environ. 86 (3), 370–384.
Wang, X., Li, Y., 2016. Predicting urban heat island circulation using CFD. Build. Environ. 99, 82–97.
Wang, T., Shi, J., Letu, H., Ma, Y., Li, X., Zheng, Y., 2019. Detection and removal of clouds and associated shadows in satellite imagery based on simulated radiance fields. J. Geophys. Res.: Atmospheres 124, 7207–7225.
Wang, C., Zhang, H., Ma, Z., Yang, H., Jia, W., 2024. Urban morphology influencing the urban heat island in the high-density city of Xi'an based on the local climate zone. Sustainability 16 (10), 3946.
Yao, R., Wang, L., Huang, X., Niu, Y., Chen, Y., Niu, Z., 2018. The influence of different data and method on estimating the surface urban heat island intensity. Ecol. Indic. 89, 45–55.
Yin, Y., He, L., Wennberg, P.O., Frankenberg, C., 2023. Unequal exposure to heatwaves in Los Angeles: Impact of uneven green spaces. Sci. Adv. 9 (17), eade8501.
Yun, G.Y., Ngarambe, J., Duhirwe, P.N., Ulpiani, G., Paolini, R., Haddad, S., Vasilakopoulou, K., Santamouris, M., 2020. Predicting the magnitude and the characteristics of the urban heat island in coastal cities in the proximity of desert landforms. The case of Sydney. Sci. Total Environ. 709, 136068.
Zhou, Xuefan, Chen, Hong, 2018. Impact of urbanization-related land use land cover changes and urban morphology changes on the urban heat island phenomenon. Sci. Total Environ. 635, 1467–1476.
Zhou, B., Rybski, D., Kropp, J.P., 2013. On the statistics of urban heat island intensity. Geophys. Res. Lett. 40 (20), 5486–5491.
Zhou, B., Rybski, D., Kropp, J.P., 2017a. The role of city size and urban form in the surface urban heat island. Sci. Rep. 7, 4791.
Zhou, Yan, Zhuang, Zhi, Yang, Feng, Yu, Yao, Xie, Xiaohong, 2017b. Urban morphology on heat island and building energy consumption. Procedia Eng. 205, 2401–2406.