# Evaluating High-Resolution Geospatial Data for Urban Microclimate Simulations: A Comparative Study Using Official Local and Open Global Datasets in Sofia

Lidia Lazarova Vitanova [1]*, Kunihiko Fujiwara [2], Ryota Karube [1,3], Evgeny Shirinyan [1], Yasuhiko Azegami [2], Quang-Van Doan [3], Hiroyuki Kusaka [4], Dessislava Petrova-Antonova [1], Sylvia Ilieva [1]

[1] GATE Institute, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria
{lidia.vitanova, ryota.karube, evgeny.shirinyan, dessislava.petrova, sylvia}@gate-ai.eu

[2] Research & Development Institute, Takenaka Corporation, Japan
{fujiwara.kunihiko, azegami.yasuhiko}@takenaka.co.jp

[3] Graduate School of Science and Technology, University of Tsukuba, Japan
s2530306@u.tsukuba.ac.jp

[4] Center for Computational Sciences, University of Tsukuba, Japan
doan.van.gb@u.tsukuba.ac.jp; kusaka@ccs.tsukuba.ac.jp

**Keywords:** VoxCity, ArcGIS, local dataset, global dataset, City-LES, urban climate modelling.

**Abstract**

Recent advances in digital technologies have enabled the development of urban tools that integrate geospatial data to support the analysis of complex urban systems and climate change. However, the reliability of such tools depends strongly on the availability and integration of high-quality geospatial data. This study evaluates the suitability of official local and open global geospatial data for high-resolution urban climate simulations through a comparative two-case approach: (1) using official local datasets (OLD) and (2) open global datasets (OGD) processed through ArcGIS Pro and VoxCity, respectively. Both datasets are incorporated into the City-LES model to simulate urban microclimate conditions. The Lozenets district in Sofia is selected as a representative case study. The simulation results from the City-LES model using the OLD and OGD datasets are compared and verified to assess the ability of local and global data to reproduce urban microclimate conditions. The results show that both OLD and OGD datasets effectively represent urban morphology and support microclimate simulations. While the simulation with the OLD provides detailed spatial resolution, the simulation with the OGD captures the overall urban structure, demonstrating complementary strengths. Differences in input data lead to variations in simulated microclimate conditions, particularly at finer spatial scales. The verified workflow demonstrates that official local and open global datasets complement each other in supporting scalable urban climate modelling, enabling urban heat assessment, climate adaptation planning, and evidence-based decision-making in data-scarce regions.

## 1. Introduction

Recent advances in digital technologies have enabled the development of urban modelling tools that integrate geospatial data to support the analysis of complex urban systems and climate change (Assenova et al., 2024; Pavlov et al., 2025; Subasinghe et al., 2022).

Tools such as ArcGIS Pro and VoxCity provide a flexible framework for integrating heterogeneous urban datasets, enabling the creation of detailed 3D city models and supporting scenario-based simulations for environmental and climate analyses (Fujiwara et al., 2026). However, the accuracy and reliability of such tools strongly depend on the availability and integration of high-quality local and global geospatial data.

Locally sourced datasets, including detailed land-use information, building and energy characteristics, meteorological observations, and urban morphology, are essential for capturing the specific conditions of a given city and ensuring realistic simulations (Koleva et al., 2024; Vitanova et al., 2025; Vitanova et al., 2023). Conversely, open global datasets provide broader spatial coverage and improved accessibility but may not adequately represent local urban characteristics (Lloyd et al., 2019).

Therefore, comparing open global datasets with high-resolution local datasets is crucial for producing reliable analyses and supporting effective, context-sensitive urban and climate mitigation strategies (Chen et al., 2020). For example, a study examines whether the global population dataset WorldPop could reliably replace local survey data for community-scale landslide risk assessment (Opdyke & Fatima, 2024). It was concluded that local survey data provided a more reliable representation of population exposure, whereas the WorldPop dataset substantially overestimated landslide risk.

Sofia city is no exception to this challenge, as the availability, completeness, and spatial accuracy of urban geospatial data vary considerably between local and global sources. Assessing how these differences affect urban climate simulations is essential to determining whether open global datasets can provide reliable alternatives when detailed local data are unavailable.

To address this research gap, the study evaluates the differences in data quality and accuracy of local and open urban geospatial data through a comparative two-case approach: (1) official local datasets (OLD) and (2) open global datasets (OGD) processed through ArcGIS Pro and VoxCity tools, respectively. Both datasets are incorporated into the City-LES model (Kusaka et al., 2024), and a high-resolution simulation is then performed. The City-LES model enables a systematic comparison of simulation outputs, which capture fine-scale atmospheric processes and are well suited to analysing the influence of urban morphology (Kusaka et al., 2024).

* Corresponding author


https://doi.org/10.5194/isprs-archives-L-4-W2-2026-245-2026 | © Author(s) 2026. CC BY 4.0 License. 

Focusing on the Lozenets district in Sofia as a representative urban case study, the research demonstrates how local and open global geospatial datasets can complement each other to support high-resolution urban climate modelling. The findings provide evidence to support the selection of the most appropriate data sources for district-scale simulations, thereby informing hybrid urban data platforms that combine heterogeneous geospatial datasets to improve the accuracy and reliability of climate-informed planning and decision-making (Petrova-Antonova & Ilieva, 2021; Vassilev et al., 2022).

This study aims to answer the following research questions:

*(1) How do City-LES simulation outputs differ between the two datasets (OLD and OGD), particularly in terms of key microclimate variables?*

*(2) To what extent can OGD provide a reliable alternative to high-resolution local datasets for representing urban morphology and supporting high-resolution urban climate simulations?*

The proposed methodology supports interoperable urban climate modelling workflows that can be applied across Sofia and extended to other cities where access to detailed official local data may be limited. The findings contribute to urban heat assessment, climate adaptation strategies, and evidence-based urban planning, while supporting the integration of open global geospatial data.

This paper is organised as follows. Section 2 describes the study area, datasets, and methodology for preparing the OLD and OGD inputs and conducting the City-LES simulations. Section 3 presents and compares the simulation results for both datasets. Finally, Section 4 concludes the paper and provides recommendations for future work.

## 2. Data and Methodology

The overall study workflow includes data collection, preprocessing, model setup, simulation analysis and verification (Figure 1).

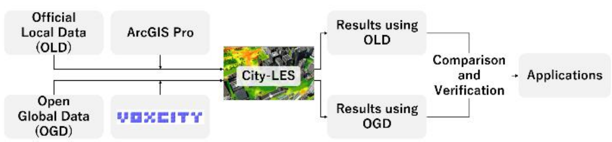


Figure 1. Study workflow.

In the application phase, the study addresses the research questions by comparing the performance of OLD and OGD datasets in high-resolution urban climate simulations of Sofia. The steps are explained in Sections 2.2-2.4.

### 2.1 Study Area

Sofia, the capital and largest city of Bulgaria, is situated in the western part of the country. Sofia covers an urbanised area of approximately 210 km² (Figure 2), with the city boundary highlighted in yellow. According to the National Statistical Institute of Bulgaria, the city had a population of 1,286,965 inhabitants in 2023 (National Statistical Institute, 2023).

The Lozenets district in Sofia is selected as a case study, indicated by the red boundary box in Figure 2. Lozenets is a well-established mixed-use urban district characterised by a diverse urban morphology comprising residential, commercial, institutional, and recreational land uses. The district contains densely built residential neighbourhoods, major transport corridors, urban parks, and tree-lined streets, resulting in a heterogeneous urban environment with varying building densities, vegetation cover, and open spaces.

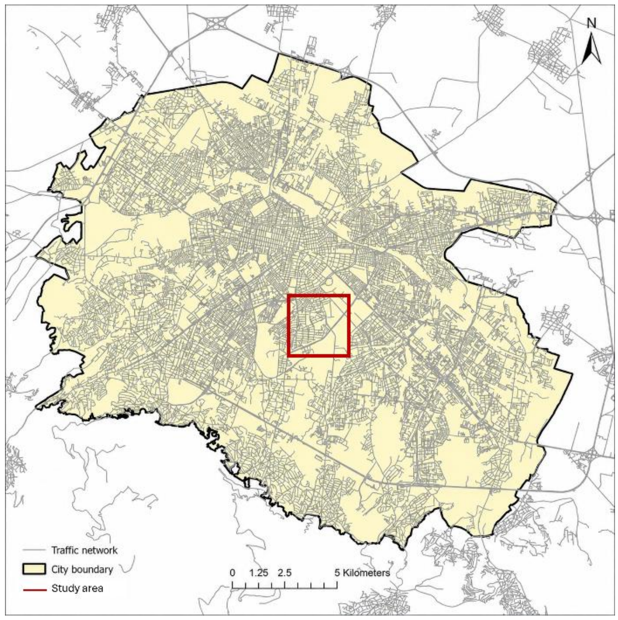


Figure 2. Sofia boundary (yellow colour) and study area (red boundary box).

The urban characteristics of the Lozenets district make it particularly suitable for evaluating the applicability of different geospatial datasets on high-resolution urban climate simulations, as it encompasses a wide range of urban forms and microclimatic conditions commonly found in European cities.

### 2.2 Data Collection and Preparation

To evaluate the applicability of geospatial data sources on urban climate simulations, two independent datasets are compiled: OLD and OGD. The datasets are selected to represent the highest-quality locally available information and globally accessible open data, respectively. Table 1 summarises the geospatial datasets used for each urban component.

| Component | OLD | OGD |
|---|---|---|
| **Land cover** | Combination of 2D vector data and orthophoto: building footprints (*Geoportal of Geodesy, Cartography and Cadastre Agency*, 2025); Street surfaces and orthophoto 10 cm/ pixel | Land cover (*OpenStreetMap*, 2026) |
| **Tree canopy** | Tree canopy derived from orthophoto and photogrammetric point cloud | High Resolution Canopy (*High Resolution 1m Global Canopy Height Maps*, 2026) |

| | | |
|---|---|---|
| **Terrain elevation** | Contour-based digital terrain model | Forest and Buildings removed Copernicus Digital Elevation Model (FABDEM) (*Forest and Buildings Removed Copernicus*, 2026) |
| **Building footprints** | Local Building data used (*Geoportal of Geodesy, Cartography and Cadastre Agency*, 2025) | Global Building data used (*OpenStreetMap*, 2026) |

Table 1. Geospatial datasets used for each component in official local data (OLD) and open global data (OGD) of Lozenets District, Sofia.

**Land cover**

For the OLD case, land cover identification was twofold. Vector data for building footprints and street surfaces helped to clearly identify corresponding land cover classes. The orthophoto with a resolution of 10 cm/pixel was used in ArcGIS Pro to classify pixels with a Support Vector Machine model and to derive pavement, grass, bare land, and asphalt areas. (Figure 3a). This approach provides a high level of spatial detail required for urban microclimate simulations.

Land cover information is obtained from OpenStreetMap (OSM) for the OGD (Figure 3b). OSM provides globally available vector data describing roads, buildings, and land-use features, allowing the automated generation of urban surface classifications suitable for large-scale applications. Although OSM provides lower thematic and spatial detail than the OLD dataset, it offers consistent global coverage and regular community-driven updates.

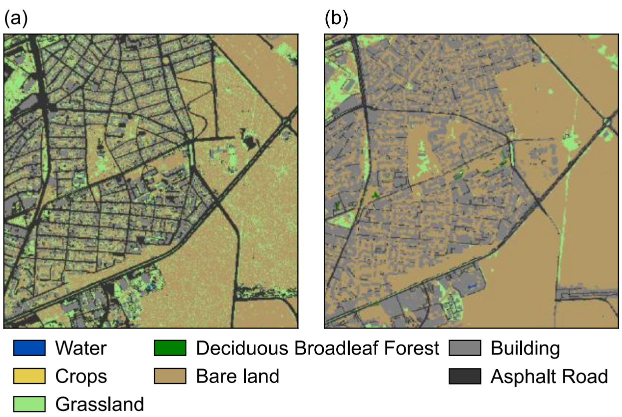


Figure 3. Differences in land cover in (a) official local data (OLD) and (b) open global data (OGD) of Lozenets District, Sofia.

**Urban Geometry and Tree Canopy**

Vegetation plays a critical role in regulating the urban microclimate by providing shading and evapotranspiration. For the OLD dataset, tree canopy information is derived from high-resolution orthophotos, enabling the identification of individual trees and accurate delineation of canopy extent (Figure 4a). Tree heights were derived from photogrammetric point clouds shared as the municipality data.

For the OGD dataset, tree canopy information is obtained from the High Resolution 1 m Global Canopy Height Maps, which provide high-resolution estimates of vegetation height derived from computer-vision-based inference applied to visible-band imagery (Figure 4b).

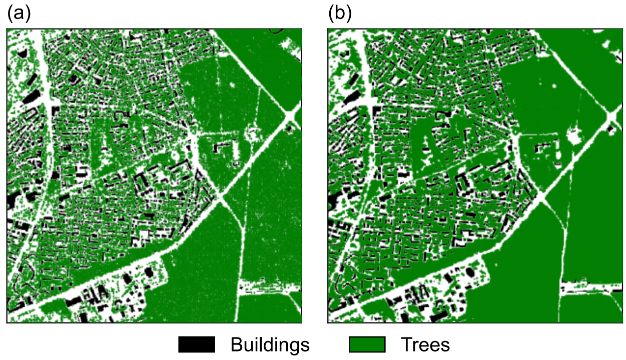


Figure 4. Differences in urban geometry and tree canopy in (a) official local data (OLD) and (b) open global data (OGD) of Lozenets District, Sofia.

**Terrain elevation**

Regarding the OLD dataset, a Digital Terrain Model (DTM) is generated from elevation contours provided as open data (Figure 5a). The contour-based model accurately represents the local topography and serves as the ground surface for the urban climate simulations.

For the OGD dataset, terrain elevation is obtained from the Forest and Buildings removed Copernicus Digital Elevation Model (FABDEM) (Figure 5b). FABDEM removes the influence of buildings and vegetation from the original Copernicus DEM, producing a bare-earth terrain model suitable for environmental and urban climate applications. Although its spatial resolution is lower than the local DTM, FABDEM provides globally consistent elevation data (Hawker et al., 2022; Tolan et al., 2024).

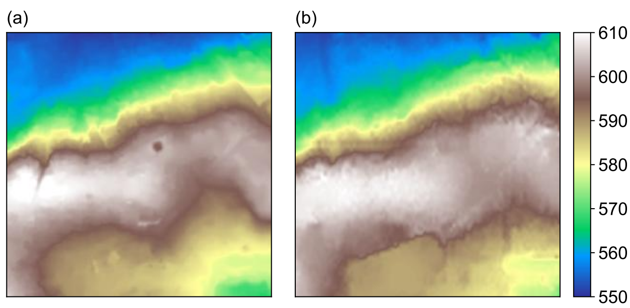


Figure 5. Differences in the elevation (m) in (a) official local data (OLD) and (b) open global data (OGD) of Lozenets District, Sofia.

**Building footprints**

Building footprints are obtained from the Geoportal of the Geodesy, Cartography, and Cadastre Agency for the OLD dataset (Figure 6a). The building heights were derived from a photogrammetric point cloud provided by . Building parts within a footprint were segmented from a rasterised point cloud using the Segment Mean Shift algorithm in ArcGIS Pro. The resulting dataset provides highly accurate representations of building footprints and urban morphology.

For the OGD dataset, building footprints are extracted from OpenStreetMap (OSM) (Figure 6b). Where required, missing building height information is supplemented through the VoxCity workflow to generate three-dimensional building models for City-LES simulations. Although the OGD dataset provides less geometric detail than the OLD dataset, it enables


https://doi.org/10.5194/isprs-archives-L-4-W2-2026-245-2026 | © Author(s) 2026. CC BY 4.0 License. 

rapid and scalable generation of urban models for cities where detailed local cadastral information is unavailable.

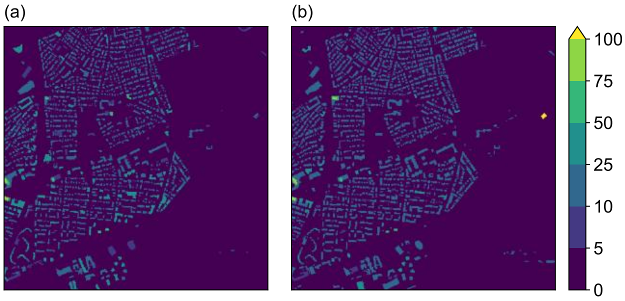


Figure 6. Differences in the building heights (m) in (a) official local data (OLD) and (b) open global data (OGD) of Lozenets District, Sofia.

Overall, the data used in the OLD and OGD are downloaded, rasterised at the target spatial resolution, and exported in a format compatible with City-LES. Both datasets are then incorporated separately into the City-LES model to simulate urban microclimate conditions.

### 2.3 Data processing

Based on the datasets and preparation described in Section 2.2, the geospatial data are processed to generate input files for City-LES.

For the OLD, the original vector and raster datasets are preprocessed using ArcGIS Pro. The datasets are harmonised in terms of coordinate reference system, spatial extent, and attribute information, clipped to the 2 km × 2 km City-LES target domain, and converted to a 5 m grid corresponding to the horizontal grid spacing of City-LES. Building footprints and building-height information are converted into gridded building-height data. Vegetation information is converted into gridded vegetation inputs, including the spatial distribution of tree-covered areas. Land cover information is classified into the land-surface categories required by City-LES, such as buildings, roads, grassland, bare land, and water surfaces.

For the OGD, the geospatial datasets are processed through the VoxCity tool. VoxCity is used to preprocess publicly available datasets, including OpenStreetMap-based urban geometry and land-use information, global canopy-height data, and global elevation data, and to export them as input files compatible with City-LES. This processing is conducted using the same target domain and horizontal grid spacing as those used in the OLD. While the overall processing workflow is similar to that executed in ArcGIS Pro, VoxCity automates the entire workflow, including data acquisition, rasterisation, and format conversion, thereby substantially improving the efficiency of data preparation.

In the simulations, terrain elevation data are not directly incorporated into the City-LES computational domain. Instead, terrain elevation is used only in the post-processing stage. Specifically, the atmospheric height of the simulation outputs is corrected using the corresponding terrain elevation at each grid point, allowing the results to be interpreted with respect to terrain-adjusted atmospheric height.

Although the case names emphasise the difference in data sources, the preprocessing workflows also differ between the two cases. The OLD is generated from local datasets using ArcGIS Pro, whereas the OGD is generated from open/global datasets using VoxCity. Therefore, the comparison in this study should be interpreted as a practical comparison between local-data-based and open-global-data-based City-LES input preparation workflows, rather than as a strict experiment that isolates the individual effects of dataset source or preprocessing software.

### 2.4 Simulation experiments

City-LES simulation experiments are conducted using the two input datasets prepared through the workflows described in Section 2.3. The same mesoscale meteorological forcing and numerical settings are applied to both cases.

The mesoscale meteorological conditions are provided by the Weather Research and Forecasting model (Skamarock et al., 2019). WRF is used to simulate regional- to urban-scale meteorological fields over Sofia, and the output from the innermost WRF domain is used as the mesoscale atmospheric forcing for City-LES. The innermost WRF domain had a horizontal grid spacing of 250 m, and the WRF simulation covered the period from 18:00 UTC on 14 July 2024 to 00:00 UTC on 22 July 2024.

The microscale simulations are conducted using City-LES version 2.0 (Kusaka et al., 2024). The City-LES domain covered a 2 km × 2 km area in the Lozenets district of Sofia, centred at 42.6739° N and 23.3306° E. The horizontal and vertical grid spacings are both set to 5 m. Buildings and trees are explicitly represented on the City-LES computational grid using the gridded input data prepared for each case. Anthropogenic heat (AH) emissions are excluded from the present experiments in order to focus on differences associated with the geospatial input data.

For each case, the City-LES simulation consisted of a surface spin-up calculation followed by a dynamic simulation. The surface spin-up is conducted from 06:00 to 19:00 EEST on 16 July 2024. During the spin-up calculation, atmospheric dynamics are not solved. Instead, near-surface meteorological variables derived from WRF, including air temperature, humidity, wind speed, downward shortwave radiation, and downward longwave radiation, are prescribed as external forcing to adjust the thermal conditions of the land surface, building surfaces, and vegetation canopy. After the surface spin-up, the dynamic simulation is initialised at 19:00 EEST and integrated for 1 h until 20:00 EEST.

The simulation date and time window are selected based on the mesoscale WRF simulation results. The 19:00–20:00 EEST period on 16 July 2024 was chosen because the WRF simulation indicated a pronounced urban heat island pattern during this period, with a clear temperature contrast between built-up and vegetated areas. This time window is therefore considered suitable for examining whether differences in geospatial input data lead to differences in fine-scale thermal variations simulated by City-LES.

The same WRF-derived initial and lateral boundary conditions are applied to both cases. Since the WRF boundary fields do not explicitly include turbulent fluctuations at the LES-resolvable scale, artificial inflow turbulence is generated using the cell perturbation method (Muñoz-Esparza et al., 2014). The same turbulence-generation settings are used in both cases, ensuring the two simulations are conducted under consistent atmospheric forcing conditions.

The simulation outputs from the two cases are compared for key microclimate variables, including near-surface air temperature, wind field, surface temperature, and spatial contrasts between built-up and vegetated areas. As discussed in Section 2.3, differences between the two cases should be interpreted as differences associated with the two practical input-data preparation workflows, rather than as the isolated effect of the data source alone.

## 3. Results

Section 3 presents the simulated microclimatic conditions for the Lozenets District in Sofia, comparing results using the OLD and OGD. The comparison focuses on four key environmental variables: 2-m air temperature, Mean Radiant Temperature (MRT), Universal Thermal Climate Index (UTCI), and 10-m wind speed and wind direction. The spatial distribution of the simulated conditions for both datasets are shown, enabling an assessment of how differences in the input geospatial data influence the representation of the urban thermal environment and pedestrian-level climatic conditions.

### 3.1 2-m Air Temperature

Figure 7 compares the simulated 2-m air temperature distributions for the OLD and OGD datasets in the Lozenets district. In both simulations, temperatures range from approximately 30.5 °C to 34.0 °C, with most areas between 32.5 °C and 33.5 °C. The highest temperatures (>33.5 °C) occur in densely built-up areas with extensive impervious surfaces, while the lowest (<31.5 °C) are found near the vegetated eastern boundary.

Both datasets reproduce similar large-scale thermal patterns, but the OLD (Figure 6a) shows greater spatial variability, with more distinct hot and cool patches due to its detailed representation of buildings and urban morphology. In contrast, the OGD (Figure 6b) produces smoother temperature gradients and more continuous warm regions, reflecting its lower spatial detail.

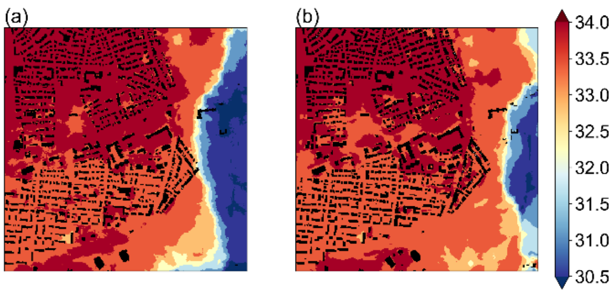


Figure 7. Differences in 2-m air temperature (°C) simulated by the City-LES model with (a) official local data (OLD) and (b) open global data (OGD) of Lozenets District, Sofia.

### 3.2 Mean Radiant Temperature

Mean Radiant Temperature (MRT) at 2m represents the combined shortwave and longwave radiation received by pedestrians and is a key indicator of outdoor thermal comfort. Higher MRT values indicate greater radiant heat exposure and increased thermal stress, particularly during hot weather.

Figure 8 shows the simulated MRT fields for the OLD and OGD datasets. MRT ranges from approximately 22 to 32 °C, with most of the study area at 26–30 °C. Higher values occur on sunlit urban surfaces, while lower values are found in areas shaded by buildings and vegetation, highlighting the strong influence of local shading and urban morphology.

The OLD simulation (Figure 8a) exhibits greater spatial variability, with more distinct low- and high-MRT zones, owing to its detailed representation of buildings and vegetation. In contrast, the OGD simulation (Figure 8b) shows smoother transitions between shaded and sunlit areas because of its simplified urban geometry. These differences are most evident along building edges, street corridors, and vegetation boundaries, where localised shading strongly influences MRT.

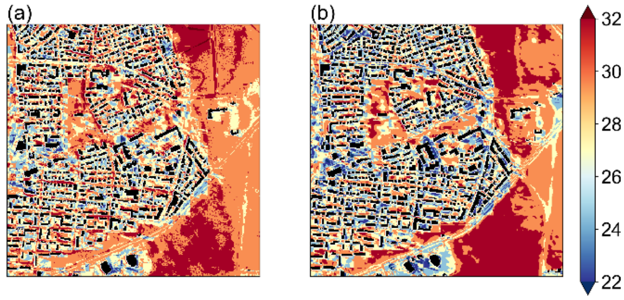


Figure 8. Differences in Mean Radiant Temperature (MRT) (°C) simulated by the City-LES model with (a) official local data (OLD) and (b) open global data (OGD) of Lozenets District, Sofia.

### 3.3 Universal Thermal Climate Index

The Universal Thermal Climate Index (UTCI) was evaluated at pedestrian level (2 m above ground), using 2 m air temperature and relative humidity, the mean radiant temperature at 2 m, and wind speed at 10 m above ground, as simulated by the City-LES model.

Figure 9 shows that UTCI values range from approximately 27.0 °C to 31.5 °C, indicating moderate to strong heat stress across the study area. Higher values (30.5–31.5°C) occur in densely urbanised areas with extensive impervious surfaces and limited vegetation, while lower values (27.0–28.5 °C) are concentrated near vegetated areas and the eastern boundary, where shading reduces thermal stress.

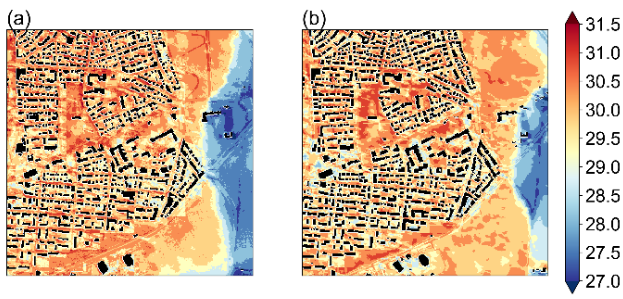


Figure 9. Differences in Universal Thermal Climate Index (UTCI) (°C) simulated by the City-LES model with (a) official local data (OLD) and (b) open global data (OGD) of Lozenets District, Sofia.

### 3.4 10-m Wind Speed and Wind Direction

Wind speed at 10 m is strongly controlled by urban form and the arrangement of buildings. Figure 10 shows that both simulations reproduce similar district-scale airflow patterns but differ at the local scale. The OGD simulation (Figure 10b) generates broader areas of moderate to high wind speeds (1.5–3.0 m $s^{-1}$), especially


https://doi.org/10.5194/isprs-archives-L-4-W2-2026-245-2026 | © Author(s) 2026. CC BY 4.0 License. 

along major streets, whereas the OLD simulation (Figure 9a) produces more localised low-wind zones (<1.0 m s$^{-1}$) within dense urban blocks.

These differences arise from the simplified building representation in the OGD dataset, which allows for more continuous airflow, whereas the higher-resolution OLD dataset captures greater spatial variability due to the complex urban geometry.

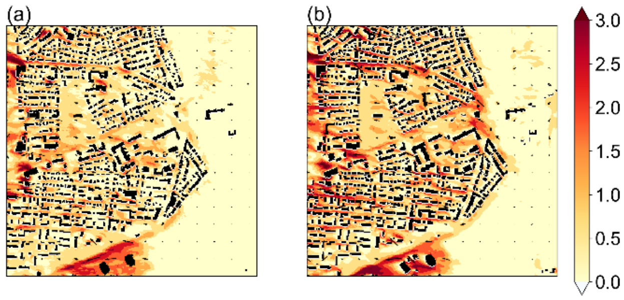


Figure 10. Differences in 10-m Wind Speed and Wind Direction (m s$^{-1}$) simulated by the City-LES model with (a) official local data (OLD) and (b) open global data (OGD) of Lozenets District, Sofia.

### 3.5 Applications

The results enable the identification of urban heat hotspots, thermally comfortable areas, and low-ventilation zones, providing valuable information for climate adaptation measures such as urban greening, street design, and heat mitigation strategies. While the simulation with OLD is better suited for detailed neighbourhood-scale analyses and urban design interventions, the simulation with OGD offers a scalable and accessible alternative for district- and city-scale assessments, particularly in regions where high-quality local datasets are unavailable. The workflow therefore supports evidence-based decision-making and can be readily applied to other cities for urban heat assessment and climate-resilient planning.

## 4. Discussion

This section compares the simulated results using OLD and OGD data sets as input, indicating similar district-scale microclimate patterns and important local-scale differences arising from variations in the representation of urban morphology (Table 2).

| Variable | OLD | OGD | Main Difference |
|---|---|---|---|
| **2-m Air Temp.** | 30.5–34.0 °C; localised hot/cool spots. | 30.5–34.0 °C; smoother temperature distribution. | OLD captures finer thermal variability through detailed urban morphology. |
| **MRT** | 22–32 °C; distinct shaded and sunlit areas. | 22–32 °C; smoother spatial transitions. | Detailed buildings and trees improve shading representation in OLD. |
| **UTCI** | 27.0–31.5 °C; stronger local variability. | 27.0–31.5 °C; similar overall pattern with reduced local contrast. | Both identify heat-stress zones, but OLD resolves thermal comfort more precisely. |
| **10-m Wind** | Localised low-wind zones (<1.0 m s$^{-1}$); complex airflow. | Broader areas of 1.5–3.0 m s$^{-1}$; smoother airflow. | The largest differences arise from the simplified building geometry in OGD, which reduces flow obstruction and street-canyon effects. |
| **Overall** | Greater spatial details; Suitable for neighbourhood-scale studies. | Reliable district-scale representation; scalable for data-scarce cities. | Both reproduce overall microclimate patterns, but OLD provides good local accuracy. |

Table 2: Microclimate patterns in official local data (OLD) and (b) open global data (OGD) of Lozenets District, Sofia.

For 2-m air temperature, the simulations using the OLD and OGD datasets produce a similar temperature range (30.5–34.0 °C), indicating that the overall thermal environment is well represented regardless of the data source. However, the simulation with OLD produces more localised hot and cool spots, whereas the simulation with the OGD generates smoother temperature gradients. These differences are primarily associated with the higher spatial resolution of the local dataset, which captures individual buildings, narrow streets, and vegetation patches more accurately. Consequently, the OLD better represents localised shading and heat storage effects that influence near-surface air temperature.

The differences become more evident for MRT and UTCI, which are highly sensitive to urban geometry and shading. Although both datasets yield comparable ranges (MRT: 22–32 °C; UTCI: 27.0–31.5 °C), the simulation with OLD exhibits sharper spatial transitions between sunlit and shaded areas. This is because the detailed building footprints and tree canopy in the OLD more accurately represent shadow patterns, while the generalised geometry in the OGD results in smoother radiation fields and less pronounced local thermal contrasts.

The largest differences are observed in the 10-m wind field. While the simulations using the OLD and OGD datasets reproduce the dominant wind direction across the study area, the wind speed distributions differ considerably. The simulation with OGD produces broader areas with wind speeds of approximately 1.5–3.0 m s$^{-1}$, whereas the simulation with OLD identifies localised low-wind zones (<1.0 m s$^{-1}$) within dense urban blocks. These differences are primarily attributed to the representation of building geometry. The simplified building shapes and street layouts in the OGD create wider and more continuous airflow pathways, allowing wind to penetrate more uniformly through the urban fabric. In contrast, the detailed building footprints in the OLD better capture complex street canyons, irregular building alignments, and local obstructions that generate flow separation, recirculation zones, and wind sheltering. As a result, the OLD provides a more realistic representation of pedestrian-level airflow, particularly in densely built environments.

In conclusion, the results demonstrate that the choice of geospatial dataset has a relatively small influence on the simulation of district-scale thermal patterns but a much greater impact on variables that are strongly controlled by urban morphology, particularly wind speed and, to a lesser extent, radiant temperature and thermal comfort. These findings suggest that open global datasets are suitable for city- or district-scale climate assessments, whereas high-resolution local datasets are essential for detailed pedestrian-level analyses of ventilation, thermal comfort, and urban design interventions.

## 5. Conclusion and Future Work

This section explains the main conclusions, limitations and proposed future work as follows.

### 5.1 Main Conclusions

The study compared the performance of OLD and OGD for high-resolution urban microclimate simulations in the Lozenets district of Sofia using the City-LES model. Both datasets successfully reproduced the overall spatial patterns of the urban microclimate, while differences were primarily observed in the representation of fine-scale urban features.

The main conclusions can be summarised as follows:

(1) The simulated 2-m air temperature ranged from 30.5–34.0 °C, with the highest temperatures (>33.5 °C) occurring in densely built-up areas and the lowest (<31.5 °C) in vegetated zones.
(2) MRT ranged from 22 to 32 °C, with most areas between 26–30 °C, while UTCI varied from 27.0–31.5 °C, indicating moderate to strong heat stress, with the highest values (30.5–31.5 °C) concentrated over impervious urban surfaces.
(3) The simulations using the OLD and OGD datasets reproduced broadly similar district-scale wind patterns. However, some differences were observed in the wind-speed distributions: the simulation using the OGD produced more extensive areas with wind speeds between 1.5 and 3.0 m $s^{-1}$, whereas the simulation using the OLD resolved more localised low-wind zones below 1.0 m $s^{-1}$ within densely built-up blocks.
(4) In conclusion, the simulation with the OLD provided a more detailed representation of localised thermal and airflow variability owing to its higher-resolution urban morphology, whereas the simulation with the OGD captured the dominant district-scale microclimate patterns.

These findings indicate that open global datasets offer a practical and scalable alternative for urban climate simulations when detailed local data are unavailable, whereas local datasets remain preferable for neighbourhood-scale analyses that require higher spatial precision.

### 5.2 Limitations and Future Work

First, the analysis was limited to a single urban district and a single summertime heat event and, therefore, may not capture seasonal variability or the influence of different urban forms under varying meteorological conditions. Second, AH emissions are excluded from the present experiments to focus on differences associated with the geospatial input data.

Future work should extend the proposed workflow to different seasons, additional districts in Sofia, and cities with diverse urban characteristics. Investigating hybrid approaches that integrate local and open geospatial datasets, together with other urban processes, would further improve the accuracy, scalability, and applicability of high-resolution urban climate modelling for digital twins and climate-resilient urban planning.

## Acknowledgements

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

Moreover, this work was partly supported by MEXT Promotion of Development of a Joint Usage/Research System Project: Coalition of Universities for Research Excellence Program (CURE) Grant Number JPMXP1323015474. This work was partly supported by JST SPRING, Grant number JPMJSP2124. This research used computational resources of Miyabi, provided by the Multidisciplinary Cooperative Research Program in Center for Computational Sciences, University of Tsukuba.

## References

Assenova, I. A., Vitanova, L. L., Petrova-Antonova, D., 2024: Urban heat islands from multiple perspectives: Trends across disciplines and interrelationships. *Urban Climate*, 56, 102075. doi.org/10.1016/J.UCLIM.2024.102075.

Chen, J., Zhao, F., Zeng, N., Oda, T., 2020: Comparing a global high-resolution downscaled fossil fuel CO2 emission dataset to local inventory-based estimates over 14 global cities. *Carbon Balance and Management*, 15(1), 9. doi.org/10.1186/s13021-020-00146-3.

Forest and Buildings removed Copernicus, 2026: https://gee-community-catalog.org/projects/fabdem/ (26 August 2026).

Fujiwara, K., Tsurumi, R., Kiyono, T., Fan, Z., Liang, X., Lei, B., Yap, W., Ito, K., Biljecki, F., 2026: VoxCity: A seamless framework for open geospatial data integration, grid-based semantic 3D city model generation, and urban environment simulation. *Computers, Environment and Urban Systems*, 123, 102366. doi.org/10.1016/j.compenvurbsys.2025.102366.

Geoportal of Geodesy, Cartography and Cadastre Agency, 2025: https://kais.cadastre.bg/en/OpenData (26 August 2026).

Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., Neal, J., 2022: A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, 17(2), 024016. doi.org/10.1088/1748-9326/ac4d4f.

High Resolution 1m Global Canopy Height Maps, 2026: https://gee-community-catalog.org/projects/meta_trees/ (26 August 2026).

Koleva, T., Vitanova, L. L., Petrova-Antonova, D., Kostadinov, A., 2024: Semantic Enrichment and Analysis of Building Energy Consumption Data for the City of Sofia. *Artificial Intelligence Applications and Innovations. AIAI 2024 IFIP WG 12.5 International Workshops*, 462–475. doi.org/10.1007/978-3-031-63227-3_33.

Kusaka, H., Ikeda, R., Sato, T., Iizuka, S., Boku, T., 2024: Development of a Multi-Scale Meteorological Large-Eddy Simulation Model for Urban Thermal Environmental Studies: The “City-LES” Model Version 2.0. *Journal of Advances in Modeling Earth Systems*, 16(10). doi.org/10.1029/2024MS004367.

Lloyd, C. T., Chamberlain, H., Kerr, D., Yetman, G., Pistolesi, L., Stevens, F. R., Gaughan, A. E., Nieves, J. J., Hornby, G., MacManus, K., Sinha, P., Bondarenko, M., Sorichetta, A., Tatem,

A. J., 2019: Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data*, 3(2), 108–139. doi.org/10.1080/20964471.2019.1625151.

Muñoz-Esparza, D., Kosović, B., Mirocha, J., van Beeck, J., 2014: Bridging the Transition from Mesoscale to Microscale Turbulence in Numerical Weather Prediction Models. *Boundary-Layer Meteorology*, 153(3), 409–440. doi.org/10.1007/s10546-014-9956-9.

National Statistical Institute, 2023: *Population and Demographic Processes in 2023*. https://www.nsi.bg/sites/default/files/files/pressreleases/Population2023_en_ZYBLHGJ.pdf (26 August 2026).

Opdyke, A., Fatima, K., 2024: Comparing the suitability of global gridded population datasets for local landslide risk assessments. *Natural Hazards*, 120(3), 2415–2432. doi.org/10.1007/s11069-023-06283-5.

OpenStreetMap, 2026: https://www.openstreetmap.org/ (26 August 2026).

Pavlov, D., Vitanova, L. L., Petrova-Antonova, D., 2025: Past Expansion and Future Prediction of Land Use and Land Cover of Sofia City. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-G-2025, 1159–1166. doi.org/10.5194/isprs-archives-XLVIII-G-2025-1159-2025.

Petrova-Antonova, D., Ilieva, S., 2021: Digital twin modeling of smart cities. *Advances in Intelligent Systems and Computing*, 1253 AISC, 384–390. doi.org/10.1007/978-3-030-55307-4_58.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., 2019: A description of the Advanced Research WRF model version 4. *NCAR Tech. Note NCAR/TN-556+STR*, 145. doi.org/10.5065/1dfh-6p97.

Subasinghe, S., Wang, R., Simwanda, M., Murayama, Y., Vitanova, L. L., 2022: Neighborhood dynamics of urban expansion based on morphological spatial pattern analysis and geospatial techniques: a case study of the Colombo metropolitan area, Sri Lanka. *Asian Geographer*, 39(2), 155–175. doi.org/10.1080/10225706.2021.1903519.

Tolan, J., Yang, H.-I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., Spore, J., Majumdar, S., Haziza, D., Vamaraju, J., Moutakanni, T., Bojanowski, P., Johns, T., White, B., Tiecke, T., Couprie, C., 2024: Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sensing of Environment*, 300, 113888. doi.org/10.1016/j.rse.2023.113888.

Vassilev, V., Ilieva, S., Krasteva, I., Pavlova, I., Petrova-Antonova, D., Sowinski-Mydlarz, W., 2022: AI-Based Hybrid Data Platforms. In: *Data Spaces*, 147–170. Springer International Publishing. doi.org/10.1007/978-3-030-98636-0_8.

Vitanova, L. L., Petrova-Antonova, D., Hristov, P. O., Shirinyan, E., 2023: TOWARDS ENERGY ATLAS OF SOFIA CITY IN BULGARIA. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W2-2023, 123–129. doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-123-2023.

Vitanova, L. L., Petrova-Antonova, D., Shirinyan, E., 2025: Urban digital twin for assessing and understanding urban Heat Island impacts. *Urban Climate*, 62, 102530. doi.org/10.1016/J.UCLIM.2025.102530.