Spatially Constrained Clustering Framework for Urban Heat Risk Zone Identification

Agata Zabuska[1], Lidia Lazarova Vitanova [1,*], Tereza K. Trendafilova [1], Dessislava Petrova Antonova[1]

[1]GATE Institute, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria
{agata.zabuska, lidia.vitanova, dessislava.petrova}@gate-ai.eu
tktrendafilova@gmail.com

Keywords: Machine Learning, Spatial Constrained Clustering, Geo Spatial Analysis, Weather Research and Forecasting Model, Wet Bulb Globe Temperature.

Abstract

This study developed a spatially constrained clustering (SCC) framework for identifying urban heat-risk zones by integrating environmental heat hazard, population exposure, socio-demographic vulnerability, and mitigation indicators. Heat risk was conceptualized as a multidimensional framework, and SCC was applied to derive geographically contiguous zones in Sofia, Bulgaria. The results revealed a dominant urban–peripheral gradient in heat risk, with the urban core characterized by the co-occurrence of elevated thermal stress, high population density, and reduced mitigation capacity. While the SCC approach successfully identified spatially coherent and interpretable risk zones, the findings also suggest that additional variables and higher-resolution data are required to capture more detailed intra-urban differences. Overall, the proposed framework demonstrates the potential of integrating multidimensional geospatial data with the SCC approach to support urban heat-risk assessment and inform climate adaptation strategies.

1. Introduction

Urbanization and climate change are intensifying heat-related risks in cities worldwide. Europe is warming at approximately twice the global average rate, resulting in more frequent and severe heatwaves (World Meteorological Organization, 2022). Urban areas are particularly vulnerable because the urban heat island effect amplifies thermal stress through dense built surfaces, increased anthropogenic heat emissions (Koleva et al., 2024), and limited vegetation (Beddow, 2022). As a result, heat exposure increasingly threatens public health, urban infrastructure, and the resilience of cities (EEA, 2024).

Assessing these socioeconomic and infrastructural vulnerabilities requires more than measuring temperature alone. The Intergovernmental Panel on Climate Change (IPCC) conceptualizes climate risk as the interaction of hazard, exposure, and vulnerability (Simpson et al., 2021), recognizing that heat-related impacts depend not only on the intensity of the hazard but also on the characteristics of exposed populations and their capacity to cope with extreme heat. This framework has therefore become the conceptual basis for integrated urban heat-risk assessment.

Applying this integrated approach is particularly urgent in Southeastern Europe, where the consequences of extreme heat have severely intensified. Bulgaria, for instance, has recorded one of the highest per-capita heat-related mortality rates in recent summers (Janoš et al., 2025), while its capital, Sofia, has experienced rapid urban expansion accompanied by increasing built-up areas and declining green space (Vitanova et al., 2026). These intersecting trends increase both heat exposure and the vulnerability of urban populations, making Sofia a critical case study for investigating the spatial patterns of urban heat risk.

While Sofia provides an ideal empirical setting for this research, existing local literature has yet to fully adopt an integrated risk perspective. Although several studies have examined urban heat islands and thermal environments in the city (Vitanova et al., 2021; Vitanova and Kusaka, 2018), most focus primarily on temperature distributions. Comparatively few studies integrate climatic, demographic, socioeconomic, and mitigation indicators within a unified spatial assessment. Moreover, urban heat vulnerability is commonly represented using composite indices, weighted overlays, or predefined vulnerability classes, which may overlook geographically coherent patterns of heat risk.

To address this gap, this study develops a geospatial framework that integrates hazard, exposure, vulnerability, and mitigation indicators using Spatially Constrained Clustering (SCC). Unlike conventional urban heat vulnerability mapping, the proposed framework identifies geographically contiguous heat-risk zones directly from multivariate data rather than predefined classes or composite indices. Specifically, the study evaluates whether SCC can identify coherent heat-risk regions during an extreme heat event and examines how these regions differ in terms of hazard, exposure, vulnerability, and mitigation characteristics.

By identifying spatially coherent heat-risk zones, the proposed framework provides a more actionable representation of urban heat risk than fragmented, variable-based assessments. Such zones can better support urban planning by informing district-scale cooling strategies, targeted greening initiatives, the placement of cooling shelters, and the development of shaded mobility corridors.

2. Data and methodology

Figure 1 outlines the workflow for implementing the SCC framework, including data collection, preprocessing, exploratory analysis, and modelling through the application of SCC algorithms. Environmental, land use, and socio-demographic data are integrated, cleaned, and standardised into a common spatial dataset. Heat risk is represented by integrating the hazard, exposure, and vulnerability components, consistent with the IPCC risk framework, while spatially constrained clustering is used to delineate contiguous urban heat-risk zones.

* Corresponding author

https://doi.org/10.5194/isprs-annals-XII-4-W1-2026-343-2026

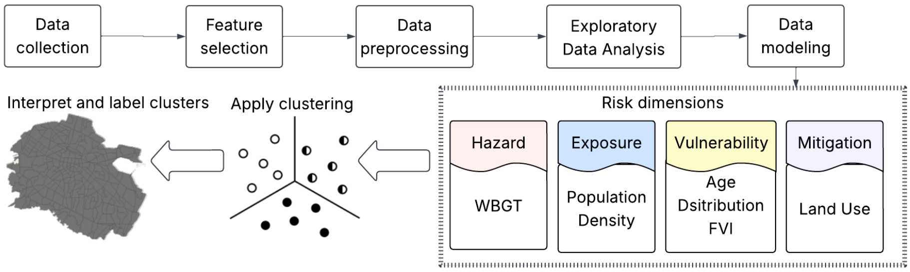

Figure 1. Study workflow.

2.1 Study area

The study area is Sofia City, the capital of Bulgaria. Over the past twenty years, Sofia's population has grown steadily, reaching nearly 1.3 million in 2024, or about one-fifth of Bulgaria's total (World Population Review, 2024). Recently, the region has seen more prolonged heatwaves, with extended periods above 32 °C (The Sofia Globe staff, 2024), and Bulgaria recorded the second-highest heat-related mortality rate per million inhabitants in summer 2023 (Gallo et al., 2024). These trends underscore the urgent need to improve urban heat-risk assessment in the region.

2.2 Data collection

The study utilizes Wet Bulb Globe Temperature (WBGT), land use, and socio-demographic datasets to represent the dimensions of urban heat risk. Three main data sources are utilized in this research: simulation outputs from the Weather Research and Forecasting (WRF) model, and datasets from Copernicus and Sofia Plan.

2.3 Feature selection

WBGT data is obtained from high-resolution simulations generated using the WRF model. The dataset is provided in netCDF format and extracted to .txt at 250 m resolution. For this analysis, WBGT values corresponding to 16 July 2024 at 16:00 are selected, representing the hottest hour of one of the most extreme summer days. The WBGT index integrates air temperature, humidity, wind speed, and radiative heat load, and is widely used as an indicator of human thermal stress (Ren et al., 2023). In the proposed framework, the index is used to represent the hazard dimension.

Land use data is derived from the Sentinel-2 Land Cover Explorer dataset Esri and Microsoft (2024). The dataset is provided in raster format and includes five urban classes with a spatial resolution of 10 meters per pixel (10 m × 10 m). The version used in this study reflects land use conditions for the year 2024. It is used to derive an urban green ratio representing local mitigation capacity.

Population density data is obtained from SofiaPlan (2024). It is expressed as percentages at the Urban Unit (GE) level. Spatial analysis and visualization are performed in ArcGIS Environmental Systems Research Institute (Esri) (2025). The data is aggregated into the common spatial unit used in this study to represent population exposure to extreme heat.

Age distribution data is also obtained from (SofiaPlan, 2024). The original demographic classification provided by SofiaPlan is structured into the following age groups: 0–2 years, 3–6 years, 7–14 years, 15–18 years, 19–65 years, and over 65 years. These indicators are used to represent demographic vulnerability to heat stress.

Financial Vulnerability Index (FVI) data is sourced from (SofiaPlan, 2013) and provided in vector format. It contains information on average income levels across spatial units and their deviation from the city-wide average income. The FVI is used to represent socio-economic vulnerability within the urban population.

2.4 Data preprocessing

Prior to analysis, all datasets are harmonized to ensure spatial and numerical consistency. First, all spatial layers are reprojected to a common coordinate reference system (WGS 84 / UTM Zone 34N, EPSG:32634). WBGT, land use, and socio-demographic datasets are then spatially aggregated and clipped to a unified set of 251 urban neighborhood units, which serve as the primary analytical spatial units in this study.

To enable comparison across variables measured in different units, all non-geometric attributes are standardized prior to modeling. Continuous indicators are transformed using z-score normalization, ensuring a zero mean and unit variance.

Some features are aggregated or transformed into indices to facilitate a more consistent and interpretable analytical process. The selected variables and indices created are described below.

For the purposes of this study, the age distribution groups are reclassified into broader analytical categories: 0–6 years, 7–18 years, 19–64 years, and 65+ years. This reclassification is introduced to better reflect age groups particularly vulnerable to heat-related illnesses (Clark et al., 2024), while ensuring consistency and comparability across spatial units.

For the purposes of this analysis, the original five-class land cover data are aggregated into a binary classification. The Built Area class is assigned to the Urban Area category, while all other classes (water, trees, crops, and rangeland) are consolidated into the Green Area category. The Green Ratio Index (GRI) is then defined as a fractional index, calculated as

$$GRI_i = \frac{A_i^{green}}{A_i^{green} + A_i^{urban}} \tag{1}$$

where A_i^{urban} and A_i^{green} represent the total areas (in m^2) of urban and green land-cover classes, respectively, within spatial unit i (Bahr, 2024).

Values approaching 0 indicate predominantly urban environments, whereas values approaching 1 indicate highly vegetated areas.

Additionally, this study calculates the Financial Vulnerability Index (FVI). The FVI is constructed following established composite indicator methodologies (OECD Publishing, 2008; Cutter et al., 2003). It is developed to quantify the relative economic disadvantage between districts by combining measures of absolute income levels and relative income position within the municipality. The index is based on the average monthly income per individual and the percentage deviation of district-level income from the municipal average, thereby capturing both material capacity and relative deprivation.

I_i denotes the average monthly income in district i, and P_i represents the percentage deviation of district income from the Sofia municipal average. Absolute income vulnerability is calculated as

$$V_i^{abs} = 1 - \frac{I_i - I_{\min}}{I_{\max} - I_{\min}} \tag{2}$$

Relative income vulnerability is calculated by retaining only negative deviations from the municipal average, as districts with income equal to or above the citywide mean are not considered financially vulnerable relative to the municipal average.

$$V_i^{rel} = \max\left(0, -\frac{P_i}{100}\right) \tag{3}$$

The FVI is defined as a weighted composite indicator:

$$FVI_i = 0.6\, V_i^{abs} + 0.4\, V_i^{rel} \tag{4}$$

where i denotes the spatial unit.

2.5 Exploratory data analysis

Exploratory data analysis is conducted to understand the dataset's structure and characteristics. The steps are outlined below.

Before clustering, summary statistics (mean, standard deviation, minimum, maximum, and quartiles) are computed for all numerical features to describe central tendency and dispersion across spatial units. Skewness is also calculated to assess distributional asymmetry not captured by the mean and standard deviation.

A correlation matrix is generated to examine relationships between pairs of features before clustering. It reports correlation coefficients that measure the strength and direction of linear relationships, helping to reveal underlying structures and assess associations between variables (Looney and Hagan, 2011).

Moran's I is used to measure spatial autocorrelation. It produces values from -1 to 1, where high positive values indicate clustering, negative values indicate dispersion, and values near zero suggest randomness. Moran's I is computed for each feature to assess spatial structure and determine whether clustering is likely to yield informative spatial groupings (Li et al., 2007).

2.6 Spatially constrained clustering (SCC)

The risk assessment follows the conceptual framework established by the Intergovernmental Panel on Climate Change (IPCC, 2014, 2022), in which risk is defined as a function of hazard, exposure, and vulnerability. Following urban adaptations of this model (Tapia et al., 2017), an additional mitigation dimension representing nature-based adaptive capacity is integrated to assess the buffering effect of green infrastructure. In this study, risk is defined based on the following components:

$$\text{Risk} = f(H, E, V, M) \tag{5}$$

where the dimensions are defined as:

$$H = \text{Wet Bulb Globe Temperature (WBGT)} \tag{6}$$
$$E = \text{Population Density} \tag{7}$$
$$M = \text{Green Space Index} \tag{8}$$
$$V = \frac{(\text{Age}_{0-6} + \text{Age}_{65+} + \text{FVI})}{3} \tag{9}$$

To ensure equal contribution of each dimension, all features are standardized using z-score normalization. The vulnerability dimension, comprising three distinct indicators, is aggregated into a composite mean index to mitigate high-dimensional bias in Euclidean distance calculations.

To establish a statistical baseline, the K-means algorithm, an unsupervised machine learning method that partitions the input data into distinct clusters (Jain et al., 1999), is first applied to the standardized features. The optimal number of clusters (k) is determined by evaluating the Silhouette Score for a range of 2 to 20 clusters. This non-spatial stage identifies the "global" risk archetypes within the dataset without geographic constraints.

To ensure geographic coherence and account for the spatial contiguity of urban units, two spatially aware approaches are implemented:

1. Spatially Weighted K-Means: To enhance clustering by explicitly incorporating spatial proximity, a hybrid K-means approach is applied. By integrating coordinate-based proximity with the IPCC risk dimensions, the method evaluates whether geographically coherent risk zones can be identified.

2. Self-Organizing Maps (SOM): The SOM projects the multidimensional input feature space onto a two-dimensional lattice (Kaski, 2010). This process preserves the topological relationships within the data, ensuring that urban units with similar risk profiles are mapped to adjacent neurons. The structural output is analyzed through component planes, which act as thematic layers illustrating the weight of individual variables across the grid. By cross-referencing these planes, the convergence of hazard, exposure, and vulnerability that defines each zone is analyzed, enabling a more granular interpretation of urban risk beyond a simple binary classification.

Following each spatial clustering process, the resulting clusters are characterized through profile analysis. Centroids for each IPCC dimension are calculated to identify the dominant risk

https://doi.org/10.5194/isprs-annals-XII-4-W1-2026-343-2026

drivers within each cluster. Furthermore, an ordinal ranking of these centroids facilitates comparative cross-cluster analysis.

The performance and stability of the clustering solutions are assessed using the Silhouette Score as the primary internal validation metric. This method evaluates the balance between intra-cluster cohesion and inter-cluster separation.

3. Results and discussions

3.1 Choropleth-based analysis

Selected features are visualized as choropleth maps to serve as a diagnostic tool for examining how WBGT, land use, and socioeconomic factors are distributed across urban units. Lower WBGT values are observed mainly in the peripheral parts of the city, as shown in Figure 2. It is important to note that even the lowest observed values still fall within the "Warning" category of the WBGT classification (Guidelines for the prevention of heat illness in daily life, 2024). In contrast, inner-city areas exhibit WBGT values ranging from approximately 29.00 to nearly 31.00 °C, corresponding to the severe warning and danger categories (Guidelines for the prevention of heat illness in daily life, 2024).

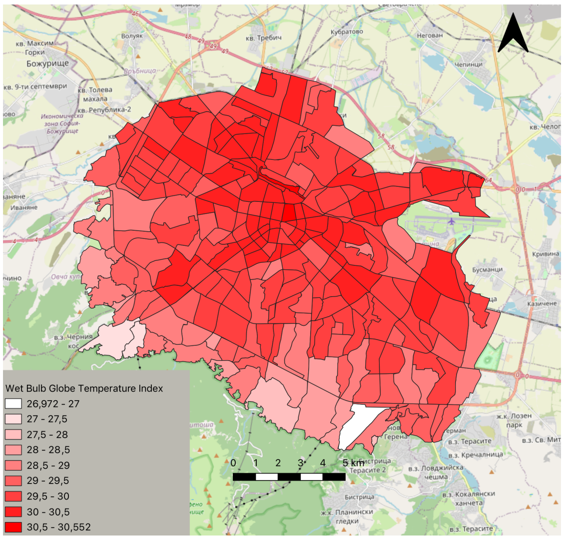

Figure 2. Wet Bulb Globe Temperature (WBGT) on 16 July 2024 at 16:00 LST.

Figure 3 visualizes how lower GRI values are concentrated in the central parts of Sofia. The mapping reveals a scattered spatial distribution of green areas throughout the city, including Borisova Gradina Park, South Park, and West Park, as well as the disproportionality between highly green and low-green areas. Out of the 251 urban units, only 31 have a GRI value higher than 0.50, while a value of 0.00 appears most frequently, indicating that most areas within the city exhibit limited capacity to mitigate thermal stress.

FVI values are visualized in Figure 4, revealing a clear north–south divide, with lower vulnerability concentrated in the southern and central parts of the city and higher values (estimated at 0.70 and above) forming more contiguous clusters in northern areas. This pattern closely aligns with previously identified poverty distributions in Sofia (Ryustem et al., 2018), suggesting a persistent spatial structuring of socio-economic

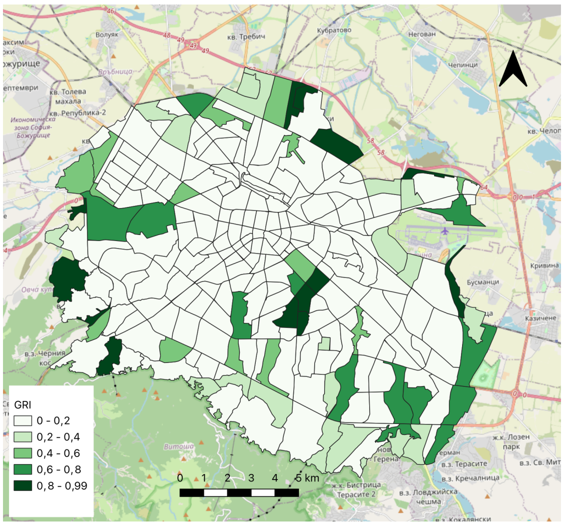

Figure 3. Green Ratio Index (GRI) in 2024.

disadvantage. The observed clustering indicates that financial vulnerability is not randomly distributed but instead reflects underlying urban inequalities.

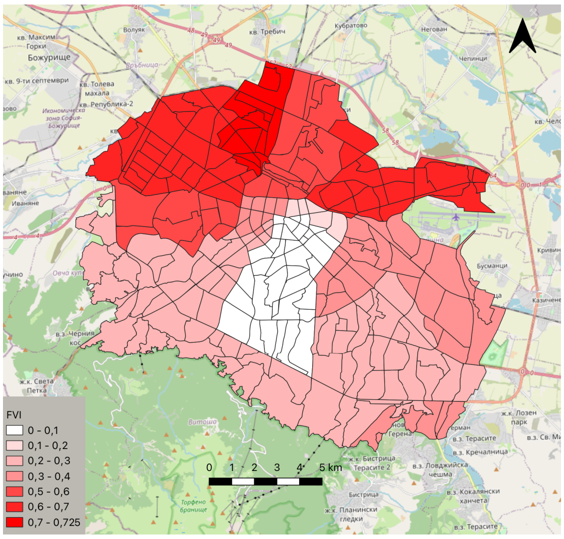

Figure 4. Financial Vulnerability Index (FVI) in 2013.

Figure 5 illustrates that the highest population densities are concentrated in the southeastern parts of the city, particularly in the Mladost and Gorublyane districts, as well as in the Lyulin area in the western part of Sofia. These high-density zones form spatially distinct clusters, indicating concentrated residential patterns. Lower densities in central and peripheral areas suggest a more heterogeneous urban fabric. From a risk perspective, the clustering of high population densities highlights areas of elevated exposure, which may increase the potential impact of environmental hazards, particularly when these zones spatially coincide with high hazard or vulnerability levels.

The spatial distribution of age groups of particular interest is also examined. Figure 6 presents the distribution of residents aged 0–6 years, a group widely recognized as particularly vulnerable to heat stress. The highest proportion of this age group (17.65%) is observed in Boyana, suggesting a localized concentration of young families in specific residential areas. More

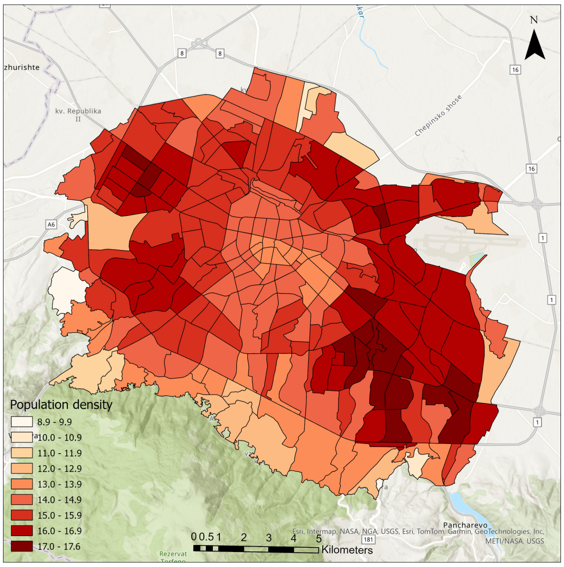

Figure 5. Population density in 2024.

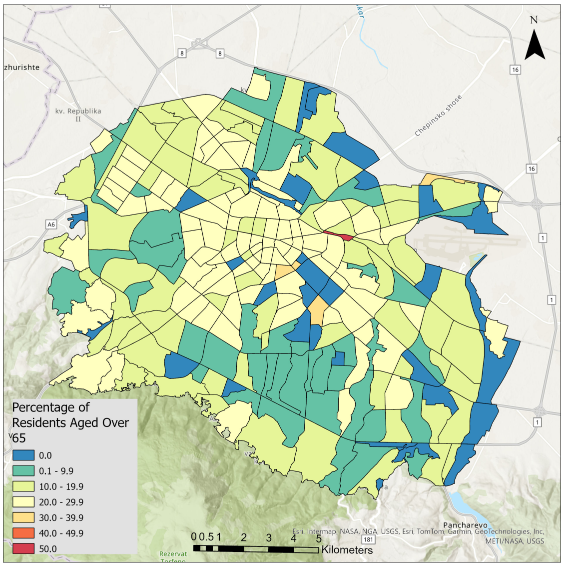

Figure 7. Percentage of people aged over 65 in 2024.

broadly, the distribution of young children appears spatially uneven, with higher values concentrated in selected districts rather than being uniformly distributed across the city.

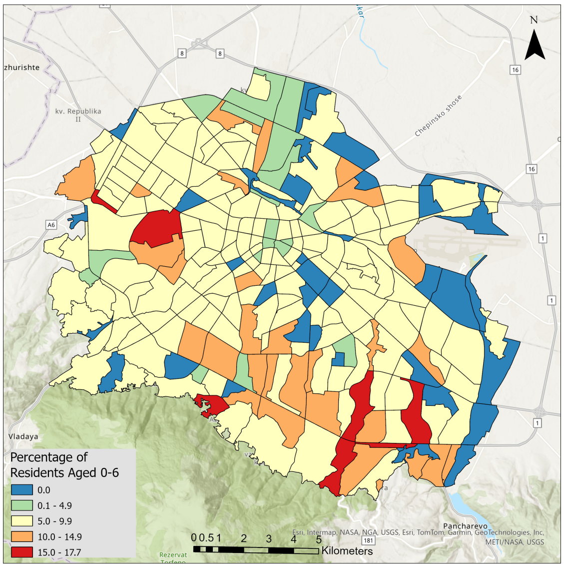

Figure 6. Percentage of people aged 6 and below in 2024.

Figure 7 illustrates the distribution of residents aged 65 years and older, a group consistently identified as the most vulnerable during extreme heat events. The highest proportion of elderly residents (50%) is recorded in Poduene, indicating a significant spatial concentration of ageing populations in this area. However, most urban units exhibit relatively low shares of elderly residents, with 155 units showing proportions below 20%, suggesting that high elderly concentrations are localized rather than widespread. From a vulnerability perspective, the spatial separation of these age groups highlights that different forms of demographic vulnerability are unevenly distributed across the city.

3.2 Descriptive statistics

The previous step provided a visual assessment of the data distribution; however, generating a statistical description enables a more quantitative evaluation. The GRI exhibits strong positive skewness (Table 1), indicating that the majority of districts have low green-space availability, while a small number of districts contain substantially higher proportions of green space. This skewness validates the observations made during the assessment of the GRI visualizations.

Conversely, WBGT displays negative skewness (Table 1) with a relatively low standard deviation, suggesting that heat exposure is comparatively homogeneous across the study area, with only a few cooler outliers. This limited variability provides insight into the potential for cluster formation, suggesting that hazard alone may not strongly differentiate spatial clusters. Demographic and FVI variables exhibit greater dispersion than WBGT, indicating that clustering outcomes may be more strongly influenced by vulnerability and mitigation dimensions than by hazard intensity alone.

	Pop.	A0–6	A65+	GRI	WBGT	FVI
Skewness	-0.88	0.17	-0.21	1.81	-1.46	-0.00

Table 1. Skewness values across variables.

3.3 Spearman Correlation Matrix

A weak positive correlation ($r = 0.19$) is observed between WBGT and population density, indicating that denser areas experience slightly higher thermal exposure, consistent with intensified urban heat in compact cities. A moderate negative correlation ($r = -0.59$) between GRI and WBGT shows that districts with more green space tend to have lower heat intensity, supporting the mitigating role of vegetation. A moderate negative association ($r = -0.22$) between GRI and population density reflects the contrast between dense urban cores and greener peripheral districts. These relationships, shown in Figure 8, provide a preliminary assessment of the strength and direction of associations among the variables.

3.4 Moran's I

The values obtained from Moran's I analysis are presented in Table 2. Population density, WBGT, and FVI exhibit positive spatial autocorrelation, indicating spatial clustering (e.g., high

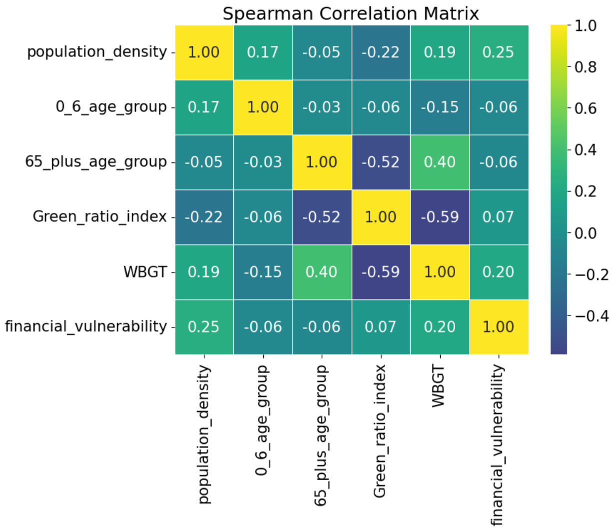

Figure 8. Spearman Correlation Matrix visualization.

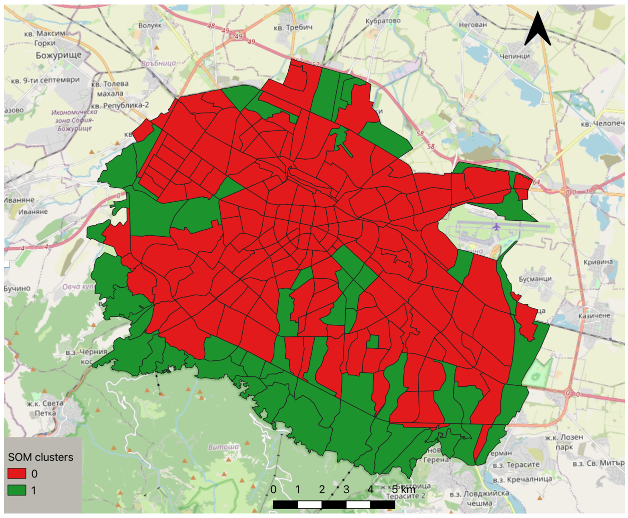

Figure 9. Self-Organizing Maps (SOM) clusters.

values near other high values). In contrast, age distribution and GRI values are lower, suggesting a more random or dispersed spatial distribution. This step provides insight into the spatial structure of the analyzed features by identifying which variables exhibit clustering and measuring the intensity of that clustering.

	Pop.	A0–6	A65+	GRI	WBGT	FVI
Moran's I	0.52	0.13	0.20	0.22	0.55	0.89

Table 2. Spatial autocorrelation (Moran's I) across variables.

3.5 Spatial modeling and clustering

A K-means algorithm is first applied to the H, E, V, and M features. The optimal number of clusters is selected using the Silhouette Score for $k = 2$ to $k = 20$, with the best score (0.34) at $k = 2$. This initial risk topology reveals a basic clustering structure without spatial constraints. To enforce geographic coherence, a Spatially Weighted K-means is then used by adding scaled longitude and latitude to the features and applying a spatial weight of 1.5 to favor geographic proximity. The distribution remains consistent, again yielding the best Silhouette score (0.35) at $k = 2$. This indicates a dominant macro-scale split between the dense inner city and peripheral areas, rather than finer-grained risk zones. Finally, a SOM algorithm is tested for $k = 2$ to $k = 6$, with the highest Silhouette score (0.34) again at $k = 2$. The resulting clusters, summarized in Table 3, show that Cluster 0 (the urban cluster, Figure 9) has the highest H, V, and E values, while Cluster 1 (the peripheral cluster) has the highest M value.

Cl.	Pop.	A0–6	A65+	GRI	WBGT	FVI
0	2.00	2.00	2.00	1.00	2.00	2.00
1	1.00	1.00	1.00	2.00	1.00	1.00

Table 3. Self-Organizing Maps (SOM) cluster characteristics.

While global clustering identifies the dominant urban–peripheral structure, SOM component planes reveal heterogeneity within these macro-scale zones, enabling the identification of localized risk hotspots within the broader urban cluster. To identify these hotspots, the SOM is queried through multidimensional risk-intersection analysis. Neurons are filtered based on a profile exhibiting values in the upper tercile (above the 70th percentile) for H, E, and V, together with below-average capacity in the M plane (Figure 10).

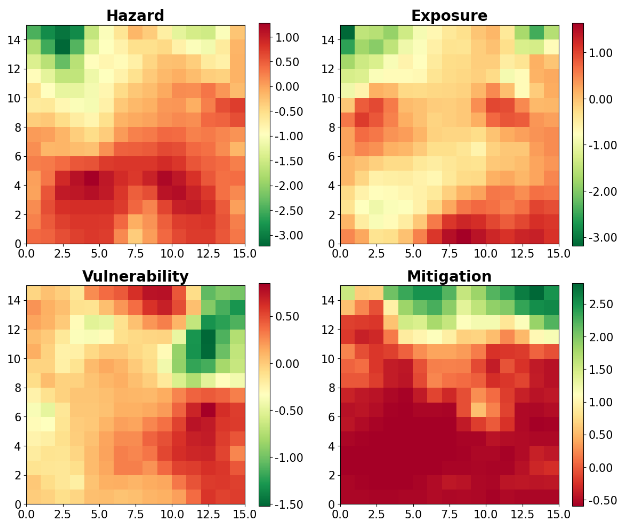

Figure 10. The Intergovernmental Panel on Climate Change (IPCC) pillars visualized as component planes.

To validate these findings geographically, a reverse-mapping procedure is employed. Coordinates from the component planes are queried to identify the specific urban units represented in the original database. This process reveals neighbourhoods (listed in Table 4) that fall under the identified high-risk classification.

Figure 11 visualizes localized risk zones that the global clustering approach is unable to capture, but which the SOM component plane analysis successfully isolates.

4. Conclusions and future work

This study demonstrates how an IPCC-inspired risk framework can be integrated with SCC approaches to analyse urban heat risk. Rather than introducing a new clustering methodology,

Coord.	H	V	E	M	Areas
(3,10)	1.16	0.62	0.59	-0.54	Kremikovtsi; Nadezhda/Vrabnitsa
(2,11)	1.01	0.55	0.52	-0.56	Vrabnitsa; Ilinden
(2,10)	0.97	0.54	0.65	-0.57	Krasna Polyana; Poduene
(3,11)	0.96	0.66	0.59	-0.58	Nadezhda
(2,12)	0.86	0.61	0.66	-0.48	Nadezhda; Serdika/Ilinden
(1,12)	0.80	0.56	0.97	-0.43	Vrabnitsa
(0,12)	0.75	0.59	1.28	-0.55	Poduene; Lyulin
(0,13)	0.54	0.54	1.31	-0.56	Poduene; Lyulin
(2,13)	0.53	0.58	0.99	-0.32	Poduene; Lyulin

Table 4. Top Self-Organizing Maps (SOM) nodes and associated areas.

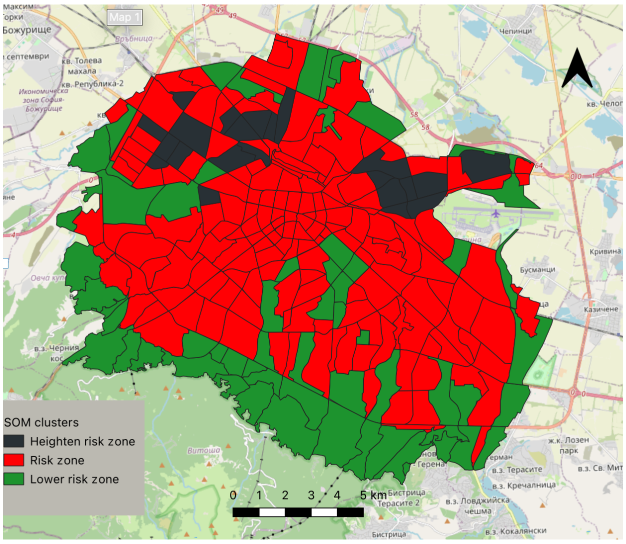

Figure 11. Heightened risk zones visualization.

it combines established SCC techniques within a unified risk assessment workflow. The heterogeneous spatial distributions of the analysed variables, evident in both the choropleth maps and descriptive statistics, highlight the multidimensional nature of urban heat risk.

All methods consistently identified $k = 2$ as the optimal number of clusters (Silhouette 0.34–0.35), revealing a dominant binary spatial pattern. The high-risk cluster shows greater hazard, exposure, and vulnerability, while the low-risk cluster has higher mitigation capacity. This agreement across methods indicates that the urban–peripheral gradient is the main spatial structure of heat risk in Sofia. The limited differentiation between clusters likely reflects the chosen variables, spatial resolution, and reliance on a single extreme heat event, which together highlight the broad urban–peripheral contrast over finer intra-urban differences. Unlike conventional composite indices, weighted overlays, or predefined classes, the proposed framework derives spatially contiguous heat-risk zones directly from the data, offering a complementary, data-driven view of urban heat vulnerability.

SOM component-plane analysis provides a more granular interpretation of heat risk. Areas including Kremikovtsi, Nadezhda, Vrabnitsa, Ilinden, Krasna Polyana, Poduene, Serdika, and Lyulin exhibit relatively highest values in dimensions H, E, and V and low M capacity. This finding suggests that although these areas are a part of a bigger risk zone, these specific neighbourhoods should be prioritized when planning heat mitigation actions.

The limited number of input variables, the use of a single extreme heat event, and the available spatial resolution constrain the ability to capture finer intra-urban variability, suggesting that additional variables and temporal depth are required for more granular risk differentiation. Although WBGT was used as the primary hazard indicator, its composite formulation incorporates air temperature, humidity, wind speed, and solar radiation, providing a robust representation of outdoor heat stress while not capturing every aspect of the urban thermal environment.

As limitations, this study does not incorporate direct health outcome data, such as emergency department visits or heat-related mortality. The availability of such data is often limited and typically lacks the spatial resolution required for integration within this framework. As a result, the identified zones represent potential heat risk rather than empirically observed health impacts. Future validation could compare the identified zones with observed heat-related health outcomes where such data become available.

Future work should focus on integrating additional variables across the hazard, exposure, vulnerability, and mitigation dimensions to improve cluster granularity. Sensitivity analyses of the spatial weighting parameter and additional cluster-validation metrics would further strengthen the framework's methodological robustness. In particular, incorporating additional climatic indicators alongside WBGT, temporal heat dynamics, and socio-demographic indicators would enhance the representation of urban risk. Expanding the framework to a multi-temporal or predictive context would further support its application in long-term urban planning and climate adaptation.

5. Acknowledgments

This research was supported by the GATE project, funded by the H2020 WIDESPREAD-2018-2020 TEAMING Phase 2 programme, under grant agreement no. 857155, the programme "Research, Innovation and Digitalisation for Smart Transformation" 2021-2027 (PRIDST) under grant agreement no. BG16RFPR002-1.014-0010-C01 and DS4SSCC–DEP project funded by the European Union Digital Europe Work Programme under grant agreement no. 101123342. We acknowledge the provided access to the Discoverer at Sofia Tech Park (Bulgaria).

References

Bahr, S., 2024. The relationship between urban greenery, mixed land use and life satisfaction: An examination using remote sensing data and deep learning. *Landscape and Urban Planning*, 251, 105174.

Beddow, H., 2022. How the urban heat island effect makes cities vulnerable to climate change. https://ukgbc.org/news/how-the-urban-heat-island-effect-makes-cities-vulnerable-to-climate-change (20 Feburary 2026).

Clark, A., Grineski, S., Curtis, D. S., Cheung, E. S. L., 2024. Identifying groups at-risk to extreme heat: Intersections of age, race/ethnicity, and socioeconomic status. *Environment International*, 191, 108988. http://dx.doi.org/10.1016/j.envint.2024.108988.

Cutter, S. L., Boruff, B. J., Shirley, W. L., 2003. Social Vulnerability to Environmental Hazards*. *Social Science Quarterly*, 84(2), 242–261. http://dx.doi.org/10.1111/1540-6237.8402002.

EEA, 2024. The impacts of heat on health. https://data.europa.eu/doi/10.2800/8167140.

Environmental Systems Research Institute (Esri), 2025. (arcgis desktop: Release 10.8). https://www.esri.com (26 Feburary 2026). Redlands, CA: Environmental Systems Research Institute.

Esri, I. O., Microsoft, 2024. Sentinel-2 Land Cover Explorer. https://livingatlas.arcgis.com/landcoverexplorer/.

Gallo, E., Quijal-Zamorano, M., Méndez Turrubiates, R. F., Tonne, C., Basagaña, X., Achebak, H., Ballester, J., 2024. Heat-related mortality in Europe during 2023 and the role of adaptation in protecting health. *Nature Medicine*, 30(11), 3101–3105. http://dx.doi.org/10.1038/s41591-024-03186-1.

Guidelines for the prevention of heat illness in daily life, 2024. Guidelines for the prevention of heat illness in daily life. *Japanese Society of Biometeorology*.

IPCC, 2014. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

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*. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Jain, A. K., Murty, M. N., Flynn, P. J., 1999. Data clustering: a review. *ACM Comput. Surv.*, 31(3), 264–323. https://doi.org/10.1145/331499.331504.

Janoš, T., Quijal-Zamorano, M., Shartova, N., Gallo, E., Méndez Turrubiates, R. F., Denisse Beltrán Barrón, N., Peyrusse, F., Ballester, J., 2025. Heat-related mortality in Europe during 2024 and health emergency forecasting to reduce preventable deaths. *Nature Medicine*, 31(12), 4065–4074. http://dx.doi.org/10.1038/s41591-025-03954-7.

Kaski, S., 2010. *Self-Organizing Maps*. Springer US, Boston, MA, 886–888.

Koleva, T., Vitanova, L., Petrova-Antonova, D., Kostadinov, A., 2024. Semantic enrichment and analysis of building energy consumption data for the city of sofia. *IFIP International Conference on Artificial Intelligence Applications and Innovations*, Springer, 462–475.

Li, H., Calder, C. A., Cressie, N., 2007. Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. *Geographical Analysis*, 39(4), 357-375. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1538-4632.2007.00708.x.

Looney, S. W., Hagan, J. L., 2011. *Statistical Methods for Assessing Biomarkers and Analyzing Biomarker Data*. Elsevier, 27–65.

OECD Publishing, 2008. *Handbook on constructing composite indicators*. Organization for Economic Co-operation and Development (OECD), Paris Cedex, France.

Ren, J., Shi, K., Li, Z., Kong, X., Zhou, H., 2023. A Review on the Impacts of Urban Heat Islands on Outdoor Thermal Comfort. *Buildings*, 13(6), 1368. http://dx.doi.org/10.3390/buildings13061368.

Ryustem, R., Dimitrova, D., Kostova, M., Corral, P., Azevedo, J. P., 2018. Poverty mapping in the republic of bulgaria. p. 44, (05 March 2026).

Simpson, N. P., Mach, K. J., Constable, A., Hess, J., Hogarth, R., Howden, M., Lawrence, J., Lempert, R. J., Muccione, V., Mackey, B., New, M. G., O'Neill, B., Otto, F., Pörtner, H.-O., Reisinger, A., Roberts, D., Schmidt, D. N., Seneviratne, S., Strongin, S., van Aalst, M., Totin, E., Trisos, C. H., 2021. A framework for complex climate change risk assessment. *One Earth*, 4(4), 489–501. http://dx.doi.org/10.1016/j.oneear.2021.03.005.

SofiaPlan, 2013. Sofiaplan open data api. https://api.sofiaplan.bg/datasets/635.

SofiaPlan, 2024. Sofiaplan open data api. https://sofiaplan.bg/api/.

Tapia, C., Abajo, B., Feliu, E., Mendizabal, M., Martinez, G., Fernández, J., Laburu, T., Lejarazu, A., 2017. Profiling urban adaptation to climate change: A set of common indicators for European cities. *International Journal of Climate Change Strategies and Management*, 9(4), 488–508.

The Sofia Globe staff, 2024. Bulgaria's july 2024 one of two hottest since 1930. https://sofiaglobe.com/2024/07/31/bulgarias-july-2024-one-of-two-hottest-since-1930-weather-bureau/ (26 Feburary 2026).

Vitanova, L. L., Kusaka, H., 2018. Study on the urban heat island in Sofia City: Numerical simulations with potential natural vegetation and present land use data. *Sustainable cities and society*, 40, 110–125.

Vitanova, L. L., Kusaka, H., Doan, Q., Subasinghe, S., 2021. How urban growth changes the heat island effect and human thermal sensations over the last 100 years and towards the future in a European city? *Meteorological Applications*, 28(4). http://dx.doi.org/10.1002/met.2019.

Vitanova, L. L., Matsumura, S., Petrova-Antonova, D., 2026. Influence of Muesmann's plan on contemporary urban planning in Sofia City. *City and Environment Interactions*, 29, 100303. http://dx.doi.org/10.1016/j.cacint.2026.100303.

World Meteorological Organization, 2022. Temperatures in europe increase more than twice global average. https://wmo.int/news/media-centre/temperatures-europe-increase-more-twice-global-average (20 Feburary 2026).

World Population Review, 2024. World population review.