Temperature Cooling Strategies for Outdoor Built Environments

Exploring tree canopy morphology for surface temperature mitigation with interpretable machine learning

Zishen Bai[1], Kening Guo[2], Jiaqi Yu[3]
[1,2]University College London [3]University of Sheffield
[1]zishen.bai.23@ucl.ac.uk [2]kening.guo.23@ucl.ac.uk [3]jyu89@sheffield.ac.uk

Landscape and urban designers strive to create urban forests to mitigate urban heat. However, research has found that the cooling effect of low or large trees varies, and tree misuse can negatively impact land surface temperature (LST). Existing research remains unclear on how tree canopy morphology controls LST at neighboring scales (below 1 km), a significant limitation in guiding design efforts. This study aims to provide designers with strategies for selecting canopy morphology to optimize the thermal environment. The study extracted data from 5.8 million trees in Greater London, including height, variability, and canopy area. The LST data (2015-2019) was acquired from Landsat 8 using Google Earth Engine. An interpretable machine learning approach was employed in the study using the XGBoost model and SHAP tool (500 m grid, n=6,079). The results explained 83% of the LST phenomena. From the training results, it was found that high trees (>8m) and reasonable canopy cover (18%-40%) could achieve the greatest benefits from tree cooling. The study also simulates future extreme heat scenarios through the proposed framework. It has been found that global warming may result in extra trees being planted in the future to achieve the cooling effect. Finally, the study discusses how the framework can be applied to practical design work and proposes a long-term development plan based on a crowdsourcing approach.

Keywords: *Tree canopy morphology, Land surface temperature, Nonlinear Relationships, Interpretable machine learning, Climate change.*

INTRODUCTION

Global warming poses a serious threat to the health of the world's population, and its effects have intensified in recent years. The World Meteorological Organisation (WMO) recently warned that the objective of the 2015 Paris Climate Change Conference to limit global warming to a 1.5°C increase is nearly unattainable (Polya, 2023). Global warming brings frequent extreme temperatures, causing increased disease burdens for the residents. Governments around the world need larger budgets each year to mitigate the threats from global warming (Lenton et al., 2023). In recent years, city managers and designers have explored low-cost strategies to mitigate urban high temperatures and reduce the of heat exposure to residents. Enhancing urban vegetation systems, particularly optimizing the urban tree layout, has become a commonly adopted approach (Tan et al., 2016). Trees in cities are core green infrastructure that considered an effective way to mitigate urban heat (He et al., 2024). Landscape and urban designers have attempted to create urban forests in their project (Gill et al., 2007). However, recent research indicates that tree canopies have a complex effect on air ventilation and heat retention (Dong et al., 2022). Dwarf and large trees have different cooling effects, and the misuse of trees can negatively affect land surface temperature (LST) (He et al., 2024). Designers estimate the trees in their proposals from past design experience. However, the lack of quantitative modelling makes it hard for trees to mitigate environmental overheating.

In past design processes, the general approach has used Computational Fluid Dynamics (CFD) models to simulate the distribution of airflow and heat storage on the site (Gülten et al., 2016). However, the large computational resources and time costs of CFD modelling significantly hinder the progress of the design work (Jurado et al., 2022). Although designers work on projects at micro to medium community scales (<1km^2), they normally need to discuss the impacts of their projects on their region or even the whole city. With current technological tools, frequent simulations are complex to implement quickly and iteratively. Simulation-based results are generally challenging when describing a real environmental situation. Research emphasizes that the city is a dynamic system with micro-climate zones composed of physical forms (such as buildings, topography, and water bodies) (Demuzere et al., 2019; Pellegatti Franco et al., 2019).

Therefore, this study aims to provide an efficient tool to assist designers with tree design in their projects to mitigate the overheating risk. Interpretable machine learning were created from 5.8 million tree morphology data from the Greater London region. The completed trained models can provide recommendations on how tree morphology can be controlled for LST at neighbourhood scales (up to 1km). The study also provides a framework for responding to future global warming and urges city managers and planners to consider what is reasonable for cities to mitigate urban overheating.

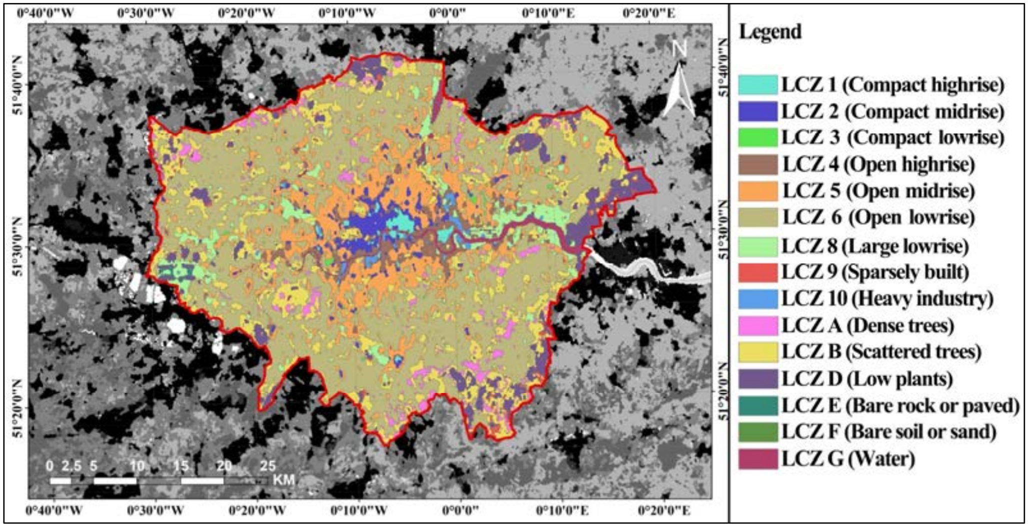

Figure 1
Map of LCZ distribution in the study area

DATA AND METHODS

Data

The core data sources of the study are urban land surface temperature and tree canopy maps. For the LST, the study obtained Landsat8 data from Google Earth Engine (GEE), which can have a spatial resolution of 30 metres (Ermida et al., 2020). The study obtained the urban tree canopy maps from the Bluesky National Tree Map (NTM) (Bluesky Ltd, 2025). The NTM data provides canopy height, area, and variability data for 5.8 million trees in Greater London using vector geographic information files (.shp) and records the geographic coordinates for each tree.

For the background environmental parameters of the study area, the Local Climate Zone (LCZ) framework proposed by Demuzere was selected for the study (Demuzere et al., 2019).

LCZs are an approach that describes the local microclimate, replacing the long-term climatic context at the macro scale (Aslam and Rana, 2022). The LCZ was calculated based on the urban form (height, density), sky view factor (SVF), and surface cover (pervious and impervious) within the study area, with a resolution of 100 m. A total of 17 categories are included in the LCZ framework, with 10 for building clusters (compact high-rise buildings, compact low-rise buildings, open low-rise buildings, heavy industry) and 7 for vegetation (dense trees, scattered trees, bare soil, water). Figure 1 shows the distribution of LCZs in the study area.

Methods

Figure 2 presents the workflow of the study, which is divided into three main steps.

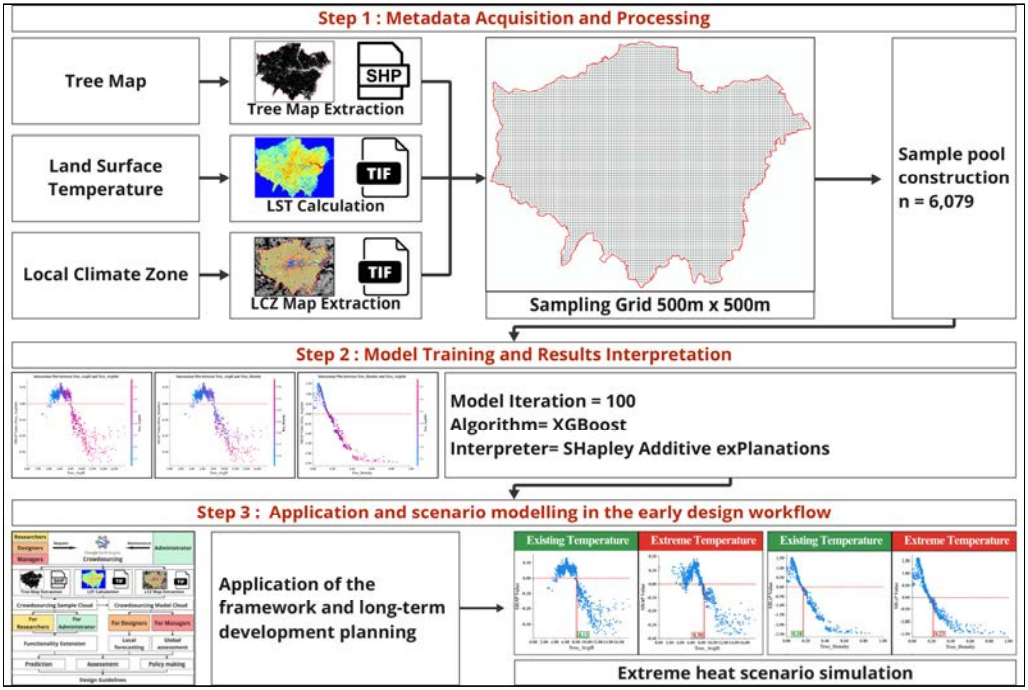

Figure 2
Workflow of the study

Step 1: Data acquisition and processing.
Firstly, the study obtains the surface temperature from remote sensing satellite data. Satellites convert the electromagnetic digits received into spectral radiance, which can detect any electromagnetic energy emitted above absolute zero (K) (Ziaul and Pal, 2018). The brightness temperatures of the Landsat thermal infrared channel are provided by the U.S. Geological Survey (USGS) to Google Earth Engine (GEE). The Ermida study provided a standardised procedure for extracting Landsat data from GEE, which just required specifying the study boundary and retrieval period (Ermida et al., 2020). However data from satellites is challenging to provide a continuous image database, which is a common issue (Tang et al., 2021). Satellite return period limitations and the cloud coverage are important factors in determining the amount of available data. Firstly, the work of satellites is to orbit the earth and take images of the area they pass through, which means that satellites cannot record the data for a long time series in an area (Loveland and Irons, 2016). Secondly, the clouds have a debilitating effect on the computational results of LST. The studies in recent years have discussed how to weaken the effect of clouds and proposed algorithms to eliminate them, but it is still controversial (Alvarez-Mendoza et al., 2019). The reliability of land surface temperature data obtained from satellites has been a ongoing discussion topic, with many approaches proposed in the past. Typical approaches include advanced cloud removal algorithms and multi-source data fusion methods. These approaches generally aim to address data loss at specific points in time. However, filling in data for long time series generally requires very high computational costs. More importantly, the reliability of current methods still needs further discussion. In our research, the focus is on exploring the global trends in surface temperature distribution across the entire city, rather than specific local values. This approach aims to avoid the influence of extreme weather events, such as extreme heatwaves. Therefore, a more conservative approach remains to collect average surface temperature data over the past five years or longer. Therefore, in this study, the average LST data for the summer of the last five years (2015-2019) were used to avoid the negative effects of missing data. Summer data was chosen because summer cloud cover is much lower than winter.

For the canopy maps, the study extracted the canopy area, mean height, and height variation standard deviation (STD) from NTM for each tree. For LCZ maps, the data provided by Demuzere were stored using raster graphics (.tif).

Step 2: Sample production and model training.
In the second step, a sample pool was created for model training. For this study, samples were defined as map tiles divided using a square grid. The feature values for each sample were canopy morphology metrics (mean height, density, variability) and LCZ configuration. The mean annual summer average surface temperature was the target value for the sample. The sample size is commonly considered the unit of analysis, which is highly relevant to the research question (Esposito et al., 2023). In past research experience, the scale of the unit of analysis (sample size) is divided into three categories: street scale (10-100 m), neighbourhood scale (100 m-1 km) and urban cluster scale (10-20 km) (Ferreira et al., 2021). Although past studies have not discussed the impact of scale on results much, the trade-off between sample size and sample quality needs to be considered. For machine learning models, an appropriate increase in sample size usually helps improve the model's accuracy (Rajput et al., 2023). However, if the study area is constant, expanding the sample size decreases the number of feature values within each sample. Therefore, based on experience considerations from past research (Anjos et al., 2020), a square grid with 500m sides (total number of grids 6,079) covering the whole of Greater London was used as the analysis unit for this study.

After sample production, the study created machine learning models using the eXtreme Gradient Boosting (XGBoost) algorithm. The SHapley Additive exPlanations (SHAP) approach was used to interpret the contribution of sample features in model training (Li, 2022). Using the SHAP interpreter coupled with the XGBoost model is a relatively common approach for recent interpretable machine learning (Vega García and Aznarte, 2020). In SHAP results, the user can access the contribution of the features to the model through their values on the Y-axis, which may be positive or negative (Ponce-Bobadilla et al., 2024). In this study, the X-axis represents a morphological indicator of the tree, such as the average height of the tree. The value on the Y-axis is a relative value that describes the contribution of the sample feature in the model (Li, 2022). In the feature dependency plot, when Y = 0 indicates that the contribution of the feature value to the model at this point will shift boundaries, resulting in a positive (Y > 0) or negative (Y < 0) transition. The study extracted the contribution of three parameters of trees in the model, including tree height, canopy cover, and tree height standard deviation.

Step 3: Application and scenario modelling in the early design workflow.

The third step discusses how the research results are involved in the daily workflow of designers and city managers. The study proposes plans for further development and how the demands of the various stakeholders can be met. Finally, the study modelled future extreme temperature scenarios and estimated tree demand. The study creates a counterfactual scenario by assuming that the average temperature in the future after global warming is the highest temperature from past summers (Chen et al., 2025). The study loaded the current tree distribution into the model and obtained the minimum demand of trees to achieve tree cooling under extreme temperatures.

RESULTS AND DISCUSSIONS

Model training results

After 100 iterations, the model achieved a predictive capability of 0.83, with the training results explaining 83% of the LST phenomenon (training R^2 = 0.83, training MAE = 0.85°C; testing R^2 = 0.83, testing MAE = 0.86°C).

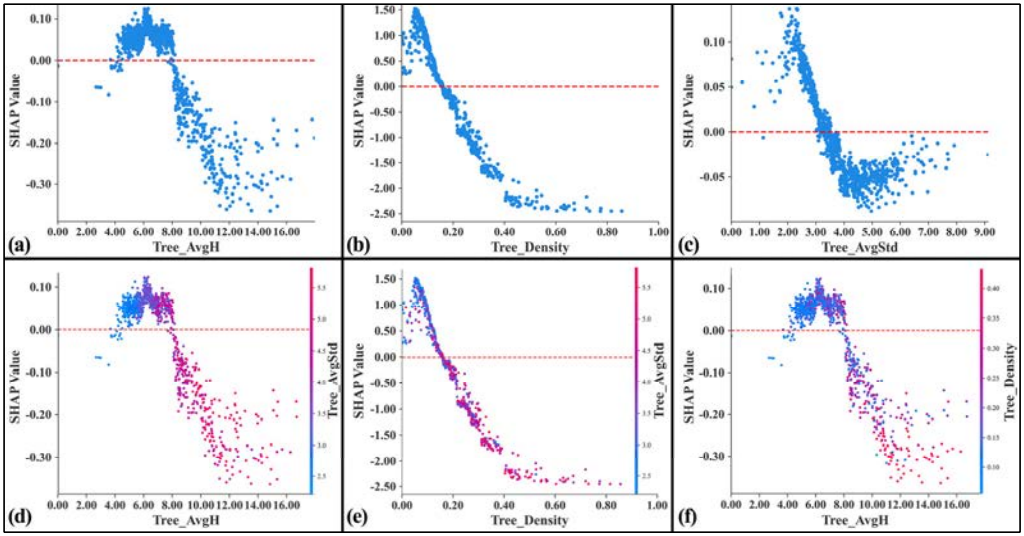

Figure 3
Feature dependency scatterplot

Figure 3 shows the feature dependency scatterplot generated by the SHAP approach to explain the variation in the contribution of feature values in the model predictions. The results (Figure 3) found that low trees did not appear to cool the site significantly but instead promoted temperature increases. The cooling effect of trees occurred when the average tree height was higher than 8 metres (Figure 3a). It was found that tree height variation and canopy coverage also have an important effect on temperature.

Figure 3b shows that a minimum of 18 % canopy coverage in the site achieves a cooling effect. However, when the canopy cover exceeds 40%, the cooling effect tends to peak.

For tree height differences, figure 3c shows that the cooling effect starts to appear when the canopy height difference is higher than 3 metres. Figures 3d,e, and f show the feature dependency plots with interaction. It was found that larger height variations commonly accompanied higher trees in the study area and showed a clustered pattern. Scattered and low trees could not contribute to temperature reduction but, instead, may exacerbate the increase in temperature.

Framework application scenarios

Framework development strategies.

The study's outcomes can provide recommendations for urban landscape policy-making, especially tree design, including current status studies and future forecasting. Green infrastructure in cities frequently involves a wide range of stakeholders, including designers, city managers, and scholars. The study recognises that the purposes for using the research results are diverse for different stakeholders.

Figure 4 presents the research development blueprint, describing how stakeholders can use and further develop the results from the research. The study proposes a crowdsourcing-based framework for the sustainable development of the study outcomes. The crowdsourcing approach is relatively common in geographic research, such as encouraging users to generate LCZs of the study area using the LCZ map generator and share them with the community (Demuzere et al., 2021). The study defined three user groups: managers, designers and researchers.

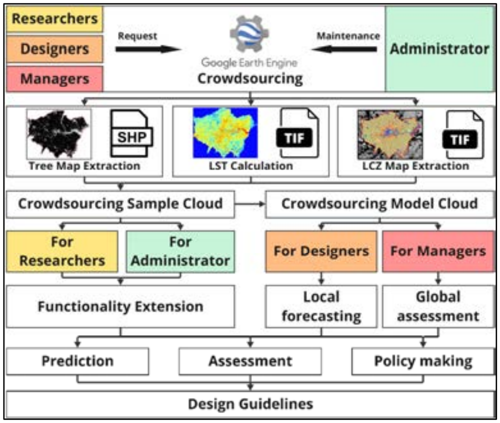

Figure 4
The research development blueprint

For urban managers, the study provides a quantitative analysis model based on a macro-scale approach, which can be used to assess the overall tree distribution and thermal environment of a city. By integrating regional-scale canopy distribution, land surface temperature (LST), and local climate zone (LCZ) data, the model can generate comprehensive green space distribution strategies and cooling potential assessments, helping managers identify high-temperature risk areas, areas with insufficient tree coverage, and the impact of urban expansion on the thermal environment. Based on this, managers can set landscaping targets for different districts and formulate long-term goals for green infrastructure development. For example, using the model's preliminary training results, managers can quantify the minimum tree coverage, and structural indicators (such as tree height and crown density) required for each district and conduct a comparative analysis with current data to scientifically formulate landscaping policies and investment priorities. Additionally, when urban data is missing or outdated, managers can utilise similar cases from other cities within the platform as references, leveraging analogy mechanisms to assist in decision-making validation.

For urban designers, the research provides scientific tools to support the entire process from conceptual design to detailed design. Designers can simulate the cooling effects of different tree quantities and forms on a 500-metre grid to determine the minimum landscaping intervention scale required to achieve a certain cooling target. This model enables designers to quickly generate green space configuration strategies in the early stages of a project for conceptual design derivation and multi-option comparison. In the later stages of design development, designers can use the model's recommended criteria to quantitatively control tree species selection, layout methods, and spatial density, thereby transitioning from 'greening recommendations' to 'implementation standards.' Additionally, the platform's crowdsourced case library helps designers draw on design experiences from sites with similar climates, topography, and functions, assisting them in optimising their current design strategies.

For researchers, this research platform also offers a highly scalable open architecture. Researchers can not only use existing models for local validation but also further expand model functionality and develop new application scenarios based on existing data. For example, counterfactual modelling and scenario simulation methods can be introduced to explore potential tree configuration needs for different cities under future global warming scenarios, or to analyse fairness issues in green space distribution under different administrative systems. Additionally, researchers can integrate their own research outcomes (such as new remote sensing processing methods, machine learning models, or urban classification systems) into the platform, participate in platform development and technological evolution, thereby establishing an iterative mechanism of 'research-development-research.' Through crowdsourcing mechanisms, researchers can also aggregate case data from different cities and scales to train models with greater universality or adaptability, thereby supporting cross-city comparative studies and the development of global urban green policies.

Although the purpose of each user group differs, it is inevitable that databases and model training are required for all users. As a link of interest for each user group, the database is crowdsourced by administrators and the necessary technical support is provided. Designers and researchers can easily access the GEE and process the data using the platform. The shared cloud data can be used as a grassroots database for further model development and will be continuously iterated. A potential advantage of crowdsourced data is that it can be used as a reference. Similar background cases can be a better reference when the study area faces extreme data loss.

Scenario simulation.

As a further demonstration, the study conducted a test from the perspectives of researchers and managers. The study simulated a future global warming scenario. Figure 5 shows the distribution of tree cooling effects for the high-temperature scenario. The study found that global warming may result in more trees requiring planting in the future to achieve the cooling effect. Relative to existing trees, in the future, global warming may result in the need for a 27.78% increase in canopy coverage, a 3.08% increase in height, and a 10.23% increase in height difference (Figure 5).

This simulation not only provides designers and managers with quantitative references but also serves as a warning signal for 'green lag': the costs of passive adaptation to climate warming will continue to rise, and early intervention in green infrastructure construction is more cost-effective than future remedial measures. For managers, these results can serve as an important basis for formulating long-term greening policies and fiscal investment plans, emphasising the necessity of 'early planning and sustained investment.' For designers, design standards under future scenarios may need to be 'dynamically upgraded,' not relying solely on current climate data but also incorporating future temperature rise trends into redundant designs to ensure the long-term adaptability and resilience of the plans.

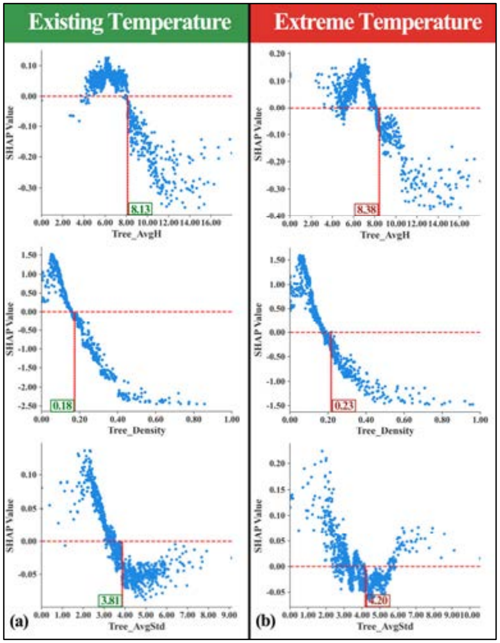

Figure 5 Comparison of tree cooling effects in existing and scenario simulations

CONCLUSIONS

This study presents a quantitative modelling of the cooling capacity of trees using an interpretable machine learning approach. It was found that tree morphology showed a non-linear relationship with environmental cooling, implying that misuse of trees may exacerbate environmental overheating. The most significant benefits of tree cooling can be achieved with high trees (>8m) and reasonable coverage (18%-40%). This study gives designers and city administrators a quick assessment tool and guides the necessary suggestions when designing trees. The outcome of the study is a methodological framework for long-term development and can be further developed in the future based on the crowdsourcing network approach. Finally, the study warns that future global warming may lead to more significant tree demand in mitigating environmental overheating risks.

REFERENCES

Alvarez-Mendoza, C.I., Teodoro, A., Ramirez-Cando, L. (2019). 'Improving NDVI by removing cirrus clouds with optical remote sensing data from Landsat-8 – A case study in Quito, Ecuador', *Remote Sensing Applications: Society and Environment*, 13, pp.257–274. Available at: https://doi.org/10.1016/j.rsase.2018.11.008 (Accessed: 10 March 2025)

Anjos, M., Targino, A.C., Krecl, P., Oukawa, G.Y., Braga, R.F. (2020). 'Analysis of the urban heat island under different synoptic patterns using local climate zones', *Building and Environment*, 185, pp.107268. Available at: https://doi.org/10.101/10.1016/j.buildenv.2020.107268 (Accessed: 10 March 2025)

Aslam, A., Rana, I.A. (2022). 'The use of local climate zones in the urban environment: A systematic review of data sources, methods, and themes', *Urban Climate*, 42, pp.101120. Available at: https://doi.org/10.1016/j.uclim.2022.101120 (Accessed: 10 March 2025)

Bluesky Ltd (2025). National Tree Map [Online]. Available at: https://bluesky-world.com/ntm/ (Accessed: 10 March 2025)

Chen, X., Li, H., Yu, H., Hou, E., Song, S., Shi, H., Chai, Y. (2025). 'Counterfactual analysis of extreme events in urban flooding scenarios', *Journal of Hydrology: Regional Studies*, 57, pp.102166. Available at: https://doi.org/10.1016/j.ejrh.2024.102166 (Accessed: 10 March 2025)

Demuzere, M., Bechtel, B., Middel, A., Mills, G. (2019). 'Mapping Europe into local climate zones', *PLoS ONE*, 14, pp.e0214474. Available at: https://doi.org/10.1371/journal.pone.0214474 (Accessed: 10 March 2025)

Demuzere, M., Kittner, J., Bechtel, B. (2021). 'LCZ Generator: A Web Application to Create Local Climate Zone Maps', *Front. Environ. Sci.*, 9, pp.637455. Available at: https://doi.org/10.3389/fenvs.2021.637455 (Accessed: 10 March 2025)

Dong, Y., Cai, Y., Li, Chong, Wang, H., Zhou, L., Sun, J., Li, Chen, Song, B., Zhou, G. (2022). 'Vertical thermal environment of subtropical broad-leaved urban forests and the influence of canopy structure', *Building and Environment*, 224, pp.109521. Available at: https://doi.org/10.1016/j.buildenv.2022.109521 (Accessed: 10 March 2025)

Ermida, S.L., Soares, P., Mantas, V., Göttsche, F.-M., Trigo, I.F. (2020). 'Google Earth Engine Open-Source Code for Land Surface Temperature Estimation from the Landsat Series', *Remote Sensing*, 12, pp.1471. Available at: https://doi.org/10.3390/rs12091471 (Accessed: 10 March 2025)

Esposito, A., Grulois, M., Pappaccogli, G., Palusci, O., Donateo, A., Salizzoni, P., Santiago, J.L., Martilli, A., Maffeis, G., Buccolieri, R. (2023). 'On the Calculation of Urban Morphological Parameters Using GIS: An Application to Italian Cities', *Atmosphere*, 14, pp.329. Available at: https://doi.org/10.3390/atmos14020329 (Accessed: 10 March 2025)

Ferreira, D.G., Diniz, C.B., Assis, E.S.D. (2021). 'Methods to calculate urban surface parameters and their relation to the LCZ classification', *Urban Climate*, 36, pp.100788. Available at: https://doi.org/10.1016/j.uclim.2021.100788 (Accessed: 10 March 2025)

Gill, S.E., Handley, J.F., Ennos, A.R., Pauleit, S. (2007). 'Adapting Cities for Climate Change: The Role of the Green Infrastructure', *built environ*, 33, pp.115–133. Available at: https://doi.org/10.2148/benv.33.1.115 (Accessed: 10 March 2025)

Gülten, A., Aksoy, U.T., Öztop, H.F. (2016). 'Influence of trees on heat island potential in an urban canyon', *Sustainable Cities and Society*, 26, pp.407–418. Available at: https://doi.org/10.1016/j.scs.2016.04.006 (Accessed: 10 March 2025)

He, T., Hu, Y., Guo, A., Chen, Y., Yang, J., Li, M., Zhang, M. (2024). 'Quantifying the impact of urban trees on land surface temperature in global cities', *ISPRS Journal of Photogrammetry and Remote Sensing*, 210, pp.69–79. Available at: https://doi.org/10.1016/j.isprsjprs.2024.03.007 (Accessed: 10 March 2025)

Jurado, X., Reiminger, N., Benmoussa, M., Vazquez, J., Wemmert, C. (2022). 'Deep learning methods evaluation to predict air quality based on Computational Fluid Dynamics', *Expert Systems with Applications*, 203, pp.117294. Available at: https://doi.org/10.1016/j.eswa.2022.117294 (Accessed: 10 March 2025)

Lenton, T.M., Xu, C., Abrams, J.F., Ghadiali, A., Loriani, S., Sakschewski, B., Zimm, C., Ebi, K.L., Dunn, R.R., Svenning, J.-C., Scheffer, M. (2023). 'Quantifying the human cost of global warming', *Nat Sustain*, 6, pp.1237–1247. Available at: https://doi.org/10.1038/s41893-023-01132-6 (Accessed: 10 March 2025)

Li, Z. (2022). 'Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost', *Computers, Environment and Urban Systems*, 96, pp.101845. Available at:

https://doi.org/10.1016/j.compenvurbsys.2022.101845 (Accessed: 10 March 2025)
Loveland, T.R., Irons, J.R. (2016). 'Landsat 8: The plans, the reality, and the legacy', *Remote Sensing of Environment*, 185, pp.1–6. Available at: https://doi.org/10.1016/j.rse.2016.07.033 (Accessed: 10 March 2025)
Pellegatti Franco, D.M., Andrade, M.D.F., Ynoue, R.Y., Ching, J. (2019). 'Effect of Local Climate Zone (LCZ) classification on ozone chemical transport model simulations in Sao Paulo, Brazil', *Urban Climate*, 27, pp.293–313. Available at: https://doi.org/10.1016/j.uclim.2018.12.007 (Accessed: 10 March 2025)
Polya, G. (2023). WMO Warning: 1.5 Degree C Warming Breach Very Soon & With Increasing Frequency [Online]. Available at: https://countercurrents.org/2023/05/wmo-warning-1-5-degree-c-warming-breach-very-soon-with-increasing-frequency-act-now/ (Accessed: 10 March 2025)
Ponce-Bobadilla, A.V., Schmitt, V., Maier, C.S., Mensing, S., Stodtmann, S. (2024). 'Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development', *Clinical Translational Sci*, 17, pp.e70056. Available at: https://doi.org/10.1111/cts.70056 (Accessed: 10 March 2025)
Rajput, D., Wang, W.-J., Chen, C.-C. (2023). 'Evaluation of a decided sample size in machine learning applications', *BMC Bioinformatics*, 24, pp.48. Available at: https://doi.org/10.1186/s12859-023-05156-9 (Accessed: 10 March 2025)
Tan, Z., Lau, K.K.-L., Ng, E. (2016). 'Urban tree design approaches for mitigating daytime urban heat island effects in a high-density urban environment', *Energy and Buildings*, 114, pp.265–274. Available at: https://doi.org/10.1016/j.enbuild.2015.06.031 (Accessed: 10 March 2025)
Tang, Z., Adhikari, H., Pellikka, P.K.E., Heiskanen, J. (2021). 'A method for predicting large-area missing observations in Landsat time series using spectral-temporal metrics', *International Journal of Applied Earth Observation and Geoinformation*, 99, pp.102319. Available at: https://doi.org/10.1016/j.jag.2021.102319 (Accessed: 10 March 2025)
Vega García, M., Aznarte, J.L. (2020). 'Shapley additive explanations for NO2 forecasting', *Ecological Informatics*, 56, pp.101039. Available at: https://doi.org/10.1016/j.ecoinf.2019.101039 (Accessed: 10 March 2025)
Ziaul, Sk., Pal, S. (2018). 'Analyzing control of respiratory particulate matter on Land Surface Temperature in local climatic zones of English Bazar Municipality and Surroundings', *Urban Climate*, 24, pp.34–50. Available at: https://doi.org/10.1016/j.uclim.2018.01.006 (Accessed: 10 March 2025)