



Decoupling the multi-drivers of urban extreme heat environment in urban agglomerations using ensemble learning

Xiaochang Liu^b, Zhiyu Liu^c, Zhiliang Zhu^d, Renlu Qiao^{a,*}

^a Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, 1239 Siping Road, Shanghai, PR China

^b School of Urban and Regional Science, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, PR China

^c Shanghai Tongji Urban Planning & Design Institute, 1111 Zhongshan Beiyi Road, Shanghai, PR China

^d College of Environmental Science and Engineering, Tongji University, 1239 Siping Road, Shanghai, PR China



ARTICLE INFO

Keywords:

Urban extreme heat
Urban morphology
Natural and climatic factors
Heterogeneity
Urban agglomerations
Ensemble learning

ABSTRACT

Urban extreme heat (UEH) has been imparting a formidable menace to various facets of sustainable urban and regional development. While increasing studies have examined the urban thermal environment, a gap exists in decoupling the nonlinear multi-drivers of UEH environment in intra-urban regions, especially across the urban agglomerations. In this study, we estimated UEH through the land surface temperature (LST) in the extended summer season exceeding a specific threshold computed through cross-regional statistics. Furthermore, an additive interpretable ensemble learning model enhanced via the Bayesian Optimization algorithm (BO) and Monte Carlo Simulation framework (MCS) was employed to dissect the intricate nonlinear interplay between UEH and intra-urban morphology, green-blue composition, and atmospheric determinants. Our results highlight the indispensably suppressive effect of green and blue factors on UEH, whereas urban morphological variables generally exhibit an opposite trend. Interestingly, we reveal heterogeneity responses of building heights, elevated urban structures mitigate UEH in temperate monsoon and subtropical monsoon zones and exhibit diminished marginal utility in temperate continental zones. Furthermore, the arid climatic characteristics are expected to exert an unexpectedly enhancing effect on water bodies' cooling capacity. These findings provide heterogeneous local guidance for landscape and urban planning towards the severe challenges of climate change.

1. Introduction

There is overwhelming evidence that urban extreme heat (UEH), one of the deadliest consequences of climate change [1], are exhibiting an alarming increase in frequency and intensity worldwide (W. [2]). These extreme heat events (EHEs) have been poised to impart a formidable menace to various facets of natural ecosystems, socio-economic stability, and human well-being, leading to irreversible ecological and societal impacts [3]. Simultaneously, urbanization has become an inevitable global trend, primarily characterized by the continuously radical expansion of impervious surfaces across urban regions (e.g., urban agglomerations) (Y. [4]), which has profoundly impacted the surface energy balance during this rapid urbanization (X. [5]), are likely to amplifies the deleterious impacts of UEH [6]. To this end, it is imperative to recognize the spatiotemporal characteristics of UEH and its influencing factors, which can provide insights for urban planners to achieve the sustainable urban and regional development.

Numerous existing studies have explored the spatiotemporal characteristics of overall land surface temperature (LST) or near surface air temperature (NSAT) under disparate seasons or the urban/regional heat islands (UHIs/RHIs) effects following an "urban-rural" dichotomy where urban areas are treated as a whole against its non-urban surroundings ([7,8]; L. [9]), but it tends to overlook the sudden UEH attributed to EHEs that occur on a daily or weekly basis [10], and ignore the heterogeneous intra-urban thermal patterns (H. M. [11]). In contrast, an atmospheric hazards like UEH and the like perspective [12], extracting diurnal LST/NSAT exceeding a certain threshold climatology of the corresponding calendar day in the extended summer season (W. [2]), thus shedding light on the inherent dynamic nature of urban temperature fluctuations, is particularly valuable for exploring the morphological factors, spatial composition, and the impact of anthropogenic heat in diverse urban and reginal land use scenarios. Moreover, considering the above limitations of existing UHIs/RHIs-based studies (D. [13]), an intra-urban perspective, regarding as a break away from the

* Corresponding author.

E-mail address: qiaorl@tju.edu.cn (R. Qiao).

"urban-rural" dichotomy, thus largely weakening the interference from the diversified cooling effects and types of rural land covers, may enable generality and difference recognition of UEH problems and provide new insights into the urbanization imprint on UEH environment. Therefore, we may utilize the various intra-urban drivers within urban agglomerations as natural laboratories to identify both problems and solutions to UEH mitigation and adaptation [14].

Most of existing UEH environment studies primarily center on using meteorological data, satellite data and basic geographic information data to estimate UEH and analyzing its spatiotemporal trends in frequency and intensity and duration ([15]; W. [16]; W. [17]). Several definitions of UEH indices have been predominantly proposed: TX95p relevant for the frequency of hot extremes and defined as the 95 percentage of hot days per month in the hottest periods; TX7d defined as the mean value of daily maximum temperatures over the hottest week per year and relevant as a measure of the intensity of hot extremes [18]; The number of consecutive temperature exceeding the threshold to calculate relevant as a measure of the duration of hot extremes. Currently, there is no uniform standardization in defining a temperature threshold to characterize a UEH. It is well known that an absolute threshold can't be used directly elsewhere, a percentiles threshold is not always representing a UEH, even for rare historical observations of maximum temperatures [12]. Thus, it is urgent to develop a cross-regional statistics to reflect both the UEH excluding the influence of abnormal values and the overall temperature of the study area.

To date, the multi-drivers of UEH across urban agglomerations studies are somewhat scarce. Existing research indicates that it is feasible to optimize the urban built environment [19], urban ecological infrastructure [20,21], and climatic and human activities characteristics (J. [22]) to alleviate persistent UHE and enhance urban thermal environment comfort. Hence, numerous antecedent scholarly inquiries have focused on the influences of plane (i.e., building density) and vertical (i.e., building height, floor area ratio, sky view factor) morphological variables and blue-green spaces [23] on LST, and these effects varied depending on the climatic region of the urban regions. However, the majority of the above explorations have been conducted in the setting of non-UEH, let alone grid-based microscale analyses of the comprehensive impacts of intra-urban multi-drivers on UEH environment. In addition, despite the extensive investigations into various drivers affecting surface temperature, many studies have primarily relied on simple linear regression techniques such as ordinary least squares (OLS), spatial autocorrelation (SAC) and geographically weighted regression (GWR) ([24]; L. [25]) to assess multiple contributors. Due to the mathematical simplicity of these models, the conventional methodologies may manifest intrinsic constraints and overlook nonlinear interplay between urban variables and surface temperature [8]. Despite some efforts to address this through artificial neural network (ANN) or partial dependence plots (PDP), the interpretation of these methods proves to be an exceedingly challenging endeavor and tends to overlook the holistic correction of variables (H. M. [11]).

To address these gaps above mentioned, the present study aims at large scale across the urban regions, and investigates the nonlinear impacts of the intra-urban multi-drivers on UEH environment. In response to this aim, it is specifically needed to addresses the following aspects: (1) developing a method to assess UEH to reflect both the extreme heat excluding the influence of abnormal values and the overall temperature of the study area. (2) architecting an ensemble learning model to decouple the intricate relationship between intra-urban multi-drivers and UEH environment. (3) utilizing Shapley Additive exPlanations (SHAP) algorithm to meticulously interpret the stress thresholds of the intra-urban multi-drivers so that their marginal utility can be captured. The remainder of this paper describes the methodology (Section 2), results (Section 3), discussions (Section 4), and conclusions (Section 5).

2. Methodology

2.1. Study area

It is well-documented that the overall risk of UEH was the highest in East China (L. [26]), the frequency of UEH was the highest in South China, but the intensity was higher in North China than in South China, which presents the spatial differences in UEH environment characteristics, the duration of UEH was the highest in South China. Typically, the frequency, intensity and duration of UEH increased significantly, accompanied by high frequency, intense intensity and long-lasting in the most urbanized East, South and North China characterized by Beijing-Tianjin-Hebei (BTH), Yangtze River Delta urban agglomerations (YRD), the Greater Bay Areas (GBA) urban agglomerations, respectively (L. [27]). Therefore, taking these three major urban agglomerations as testbeds will acquire a more significantly profound understanding of challenges, mechanisms and solutions associated with UEH.

The geographical scope of this study encompasses the 3 major urban agglomerations in China. Notably, the spatial expansion rate of official urban agglomerations has been largely under-estimated, the definition of the boundaries of the 3 major urban agglomerations in this study is identified via natural cities method (H. M. [11,28]) (Fig. 1). Leveraging land cover classification maps sourced from the European Space Agency (ESA) [29], the delineation of intra-urban built-up area boundaries for the three major urban agglomerations was conducted based on the ESA class of built-up areas (<https://www.esa.int/>), including the core areas (CA) and fringe areas (FA) as shown in Fig. 2 b. Among them, CA is the most contiguous group of built-up pixels ($\geq 5 \text{ km}^2$) in which at least 50 % of the 5 km circle surrounding neighborhood is built-up area. FA consists of the built-up pixels surrounding the CA with neighborhoods that are 20–50 % built-up areas, and the non-built-up pixels surrounding the CA with neighborhoods that are 20–50 % built-up areas ([30]; Z. [31]).

Subsequently, within the demarcated intra-urban built-up areas across the three major urban agglomerations, a fishnet system was established and the grid unit was set at $1\text{km} \times 1\text{ km}$. Eventually, based on Köppen-Geiger climate classification maps at 1 km resolution [32], the BTH (10 cities, comprising 51135 grid units) spans both temperate monsoon climate and temperate continental climate, the YRD (15 cities, comprising 67730 grid units) encompasses both temperate monsoon climate and subtropical monsoon climate, and the GBA (11 cities, comprising 37635 grid units) has a subtropical monsoon climate.

2.2. Data

2.2.1. Estimate of UEH

Aiming at a cross-regional statistics, the methodologies of both TX95p and TX7d have been integrated to estimate UEH through calculating the mean of LSTs exceeding a certain threshold in this study: (1) Initially, utilizing the method of TX95p, we considered extreme heat as the occurrences wherein the maximum LST of a given of a specific sampling grid unit above the calendar day 95th percentile for the baseline period, defined in this study as the first 18 years varying from 2000 to 2017 of the datasets (L. [31]; W. [2]), to reflect the highest temperature in the study area and exclude the influence of abnormal values; Remarkably, the 95th percentile holds paramount significance, as numerous empirical studies have substantiated a notable increase in population mortality rates when temperatures surpass it [33–36]. Notably, various thresholds could validate the robustness of the findings, as illustrated in the Appendix. Supplementary materials. (2) Concurrently, employing the method of TX7d, we computed the mean value of daily maximum LST observed in the hottest week per year for the baseline period to reflect the overall temperature of the study area. (3) The mean of (1) and (2) was defined as a threshold of UEH and was calculated separately in the June, July and August, the extended summer season experiencing the most frequent occurrences of extreme heat regarded as the urban extreme heat quarter (UEHQ) on the basis of

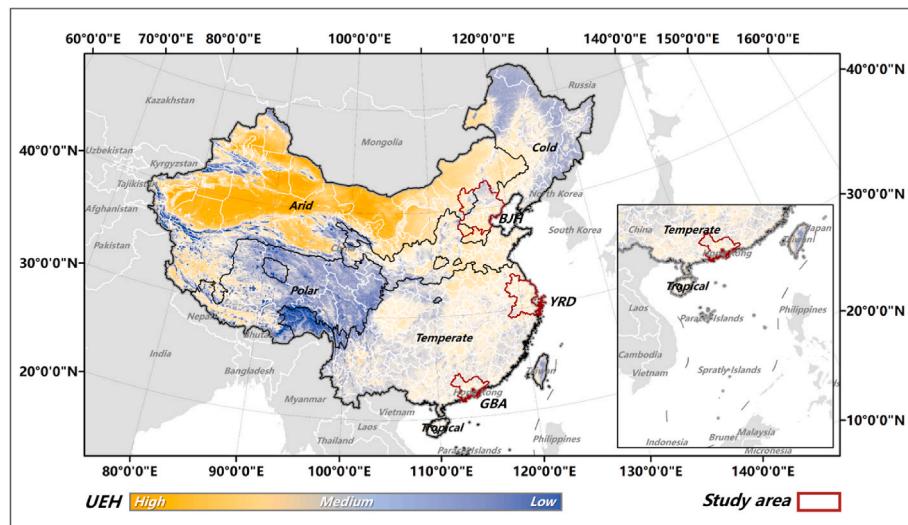


Fig. 1. The geographical features of the study area, the three main urban agglomerations.

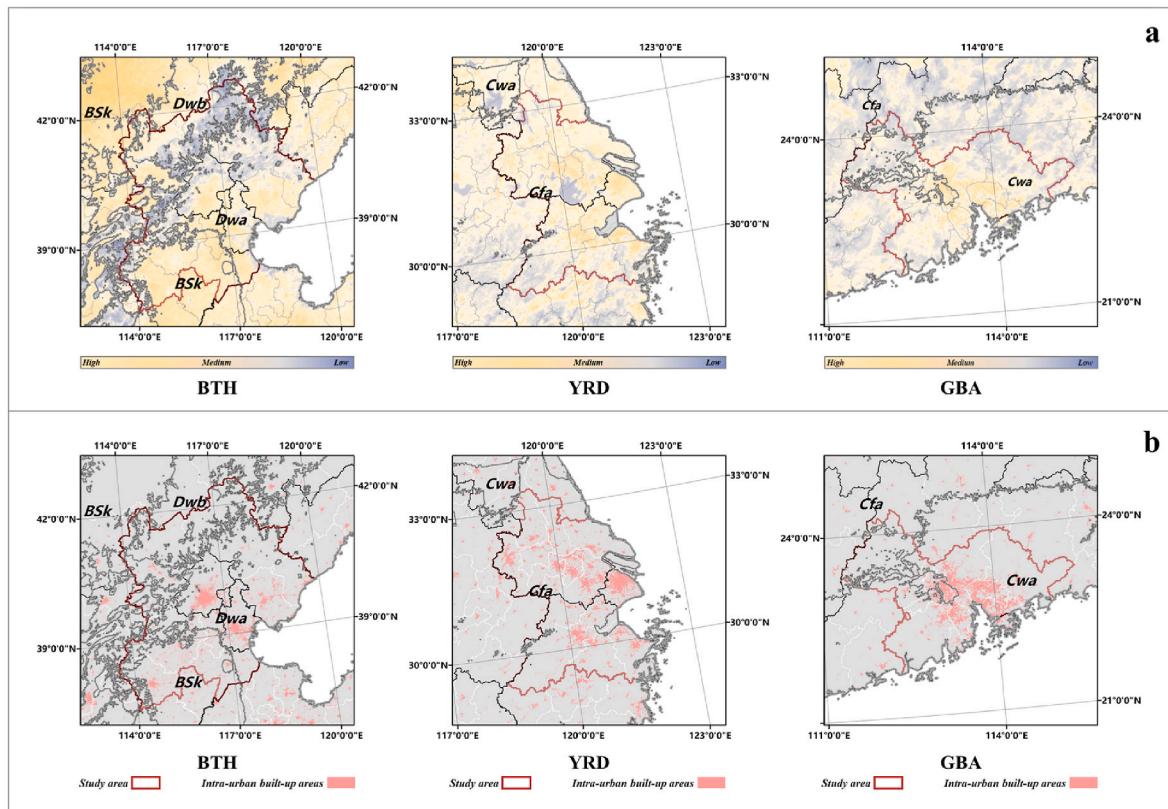


Fig. 2. The intra-urban built-up areas of the three main urban agglomerations with sub-climate classifications and UEHs. a-b. the UEH spatial distribution and the intra-urban built-up areas in BTH, YRD and GBA with sub-climate classifications.

observations, during the 5-years temporal scope spanning from 2018 to 2022 (Fig. 2_a).

This research utilized the EOS-Aqua-MODIS V5 composite product (MOD11A2) to guarantee uniformity in observational data. This product offers continuous diurnal LST data, featuring a spatial resolution of 1 km. Significantly, this remote sensing tool has improved the accuracy of LST measurements. It achieves this by correcting inconsistencies caused by various factors, including cloud interference, topographical differences, and fluctuations in the zenith angle.

2.2.2. Urban morphology

Several previous studies have demonstrated that urban morphology has a substantial impact on the urban thermal environment (H. M. [8,11, 19]). This study leverages the European Commission Global Human Settlement Layer (GHSL) dataset (<https://ghsl.jrc.ec.europa.eu/>), which provides application programming interfaces for retrieving vector data of built-up environment factors such as building footprints at a spatial resolution of 100 m. The urban morphology indicators consist of 2D and 3D indexes, including building densities (BD), building height (BH) and floor area ratio (FAR) [8]. Among that, BD determines how crowded or

built-up a neighborhood appears, which is measured by building square footage divided by the unit grid area. BH refers to the vertical distance measured from the ground to the highest point of the roof surface, which is the average building height in the unit grid area. FAR is the ratio of a building's total floor area to the size of the piece of land upon which it is built [37], indirectly reflecting the extent of various man-made functions implemented on a unit of land, i.e., the development intensity of the land (C. [4]). The calculation of FAR is articulated as follows:

$$FAR_i = \frac{\sum_{j=1}^n (BSF_j \times NF_j)}{UA} \quad (2-1)$$

where BSF_j is the building square footage of building j in the grid unit i ; NF_j is the number of floors of building j in the grid unit i ; n is the numbers of buildings in the grid unit i ; UA is the unit area of the fishnet in the study.

2.2.3. Natural environment

Extant literature has emphasized the decisive role of urban natural environmental factors in ameliorating urban microclimate environment, with a particularly noteworthy emphasis on the significantly enhanced capacity of green and blue spaces [8,38,39].

For the green spaces, the normalized difference vegetation index (NDVI) derived from MODIS dataset (MOD13A3) has been extensively employed in numerous statistical studies to quantify vegetation greenness and is useful in understanding vegetation density, given that the cooling effects of vegetation transpiration and canopy shading from intense sunlight radiation can contribute to the improvement of the urban thermal environment (Z. [40]). We computed the median values of NDVI in the grid unit, thus encapsulating both long-term and short-term greening dynamics. For the blue spaces, water bodies are considered an effective factor in mitigating RHIs effects in the urban regions [41]. We analyzed the potential influence of the cooling effects of water bodies on the LST based on the water-bodies density (WD) (X. [42]), which refers to the proportion of water-bodies area within the assessed grid unit, and data is derived from the ESA's land cover datasets (<https://www.esa.int/>) including the classification of water-bodies.

2.2.4. Climatic and human activities characteristics

Besides the green and blue factors, several instrumental variables characterizing climatic condition of urban agglomerations are equally crucial. Therefore, we adopted the TerraClimate (<https://www.climateologylab.org/terraclimate.html>) to acquire data of precipitation (Prcp), evapotranspiration (ETa) and wind speed (WS). TerraClimate uses climatically aided interpolation, combining high-spatial resolution climatological normals from the WorldClim dataset, with coarser spatial resolution, but time-varying data from CRU Ts4.0 and the Japanese 55-year Reanalysis (JRA55). Conceptually, the procedure applies interpolated time-varying anomalies from CRU Ts4.0/JRA55 to the high-spatial resolution climatology of WorldClim to create a high-spatial resolution dataset [43].

Climate change is caused by human activities and threatens life on earth as we know it [1]. Urbanization, regarded as the only a pivotal of human activities, contributes to more than one third of the increase of intensity of UEH in the region (X. [44,45]). Therefore, we designated urbanization rate (Urb.) as a covariate, which was sourced from the China City Statistical Yearbook (<https://kns.cnki.net/>).

2.2.5. Data extraction with ArcGIS and processing with fixed effect model

It is worth noting that we used ArcGIS to establish a fishnet covering all study areas to ensure the consistency of the spatial framework, and the unit area of the fishnet is set at $1\text{km} \times 1\text{ km}$. Through this, those rasters of 10 variables in this study as illustrated in Table 1 were resampled to align with our established sampling fishnet.

Two-way fixed effects model was adopted to adjust for unobserved

Table 1

The description of the variables in this study.

Variables	Description	Unit
UEH	urban extreme heat	°C
Urban morphology		
BD	The building density of the sampling grid	%
BH	The average building height of the sampling grid	m
FAR	The floor area ratio of the sampling grid	-
Natural environment		
NDVI	The median of NDVI values within the sampling grid	-
WD	The water-bodies density of the sampling grid	%
Climatic and human activities characteristics		
Prcp	The median of precipitation values within the sampling grid	mm
ETa	The median of evapotranspiration values within the sampling grid	mm
WS	The median of wind speed values within the sampling grid	m/s
Urb	The urbanization rate within the sampling grid	%

unit-specific and time-specific confounders at the same time in this study [46]. Among them, one of the most influential factors in the UEH environment is the climatic background of a region. The heterogeneity of climates introduces a layer of complexity when analyzing the multi-drivers of UEH, as it means that urban spaces, even if they share similar ecological or built-up characteristics, can exhibit vastly different impacts based on their climatic context [47]. Therefore, considering the climatic classifications as a unit-specific, temporally invariant characteristics across the three urban agglomerations (Fig. 2_b), we transformed the individual grid observations of all variables into deviations from the mean within their respective climatic classification groups, including Arid-steppe-cold (BSk), Cold-dry winter-warm summer (Dwb), Cold-dry winter-hot summer (Dwa), Temperate-dry winter-hot summer (Cwa), and Temperate-no dry season-hot summer (Cfa). This procedure was implemented to mitigate endogeneity issues arising from unit-specific fixed effects, thereby achieving comparability among different urban regions.

2.3. Methods

The methodology flowchart describes the analysis of spatial features of UEH and the use of an interpretable ensemble learning model optimized via BO algorithm and MCS framework to understand the impact of multi-drivers on UEH. Fig. 3 shows the overall flow of the study.

2.3.1. Spatial features of UEH

Several preceding studies have consistently illustrated that LST adheres to the Waldo Tobler's first law of geography regarded as the fundamental assumption used in all spatial analysis [48]. The spatial association magnitude between urban thermal environment and various factors may exhibit distinct patterns across adjacent grid units. Global Moran's I and local Moran's I are widely employed methodologies for quantifying the extent of this spatial interdependence (R. [49]).

2.3.2. Decoupling model of UEH

UEH are influenced by a highly complex set of factors, such as urban morphology, natural environment, climatic and human activities characteristics. It is challenging to adequately elucidate the intricate multi-driving mechanism of UEH with a basic weak learner. UEH data usually consists of only a few thousand observations, exacerbating the problem of the curse of dimensionality. This phenomenon occurs when models lose accuracy as the number of variables increases [8]. We used an ensemble learning algorithm to obtain a driver model that can overcome the aforementioned limitation. The ensemble learning algorithm combines the predictions of multiple weak learners to improve the model's accuracy and robustness. This approach results in better learning performance than a single weak learner, and the model has excellent generalization performance and is less prone to overfitting. Additionally, the model is insensitive to outliers and has good noise immunity, making it suitable for analyzing complex datasets with high

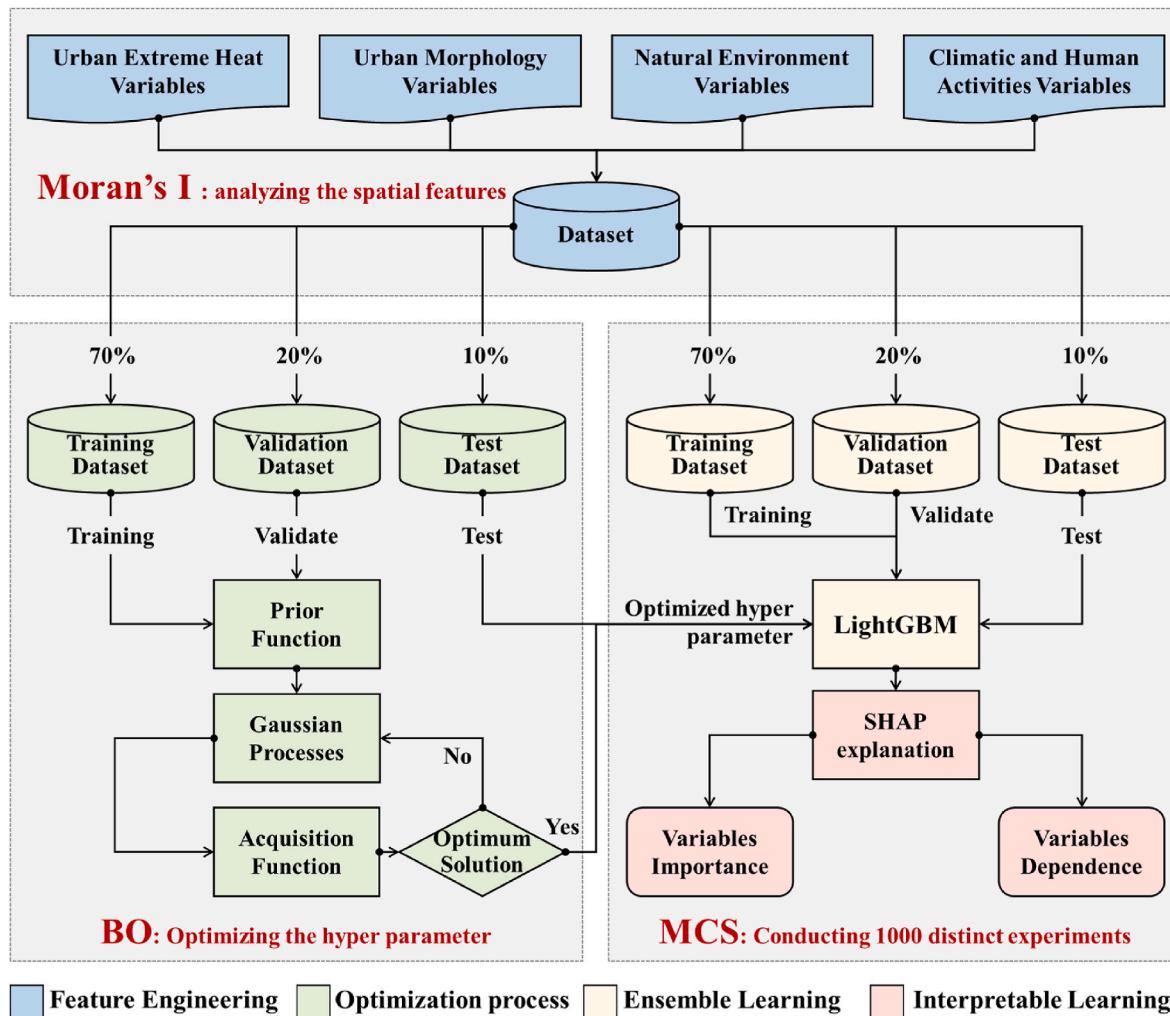


Fig. 3. The methodology flowchart of this study.

variability ([50]; R. [51]; R. [8,49]).

Gradient Boosting Decision Tree (GBDT) stands as a typical model in the field of machine learning ensemble techniques. Its primary strategy involves fitting data using numerous weak learners and decision trees, ultimately leading to the highest model quality. However, GBDT necessitates multiple passes through the dataset at each stage of the fitting procedure. Therefore, GBDT may encounter constraints and fail to reach its full potential in the context of extensive datasets related to UEH [52].

In response to the rapid expansion of data volume and dimensionality, the widely acknowledged and highly robust Light Gradient Boosting Machine (LightGBM) has emerged as an evolution of the Gradient Boosting Decision Trees (GBDT) [52]. Compared to GBDT, the LightGBM regarded as GOSS_EFB_GBDT introduces three salient enhancements.

a) Gradient-based One-Side Sampling (GOSS) algorithm

GOSS is a method that leverages the fact that there is no native weight for data instance in GBDT. Since data instances with different gradients play different roles in the computation of information gain, the instances with larger gradients will contribute more to the information gain. To retain the accuracy of the information, GOSS keeps the instances with large gradients and randomly drops the instances with small gradients [52].

b) Exclusive Feature Bundling (EFB) algorithm

EFB is a near-lossless method to reduce the number of effective features. In a sparse feature space, many features are nearly exclusive, implying they rarely take nonzero values simultaneously. One-hot encoded features are a perfect example of exclusive features. EFB bundles these features, reducing dimensionality to improve efficiency while maintaining a high level of accuracy. The bundle of exclusive features into a single feature is called an exclusive feature bundle [52].

c) Histogram Technique

This approach bolsters model generalization by transmuting continuous data splits into their discrete counterparts.

The ensemble learning approach enables the model to learn from multiple individual weak learners and integrate their results, leading to better generalization performance and reducing the risk of overfitting. The model is also robust to outliers and noise in the data, making it a reliable tool for decoupling the multi-drivers of UEH.

2.3.3. Decoupling model optimization of UEH

LightGBM possesses numerous hyperparameters, making it extremely challenging to optimize the model's parameters through the grid search method, which could potentially waste a substantial amount of computational power. As a result, we employ the Bayesian optimization (BO) to adjust hyperparameters of LightGBM. Prior function (PF) and acquisition function (AF) are the two essential indications of the process [53]. The Gaussian processes are commonly employed to

construct the PF by computing the mean and variance of each hyper-parameter at each observation point. In optimization, selecting points with a higher mean value is termed exploitation, while choosing points with a higher variance value is referred to as exploration. The objective of the AF is to weigh and evaluate the balance between exploitation and exploration. The calculation of the AF is articulated as follows:

$$AF(x) = \begin{cases} (\mu(x) - f(x^+))\Phi(Z) + \sigma(x)\varphi(Z), & \sigma < 0 \\ 0, & \sigma \geq 0 \end{cases} \quad (2-1)$$

$$Z = \frac{\mu_x - f(x^+)}{\sigma} \quad (2-2)$$

In this calculation, x is the observation point; $\mu(x)$ is the mean value of observation points; $\sigma(x)$ is the variance value of observation points; $f(x^+)$ is the present maximum value of the function; $\Phi(Z)$ is the cumulative distribution function (CDF) of the standard normal distribution; $\varphi(Z)$ is the probability density function (PDF) of the standard normal distribution.

2.3.4. Interpretation model of UEH

To clarify the predictive outcomes of the BO-based ensemble learning model, we utilized the Shapley Additive Explanation (SHAP) algorithm. The goal of SHAP is to explain the prediction of an instance by computing the contribution of each feature to the prediction. The SHAP explanation method computes Shapley values from coalitional game theory. The feature values of a data instance act as players in a coalition. Shapley values tell us how to fairly distribute the "payout" (the prediction) among the features. A player can be an individual feature value, e.g. for tabular data. A player can also be a group of feature values. For example, to explain an image, pixels can be grouped to super pixels and the prediction distributed among them [54]. One innovation that SHAP brings to the table is that the Shapley value explanation is represented as an additive feature attribution method, a linear model. That view connects LIME and Shapley values. The calculation is articulated as follows:

$$y_i = y_{base} + \sum_{j=1}^k SHAP(x_{i,j}) \quad (2-3)$$

In the calculation, y_i is the predictive value of the model at sample i ; y_{base} is the mean value of the predictive value at other samples; $SHAP(x_{i,j})$ is the SHAP value of the feature j at sample i ; k is the number of features. Hence the positive and negative of $SHAP(x_{i,j})$ characterize the specific influence between the predominant factors and UEH at per-sample points in the study area.

2.3.5. Enhancement of interpretation reliability

This study employs a decoupling model constructed by using ensemble learning and explainable algorithms, providing an effective approach to elucidate the nonlinear multi-driving mechanisms of UEH. However, it is crucial to acknowledge that, like all machine learning models, its predictions may be influenced at times by inherent uncertainties in the data or model [55]. Therefore, we employed Monte Carlo Simulation (MCS) to enhance the robustness and reliability of the interpretations. MCS methods are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results. The underlying concept is to use randomness to solve problems that might be deterministic in principle [56]. Hence, we conducted 1000 distinct experiments on the decoupling model, maintaining consistent parameters for each experiment but varying the partitions in their respective training and testing datasets. Subsequently, we aggregated predictions from all models and utilized the SHAP algorithm to generate more stable and compelling interpretive outcomes.

3. Results

3.1. UEH spatiotemporal pattern

Based on LST data in the UEHQ spanning from 2018 to 2022, we comprehensively estimated the spatial distribution of UEH across the 3 major urban agglomerations. Upon employing the global Moran's I test to scrutinize the spatial distribution features of UEH further, a significant spatial autocorrelation was discerned. The global Moran's I value was 0.92 in BJH, 0.85 in YRD, and 0.87 in GBA, and the associated p-values were all below 0.01 in three clusters, these observational findings substantiated the intrinsic spatial clustering of UEH.

Upon adopting the more granular local Moran's I test, the study area was segmented into five clusters: High-High (HH), High-Low (HL), Low-High (LH), Low-Low (LL), and Not-Significant (NS) regions. Especially, HH, LL, NS were three predominant ones across three agglomerations (Fig. 4_a). In the case of BJH with significant north and south divergence, the HH was predominantly situated in the southern North China Plain, and the northern exhibited distinctive characteristics of LL. As for YRD, the HH exhibited a zigzag pattern, which intriguingly correspond to the distinctive Z-shaped distribution with the highest economic and demographic density, and the LL regions were predominantly situated within the southern and western topographically undulating terrain characterized by high NDVI. Besides, GBA was marked by central HH and peripheral LL.

3.2. UEH multi-drivers and their nonlinear impacts

UEH, regarded as an urban thermal environment metric, is subject to the influence of multiple determinants with a complex interplay, constituting both natural and anthropogenic features. In order to decouple the complex interplay, this study constructs BO_LightGBM ensemble learning model, thereby elucidating the intricate couplings between UEH and its natural or anthropogenic factors. Furthermore, we harness the SHAP algorithm to attain the nonlinear mechanisms of precisely quantitative interpretation. The results of the nonlinear decoupling indicate that those metrics of intra-urban built-up characteristics contribute significantly and conspicuously to the model's efficacy (Fig. 5_j).

Among the morphological parameters intra-urban built-up areas accounting for 30.28 %, FAR is the most prominent determinant, followed by BD, with BH being the ultimate factor, respectively. FAR exerts a significantly positive facilitating effect on UEH within the range of -0.79 to 0.5 . With each 1.0 increase in FAR, there is a significant surge of 0.89 °C in the associated UEH, which underscoring its role as an intra-urban built-up morphological dominant determinant. Once the mean FAR surpasses the 0.5 threshold, there is a significant diminishing marginal utility on UEH. In contrast to FAR, the positive trend of BD is somewhat attenuated, mainly manifesting as a progressively increasing return with some pronounced fluctuations, there is a pronounced surge of 0.14 °C in the associated UEH with each 5.0 % increment spanning -15.0 %– 12.0 % (Fig. 5_b). As for the BH variable, the impact exhibits an overall inverted U-shaped pattern. Intriguingly, once the mean BH surpasses the -1.43 m threshold, there is a significant negative impact on UEH (Fig. 5_c). On the whole, the overall influence of urban morphological elements on UEH appears relatively weaker than natural, climatic and human activities characteristics. This form of disparateness may be attributed to the predominant role of urban regional ecological and atmospheric determinants, which often overshadow the cumulative effects at the macro scale (Z. [57]).

Notably, the collective relative importance of NDVI and WD stands at 37.42 % (as illustrated in Fig. 5_j), which underscores the pronounced influence of intra-urban natural environment factors, precisely reflecting the mechanisms of UEH. Evidently, NDVI stands out as the preeminent determinant on UEH, corroborating with previous studies and indicating that green spaces can effectively reduce land surface

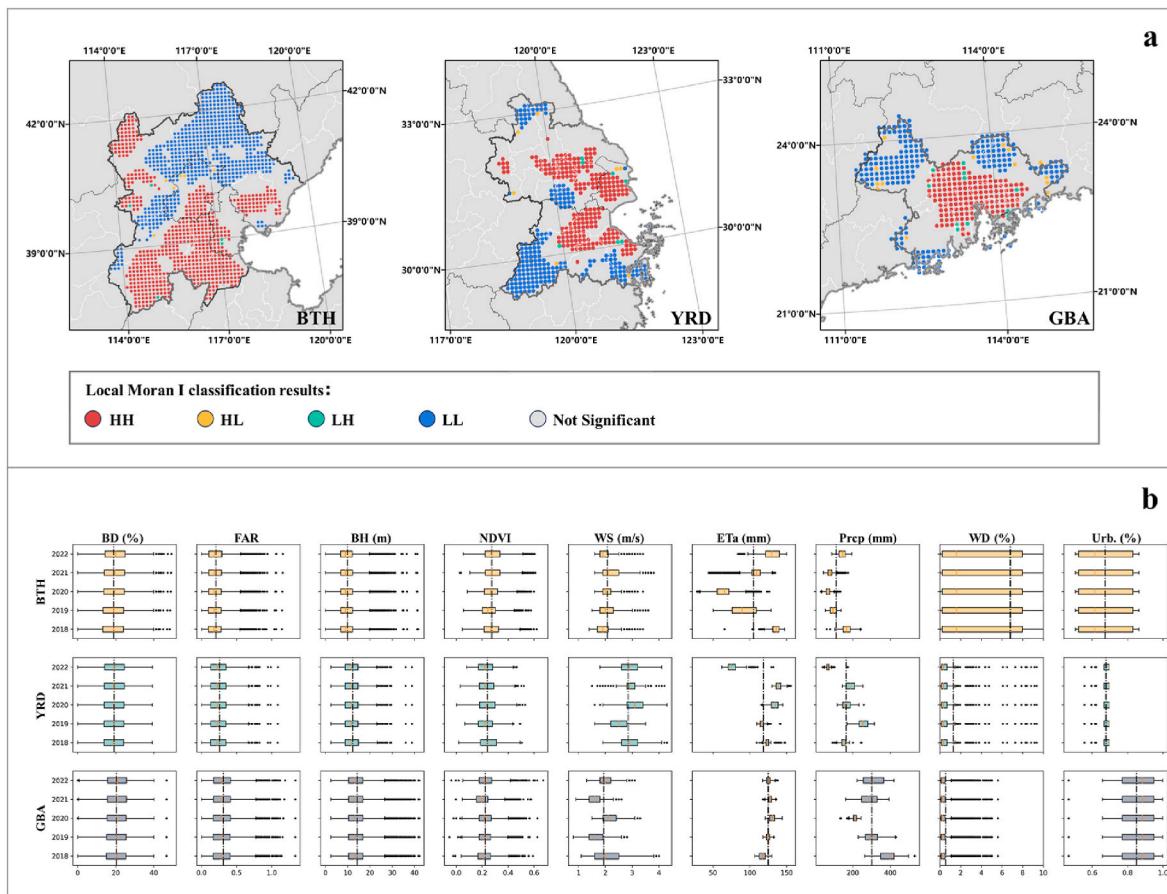


Fig. 4. The spatiotemporal distribution of UEH factors across three urban agglomerations

a. the UEH spatial distribution of Local Moran's I in BTH, YRD and GBA; b. the temporal distribution of UEH determinants.

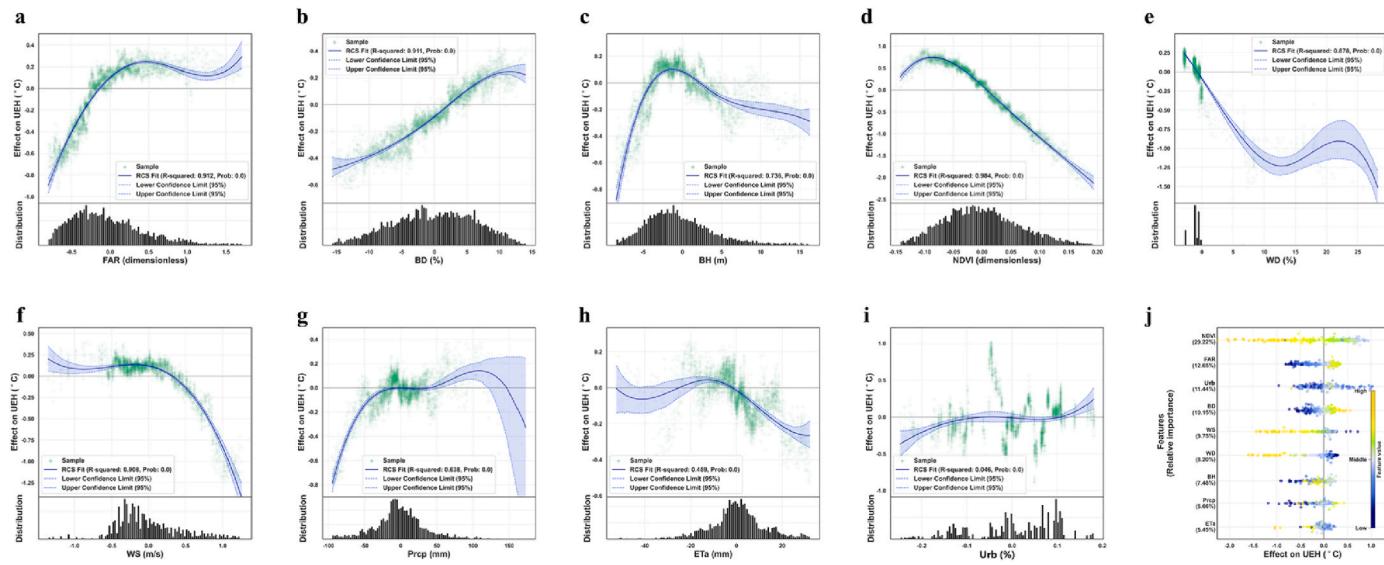


Fig. 5. The effect of per feature on UEH

a. FAR; b. BD; c. BH; d. NDVI; e. WD; f. WS; g. Prep; h. ETa; i. Urb.; j. the overall ranking of each feature's effect (SHAP values) on UEH. Note: in Subfigures a-i, the X-axis shows the sample feature value after within-group deviation, and the Y-axis shows the SHAP value of the sample feature, i.e., the local effect on UEH. In Subfigure j, the relative importance of each feature expressed as the ratio of each feature's mean absolute SHAP value to their total value.

temperature extremes [6,58]. Precisely, as inferred from the regression analysis, in the range of NDVI from -0.08 to 0.2 , a surge in NDVI by one single 0.1 results in a swift reduction of 1.07°C in the observed UEH

(Fig. 5-d). Although WD exhibits a trajectory analogous to NDVI, the transition of WD from -3.2% to 13.1% witnessing a rapid UEH decrease from 0.25°C to -1.24°C , slightly diverging, the returns exhibit

diminishing returns as WD increases, upon reaching the threshold of +13.1 % (Fig. 5_e).

Simultaneously, WS exhibits a trend akin to that of NDVI, indicating that its augmentation can significantly mitigate local UEH. Upon exceeding -0.17 m/s, each subsequent 1 m/s increase results in a corresponding reduction of 1.00 °C in UEH (Fig. 5_f). It is worth mentioning that ETa surpassing the -10 mm and UEH exhibit a pronounced negative correlation within the 95 % confidence intervals (Fig. 5_h), aligning with numerous studies, the urban landscape evaporation capacity ascending is a primary driver of observed urban temperature cooling [47]. Meanwhile, Prcp surpassing the 100 mm exhibits a tendency analogous to ETa, showing that its amplification may mitigate local instances of UEH (Fig. 5_g). Urb., serving as a covariate in this study, contributes 11.44 % to the decoupling model, thus emerging as a significant factor, consistent with prior research findings (X. [44]).

3.3. Three urban agglomerations heterogeneity of UEH

The intricate couplings between different urban agglomerations and

UEH manifest unique characteristics. To unveil the underlying multi-drivers of these distinctions, we architect three decoupling models of UEH, each grounded in the methodology described earlier, to elucidate the heterogeneity features of BTH, YRD, and GBA.

As shown in Fig. 6, across distinct urban agglomerations, the ameliorative regulatory role of green and blue determinants on UEH manifests conspicuous heterogeneity. As a significant factor, the relative importance of the variable NDVI surpasses the threshold of 20 % across the three regional groups and consistently holds the top position in the hierarchy of contributions among all factors.

In contrast to NDVI, the influence exerted by variable WD on UEH appears comparatively subdued, with its relative importance typically ranges from 8 % to 10 %. However, an exception to this trend is observed in the BTH characterized by a pronounced scarcity of water resources [8], where its relative importance roughly doubling this benchmark, reaching 16.10 %. Intriguingly, as the WD continues to increase, surpassing a certain range, the declining trend of intra-water body temperature gradually becomes gentle, exhibiting conspicuous oscillations in its impact on UEH. From the point of view of cooling benefit, it is

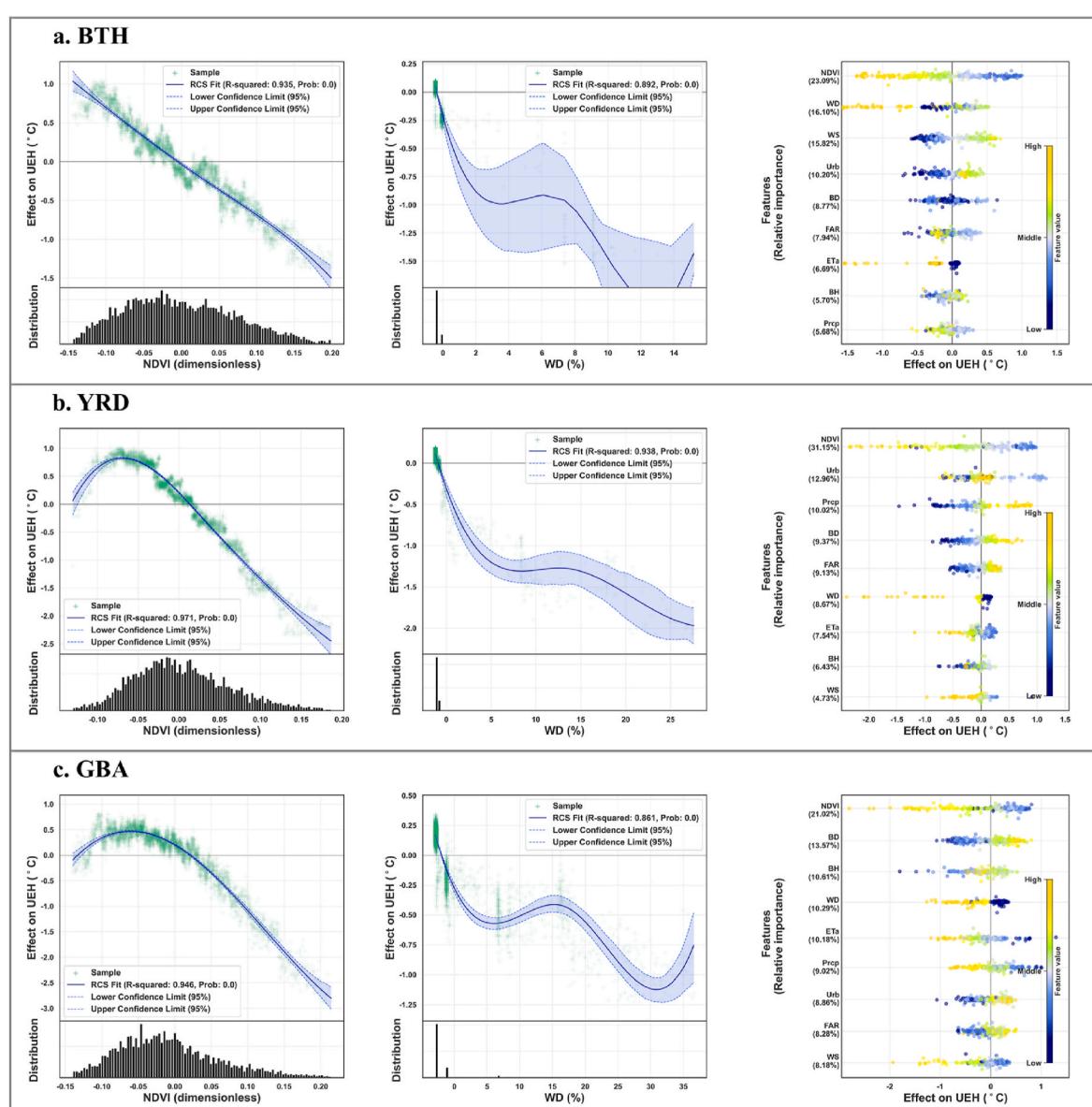


Fig. 6. The effect of per urban natural feature on UEH
a-c. the SHAP values of urban natural features in BTH, YRD and GBA.

reasonable to control the WD within a certain threshold to effectively increase the cooling effect intensity. The thresholds vary for each urban agglomeration, with BTH, YRD, and GBA set at $[-0.5\%, +13.6\%]$, $[-1.7\%, +27.5\%]$, and $[-0.3\%, +31.3\%]$ respectively.

As illustrated in Fig. 7, there exist subtle variations in the driving patterns of intra-urban built-up factors among different urban agglomerations. Among that, while there are some fluctuations in the impact curve of BD on UEH, the trends are generally consistent. With regard to variable FAR, analogous observations are applicable. It is imperative to underscore that BD emerges as a highly significant urban morphological determinant in the three decoupling models, which is particularly striking. The feature importance of BD is relatively weaker in BTH, but stronger in YRD and GBA, even surpassing that of natural and climatic characteristics such as WD, ET_a, and WS. More precisely, for every increment of 5 % in BD from to -10.0% to $+15.0\%$, the UEH of YRD and GBA both undergoes an average augmentation of $0.24\text{ }^{\circ}\text{C}$. Excessive BD results in poor ventilation, thereby hindering the efficient dispersion of heat (S. [11]). The positive trend of FAR is somewhat attenuated, mainly manifesting as a progressively increasing return with some

pronounced fluctuations within the $[-0.7, 0.5]$. Here, in the range of FAR between -0.7 and 0.5 , there is a similar elevation in $0.22\text{ }^{\circ}\text{C}$ for every 0.5 increase of FAR among the three urban agglomerations.

During the investigation into the impact of BH on UEH across the three urban agglomerations, conspicuous variations in heterogeneity emerge. The feature relative importance of BH is highest in GBA, reaching 10.61% , followed by YRD at 6.43% , and finally in BTH, amounting to 5.70% . More precisely, when the building elevation exceeds the average in BTH, the impact on UEH is negligible, exhibiting neither enhancement nor attenuation but oscillating around the zero Shapley value (Fig. 7_a). However, in both the region of YRD and GBA, a completely opposite trend is observed. Specifically, an increment of 1 m in BH is respectively associated with an average decrease in UEH by $0.12\text{ }^{\circ}\text{C}$ and $0.10\text{ }^{\circ}\text{C}$ respectively, when BH exceeds the average value of intra-group (Fig. 7_b, Fig. 7_c).

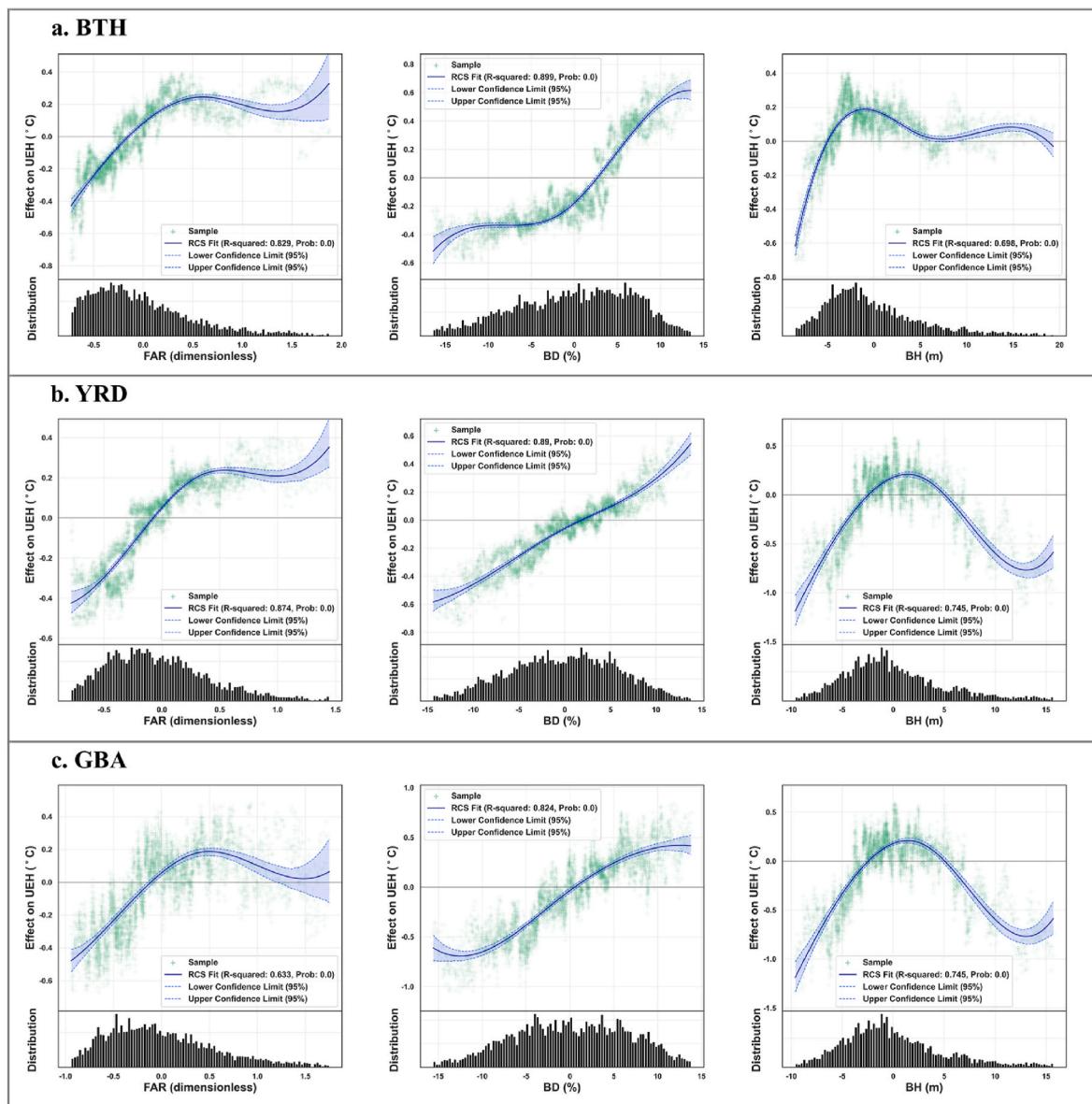


Fig. 7. The effect of per urban morphological feature on UEH
a-c. the SHAP values of urban morphological features in BTH, YRD and GBA.

4. Discussion

4.1. Heterogeneity of BH among urban agglomerations

BH, standing as one of the most pivotal urban morphological factors influencing the UEH environment, exhibits heterogeneity across the three individuals. As it implies that buildings, even if they share similar ecological or climatic features, may manifest markedly different impacts on UEH based on their varying heights. The taller the building, the larger the shadow area, the higher the thermal inertia, and the more robust mechanical turbulence to enhance the convective heat dissipation, providing a significant cooling effect in the urban area. Therefore, the underlying rationale for the discrepancy may be twofold: air flow rate and shading effect.

Firstly, as shown in Fig. 8, BH in YRD and GBA exceeding the intra-group mean exhibits a larger median WS, measuring 3.12 m/s (YRD_H) and 2.14 m/s (GBA_H) respectively, otherwise in opposite, at 3.07 m/s (YRD_L) and 1.91 m/s (GBA_L), separately. When air flows through a narrow passage formed by tall buildings, resembling a canyon, the restricted channel leads to an increased airflow velocity, i.e. "Venturi Effect" and "Downwash Effect", generating a canyon effect, i.e. a narrow sky view factor (SVF). Consequently, temperature decreases, alleviating UEH to some extent [59]. However, the scenario is entirely reversed for BTH, whose BH surpassing the intra-group mean manifests a smaller mean WS at 2.17 m/s (BTH_H), otherwise in opposite with 2.23 m/s (BTH_L). This counterintuitive phenomenon elucidates why, when the BH surpasses a certain threshold, its effect on UEH of BTH exhibits diminishing marginal utility rather than the significant cooling trend that occurs in YRD and GBA (Fig. 8).

Secondly, tall structures partially obstruct sunlight, casting extensive building shadows and leading to a lower rise in LST, thereby mitigating their impact on UEH, especially in YRD and GBA where solar radiation is relatively intense [60]. As indicated in the previous studies, solar radiation must be alleviated by building shadows [61]. As a special urban surface component that is closely related to the building structure and distribution, building shadows have effectively improved UEH environment by preventing solar radiation from penetrating into the street canyon [62]. Given BTH's higher latitude and the resulting smaller solar elevation angle, the building shadows witness a considerable expansion.

Consequently, the gain in mitigating UEH through BH is diminished, exhibiting a marginal diminishing utility. Contrastingly, the situation is reversed for YRD and GBA. This is also why, when BH surpasses the intra-group average, it exhibits a significant facilitating effect on mitigating UEH. Therefore, under identical BD and FAR, the recommendation leans towards more relatively elevated buildings as a measure to alleviate UEH in YRD and GBA.

4.2. Homogeneity of BD, FAR, and NDVI

A comparative investigation, encompassing both overarching trends and minute variations, unveils noteworthy conclusions. There is a significantly similar correlation between the enhancement of BD and FAR and the bolstered heat extremes among BTH, YRD, and GBA. Densely and intensively constructed urban terrains appear to attenuate LST cooling, which could be because buildings, especially those made of materials like concrete, exhibits a higher absorptance and a lower heat capacity, contributing a swift temperature elevation and substantial radiative emission to the surrounding environment and atmosphere [63]. Furthermore, elevated BD and FAR lead to intense diffuse reflection and reduced wind speed, hindering the efficient dispersion of heat [64–66]. In addition, the study shows that building footprints is the important factors in intensifying the UEH. The high-density and high-intensity buildings will mean more artificial surface and more thermal radiation to heat air [8].

One of the most influential natural factors in the urban thermal environment is the green background of urban regions [67]. Without exception, NDVI emerges as the paramount factor for all three major urban agglomerations in China. The cooling potential of vegetation arises out of the interactions between three main processes, namely ETa, moisture availability, and shading ([68]; Z.-H. [69]). These processes are intricately facilitated by modifying the vegetation ratio ([70,71]; C. [72]). These findings underscore the significance of urban morphological and natural characteristics in UEH mitigation that can inform strategic urban planning with spatial arrangements of buildings and green infrastructures [73].

4.3. Arid climatic characteristics enhancing water cooling effect

As previously observed, water plays a pivotal role in ameliorating the regional thermal environment. Moisture, an integral component of the atmosphere, possesses a significant heat capacity, allowing it to store and release substantial amounts of latent heat, especially during phase transitions such as evaporation and condensation [74].

However, there is heterogeneity in the mechanism of WD's impact on UEH across three groups. In comparison to YRD and GBA, BTH possesses significantly greater importance of the WD variable, likely attributed to its climatic characteristics. Here, it is located in both a temperate monsoon climate and a temperate continental climate, characterized by aridity and scarce precipitation. As shown in Fig. 4_b, such an assertion finds resonance in the data distribution outcomes of this study. BTH, considered one of the most domestic water deficient urban regions in China [8], exhibits a WD of 16.10 % importance, merely 8.67 % importance in YRD and 10.29 % importance in GBA. Furthermore, the mean ETa (95.32 mm) and Prcp (123.58 mm) in BTH are significantly lower than those in both YRD (111.82 mm, 163.19 mm) and GBA (123.13 mm, 292.48 mm). In relatively humid urban regions such as YRD and GBA, an excess of WD may lead to moisture saturation, and the additional water bodies become challenging to absorb more heat through evaporation and transpiration processes, thereby resulting in a relatively minor impact on cooling (Y. [75]). Conversely, in arid urban regions with insufficient precipitation like BTH, the introduction of additional moisture can more efficiently absorb extreme heat, consequently exerting a more substantial influence on temperature reduction [76]. Waterbody patches had significant cooling intensity, which could maximally cool the extreme heat of the surroundings down by 1.86 °C,

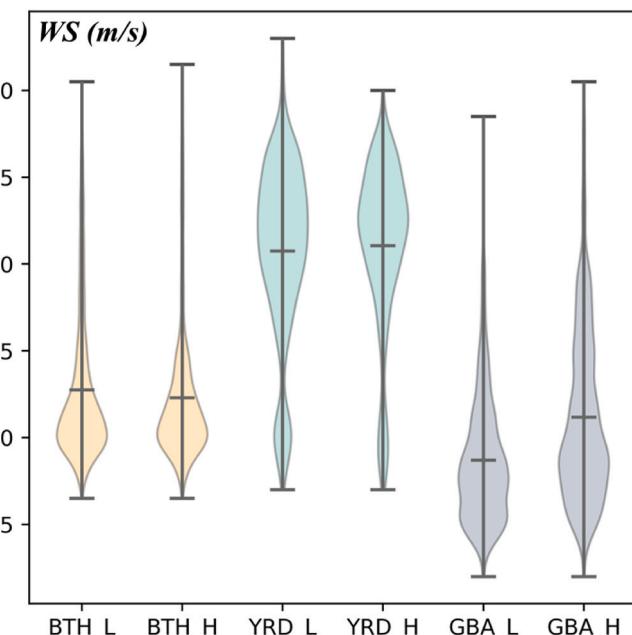


Fig. 8. WS in high-rise (H) and low-rise (L) areas vary across different urban agglomerations.

and the mean cooling intensity was as high as 1.01 °C. It is reasonable to control the WD within a certain threshold to effectively increase the cooling effect intensity.

It is widely proved that the cooling intensity of water bodies increases along with increasing their density, due to the enhanced heat exchange between waterbody and its local environment (G. [77]). However, the logarithmic relationship identified in this study implies that the augmentation of cooling intensity cannot be linearly escalated. In other words, as WD increases, the rate of augmentation in cooling intensity will gradually diminish and eventually reach a certain limit. Furthermore, given the scarcity of urban land resources, perpetually increasing WD is neither sustainable nor practical [78]. Thus, the cooling efficiency of water bodies should be focused on sustainable urban planning, especially for strategically designing and managing blue landscapes.

4.4. Metropolitan planning

Controlling and regulating the design strategies of morphological and natural environmental factors are crucial to improve UEH in Metropolitan planning. Exploring the contribution of various urban characteristics to the decoupling model reveals that the building footprints and building elevations are not the most predominant factors in raising UEH. The deterioration of the UEH environment is primarily attributed to the degradation of vegetation coverage, more than +0.18 NDVI is required to alleviate the urban high temperatures. Furthermore, increasing the waterbodies density to at least +12.5 % intra-urban built-up areas is always an effective strategy for mitigating UEH. Alternatively, slowing down the urbanization in a fast-rising metropolitan like BTH, YRD, and GBA is impractical. As a result, the development of metropolitan areas should appropriately increase the height of buildings to meet various urban functionalities, reduce building density and building intensity, and increase urban blue-green space area to alleviate the hazards of UEH environment.

4.5. Limitations and prospects

While this study decoupled the multi-drivers of UEH environment in urban agglomerations using an ensemble learning, it is important to acknowledge several limitations that call for further investigation.

Initially, while we have endeavoured to incorporate a comprehensive dataset spanning multiple characteristics and years, the quantity of indicators may still be insufficient to capture the full complexity and variability inherent in UEH. This limitation underscores the need to broaden and update the existing metrics, such as Sky View Factors (SVF) and Leaf Area Index (LAI) ([79]; Q. [80,81]), which this study has not directly explored.

Besides, although the MCS-based interpretable ensemble learning model has exhibited superior fitting performance with the R^2 of trainset ($99.89 \pm 0.05\%$) and test set ($97.62 \pm 0.10\%$) in the 95 % confidence interval in this study, differences in the fitting of various models still persist, emphasizing the need to involve a comparative study of multiple models' performance (e.g. RF, XGBoost, LightGBM, MLP, DBN) through measuring the goodness of fit.

Moreover, nighttime UEH has greater impacts on human society, it reduces people's ability to cool off and prevents the human body recovering from daytime UEH exposure, thereby increasing the risks of heat illnesses and deaths [82]. In the future, we will deeply delve into the exploration of dissecting the multi-drivers of nighttime UEH, as well as daytime-nighttime compound UEH, to gain a deeper understanding of UEH.

5. Conclusions

In this study, we delve into the intricate nonlinear impact mechanisms of urban morphology, natural environment and climatic

characteristics on UEH, elucidating their divergences across different urban agglomerations. Paradigmatically, our analytical approach is innovative, quantifying the UEH intra-urban built-up area during the UEHQ via a robust cross-regional statistics, reflecting both the UEH excluding the influence of abnormal values and the overall temperature of the study area, which addresses the limitations inherent in preceding studies that predominantly focused on the disparity between urban and rural thermal extreme. Methodologically, we adopt a Bayesian optimization-based ensemble learning model enhanced by the Monte Carlo Simulation framework to decouple the intricate nonlinear relationship between intra-urban multi-drivers and UEH environment, ensuring the robustness and reliability of the results. The SHAP additive interpretation algorithm is also used to meticulously interpret the stress thresholds of the intra-urban multi-drivers so that their marginal utility can be captured.

The findings underscore the pivotal role of natural and climatic factors in mitigating UEH, juxtaposed with the pervasive facilitating effect of urban morphological variables. It is noteworthy that these ecological and anthropogenic determinants exhibit variations in the importance, minimum effective dose, maximum ineffective dose, marginal utility threshold range, intensity, and even directional trends across different urban regions. The impact of building elevation on UEH in BTH exhibits diminishing marginal utility rather than the significant cooling trend that occurs in YRD and GBA. In comparison to YRD and GBA, BTH possesses significantly greater implications of the water bodies on UEH, likely attributed to its climatic characteristics. Relatively smaller waterbodies with varying thresholds across three urban agglomerations equally distributed intra-urban built-up areas represent a more efficient approach to UEH reduction. The heterogeneity of multi-driving mechanisms underscores the necessity of making policies adapted to local conditions, to accommodate the unique idiosyncrasies of individuals, thereby providing scientifically quantitative guidance for urban morphology and landscape planning towards the severe challenges of climate change.

CRediT authorship contribution statement

Xiaochang Liu: Writing – original draft, Resources, Methodology, Funding acquisition, Conceptualization. **Zhiyu Liu:** Data curation. **Zhiliang Zhu:** Writing – review & editing, Supervision. **Renlu Qiao:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by Shanghai University of Finance and Economics Fund (NO. 2023110783).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.buildenv.2024.111618>.

References

- [1] P. Stott, *How climate change affects extreme weather events* Research can increasingly determine the contribution of climate change to extreme events such as droughts, *Science* 352 (6293) (2016) 1517–1518.
- [2] W. Wang, K. Shi, X. Wang, Y. Zhang, B. Qin, Y. Zhang, et al., The impact of extreme heat on lake warming in China, *Nat. Commun.* 15 (1) (2024) 70, 70.
- [3] P. Luo, B. Yu, P. Li, P. Liang, Y. Liang, L. Yang, How 2D and 3D built environments impact urban surface temperature under extreme heat: a study in Chengdu, China, *Build. Environ.* 231 (2023).
- [4] C. Li, Y. Song, N. Kaza, R. Burghardt, Explaining spatial variations in residential energy usage intensity in Chicago: the role of urban form and geomorphometry, *J. Plann. Educ. Res.* 43 (2) (2023) 317–331.
- [5] X. Li, Y. Zhou, G.R. Asrar, M. Imhoff, X. Li, The surface urban heat island response to urban expansion: a panel analysis for the conterminous United States, *Sci. Total Environ.* 605 (2017) 426–435.
- [6] E. Massaro, R. Schifanella, M. Piccardo, L. Caporaso, H. Taubenbock, A. Cescatti, et al., Spatially-optimized urban greening for reduction of population exposure to land surface temperature extremes (vol 14, 2903, 2023), *Nat. Commun.* 14 (1) (2023).
- [7] D. Hidalgo-Garcia, J. Arco-Diaz, Modeling the Surface Urban Heat Island (SUHI) to study of its relationship with variations in the thermal field and with the indices of land use in the metropolitan area of Granada (Spain), *Sustain. Cities Soc.* 87 (2022).
- [8] Z. Wu, X. Liu, Q. Liu, R. Qiao, Z. Liu, T. Yang, et al., Urban development strategies in China based on water resource constraints, *Strategic Study of CAE* 24 (5) (2022) 75–88.
- [9] L. Zhou, B. Yuan, F. Hu, C. Wei, X. Dang, D. Sun, Understanding the effects of 2D/3D urban morphology on land surface temperature based on local climate zones, *Build. Environ.* 208 (2022).
- [10] X. Gu, Z. Wu, X. Liu, R. Qiao, Q. Jiang, Exploring the nonlinear interplay between urban morphology and nighttime thermal environment, *Sustain. Cities Soc.* 101 (2024).
- [11] H.M. Liu, B. Huang, S.H. Gao, J. Wang, C. Yang, R.R. Li, Impacts of the evolving urban development on intra-urban surface thermal environment: evidence from 323 Chinese cities, *Sci. Total Environ.* 771 (2021).
- [12] B. Jones, B.C. O'Neill, L. McDaniel, S. McGinnis, L.O. Mearns, C. Tebaldi, Future population exposure to US heat extremes, *Nat. Clim. Change* 5 (7) (2015) 652–655.
- [13] D. Zhou, J. Xiao, S. Bonafoni, C. Berger, K. Deilami, Y. Zhou, et al., Satellite remote sensing of surface urban heat islands: progress, challenges, and perspectives, *Rem. Sens.* 11 (1) (2019).
- [14] W. Jia, S. Zhao, Trends and drivers of land surface temperature along the urban-rural gradients in the largest urban agglomeration of China, *Sci. Total Environ.* 711 (2020).
- [15] L. Jiang, Y. Yang, Q. Wu, L. Yang, Z. Yang, Hotter days, dirtier air: the impact of extreme heat on energy and pollution intensity in China, *Energy Econ.* 130 (2024).
- [16] W. Wang, K. Shi, X. Wang, S. Wang, D. Zhang, Y. Peng, et al., A record-breaking extreme heat event caused unprecedented warming of lakes in China, *Sci. Bull.* 68 (6) (2023) 578–582.
- [17] W. Wang, K. Shi, X. Wang, Y. Zhang, B. Qin, Y. Zhang, et al., The impact of extreme heat on lake warming in China (vol 15, 70 , 2024), *Nat. Commun.* 15 (1) (2024).
- [18] A. Garcia-Garcia, F.J. Cuesta-Valero, D.G. Miralles, M.D. Mahecha, J. Quasas, M. Reichstein, et al., Soil heat extremes can outpace air temperature extremes, *Nat. Clim. Change* 13 (11) (2023) 1237–+.
- [19] B. Stone, J.J. Hess, H. Frumkin, Urban form and extreme heat events: are sprawling cities more vulnerable to climate change than compact cities? *Environ. Health Perspect.* 118 (10) (2010) 1425–1428.
- [20] H. Du, W. Cai, Y. Xu, Z. Wang, Y. Wang, Y. Cai, Quantifying the cool island effects of urban green spaces using remote sensing Data, *Urban For. Urban Green.* 27 (2017) 24–31.
- [21] N.I. Syafii, M. Ichinose, E. Kumakura, S.K. Jusuf, K. Chigusa, N.H. Wong, Thermal environment assessment around bodies of water in urban canyons: a scale model study, *Sustain. Cities Soc.* 34 (2017) 79–89.
- [22] J. Wang, Y. Chen, W. Liao, G. He, S.F. Tett, Z. Yan, et al., Anthropogenic emissions and urbanization increase risk of compound hot extremes in cities 11 (12) (2021) 1084–1089.
- [23] A. Chatzimitor, E. Apostolopoulou, A.D. Mazaris, A review of green infrastructure research in Europe: challenges and opportunities, *Lands. Urban Plann.* 198 (2020).
- [24] Y. Gao, J. Zhao, K. Yu, Effects of block morphology on the surface thermal environment and the corresponding planning strategy using the geographically weighted regression model, *Build. Environ.* 216 (2022).
- [25] L. Zhang, M. Nikolopoulou, S. Guo, D. Song, Impact of LCZs spatial pattern on urban heat island: a case study in Wuhan, China, *Build. Environ.* 226 (2022).
- [26] L. Li, Y. Zha, Mapping relative humidity, average and extreme temperature in hot summer over China, *Sci. Total Environ.* 615 (2018) 875–881.
- [27] L. Li, Y. Zha, Population exposure to extreme heat in China: frequency, intensity, duration and temporal trends, *Sustain. Cities Soc.* 60 (2020).
- [28] Y. Long, W.X. Zhai, Y. Shen, X.Y. Ye, Understanding uneven urban expansion with natural cities using open data, *Lands. Urban Plann.* 177 (2018) 281–293.
- [29] K.L. Harper, C. Lamarche, A. Hartley, P. Peylin, C. Ottle, V. Bastrikov, et al., A 29-year time series of annual 300 m resolution plant-functional-type maps for climate models, *Earth Syst. Sci. Data* 15 (3) (2023) 1465–1499.
- [30] P.A.N. Jinghu, S.H.I. Peiji, D. Xiaofeng, Measurements for urban hinterland area of cities at prefecture level or above in China, *Acta Geograph. Sin.* 63 (6) (2008) 635–645.
- [31] L. Li, N. Yao, Y. Li, D. Li Liu, B. Wang, O.O.J.A.R. Ayantobo, Future projections of extreme temperature events in different sub-regions of China 217 (2019) 150–164.
- [32] H.E. Beck, N.E. Zimmermann, T.R. McVicar, N. Vergopolan, A. Berg, E.F. Wood, Present and future Koppen-Geiger climate classification maps at 1-km resolution, *Sci. Data* 5 (2018).
- [33] S.A.M. Khatana, R.M. Werner, P.W. Groeneveld, Association of extreme heat and cardiovascular mortality in the United States: a county-level longitudinal analysis from 2008 to 2017, *Circulation* 146 (3) (2022) 249–261.
- [34] S.A.M. Khatana, R.M. Werner, P.W. Groeneveld, Association of extreme heat with all-cause mortality in the contiguous US, 2008–2017, *JAMA Netw. Open* 5 (5) (2022).
- [35] A. Tobias, B. Armstrong, A. Gasparini, J. Diaz, Effects of high summer temperatures on mortality in 50 Spanish cities, *Environ. Health* 13 (2014).
- [36] C.R. Upperman, J. Parker, C. Jiang, X. He, R. Murtugudde, A. Sapkota, Frequency of extreme heat event as a surrogate exposure metric for examining the human health effects of climate change, *PLoS One* 10 (12) (2015).
- [37] K.K. Joshi, T. Kono, Optimization of floor area ratio regulation in a growing city, *Reg. Sci. Urban Econ.* 39 (4) (2009) 502–511.
- [38] S.W. Myint, E.A. Wentz, A.J. Brazel, D.A. Quattrochi, The impact of distinct anthropogenic and vegetation features on urban warming, *Lands. Ecol.* 28 (5) (2013) 959–978.
- [39] Z. Wu, R. Qiao, S. Zhao, X. Liu, S. Gao, Z. Liu, et al., Nonlinear forces in urban thermal environment using Bayesian optimization-based ensemble learning, *Sci. Total Environ.* 838 (2022).
- [40] Z. Qiao, G. Tian, L. Xiao, Diurnal and seasonal impacts of urbanization on the urban thermal environment: a case study of Beijing using MODIS data, *ISPRS J. Photogrammetry Remote Sens.* 85 (2013) 93–101.
- [41] A. Azhdari, A. Soltani, M. Alidadi, Urban morphology and landscape structure effect on land surface temperature: evidence from Shiraz, a semi-arid city, *Sustain. Cities Soc.* 41 (2018) 853–864.
- [42] X. Zhang, R.C. Estoque, Y. Murayama, An urban heat island study in Nanchang City, China based on land surface temperature and social-ecological variables, *Sustain. Cities Soc.* 32 (2017) 557–568.
- [43] J.T. Abatzoglou, S.Z. Dobrowski, S.A. Parks, K.C. Hegewisch, Data Descriptor: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015, *Sci. Data* 5 (2018).
- [44] X. Yang, L.R. Leung, N. Zhao, C. Zhao, Y. Qian, K. Hu, et al., Contribution of urbanization to the increase of extreme heat events in an urban agglomeration in east China, *Geophys. Res. Lett.* 44 (13) (2017) 6940–6950.
- [45] D.C. Zheng, W.X. Liu, X.X. Li, Z.Y. Lin, H. Jiang, Research on carbon emission diversity from the perspective of urbanization, *Appl. Ecol. Environ. Res.* 16 (5) (2018) 6643–6654.
- [46] K. Imai, I.S. Kim, On the use of two-way fixed effects regression models for causal inference with panel data, *Polit. Anal.* 29 (3) (2021) 405–415.
- [47] L. Zhao, X. Lee, R.B. Smith, K. Oleson, Strong contributions of local background climate to urban heat islands, *Nature* 511 (7508) (2014) 216–219.
- [48] H.J. Miller, Tobler's First Law and spatial analysis, *Ann. Assoc. Am. Geogr.* 94 (2) (2004) 284–289.
- [49] R. Qiao, X. Liu, S. Gao, D. Liang, G. GesangYangji, L. Xia, et al., Industrialization, Urbanization, and Innovation: Nonlinear Drivers of Carbon Emissions in Chinese Cities, 358, 2024 122598.
- [50] T.G. Dietterich, Ensemble methods in machine learning, in: J. Kittler, F. Roli (Eds.), *Multiple Classifier Systems*, 1857, 2000, pp. 1–15.
- [51] R. Qiao, X. Li, S. Gao, X. Ma, Improvement of thermal comfort for underground space: data enhancement using variational autoencoder, *Build. Environ.* 207 (2022).
- [52] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, et al., LightGBM: a highly efficient gradient boosting decision tree, 2017 Dec 04–09, in: Paper Presented at the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, 2017.
- [53] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, N. de Freitas, Taking the human out of the loop: a review of bayesian optimization, *Proc. IEEE* 104 (1) (2016) 148–175.
- [54] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, 2017 Dec 04–09, in: Paper Presented at the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, 2017.
- [55] H. Peng, Y. Nishiyama, K. Sezaki, Assessing environmental benefits from shared micromobility systems using machine learning algorithms and Monte Carlo simulation, *Sustain. Cities Soc.* 87 (2022).
- [56] K.L. Nylund, T. Asparoutiou, B.O. Muthen, Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study, *Structural Equation Modeling-a Multidisciplinary Journal* 14 (4) (2007) 535–569.
- [57] Z. Zhang, W. Luan, J. Yang, A. Guo, M. Su, C. Tian, The influences of 2D/3D urban morphology on land surface temperature at the block scale in Chinese megalopolises, *Urban Clim.* 49 (2023).
- [58] M. Masoudi, P.Y. Tan, Multi-year comparison of the effects of spatial pattern of urban green spaces on urban land surface temperature, *Lands. Urban Plann.* 184 (2019) 44–58.
- [59] F. Ali-Toudert, H. Mayer, Numerical study on the effects of aspect ratio and orientation of an urban street canyon on outdoor thermal comfort in hot and dry climate, *Build. Environ.* 41 (2) (2006) 94–108.
- [60] B. Slee, T. Parkinson, R. Hyde, Quantifying useful thermal mass: how much thermal mass do you need? *Architect. Sci. Rev.* 57 (4) (2014) 271–285.
- [61] T. Ichinose, L. Lei, Y. Lin, Impacts of shading effect from nearby buildings on heating and cooling energy consumption in hot summer and cold winter zone of China, *Energy Build.* 136 (2017) 199–210.

- [62] L. Martinelli, T.-P. Lin, A. Matzarakis, Assessment of the influence of daily shadings pattern on human thermal comfort and attendance in Rome during summer period, *Build. Environ.* 92 (2015) 30–38.
- [63] A. Mohajerani, J. Bakaric, T. Jeffrey-Bailey, The urban heat island effect, its causes, and mitigation, with reference to the thermal properties of asphalt concrete, *J. Environ. Manag.* 197 (2017) 522–538.
- [64] B. Chun, J.M. Guldmann, Spatial statistical analysis and simulation of the urban heat island in high-density central cities, *Landscl. Urban Plann.* 125 (2014) 76–88.
- [65] A. Guo, J. Yang, X. Xiao, J. Xia, C. Jin, X. Li, Influences of urban spatial form on urban heat island effects at the community level in China, *Sustain. Cities Soc.* 53 (2020).
- [66] K. Perini, A. Magliocco, Effects of vegetation, urban density, building height, and atmospheric conditions on local temperatures and thermal comfort, *Urban For. Urban Green.* 13 (3) (2014) 495–506.
- [67] N.H. Wong, C.L. Tan, D.D. Kolokotsa, H. Takebayashi, Greenery as a mitigation and adaptation strategy to urban heat, *Nat. Rev. Earth Environ.* 2 (3) (2021) 166–181.
- [68] H.J.E. Taha, buildings, Urban climates and heat islands: albedo, evapotranspiration, and anthropogenic heat 25 (2) (1997) 99–103.
- [69] Z.-H. Wang, X. Zhao, J. Yang, J. Song, Cooling and energy saving potentials of shade trees and urban lawns in a desert city, *Appl. Energy* 161 (2016) 437–444.
- [70] H.M. Imran, J. Kala, A.W.M. Ng, S. Muthukumaran, Effectiveness of vegetated patches as Green Infrastructure in mitigating Urban Heat Island effects during a heatwave event in the city of Melbourne, *Weather Clim. Extrem.* 25 (2019).
- [71] D.R. Richards, T.K. Fung, R.N. Belcher, P.J. Edwards, Differential air temperature cooling performance of urban vegetation types in the tropics, *Urban For. Urban Green.* 50 (2020).
- [72] C. Zhou, Y. Zhang, L. Fu, Y. Xue, Z. Wang, Assessing mini-park installation priority for regreening planning in densely populated cities, *Sustain. Cities Soc.* 67 (2021).
- [73] P. Herath, M. Thatcher, H. Jin, X. Bai, Effectiveness of urban surface characteristics as mitigation strategies for the excessive summer heat in cities, *Sustain. Cities Soc.* 72 (2021).
- [74] P. Ramamurthy, E. Bou-Zeid, Contribution of impervious surfaces to urban evaporation, *Water Resour. Res.* 50 (4) (2014) 2889–2902.
- [75] Y. Liu, Q. Li, L. Yang, K. Mu, M. Zhang, J. Liu, Urban heat island effects of various urban morphologies under regional climate conditions, *Sci. Total Environ.* 743 (2020).
- [76] C. Yan, Q. Guo, H. Li, L. Li, G.Y. Qiu, Quantifying the cooling effect of urban vegetation by mobile traverse method: a local-scale urban heat island study in a subtropical megalacity, *Build. Environ.* 169 (2020).
- [77] G. Yang, Z. Yu, G. Jorgensen, H. Vejre, How can urban blue-green space be planned for climate adaption in high-latitude cities? A seasonal perspective, *Sustain. Cities Soc.* 53 (2020).
- [78] Z. Yu, G. Yang, S. Zuo, G. Jorgensen, M. Koga, H. Vejre, Critical review on the cooling effect of urban blue-green space: a threshold-size perspective, *Urban For. Urban Green.* 49 (2020).
- [79] Y. Kang, M. Ozdogan, S.C. Zipper, M.O. Roman, J. Walker, S.Y. Hong, et al., How universal is the relationship between remotely sensed vegetation indices and crop leaf area index? A global assessment, *Rem. Sens.* 8 (7) (2016).
- [80] Q. Wang, S. Adiku, J. Tenhunen, A. Granier, On the relationship of NDVI with leaf area index in a deciduous forest site, *Rem. Sens. Environ.* 94 (2) (2005) 244–255.
- [81] J.L. Zarate-Valdez, M.L. Whiting, B.D. Lampinen, S. Metcalf, S.L. Ustin, P. H. Brown, Prediction of leaf area index in almonds by vegetation indexes, *Comput. Electron. Agric.* 85 (2012) 24–32.
- [82] Z. Shi, X. Xu, G. Jia, Urbanization magnified nighttime heat waves in China, *Geophys. Res. Lett.* 48 (15) (2021).