Detection and Simulation of Urban Heat Islands Using a Fine-Tuned Geospatial Foundation Model for Microclimate Impact Prediction

Jannis Fleckenstein*
IBM Research – Europe
Jannis.Fleckenstein@ibm.com

David Kreismann*
IBM Research – Europe
David.Kreismann@ibm.com

Tamara Rosemary Govindasamy
IBM Research – Africa
tamara.govindasamy@ibm.com

Thomas Brunschwiler
IBM Research – Europe
tbr@zurich.ibm.com

Etienne Vos
IBM Research – Africa
etienne.vos@ibm.com

Mattia Rigotti
IBM Research – Europe
mrg@zurich.ibm.com

Abstract

As urbanization and climate change progress, urban heat island effects are becoming more frequent and severe. To formulate effective mitigation plans, cities require detailed air temperature data, yet conventional machine learning models with limited data often produce inaccurate predictions, particularly in underserved areas. Geospatial foundation models trained on global unstructured data offer a promising alternative by demonstrating strong generalization and requiring only minimal fine-tuning. In this study, an empirical ground truth of urban heat patterns is established by quantifying cooling effects from green spaces and benchmarking them against model predictions to evaluate the model's accuracy. The foundation model is subsequently fine-tuned to predict land surface temperatures under future climate scenarios, and its practical value is demonstrated through a simulated in-painting that highlights its role for mitigation support. The results indicate that foundation models offer a powerful way for evaluating urban heat island mitigation strategies in data-scarce regions to support more climate-resilient cities.

1 Introduction

As cities grow, increasing building and population densities drive land use changes that impact the local energy balance and shape the microclimate conditions of urban areas [1]. With over 70% of the global population projected to live in cities by 2050, the urban heat island (UHI) effect has become a major concern [2, 3]. Factors such as heat-retaining materials, vegetation loss and dense structures can cause cities to be over 5 °C warmer than their surrounding areas [4]. These temperature increases lead to greater energy consumption, more heat-related health issues, and declining air quality [5, 6]. One major challenge in addressing UHI impacts is the lack of high-resolution, timely air temperature data, which is essential for early warnings and effective heat risk mitigation plans [7, 8]. Current forecasting approaches often require large amounts of input data, specialized expertise, and significant computational resources, making real-time forecasting at fine scales largely inaccessible [9, 10].

*Equal contribution.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2025.

Data limitations from sparse stations and satellite revisit intervals further constrain accuracy [11, 12]. While artificial intelligence (AI)-based models offer increased flexibility, they are often limited by their dependence on large labeled datasets and poor generalization across domains [13].

In response to these limitations, recent developments in geospatial foundation models (GFMs), show strong generalization across spatial resolutions and regions with minimal fine-tuning, making them a promising alternative for urban climate analysis [14]. Building on this potential, this study explores the use of a fine-tuned GFM by establishing an empirical ground truth of how urban green spaces affect temperature through internal and spillover cooling in UHI areas [15, 16]. The model is validated against these cooling effects to assess its ability to replicate observed processes. It is then used to forecast urban temperatures under future climate scenarios and to simulate targeted in-painting as a strategy to mitigate UHI impacts. The result is a unified workflow that integrates data analysis, forecasting, and simulation to inform climate-resilient urban planning.

2 Methodology

GFMs have demonstrated strong performance in geospatial tasks, with promising generalization across space and time, making them suitable for UHI applications in the context of increasing urbanization and climate-induced temperature extremes [17]. However, existing studies rarely conduct in-depth evaluations of physical realism for GFMs, such as examining how urban green spaces influence modelled UHI dynamics through simulations like inpainting or future-scenario forecasting, which are essential for informed urban heat mitigation and planning [18]. To address this gap, this study follows a three-phase experimental workflow.

First, an empirical baseline of how green spaces influence temperature is established by overlaying land-use/land-cover (LULC) maps with high-resolution land surface temperature (LST) imagery to extract park areas and analyse how urban temperature gradients evolve with distance from these green spaces, thereby quantifying internal cooling within parks and spillover cooling into surrounding built-up areas, with all results aggregated over summer daytime data from 2017 to 2025 across all cities, and assessing how well the GFM is able to replicate these effects for mitigation planning.
Second, the model's extrapolation ability under future climate conditions is assessed through a focused experiment in Brașov, Romania, a city intentionally selected because it was not part of the original training set. Although GFMs are designed to generalize across regions, isolating a single unseen city allows for a clearer evaluation of performance and highlights the feasibility of future climate projections. In this setting, the data are ordered by temperature and the model is fine-tuned only on the cooler 90% of observations, while the hottest 10% are withheld to test generalization to extreme heat in Brașov. The difference between predicted and observed values in this upper percentile serves as an estimate of how far the model can reliably extrapolate beyond its training range. Under the assumption that this 10% threshold represents its extrapolation limit, the model is then used to project future UHI conditions by replacing ERA5 with EURO-CORDEX climate inputs under different future Representative Concentration Pathway (RCP) scenarios.
Third, a hypothetical urban greening intervention is simulated by guided inpainting of the satellite image, replacing built-up pixels with those representing urban green areas. The input spectral indices are therefore adjusted accordingly, and the model is run to predict the resulting thermal impact to support simulated UHI mitigation planning.

2.1 Training procedure and geospatial foundation model

Building on the work of Bhamjee et al. [19], the Granite-Geospatial-Land-Surface-Temperature (Granite-GFM) model [20] is used as a GFM for predicting LST at high spatial and temporal resolution. The model is evaluated in two configurations, V1 which is fine-tuned using Harmonised Landsat Sentinel-2 (HLS L30 [21]) imagery and ERA5-Land [22] 2 m air temperature statistics from 2013–2023 across 28 cities, and V2 which is fine-tuned on an extended dataset covering 52 cities across a wider range of hydro-climatic zones. Granite-GFM incorporates a Shifted Window (SWIN) Transformer architecture and builds on the Prithvi-SWIN-L Earth Observation foundation model [23, 24, 17]. While state of the art approaches for LST prediction focus on single regions of interest, the Granite-GFM model enables the estimation of LST at a 30m spatial and hourly temporal scale, for any city of interest using fewer input samples.

2.2 Study regions and data

Our analysis covers thirteen European cities that were not part of the training data, selected to test the model's out-of-sample generalization across diverse urban environments. A regional map of the selected cities is provided in Appendix 5. Twelve of these cities, marked in red, are used in the first workflow step to assess internal and spillover cooling effects across varying hydro-climatic settings. For the second step, Brașov (marked in blue) is used as an unseen evaluation city. In the third intervention step, Prague is selected for in-painting. While these cities are highlighted, the workflow can in principle be applied to any urban area, which reflects the core advantage of geospatial foundation models. The methodology is based on four primary data sources. (1) Harmonised Landsat Sentinel-2 imagery providing 30 m multispectral bands to derive LST using a split-window algorithm [21, 25], (2) Impact Observatory 10 m LULC data for identifying urban and green spaces [26], (3) ERA5-Land reanalysis providing continuous atmospheric context, with near-surface air temperature stacked onto the HLS imagery [22], (4) EURO-CORDEX regional climate projections providing future forcing under multiple RCPs to represent different greenhouse-gas emission trajectories [27].

3 Experiments and results

This section presents the experimental evaluation of our three-phase workflow with Model V1 and Model V2. The model's predictive fidelity is assessed through the cooling anomaly (ΔT), defined as the temperature difference between a point of interest and the surrounding built-up baseline. This isolates cooling effects from background variation and measures how well the model captures green space cooling. The difference between ground truth and predicted anomalies is quantified using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Bias Error (MBE). Forecasting performance is separately evaluated based on the model's ability to extrapolate to extreme conditions, using MAE, Mean Square Error (MSE), and RMSE for the LST of the hottest 10% of unseen data.

3.1 Ground truth internal and spillover cooling patterns

The ground truth analysis reveals clear thermal patterns shaped by urban green spaces. At the macro-scale, the urban gradient shows an average UHI of $+3.3°\mathrm{C}$ in dense cores, driven by the lack of vegetation, as can be seen in Figure 1a. The heat source-sink analysis reinforces this link, with all tree cover located in the city's coolest areas, as shown in Figure 1b. At the micro-scale, parks exhibit strong internal cooling, with temperatures dropping up to $-2.6°\mathrm{C}$ within 200 meters of the edge. Beyond their boundaries, spillover cooling lowers nearby temperatures by as much as $-3.5°\mathrm{C}$ close to parks, decreasing to about $-1°\mathrm{C}$ at 150 meters distance.

(a) Urban gradient experiment

(b) Heat source-sink experiment

Figure 1: Baseline UHI characterization for Prague based on gradient and source–sink experiments.

3.2 Model performance evaluation and future climate forecasting

Although both models achieve high accuracy in replicating the internal cooling gradient, model V2 exhibits a substantial improvement in representing the more complex spillover cooling phenomenon. Specifically, the MAE decreases from $0.302°\mathrm{C}$ in V1 to $0.199°\mathrm{C}$ in V2. As shown in Figure 2 and summarized in Table 1 in the Appendix, this demonstrates a statistically significant enhancement in performance attributable to the more diverse training of model V2.

Furthermore, the model shows strong extrapolation performance, achieving an MAE of 1.74 °C on the upper 10% of unseen temperature data using model V2. This improved accuracy, particularly in extrapolating to extreme conditions, provides strong confidence in its predictive capabilities

(a) Internal cooling gradient

(b) Spillover cooling gradient

Figure 2: Comparison of ground truth cooling phenomena with model predictions.

for forecasting and simulation tasks. The climate projections in Figure 3 indicate a substantial intensification of urban heat islands under RCP 4.5 and 8.5 and show the development of UHI in Brașov for 2030, 2050, and 2100 under RCP 2.6, 4.5, and 8.5, based on EURO-CORDEX data.

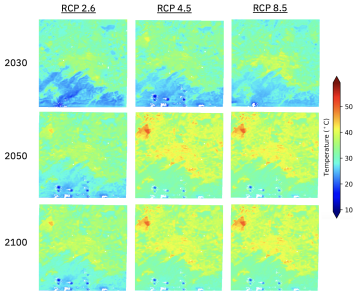


Figure 3: Projected UHI extent under RCP 2.6, 4.5, and 8.5 for 2030, 2050, and 2100 in Brașov.

3.3 Urban greening simulation through inpainting

After establishing the model's capacity to replicate real-world cooling effects and forecast scenarios with high accuracy, the effectiveness of Model V2 is demonstrated through an urban greening scenario. The simulation provides a "before and after" comparison of the intervention area across RGB-, normalized difference vegetation index (NDVI)-, and LST imagery, as can be seen in Figure 4, together with a transect showing the LST profile through the intervention area, confirming the land cover change and its cooling effect. Building on this, the simulation further quantifies the internal cooling within the new park and the spillover into the surrounding neighborhood, as shown in Figure 6 in the Appendix, demonstrating the model's ability to provide evidence-based metrics for urban planning decisions.

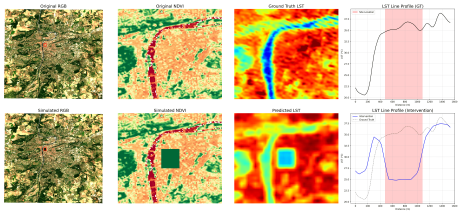


Figure 4: Comparison of inpainting results against ground truth data in Prague.

4 Conclusion

This paper demonstrates a comprehensive workflow using a fine-tuned GFM to analyze, forecast, and simulate mitigation strategies for UHI. We have shown that by first establishing an empirical ground truth, we can effectively validate a GFM's ability to replicate complex thermal phenomena. The validated model can then be applied to forecast future climate scenarios and to perform in-painting for assessing the impact of potential interventions. This approach transforms the GFM from an assessment tool into an interactive simulation platform for building climate-resilient cities.

Acknowledgments and Disclosure of Funding

This work was supported by funding from the European Union's Horizon 2020 research and innovation programme for the project NATURE-DEMO under grant agreement No. 101157448, and by the Swiss State Secretariat for Education, Research and Innovation (SERI) contract 24.00279 (REF-1131-52105).

References

[1] Atik Nurwanda and Tsuyoshi Honjo. The prediction of city expansion and land surface temperature in bogor city, indonesia. *Sustainable Cities and Society*, 52:101772, 2020. ISSN 2210-6707. doi: https://doi.org/10.1016/j.scs.2019.101772. URL https://www.sciencedirect.com/science/article/pii/S221067071832540X.

[2] Jinyao Lin, Suixuan Qiu, Xiujuan Tan, and Yaye Zhuang. Measuring the relationship between morphological spatial pattern of green space and urban heat island using machine learning methods. *Building and Environment*, 228:109910, 2023. ISSN 0360-1323. doi: https://doi.org/10.1016/j.buildenv.2022.109910. URL https://www.sciencedirect.com/science/article/pii/S0360132322011404.

[3] Awais Piracha and Muhammad Tariq Chaudhary. Urban air pollution, urban heat island and human health: A review of the literature. *Sustainability*, 14(15), 2022. ISSN 2071-1050. doi: 10.3390/su14159234. URL https://www.mdpi.com/2071-1050/14/15/9234.

[4] M. Santamouri, D. Asimakopoulos, V.D. Assimakopoulos, N. Chrisomallidou, N. Klitsikas, D. Mangold, P. Michel, Mat Santamouris, and Aris Tsangrassoulis. *Energy and Climate in the Urban Built Environment*. June 2013. ISBN 978-1-134-25790-4. doi: 10.4324/9781315073774.

[5] M. Santamouris, C. Cartalis, A. Synnefa, and D. Kolokotsa. On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—A review. *Energy and Buildings*, 98:119–124, 2015. ISSN 0378-7788. doi: 10.1016/j.enbuild.2014.09.052.

[6] Jianguo Tan, Youfei Zheng, Xu Tang, Changyi Guo, Liping Li, Guixiang Song, Xinrong Zhen, Dong Yuan, Adam J. Kalkstein, Furong Li, and Heng Chen. The urban heat island and its impact on heat waves and human health in Shanghai. *International Journal of Biometeorology*, 54(1):75–84, January 2010. ISSN 1432-1254. doi: 10.1007/s00484-009-0256-x.

[7] Qian Cao, He Huang, Yuning Hong, Xin Huang, Shaoqiang Wang, Lizhe Wang, and Lunche Wang. Modeling intra-urban differences in thermal environments and heat stress based on local climate zones in central Wuhan. *Building and Environment*, 225:109625, November 2022. ISSN 0360-1323. doi: 10.1016/j.buildenv.2022.109625.

[8] Ioannis Kousis, Ilaria Pigliautile, and Anna Laura Pisello. Intra-urban microclimate investigation in urban heat island through a novel mobile monitoring system. *Scientific Reports*, 11(1):9732, May 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-88344-y.

[9] Koen De Ridder, Dirk Lauwaet, and Bino Maiheu. UrbClim – A fast urban boundary layer climate model. *Urban Climate*, 12:21–48, June 2015. ISSN 2212-0955. doi: 10.1016/j.uclim.2015.01.001.

[10] Jordan G. Powers, Joseph B. Klemp, William C. Skamarock, Christopher A. Davis, Jimy Dudhia, David O. Gill, Janice L. Coen, David J. Gochis, Ravan Ahmadov, Steven E. Peckham, Georg A. Grell, John Michalakes, Samuel Trahan, Stanley G. Benjamin, Curtis R. Alexander, Geoffrey J. Dimego, Wei Wang, Craig S. Schwartz, Glen S. Romine, Zhiquan Liu, Chris Snyder, Fei Chen, Michael J. Barlage, Wei Yu, and Michael G. Duda. The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions. *Bulletin of the American Meteorological Society*, 98(8):1717–1737, August 2017. doi: 10.1175/BAMS-D-15-00308.1.

[11] Mikhail Varentsov, Mikhail Krinitskiy, and Victor Stepanenko. Machine Learning for Simulation of Urban Heat Island Dynamics Based on Large-Scale Meteorological Conditions. *Climate*, 11(10), 2023. ISSN 2225-1154. doi: 10.3390/cli11100200.

[12] Qiang Zhao, Le Yu, Zhenrong Du, Dailiang Peng, Pengyu Hao, Yongguang Zhang, and Peng Gong. An Overview of the Applications of Earth Observation Satellite Data: Impacts and Future Trends. *Remote Sensing*, 14(8), 2022. ISSN 2072-4292. doi: 10.3390/rs14081863.

[13] Abolfazl Farahani, Sahar Voghoei, Khaled Rasheed, and Hamid R. Arabnia. A brief review of domain adaptation. In Robert Stahlbock, Gary M. Weiss, Mahmoud Abou-Nasr, Cheng-Ying Yang, Hamid R. Arabnia, and Leonidas Deligiannidis, editors, *Advances in Data Science and Information Engineering*, pages 877–894, Cham, 2021. Springer International Publishing. ISBN 978-3-030-71704-9.

[14] Gengchen Mai, Weiming Huang, Jin Sun, Suhang Song, Deepak Mishra, Ninghao Liu, Song Gao, Tianming Liu, Gao Cong, Yingjie Hu, Chris Cundy, Ziyuan Li, Rui Zhu, and Ni Lao. On the opportunities and challenges of foundation models for geospatial artificial intelligence, 2023. URL https://arxiv.org/abs/2304.06798.

[15] Maryam Norouzi, Hing-Wah Chau, and Elmira Jamei. Design and site-related factors impacting the cooling performance of urban parks in different climate zones: A systematic review. *Land*, 13(12), 2024. ISSN 2073-445X. doi: 10.3390/land13122175. URL https://www.mdpi.com/2073-445X/13/12/2175.

[16] Yu Gao, Huifeng Pan, and Li Tian. Analysis of the spillover characteristics of cooling effect in an urban park: A case study in zhengzhou city. *Frontiers in Earth Science*, Volume 11 - 2023, 2023. ISSN 2296-6463. doi: 10.3389/feart.2023.1133901. URL https://www.frontiersin.org/journals/earth-science/articles/10.3389/feart.2023.1133901.

[17] Johannes Jakubik, Sujit Roy, C. E. Phillips, Paolo Fraccaro, Denys Godwin, Bianca Zadrozny, Daniela Szwarcman, Carlos Gomes, Gabby Nyirjesy, Blair Edwards, Daiki Kimura, Naomi Simumba, Linsong Chu, S. Karthik Mukkavilli, Devyani Lambhate, Kamal Das, Ranjini Bangalore, Dario Oliveira, Michal Muszynski, Kumar Ankur, Muthukumaran Ramasubramanian, Iksha Gurung, Sam Khallaghi, Hanxi, Li, Michael Cecil, Maryam Ahmadi, Fatemeh Kordi, Hamed Alemohammad, Manil Maskey, Raghu Ganti, Kommy Weldemariam, and Rahul Ramachandran. Foundation models for generalist geospatial artificial intelligence, 2023. URL https://arxiv.org/abs/2310.18660.

[18] Shun Fu, Lufeng Wang, Umer Khalil, Ali Hassan Cheema, Israr Ullah, Bilal Aslam, Aqil Tariq, Muhammad Aslam, and Saad S. Alarifi. Prediction of surface urban heat island based on predicted consequences of urban sprawl using deep learning: A way forward for a sustainable environment. *Physics and Chemistry of the Earth, Parts A/B/C*, 135:103682, October 2024. ISSN 1474-7065. doi: 10.1016/j.pce.2024.103682.

[19] Muaaz Bhamjee, Hiyam Debary, Zaheed Gaffoor, Tamara Govindasamy, Craig Mahlasi, Mustansar Fiaz, Etienne Vos, Levente Klein, Sibusisiwe Makhanya, Campbell Watson, and Julian Kuehnert. Detection and characterization of urban heat islands with machine learning. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 1693–1699, 2024. doi: 10.1109/IGARSS53475.2024.10641750.

[20] Muaaz Bhamjee, Zaheed Gaffoor, Tamara Govindasamy, Craig Mahlasi, Etienne Vos, Mangaliso Mngomezulu, Gciniwe Baloyi, and Sibusisiwe Makhanya. Hugging Face: Granite Geospatial Land Surface Temperature. https://huggingface.co/ibm-granite/granite-geospatial-land-surface-temperature.

[21] J. Masek, J. Ju, J.-C. Roger, S. Skakun, E. Vermote, M. Claverie, J. Dungan, Z. Yin, B. Freitag, and C. Justice. Hls sentinel-2 multi-spectral instrument surface reflectance daily global 30m v2.0, 2021. URL https://doi.org/10.5067/HLS/HLSS30.002. Accessed: 2025-08-19.

[22] Joaquín Muñoz Sabater. ERA5-land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2019. URL https://doi.org/10.24381/cds.e2161bac. Accessed: 2025-04-26.

[23] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022, October 2021.

[24] Johannes Jakubik, Linsong Chu, Paolo Fraccaro, Carlos Gomes, Gabby Nyirjesy, Ranjini Bangalore, Devyani Lambhate, Kamal Das, Dario Oliveira Borges, Daiki Kimura, Naomi Simumba, Daniela Szwarcman, Michal Muszynski, Kommy Weldemariam, Bianca Zadrozny, Raghu Ganti, Carlos Costa, Campbell Edwards, Blair Watson, Karthik Mukkavilli, Hendrik Schmude, Johannes Hamann, Parkin Robert, Sujit Roy, Christopher Phillips, Kumar Ankur, Muthukumaran Ramasubramanian, Iksha Gurung, Wei Ji Leong, Ryan Avery, Rahul Ramachandran, Manil Maskey, Pontus Olofossen, Elizabeth Fancher, Tsengdar Lee, Kevin Murphy, Dan Duffy, Mike Little, Hamed Alemohammad, Michael Cecil, Steve Li, Sam Khallaghi, Denys Godwin, Maryam Ahmadi, Fatemeh Kordi, Bertrand Saux, Neal Pastick, Peter Doucette, Rylie Fleckenstein, Dalton Luanga, Alex Corvin, and Erwan Granger. Prithvi-EO-1.0-100M, August 2023. URL https://huggingface.co/ibm-nasa-geospatial/Prithvi-EO-1.0-100M.

[25] Chen Du, Huazhong Ren, Qiming Qin, Jinjie Meng, and Shaohua Zhao. A Practical Split-Window Algorithm for Estimating Land Surface Temperature from Landsat 8 Data. *Remote Sensing*, 7(1):647–665, 2015. ISSN 2072-4292. doi: 10.3390/rs70100647.

[26] Krishna Karra, Caitlin Kontgis, Zoe Statman-Weil, Joseph C Mazzariello, Mark Mathis, and Steven P Brumby. Global land use/land cover with sentinel 2 and deep learning. In *2021 IEEE international geoscience and remote sensing symposium IGARSS*, pages 4704–4707. IEEE, 2021.

[27] Daniela Jacob, Juliane Petersen, Bastian Eggert, Antoinette Alias, Ole Bøssing Christensen, Laurens M. Bouwer, Alain Braun, Augustin Colette, Michel Déqué, Goran Georgievski, Elena Georgopoulou, Andreas Gobiet, Laurent Menut, Grigory Nikulin, Andreas Haensler, Nils Hempelmann, Colin Jones, Klaus Keuler, Sari Kovats, Nico Kröner, Sven Kotlarski, Arne Kriegsmann, Eric Martin, Erik van Meijgaard, Christopher Moseley, Susanne Pfeifer, Swantje Preuschmann, Christine Radermacher, Kai Radtke, Diana Rechid, Mark Rounsevell, Patrick Samuelsson, Samuel Somot, Jean-Francois Soussana, Claas Teichmann, Riccardo Valentini, Robert Vautard, Björn Weber, and Pascal Yiou. EURO-CORDEX: New high-resolution climate change projections for European impact research. *Regional Environmental Change*, 14(2):563–578, April 2014. ISSN 1436-378X. doi: 10.1007/s10113-013-0499-2.

A Appendix

A.1 Overview of selected study regions

For the experiments, twelve European cities were selected as study regions. For each city, Harmonized Landsat Sentinel-2 (HLS) imagery, ERA5-Land atmospheric reanalysis data, ESRI land use/land cover classifications, and EURO-CORDEX climate scenario data were downloaded and processed, covering the historical period from 2017 to 2025 and future projections for 2030, 2050, and 2100. Together, these datasets ensure consistent coverage, high resolution, and comparability across cities, supporting both ground-truth analysis and predictive modeling.

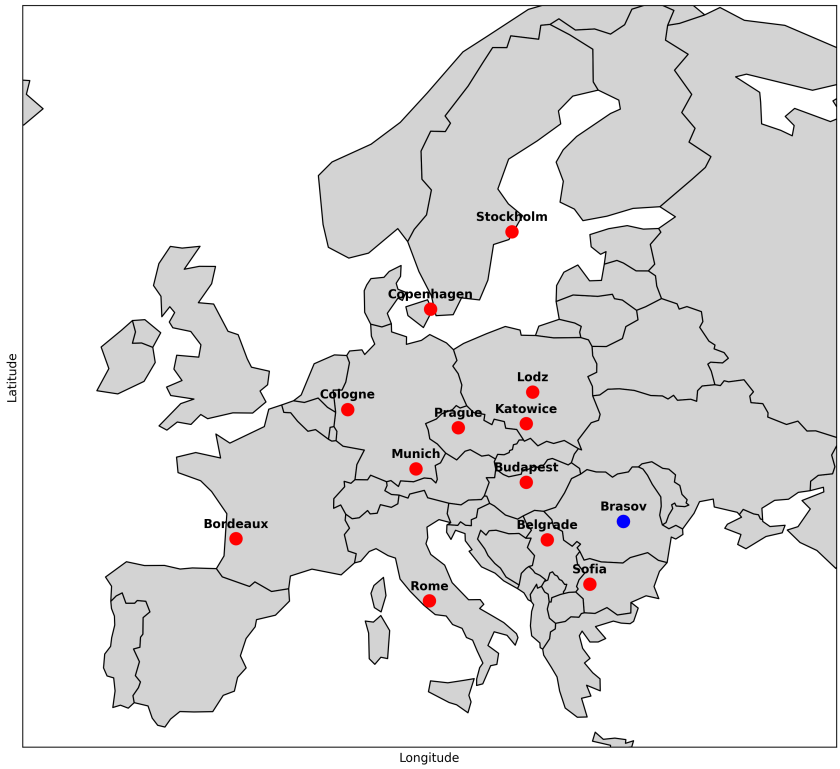


Figure 5: Map of the twelve European study cities (red dots), selected for their diverse urban characteristics, hydro-climatic settings, and suitability within the HLS tiling system, which were further analyzed for their internal and spillover cooling effects. Brașov (blue dot) was chosen as an extrapolation site to test model generalization.

A.2 Evaluation of urban green area cooling effects using a geospatial foundation model

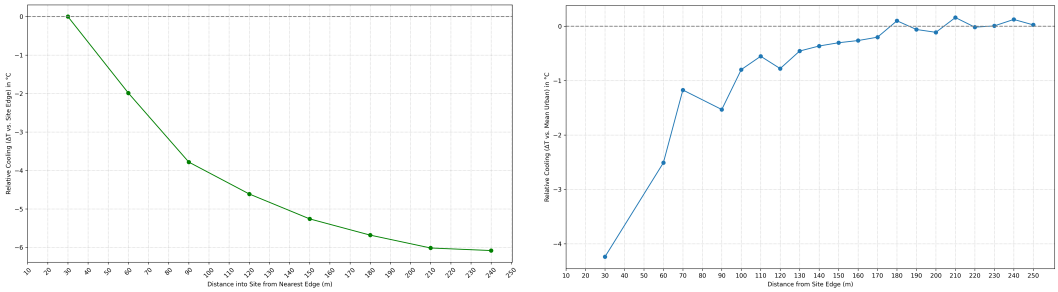
The predictive framework's ability to capture urban cooling effects was assessed by benchmarking the two model variants, V1 and V2, based on their performance in representing internal cooling gradients within parks and the spillover of cooling into surrounding urban areas. Model accuracy was assessed using MAE, RMSE, and MBE, enabling direct comparison of predictive performance across both experiments.

Table 1: Performance comparison of model variants for the two key cooling experiments. The best value per metric between V1 and V2 is underlined.

	V1			V2		
Experiment	MAE	RMSE	MBE	MAE	RMSE	MBE
Internal Cooling	0.240	0.257	+0.240	<u>0.231</u>	<u>0.249</u>	<u>+0.231</u>
Spillover Cooling	0.302	0.339	+0.302	<u>0.199</u>	<u>0.243</u>	<u>+0.199</u>

A.3 Analysis of cooling gradients for urban green areas using a geospatial foundation model

The thermal impact of the simulated greening intervention was assessed using Model V2 by analyzing both internal and external cooling responses. The internal cooling gradient captures the temperature decrease from the park edge toward its center, while the spillover gradient reflects how far this effect extends into surrounding built-up areas. Together, these analyses demonstrate both local and neighborhood-scale cooling benefits.



(a) Internal Cooling Gradient (b) Spillover Cooling Gradient

Figure 6: Cooling phenomena of the proposed intervention area.

A.4 Comparison of model variants for extrapolation capability

Table 2 presents the test set results for Model V1 and Model V2, both fine-tuned on Brașov-specific data using identical hyperparameters. Three evaluation setups were applied. The *Baseline* reflects zero-shot performance without Brașov-specific fine-tuning. The *Random Data Split* follows a conventional 72/18/10 train/validation/test split. The *High Heat Scenario* implements an extrapolation experiment by restricting training and validation to the lower 90% of temperature observations and reserving the upper 10% for testing, thereby assessing how well the models predict land surface temperatures beyond their training range.

Table 2: Performance comparison of model variants using MAE, MSE, and RMSE. Bold indicates the best value and underline indicates the second-best.

	V1			V2		
Model Variants	MAE	MSE	RMSE	MAE	MSE	RMSE
Baseline	1.95	7.25	2.69	2.81	12.94	3.60
Random Data Split	1.80	<u>6.26</u>	<u>2.50</u>	<u>1.77</u>	**5.80**	**2.41**
High-Heat Scenario (90th)	1.96	7.05	2.66	**1.74**	6.37	2.52

The results indicate that the *High Heat Scenario* models perform comparably to the random split across all metrics, while consistently exceeding the *Baseline*, except for MAE in V1. Notably, V2 achieves lower MAE in the *High Heat Scenario* than in the random split, demonstrating a stronger capacity for generalization. Across all settings and model variants, differences remain below 0.60 °C, showing that GFMs can extrapolate beyond training data and capture high-heat extremes relevant for UHI dynamics.

For Brașov, the 90th percentile temperature was 23.29 °C, with the GFM extrapolating up to 26.91 °C and reaching an MAE of 1.74 °C. This represents a successful prediction of 3.62 °C beyond the training limit, which provides an estimate of the model's extrapolation limit, forming the basis for projecting future UHI intensity under EURO-CORDEX climate scenarios.