

Heterogeneous effects of anthropogenic heat on urban thermal environment: An interpretable ML approach

Yonghang Xie¹, Cheng Fan^{1,2,3*}, Yutian Lei¹, Jiena Cai¹

¹ Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, PR China

² Key Laboratory for Resilient Infrastructures of Coastal Cities, Ministry of Education, Shenzhen University, Shenzhen, PR China

³ Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, PR China

(Corresponding Author: fancheng@szu.edu.cn)

ABSTRACT

In high-density cities, anthropogenic heat (AH) emissions are an important factor affecting the urban heat island effect. Especially in summer, the heat stress on the outdoor thermal environment caused by AH emissions cannot be ignored. Due to the oversimplification of AH emissions in previous studies, the spatial differences in the role of different types of AH emissions in the urban thermal environment are not well understood. Therefore, this study proposes an interpretable machine learning framework to model and predict AH emissions and their thermal impacts in dense urban areas, with a focus on quantifying the differential responses of land surface temperature (LST) to sector-specific AH emissions (e.g., buildings, industry, transportation). The results demonstrate that the ExtraTreesRegressor model achieves optimal performance, with a coefficient of determination (R^2) of 0.8. The Shap analysis reveals that, among the seven types of selected urban factors, NDVI, Building-related AH, and Industry-related AH are the key driving factors of LST across Shenzhen, China. Sensitivity experiments demonstrate that any reduction in the AHE intensity results in a corresponding decrease in the spatially averaged LST. Under an 80% reduction in AHE, the mean LST in Shenzhen decreased by 0.57°C (buildings), 0.46°C (industry), and 0.23°C (transportation). Furthermore, the cooling response to AH reduction varied spatially across Shenzhen. A reduction in building or industry-related AH emissions induced significant cooling areas in inland districts, including Guangming, Longhua, Pingshan, and parts of Futian. In contrast, coastal districts such as Nanshan and Futian exhibited more pronounced cooling effects when transportation-related AH emissions were reduced. This study elucidates the mechanistic response of urban thermal environments to spatially heterogeneous and source-specific AH emissions,

providing novel insights and a theoretical foundation for developing targeted urban heat mitigation strategies.

Keywords: anthropogenic heat emission, urban heat island, interpretable ML approach, sensitive analysis, urban heat mitigation

NONMENCLATURE

Abbreviations	explanation
AH	anthropogenic heat
LST	land surface temperature
UHI	Urban Heat Island
LCZ	Local climate zone
ABH	Average building height
BCR	Building cover ratio
NDVI	Normalized difference vegetation index
DEM	Digital elevation model
SHAP	Shapley Additive exPlanations
BuildingAH	Building-related anthropogenic heat
IndustryAH	Industry-related anthropogenic heat
TransportAH	Transportan-related anthropogenic heat
MetabolismAH	Metabolism related anthropogenic heat

1. INTRODUCTION

As global urbanization accelerates, the UHI effect in high-density cities is becoming a growing problem, significantly threatening the health of residents and energy sustainability[1]. As a key driver of UHI, AH is mainly emitted from the building, industrial, and transportation sectors, and its heat release further contributes to the rise of LST. However, traditional studies have mostly regarded AH emission as a homogeneous variable, ignoring the heterogeneity of its type and spatial distribution, resulting in the lack of precision in emission reduction strategies[2]. Especially in summer, the differences in the contribution of AH to

This is a paper for the 11th Applied Energy Symposium: Low Carbon Cities & Urban Energy Systems (CUE2025), July 18-22, 2024, Kitakyushu, Japan.

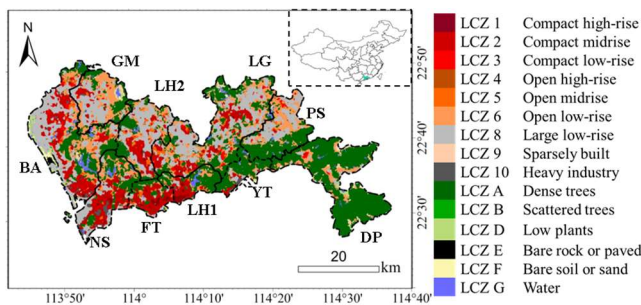
the thermal environment from different regions and sources may significantly affect the effectiveness of cooling policies. Therefore, there is an urgent need for a method that can quantify the spatially differentiated characteristics of multiple types of AH emissions and their impacts on the differentiation of LST.

Using Shenzhen, a high-density mega-city in China, as a case study, this study combines multi-source geospatial data with an interpretable ML method to quantify the differences in the contributions of seven categories of urban factors, and investigates the sensitivity and spatial heterogeneity of different types of AH emissions to the LST through a scenario-based experimental analysis.

The innovations of this study are (1) the first application of interpretable ML to the analysis of spatial heterogeneity of multiple types of AH emissions, which overcomes the limitation of the linearity assumption of the traditional statistical methods, and (2) the sensitivity experiments to quantify the differential cooling effects of different levels of AH emissions from the building, industrial, and transportation sectors on LST.

2. STUDY AREA

As a leading socio-economic hub in southern China (Fig.1), Shenzhen's rapid urbanization has exacerbated challenges associated with the deterioration of its urban thermal environment[3]. Up till 2025, its population exceeded 17 million, reflecting rapid urbanization, with an urbanization level of over 85%. August is typically the hottest month when heatwaves and temperatures exceeding 35 °C are common, and combined heat and humidity can make it feel even hotter. The LCZ map of Shenzhen[4] shows high-density urban cores (primarily FT, NS) are dominated by LCZ 1 and LCZ 2, and peripheral inland districts (e.g., LH, PS) feature LCZ 3 and LCZ 5 areas, indicating distinct spatial heterogeneity in urban land cover.



GM: Guangming; BA: Baoan; NS: Nanshan; FT: Futian; LH1: Luohu; YT: Yantian; LH2: Longhua; LG: Longgang; PS: Pingshan; DP: Dapeng

Fig. 1 The location and LCZ map across Shenzhen.

3. MATERIAL AND METHODS

3.1 Multi-source GIS data collection and processing

This study investigates the understudied relationships between urban factors and LST variations cross Shenzhen, China. We systematically categorized urban factors into four primary classes: (1) urban morphological parameters (ABH and BCR), (2) land cover characteristics (Albedo and NDVI), (3) AH indicators (AH), and (4) topographic features (DEM). Table 1 presents the comprehensive multivariate geographic dataset employed in our analysis, along with detailed descriptions of each variable's derivation methodology.

Table 1 Geospatial data and processing approach

Indicators	Calculation formula	Instructions
ABH [m]	$\frac{\sum_{i=1}^n (BH_i \times BA_i)}{\sum_{i=1}^n BA_i} \quad (1)$	BH_i and BA_i represent the height and footprint area of each building in the zone respectively, and n means the total number of buildings in the corresponding zone[5].
BCR [-]	$\frac{\sum_{i=1}^n (BA_i)}{1000^2} \quad (2)$	$\alpha_2, \alpha_4, \alpha_5, \alpha_6, \alpha_7$ is the TOA value of band i of the Landsat 8 C2 L1 Science Products [6].
Albedo [-]	$0.356\alpha_2 + 0.130\alpha_4 + 0.373\alpha_5 + 0.085\alpha_6 + 0.072\alpha_7 - 0.0018 \quad (3)$	α_4 and α_5 are the top atmosphere reflectance values of the red band and near-infrared band of Landsat 8 C2 L1 Science Products, respectively [6].
NDVI [-]	$\frac{\alpha_5 - \alpha_4}{\alpha_5 + \alpha_4} \quad (4)$	A raster dataset and the pixel's values represent AH emissions [7]. A raster dataset and the pixel's values represent the elevation [8].
BuildingAH [W/m ²]		
IndustryAH [W/m ²]		
TransportAH [W/m ²]		
MetabolismAH [W/m ²]		
DEM [-]		

3.2 ML model construction

In this study, we implemented four distinct machine learning algorithms to model LST dynamics using a comprehensive suite of urban environmental and socio-economic predictors. The predictive modeling framework employed a standardized preprocessing and training protocol across all machine learning algorithms. Each model was instantiated with 100 base estimators and a 42-times fixed random state to guarantee reproducible results. The training process involves fitting each model to the feature matrix X and target variable y , where the algorithms automatically learn the complex relationships between urban environmental factors and LST. Then, We quantitatively compared model performance using three rigorously validated metrics: R^2 Score, RMSE, MAE. Finally, through comprehensive benchmarking, the best performing model was decided.

3.3 ML models interpretation method

To enhance model interpretability, we employed SHAP[9], a game-theoretic approach that quantifies feature contributions by fairly distributing prediction differences from a baseline value according to each feature's marginal impact. The SHAP framework provides theoretically grounded, model-agnostic explanations through two complementary perspectives: (1) global interpretability, where mean absolute SHAP values reveal dominant predictors across the dataset, and (2) local interpretability, decomposing individual predictions into additive feature contributions. This method's mathematical rigor, derived from cooperative game theory, ensures consistent feature attribution while naturally capturing nonlinear interactions - particularly valuable for analyzing complex urban thermal dynamics. In our study, SHAP analysis elucidated both the individual effects and synergistic interactions of key urban factors on LST variations, providing critical insights into anthropogenic and natural drivers of thermal patterns.

3.4 ML model validation

To evaluate the predictive performance of the ML models in estimating LST, we employed a standard hold-out validation approach. The dataset was randomly split into training (80%) and testing (20%) subsets using a fixed random seed (42) to ensure reproducibility. The model was trained using the training subset, and predictions were subsequently generated on the test subset. Three commonly used evaluation metrics—coefficient of determination (R^2), RMSE, and MAE—were adopted to assess model accuracy and the magnitude of prediction errors. The calculation of these metrics are summarized in Table 2.

In these equations, a_i and b_i denote the i -th predicted and observed values, respectively, b is the mean of the observed values, and m is the number of test samples.

Table 2 Formulas of metrics of model validation

Metrics	Formulas
R^2 Score	$R^2 = \frac{\sum_i (a_i - b_i)^2}{\sum_i (b_i - b)^2} \quad (5)$
RMSE	$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (a_i - b_i)^2} \quad (6)$
MAE	$MAE = \frac{1}{m} \sum_{i=1}^m a_i - b_i \quad (7)$

3.5 AH reduction scenario settings

To analyze the sensitivity of LST to different AHE sources, we conducted scenario-based simulations by systematically reducing building-related (buildingAH), industrial (IndustryAH), and transportation-related (TransportAH) heat emissions by 20%, 50%, and 80%. These mitigation scenarios enabled quantitative assessment of LST response patterns across the urban landscape. Comparative analysis between simulated and original LST values revealed spatially heterogeneous temperature differentials, allowing identification of (1) cooling zones exhibiting LST decrease to emission reductions and (2) warming areas with increase LST. The spatial distribution of these sensitivity classes provides critical insights for targeted heat mitigation planning, particularly in identifying priority intervention areas where anthropogenic heat reduction would yield maximum thermal benefits. The results particularly highlight the differential effectiveness of sector-specific heat mitigation measures across various urban morphological contexts.

4. RESULTS AND DISCUSSION

4.1 Model performance

In this study, we evaluated the performance of four machine learning models—XGBoost, LightGBM, CatBoost, and ExtraTreesRegressor—using three key metrics: R^2 Score, RMSE, and MAE.

As shown in Table 3, CatBoost and ExtraTreesRegressor achieved the highest scores (0.80). The comparative assessment identified ExtraTreesRegressor and CatBoost as the top-performing models, with ExtraTreesRegressor showing a slight edge in RMSE, while CatBoost led in MAE. Given their balanced performance, ExtraTreesRegressor was selected for subsequent experiments due to its lower sensitivity to

outliers (RMSE). Further experiments leveraging the optimal model (ExtraTreesRegressor) were conducted, focusing on hyperparameter tuning and validation on extended datasets to enhance generalizability. Detailed results are discussed in the following sections.

Table 3 Validation metrics for ML Models

Model	R ² Score	RMSE	MAE
XGBoost	0.77	1.24	0.96
LightGBM	0.79	1.21	0.91
CatBoost	0.80	1.17	0.87
ExtraTreesRegressor	0.80	1.16	0.88

4.2 Individual impacts of urban factors on LST

To quantitatively evaluate the relative contributions of urban factors to LST variations, we employed SHAP analysis derived from the ExtraTreesRegressor model outputs. This method facilitates: (1) global interpretation by quantifying the overall importance of each predictor variable (Fig. 2 (a)), and (2) local interpretation by elucidating how specific feature values influence individual LST predictions (Fig. 2 (b)).

Fig. 2(a) demonstrates the relative importance of urban factors in driving daytime LST variations, quantified by mean absolute SHAP values. NDVI emerges as the most influential predictor (mean SHAP = 0.66), followed by building AH emissions (buildingAH: 0.49) and industrial AH emissions (IndustryAH: 0.35). Metabolic AH (metabolismAH: 0.32) and building coverage ratio (BCR: 0.20) exhibit moderate contributions, while topographic (DEM: 0.18) and other anthropogenic factors (TransportAH: 0.16, Albedo: 0.11, ABH: 0.09) show comparatively weaker impacts.

Fig. 2(b) delineates the directional relationships between feature values and LST perturbations. High NDVI and DEM correlate strongly with negative SHAP values, underscoring the thermal mitigation capacity of vegetation and higher elevations. Conversely, positive SHAP values are associated with high buildingAH and

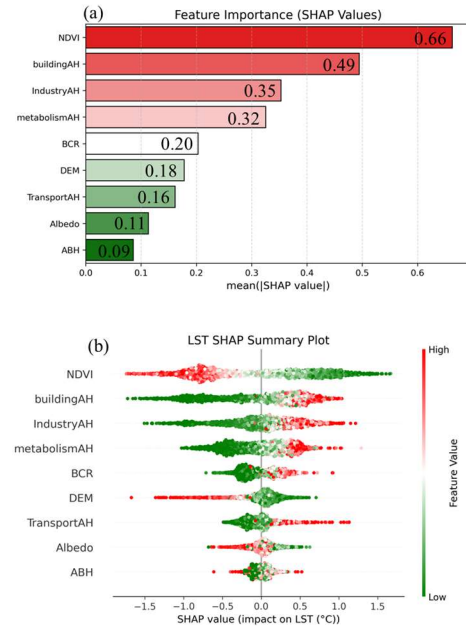


Fig. 2 Relative importance of urban factors to LST.

(a) mean absolute SHAP value and (b) local distribution of SHAP value.

IndustryAH, revealing their pronounced roles in exacerbating urban heat. These findings advocate for integrated LST management prioritizing reduction of AH emissions and strategic expansion of urban greenery, particularly in low-vegetation, high-heat-emission zones.

4.3 AH reduction scenario simulations

4.3.1 Global comparison of LST changes in each AH reduction scenario

Fig.3 presents violin plots quantifying the LST reduction across Shenzhen under progressive mitigation scenarios for three AH sources: buildingAH, IndustryAH, and TransportAH emissions. Any type of AH reduction can result in a decrease in the overall average LST cross Shenzhen City. Industrial AH reductions generated the most consistent citywide cooling and the 80% reduction scenario showed the most cooling effect.

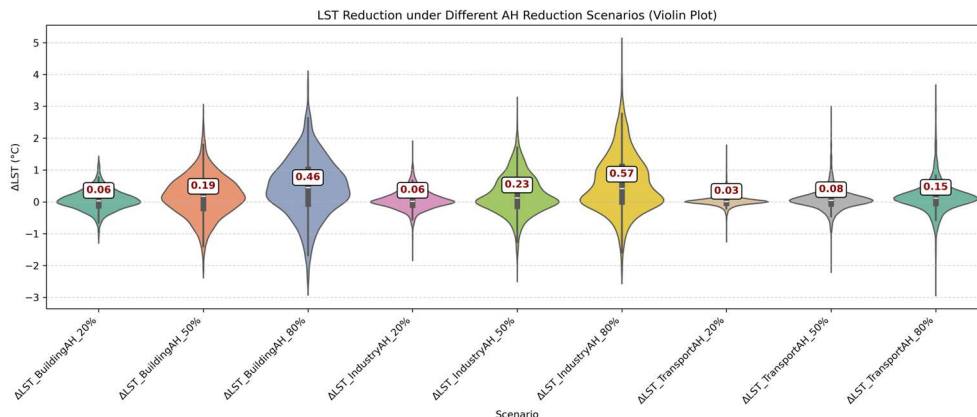


Fig. 3 The violin plot of LST decrease under different AH reduction scenarios.

Specifically, industryAH reductions produced the most substantial cooling effects, with averaged LST reduction reaching 0.57°C. BuildingAH and TransportAH mitigation showed averaged LST reduction of 0.46°C and 0.23°C, respectively. These results quantitatively demonstrate that sector-specific mitigation strategies yield distinct thermal benefits.

4.3.2 Statistical of the cooling and warming area in each district

we further analyzed the cooling and warming results of 10 districts in Shenzhen City. According to Table 4, LST in each zone showed different sensitivities to the reduction of AH. Among the cooling effects of

reductions in AH, Futian District was the most sensitive to buildingAH and transportAH (greatest cooling at 20% reduction), Guangming District responded significantly to industryAH and transportAH (greatest at 20% and 50% reductions), and Longhua District had the widest area of cooling at significant reductions in buildingAH and industrial AH (50%). Yantian District (buildings and transportation) and Dapeng New District (industry) were consistently the least sensitive, with less than 35% cooling at 80% heat source reduction. Highly sensitive areas ($\geq 80\%$) are concentrated in Futian, Guangming, and Longhua, while transportAH regulation is most effective for Guangming and Longhua.

Table 4 The percentile of cooling and warming areas under different AH reduction scenarios

District Name	BuildingAH						IndustryAH						TransportAH					
	20%		50%		80%		20%		50%		80%		20%		50%		80%	
	cooling	warming	cooling	warming	cooling	warming	cooling	warming	cooling	warming	cooling	warming	cooling	warming	cooling	warming	cooling	warming
Baoan	52.2	45.0	58.5	40.9	69.5	30.3	54.2	45.0	62.0	37.5	79.5	20.2	50.4	42.1	55.6	41.2	67.4	29.4
Dapeng	37.4	60.5	46.2	51.7	54.2	44.8	15.4	46.5	17.8	44.4	25.9	37.1	37.8	51.7	42.3	54.2	52.8	45.5
Futian	74.7	24.1	75.9	22.8	84.8	15.2	68.4	30.4	79.7	19.0	88.6	11.4	70.9	27.8	72.2	27.8	72.2	27.8
Guangming	69.6	30.4	70.2	29.8	83.9	15.5	70.2	29.8	78.9	21.1	93.8	6.2	64.6	28	74.5	21.1	85.7	13.0
Longgang	62.5	36.7	71.5	28.0	82.1	17.9	63.0	35.2	69.5	29.3	77.7	22.1	63	30	71.2	26.3	76.2	22.6
Longhua	72.4	23.8	83.2	16.2	89.7	9.7	68.1	28.6	81.1	18.9	88.6	9.7	65.4	29.2	73.5	23.8	84.3	13.5
Luohu	30.0	70.0	38.8	58.8	51.3	48.8	38.8	58.8	37.5	61.3	50.0	50.0	27.5	68.8	27.5	70.0	30.0	70.0
Nanshan	43.5	55.8	46.8	53.2	53.9	45.5	44.2	51.9	50.6	49.4	60.4	39.6	48.1	48.1	57.1	40.9	62.3	37.0
Pingshan	62.4	37.6	72.8	27.2	83.2	16.8	60.7	35.8	63.6	31.2	75.1	21.4	62.4	31.8	73.4	24.9	79.8	20.2
Yantian	9.6	87.7	20.5	78.1	32.9	65.8	26.0	68.5	28.8	67.1	29.4	65.9	16.4	82.2	17.8	82.2	19.2	78.1

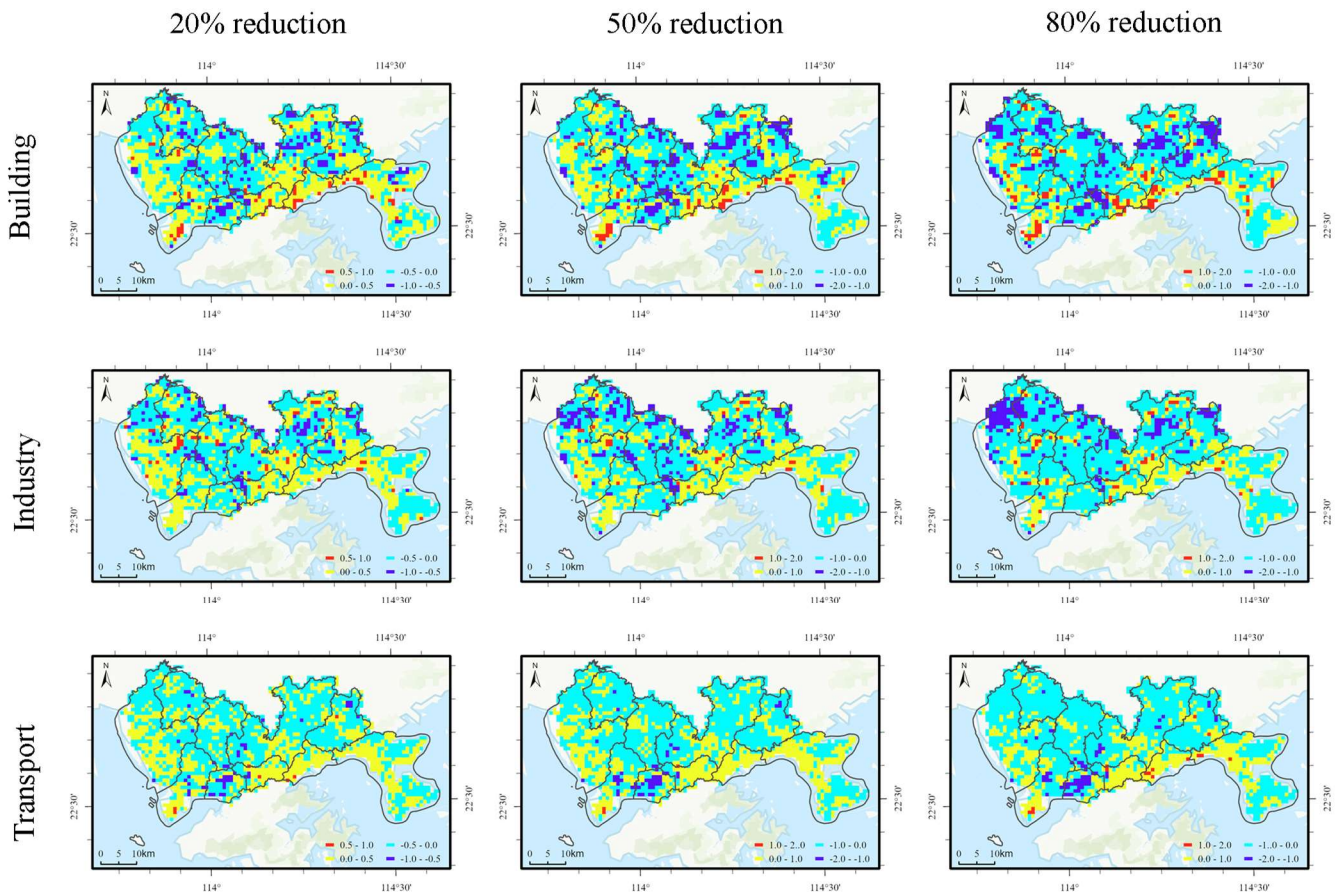


Fig. 4 Spatial distributions of LST reduction under different AH reduction scenarios.

4.4 Spatial distribution of LST reduction under varying levels of AH emission mitigation

Fig. 5 shows the spatial characteristics of LST responses to different types and magnitudes of AH reductions. In Fig. 5, each row corresponds to a specific AH emission source—building (top), industrial (middle), and traffic (bottom)—while each column represents a reduction level of 20% (left), 50% (middle), and 80%.

All scenarios exhibit a spatially heterogeneous pattern of LST change. With the reduction of AH from buildings and industries, inland areas, such as Guangming, Longhua and Pingshan, as well as areas in Futian District that are located inland, show a significant cooling trend. In the inland areas, there are mainly high-density low-rise residential areas and some industrial factory sites, indicating that reducing the AH emissions from buildings and industries in the inland areas can alleviate the heat pressure in the modified areas. In coastal areas such as Baoan and Nanshan districts, the reduction of building or industrial AH increases up to 80% before the cooling area increases slightly. It indicates that the region is less sensitive to the reduction of emissions through the reduction of building AH and industry AH. The more obvious areas where LST reductions are achieved by reducing transport AH emissions are Futian and Nanshan District. This may be related to the high traffic density and high energy consumption of transportation in this region.

5. CONCLUSIONS

This study systematically analyzes the heterogeneous effects of different types of AH emissions on LST in high-density cities based on an interpretable machine learning approach, and conducts a sensitivity experiment in Shenzhen City. The main conclusions are as follows:

AH emissions have a significant impact on the urban thermal environment, and the contributions of different types of AH to LST are significantly different.

The reduction of AH can effectively mitigate the UHI effect. Sensitivity experiments show that the average LST in Shenzhen can be reduced by 0.57°C, 0.46°C and 0.23°C when the Building AH, Industry AH and Transport AH are reduced by 80% respectively.

The response of different regions to AH reduction is spatially heterogeneous. Inland areas (e.g., Guangming, Longhua, Pingshan, and part of Futian) showed stronger cooling effects to the building AH and industry AH reduction, while coastal areas (e.g., Nanshan and the coastal part of Futian) were more sensitive to the reduction of transport AH.

This study provides a scientific basis for urban thermal environment management. Future research will integrate urban planning policies to optimize AH emission management in different zones for more effective UHI mitigation.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of this research by the National Natural Science Foundation of China (No. 52278117), Guangdong Basic and Applied Basic Research Foundation (No. 2024A1515011549), Shenzhen Science and Technology Program (No. 20220531101800001, 20240813143330039 and 20220810160221001), and the Philosophical and Social Science Program of Shenzhen (SZ2024C006).

REFERENCE

- [1] Hofstra N, Vermeulen LC. Impacts of population growth, urbanisation and sanitation changes on global human Cryptosporidium emissions to surface water. *International Journal of Hygiene and Environmental Health* 2016;219:599–605.
- [2] Yuan C, Adelia AS, Mei S, He W, Li X-X, Norford L. Mitigating intensity of urban heat island by better understanding on urban morphology and anthropogenic heat dispersion. *Building and Environment*
- [3] Zhao D, Zha J, Wu J. Comparisons of urban-related warming for Shenzhen and Guangzhou. *Atmospheric and Oceanic Science Letters* 2018;11:330–7.
- [4] Demuzere M, Kittner J, Martilli A, Mills G, Moede C, Stewart ID, et al. A global map of local climate zones to support earth system modelling and urban-scale environmental science. *Earth Syst Sci Data* 2022;14:3835–73.
- [5] Sun X, Huang X, Mao Y, Sheng T, Li J, Wang Z, et al. GABLE: A first fine-grained 3D building model of China on a national scale from very high resolution satellite imagery. *Remote Sensing of Environment* 2024;305:114057.
- [6] United States Geological Survey (USGS). Landsat 8-9 Collection 2 Level-1 Data 2021.
- [7] Qian J, Zhang L, Schlink U, Meng Q, Liu X, Jansc  T. High spatial and temporal resolution multi-source anthropogenic heat estimation for China. *Resources, Conservation and Recycling* 2024;203:107451.
- [8] United States Geological Survey. NASA Shuttle Radar Topography Mission Global 1 arc second [Data set] 2013.
- [9] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;2:56–67.