Spatiotemporal Evolution and Nonlinear Effects of Urban Morphology on Land Surface Temperature Under the Context of Heatwaves

Ling Li and Mingyi Du *

School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 100044, China;
2108570023150@stu.bucea.edu.cn
* Correspondence: dumingyi@bucea.edu.cn

Content

Figure. S1 Spatial distribution of geographic datasets used in the study: (a) Digital Surface Model (DSM), (b) building height, (c) transportation network, and (d) Spatial distribution of station data and land use classification....3
Figure. S2 Gradient distribution and transect selection in the study area....4
Figure S3 Two-dimensional urban morphology parameters....4
Figure S4 Three-dimensional urban morphology parameters....5
Figure. S5 Accuracy assessment between model-predicted temperatures (Landsat-like LST) and retrieved land surface temperatures (2018.10.17): (a) Actual Landsat LST; (b) LCZ-FSDAF prediction; (c) STARFM predicted temperature; (d) ESTARFM predicted temperature....5
Figure. S6 Accuracy assessment of model-predicted nighttime temperature (Landsat-like LST) vs. retrieved land surface temperature (2023.09.16): (a) Observed ASTER LST; (b) LCZ-FSDAF predicted LST; (c) STARFM predicted temperature; (d) ESTARFM predicted temperature....6
Figure. S7 Identification of high-temperature heatwaves from 2015 to 2024....7
Figure S8. Location of the four meteorological validation stations and surrounding land surface characteristics: (a) Land cover (LULC) types;(b) Spatial distribution of Impervious Surface Fraction (ISC); (c) NDVI spatial distribution....8
Figure. S9 Distribution of investigated plots within sixth road area in Beijing....8
Figure. S10 Response thresholds of LST spatial heterogeneity to urbanization intensity: (a) mean LST, (b) maximum LST, (c) minimum LST, (d) mean daytime temperature, (e) mean nighttime temperature, and (f) difference between maximum and minimum temperature....8
Figure. S11 SHAP value trends for NDVI: (a) mean LST, (b) maximum LST, (c) minimum LST, (d) mean daytime temperature, (e) mean nighttime temperature, and (f) difference between maximum and minimum temperature....9
Figure. S12 Influence of MBH on LST indicators based on SHAP value trends. (a) Mean LST; (b) Maximum LST; (c) Minimum LST; (d) Mean daytime LST; (e) Mean nighttime LST; and (f) Diurnal temperature range (difference between maximum and minimum LST)....9
Figure. S13 SHAP value trends for SVF: (a) mean LST, (b) maximum LST, (c) minimum LST, (d) mean daytime temperature, (e) mean nighttime temperature, and (f) difference between maximum and minimum temperature....10
Figure. S14. Schematic diagram of LCZ-FSDAF utilizing land satellite Landsat-8, MODIS, and geostationary satellite FY-2F data. The vertical axis represents observation days, the horizontal axis to the right indicates observation hours, the left forward axis represents spatial resolution, and the right forward axis represents temporal resolution. The predicted LST (dashed box) has a spatial resolution of 30 m (Landsat-8) and a temporal resolution of 1 hour (FY-2F)....10
Table S1 Detailed information of each measured station in the study area....11
Table S2 Metrics for classifying the urban-rural gradient....11
Table S3. Multicollinearity Test Table....11
Table S4. Selection Dates of In-Situ Station Data....11
Table S5 The patch distribution of urban forest in different rings in Beijing....12
Table S6 The landscape indices of urban forest within fifth ring in Beijing....12
Table S7 The landscape indices of urban forest in different rings in Beijing....13
Table S8 A summary of ZY-3 stereo image....13

Table S9 Parameter settings for the experiment..... 13
Table S10. Land surface characteristics and model performance at the four validation stations (based on LUCU, ISC and NDVI data)**.** 13
Supplementary S1. LCZ-Based Flexible Spatiotemporal Data Fusion Model for Temperature (LCZ-FSDAF) 14
Supplementary S2 LST retrieval 19

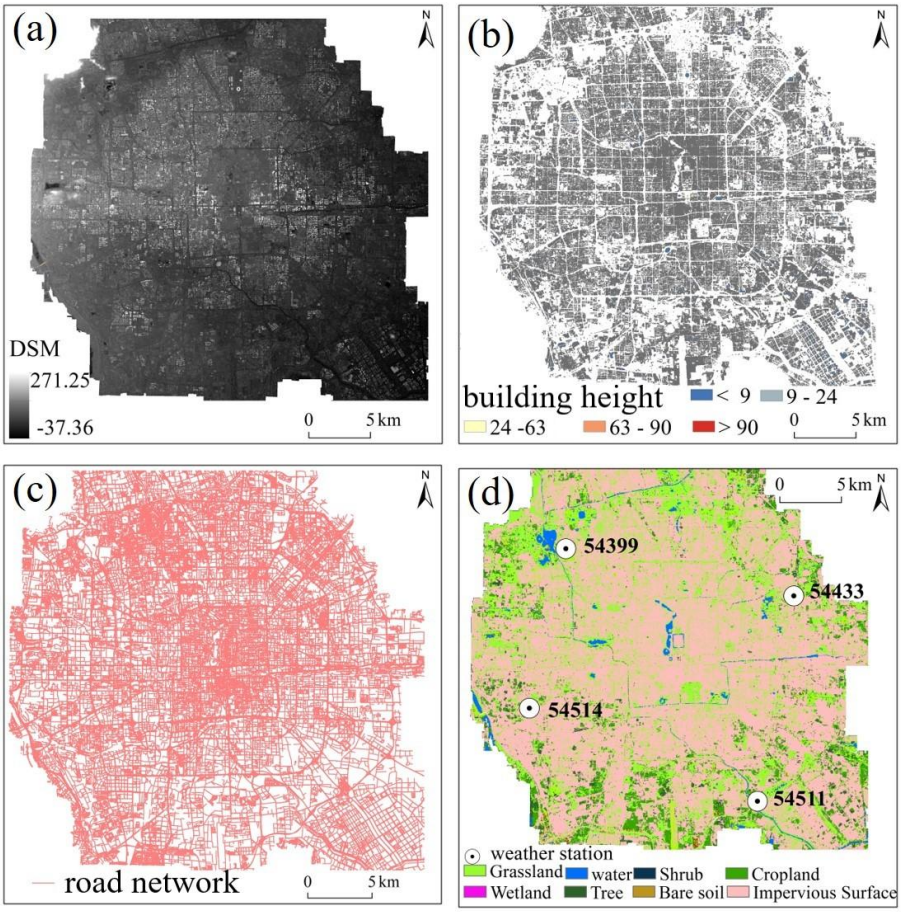


Figure. S1 Spatial distribution of geographic datasets used in the study: (a) Digital Surface Model (DSM), (b) building height, (c) transportation network, and (d) Spatial distribution of station data and land use classification.

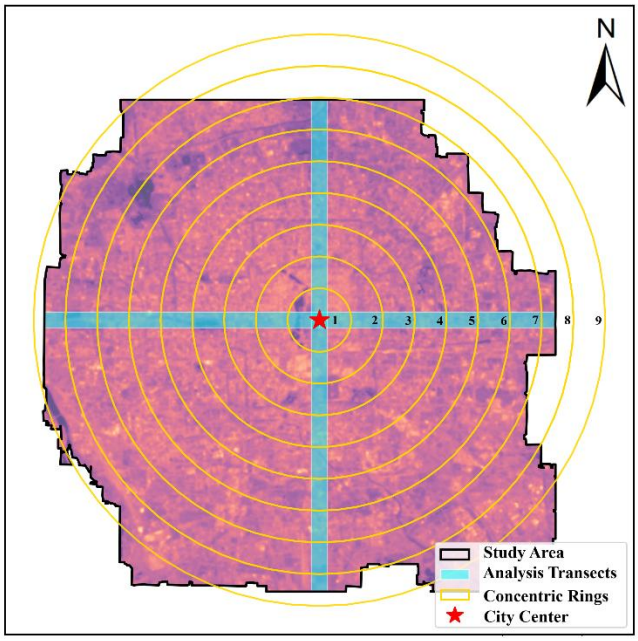


Figure S2. Gradient distribution and transect selection in the study area.

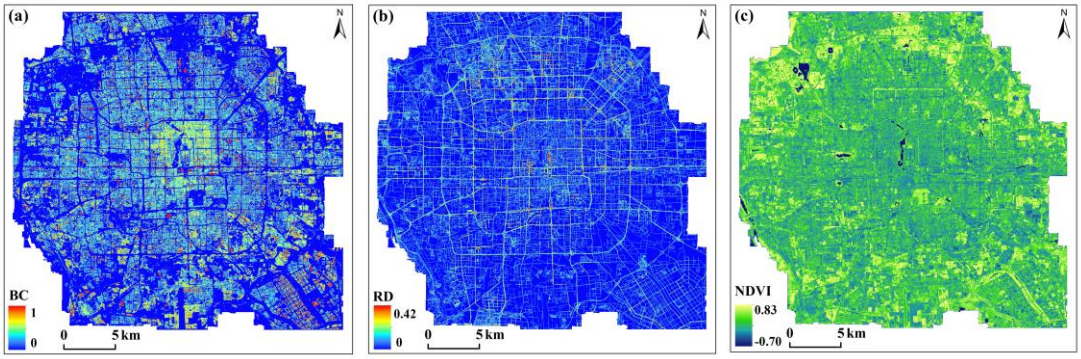


Figure S3. Two-dimensional urban morphology parameters.

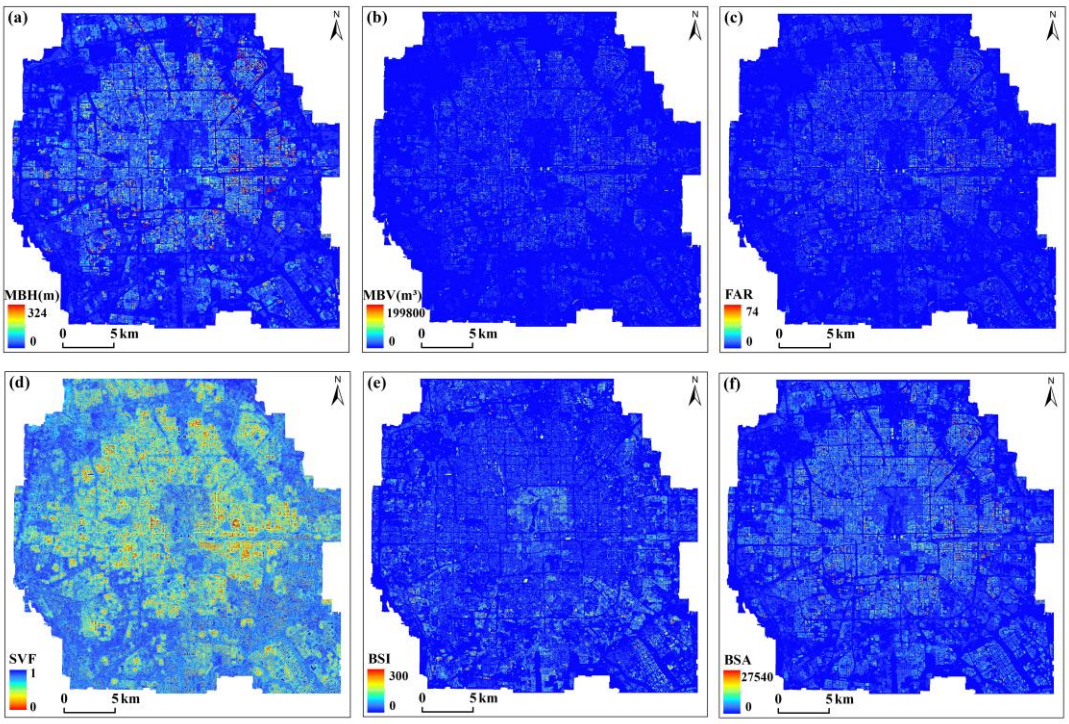


Figure S4. Three-dimensional urban morphology parameters.

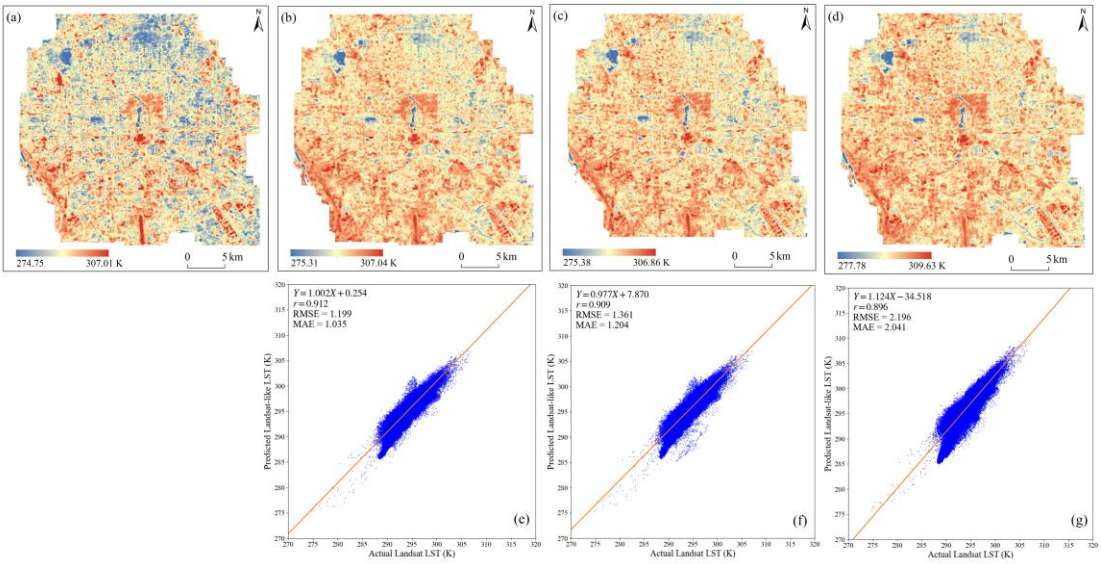


Figure. S5. Accuracy assessment between model-predicted temperatures (Landsat-like LST) and retrieved land surface temperatures (2018.10.17): (a) Actual Landsat LST; (b) LCZ-FSDAF prediction; (c) STARFM predicted temperature; (d) ESTARFM predicted temperature.

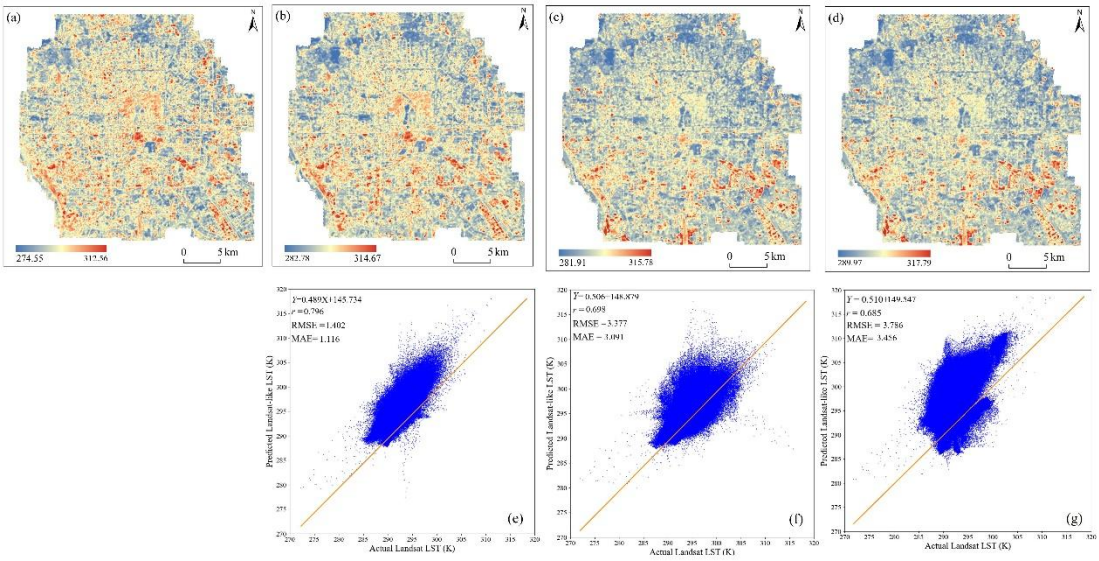


Figure S6. Accuracy assessment of model-predicted nighttime temperature (Landsat-like LST) vs. retrieved land surface temperature (2023.09.16): (a) Observed ASTER LST; (b) LCZ-FSDAF predicted LST; (c) STARFM predicted temperature; (d) ESTARFM predicted temperature.

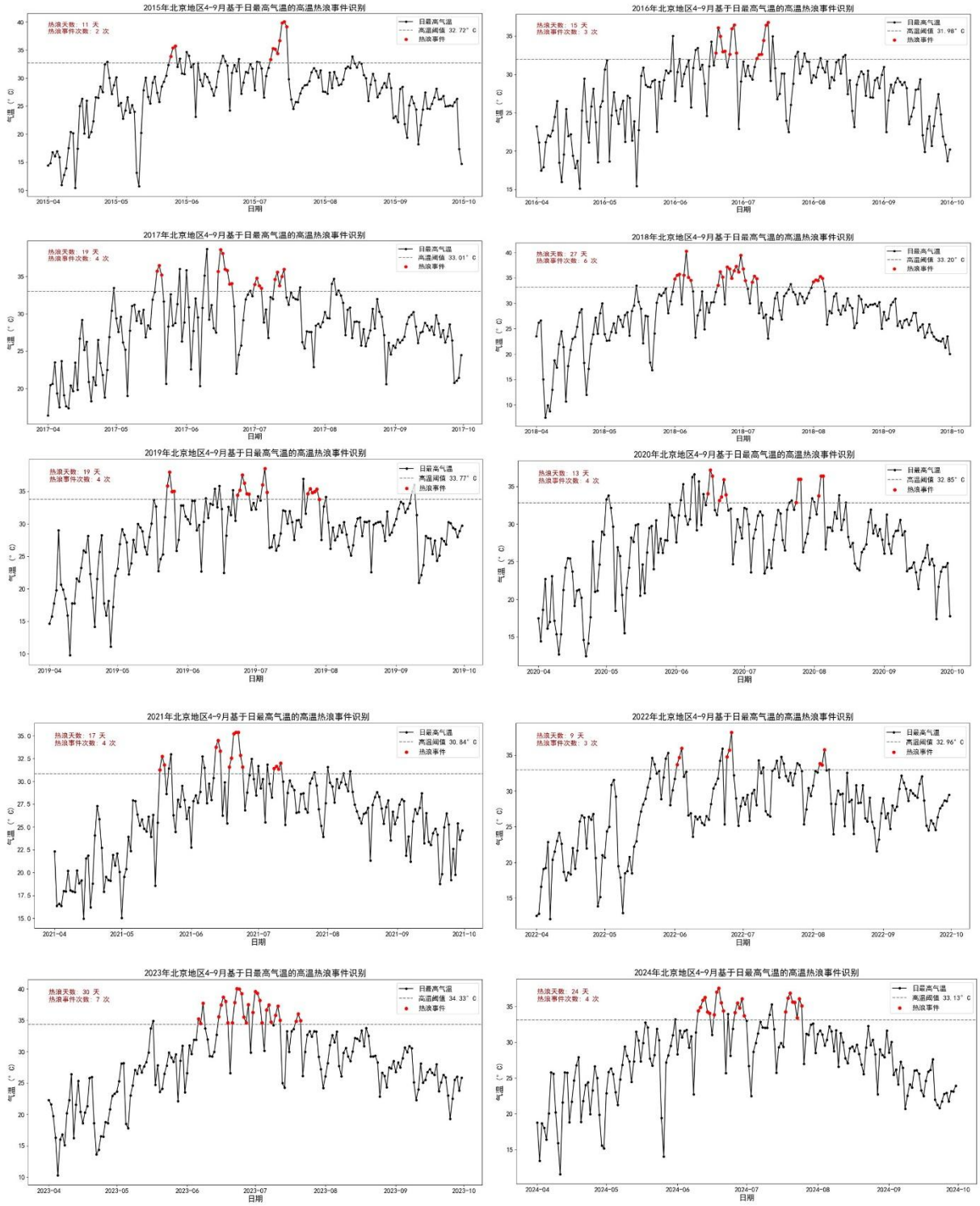


Figure S7. Identification of high-temperature heatwaves from 2015 to 2024.

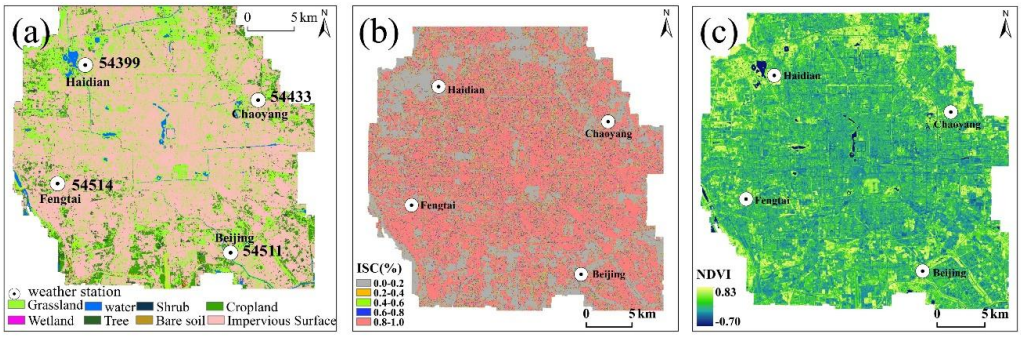


Figure S8. Location of the four meteorological validation stations and surrounding land surface characteristics: (a) Land cover (LULC) types;(b) Spatial distribution of Impervious Surface Fraction (ISC); (c) NDVI spatial distribution.

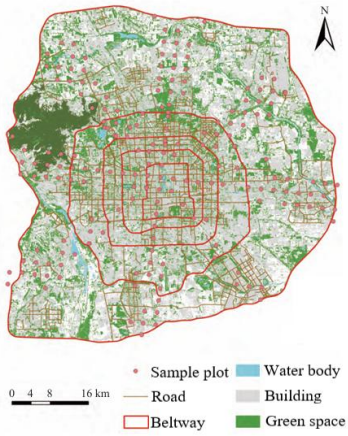


Figure S9. Distribution of investigated plots within sixth road area in Beijing.

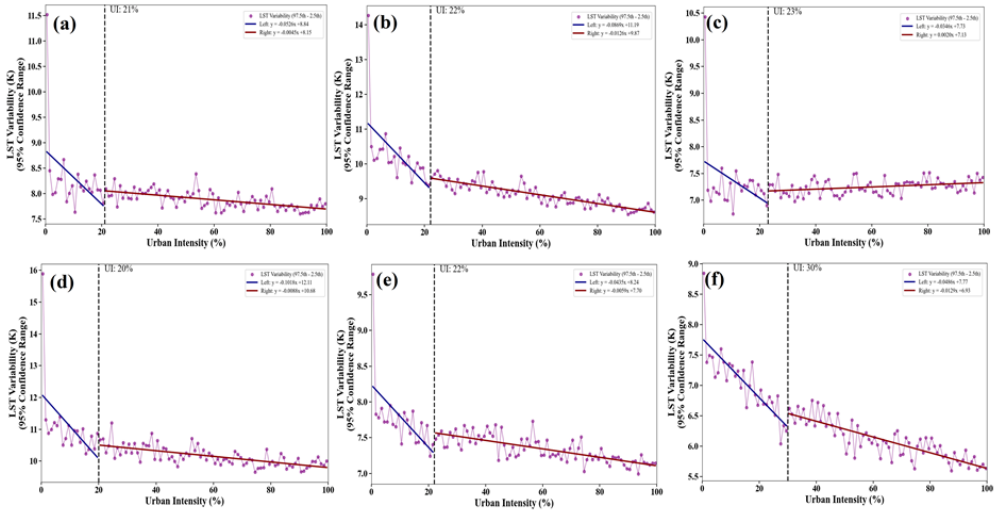


Figure. S10 Response thresholds of LST spatial heterogeneity to urbanization intensity: (a) mean LST, (b) maximum LST, (c) minimum LST, (d) mean daytime temperature, (e) mean nighttime temperature, and (f) difference between maximum and minimum temperature.

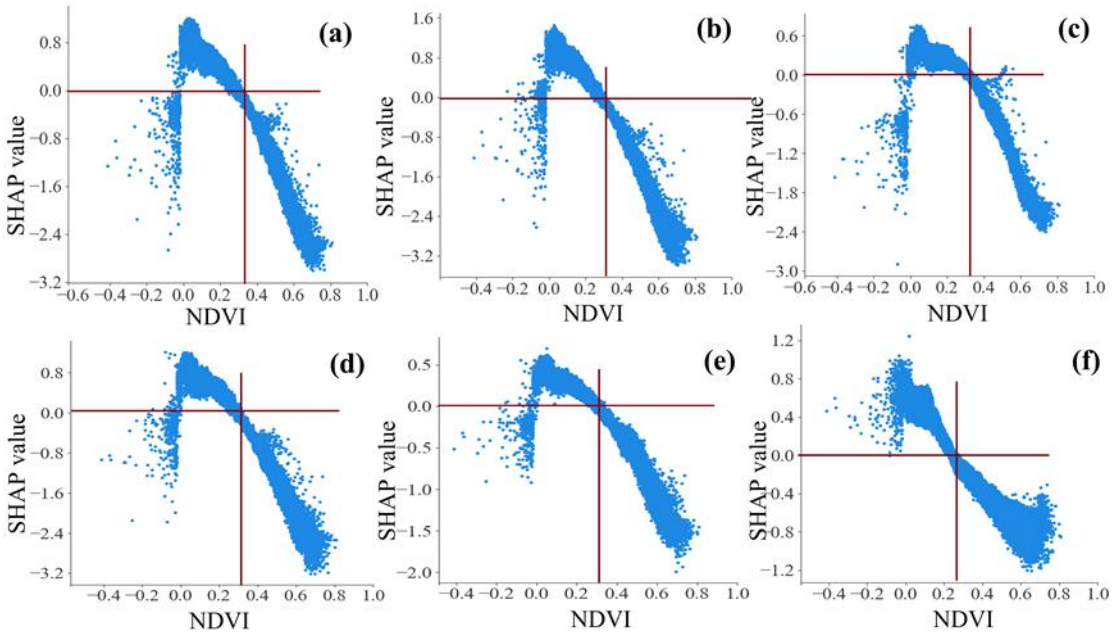


Figure S11. SHAP value trends for NDVI: (a) mean LST, (b) maximum LST, (c) minimum LST, (d) mean daytime temperature, (e) mean nighttime temperature, and (f) difference between maximum and minimum temperature.

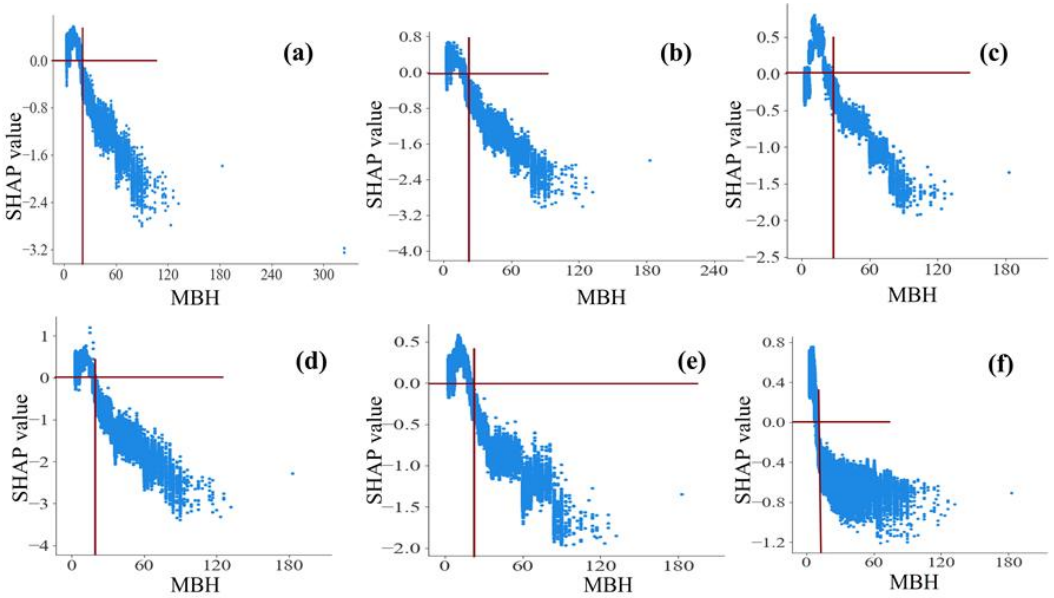


Figure S12. Influence of MBH on LST indicators based on SHAP value trends. (a) Mean LST; (b) Maximum LST; (c) Minimum LST; (d) Mean daytime LST; (e) Mean nighttime LST; and (f) Diurnal temperature range (difference between maximum and minimum LST).

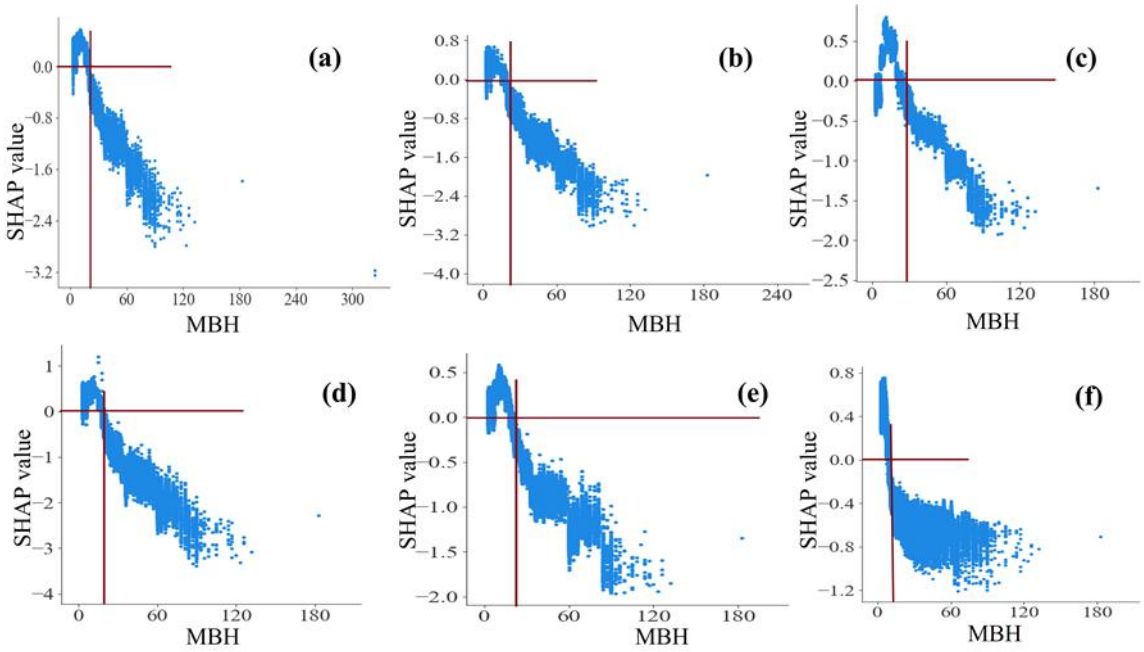


Figure S13. SHAP value trends for SVF: (a) mean LST, (b) maximum LST, (c) minimum LST, (d) mean daytime temperature, (e) mean nighttime temperature, and (f) difference between maximum and minimum temperature.

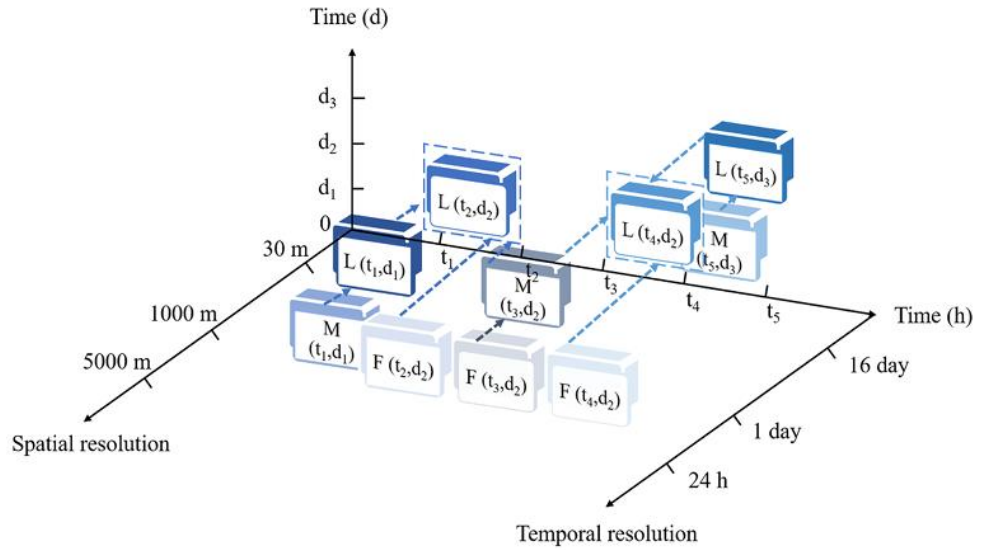


Figure S14. Schematic diagram of LCZ-FSDAF utilizing land satellite Landsat-8, MODIS, and geostationary satellite FY-2F data. The vertical axis represents observation days, the horizontal axis to the right indicates observation hours, the left forward axis represents spatial resolution, and the right forward axis represents temporal resolution. The predicted LST (dashed box) has a spatial resolution of 30 m (Landsat-8) and a temporal resolution of 1 hour (FY-2F).

Table S1. Detailed information of each measured station in the study area.

No.	Station Name	Station ID	Longitude (°E)	Latitude (°N)	Elevation (m)
1	Haidian	54399	116.28	39.98	45.8
2	Chaoyang	54433	116.5	39.95	35.3
3	Beijing	54511	116.47	39.8	31.3
4	Fengtai	54514	116.25	39.86	55.2

Table S2. Metrics for classifying the urban-rural gradient.

Metric	Abbreviation	Description
Population Density	PD	Population statistics based on administrative divisions or grid cells; distinguishes central urban areas, suburbs, and rural areas by density.
Impervious Surface Area	ISA	Uses remote sensing imagery to extract urban impervious surface or built-up land coverage to quantify the intensity of human activity.
Distance to City Center	DCC	Uses the city center (e.g., municipal government, CBD) as the origin; constructs a concentric urban-rural gradient through equidistant or variable-distance buffers.
Urban Land Coverage	ULC	Conducts buffer analysis based on land use classification (e.g., cropland, forest, built-up land) to distinguish urban-rural transition zones.
Housing Density	HD	Total number of housing units estimated based on local administrative district electoral levels; expressed as units/km².

Table S3. Multicollinearity Test Table.

UPMS	Tolerance	VIF	UPMS	Tolerance	VIF	UPMS	Tolerance	VIF
BC	0.300	3.339	MBH	0.261	3.370	SVF	0.826	1.210
NDVI	0.578	1.087	MBV	0.559	1.732	BSI	0.449	2.229
RD	0.575	1.058	FAR	0.222	3.571	BSA	0.740	1.173

Table S4. Selection Dates of In-Situ Station Data.

Haidian	Chaoyang	Beijing	Fengtai
2017-09-28 00：00	2018-10-17 00：00	2018-09-23 00：00	2018-04-25 00：00
2017-09-28 01：00	2018-10-17 01：00	2018-09-23 01：00	2018-04-25 01：00
2017-09-28 02：00	2018-10-17 02：00	2018-09-23 02：00	2018-04-25 02：00

2017-09-28 03：00	2018-10-17 03：00	2018-09-23 03：00	2018-04-25 03：00
2017-09-28 04：00	2018-10-17 04：00	2018-09-23 04：00	2018-04-25 04：00
2017-09-28 05：00	2018-10-17 05：00	2018-09-23 05：00	2018-04-25 05：00
2017-09-28 06：00	2018-10-17 06：00	2018-09-23 06：00	2018-04-25 06：00
2017-09-28 07：00	2018-10-17 07：00	2018-09-23 07：00	2018-04-25 07：00
2017-09-28 08：00	2018-10-17 08：00	2018-09-23 08：00	2018-04-25 08：00
2017-09-28 09：00	2018-10-17 09：00	2018-09-23 09：00	2018-04-25 09：00
2017-09-28 10：00	2018-10-17 10：00	2018-09-23 10：00	2018-04-25 10：00
2017-09-28 11：00	2018-10-17 11：00	2018-09-23 11：00	2018-04-25 11：00
2017-09-28 12：00	2018-10-17 12：00	2018-09-23 12：00	2018-04-25 12：00
2017-09-28 13：00	2018-10-17 13：00	2018-09-23 13：00	2018-04-25 13：00
2017-09-28 14：00	2018-10-17 14：00	2018-09-23 14：00	2018-04-25 14：00
2017-09-28 15：00	2018-10-17 15：00	2018-09-23 15：00	2018-04-25 15：00
2017-09-28 16：00	2018-10-17 16：00	2018-09-23 16：00	2018-04-25 16：00
2017-09-28 17：00	2018-10-17 17：00	2018-09-23 17：00	2018-04-25 17：00
2017-09-28 18：00	2018-10-17 18：00	2018-09-23 18：00	2018-04-25 18：00
2017-09-28 19：00	2018-10-17 19：00	2018-09-23 19：00	2018-04-25 19：00
2017-09-28 10：00	2018-10-17 10：00	2018-09-23 10：00	2018-04-25 10：00
2017-09-28 21：00	2018-10-17 21：00	2018-09-23 21：00	2018-04-25 21：00
2017-09-28 22：00	2018-10-17 22：00	2018-09-23 22：00	2018-04-25 22：00
2017-09-28 23：00	2018-10-17 23：00	2018-09-23 23：00	2018-04-25 23：00

Table S5. The patch distribution of urban forest in different rings in Beijing.

Ring Area	Patches (Number)	Patches (%)	Area (hm2)	Area (%)	Forest Coverage (%)
Within 2nd Ring	3175	11.55	1730.25	7.68	26.79
2nd - 3rd Ring	4528	16.47	2957.31	13.13	30.18
3rd -4th Ring	6234	22.67	4748.34	21.09	30.68
4th - 5th Ring	13557	49.31	13078.89	58.09	34.48
Total	27494	100	22514.79	100	32.35

Table S6. The landscape indices of urban forest within fifth ring in Beijing.

Orientation	CA (hm2)	AI (%)
Northeast	5041.71	54.63
Northwest	6876.81	57.04
Southeast	4219.29	54.78
Southwest	6132.69	51.52

Table S7. The landscape indices of urban forest in different rings in Beijing.

Ring Road Area	FN	LSI	FRAC	SIDI	SIEI	AI
Within 2nd Ring	183.5	67.95	1.55	0.59	0.88	47.66
2nd to 3rd Ring	153.11	89.49	1.56	0.49	0.73	48.77
3rd to 4th Ring	138.28	106.3	1.57	0.47	0.62	50.32
4th to 5th Ring	110.63	147.17	1.59	0.63	0.84	57.9

Table S8. A summary of ZY-3 stereo image.

Image view	Acauisition date	Roll satellite angular (deg)	Pitch satellite angular (deg)	Yaw satellite angular (deg)	Inclination telescope angular (deg)	Spatial resolution (m)
Forward	May 21,2020	-1.57	1.36	-1.80	23.5	3.5
Backward	May 21,2020	-1.77	1.49	-1.76	-23.5	3.5
Nadir 01	May 21,2020	-1.67	1.40	-1.79	0	2.1
Nadir 02	September 03,2020	-6.61	-6.16	-14.89	0	2.1
Nadir 03	September 03,2020	1.19	0.06	2.83	0	21.

Table S9. Parameter settings for the experiment.

Parameter	Value	Definition
Window Size	15	Half of the window size relative to the pixel size
Similar Pixels	16	Number of similar pixels to be selected
Minimum Classes	6	Minimum number of LCZ / land cover categories
Maximum Classes	16	Maximum number of LCZ / land cover categories
Scaling Factor	34	Ratio of coarse spatial resolution to fine spatial resolution
Pixel Size	30	Pixel resolution setting (based on high-resolution data)
Background Band	1	Number of spectral bands used

Table S10. Land surface characteristics and model performance at the four validation stations (based on LUCU, ISC and NDVI data**).**

Station Name (ID)	ISC Range	NDVI Range	Dominant Land Cover	RMSE (K)	R2
Haidian (54399)	0.2–0.4	0.6–0.83	Trees, grassland	0.53	0.986
Chaoyang (54433)	0.6–0.8	< 0.2	impervious surfaces	1.4	0.91

Beijing (54511)	0.6–0.8	0.1–0.2	Impervious surfaces with sparse trees	1.0	0.97
Fengtai (54514)	0.4–0.6	0.2–0.4	Cropland, impervious surfaces, and shrubs	0.6	0.98

Supplementary S1. LCZ-Based Flexible Spatiotemporal Data Fusion Model for Temperature (LCZ-FSDAF)

Step 1: Prediction of $T_{L'}(t_2, d_2)$

1.1 High-Resolution Image Classification and LCZ Fraction Calculation

In this model, the Local Climate Zone (LCZ) classification is introduced to replace the traditional unsupervised classification method (ISODATA). First, the high-resolution U-LST data at time t_1, together with the corresponding urban morphological information, are used to perform LCZ classification. Subsequently, for each coarse-resolution pixel (x_i, y_i), the proportion f_z of each LCZ class z within the pixel is calculated

$$f_z(x_i, y_i) = N_z(x_i, y_i) / m \tag{S1}$$

Where, $f_z(x_i, y_i)$ denotes the areal proportion (weight) of LCZ class z within the coarse-resolution pixel (x_i, y_i); $N_z(x_i, y_i)$ represents the number of high-resolution pixels belonging to LCZ class z; and mmm is the total number of high-resolution pixels contained within the coarse-resolution pixel.

1.2 Calculation of Temporal Change in Coarse-Resolution Pixels

The temperature change of coarse-resolution pixels from t_1 to t_2, denoted as ΔT_F, is calculated as:

$$\Delta T_F(x_i, y_i) = T_F(x_i, y_i, t_2, d_2) - T_F(x_i, y_i, t_1, d_1) \tag{S2}$$

ΔT_F represents the observed temperature difference of the low-resolution pixel between the prediction time t_2 and the reference time t_1. $T_F$denotes the observed land surface temperature at coarse resolution, while d_1,d_2 represent different dates, respectively.

1.3 Estimation of Temporal Changes for Each LCZ Class

Based on the theory of spectral linear mixture, n coarse-resolution pixels whose LCZ types remain unchanged from t_1 to $t_2$are selected. The temperature change for each LCZ class, denoted as $\Delta T_L(z)$, is then solved using the following equation:

$$\Delta T_F(x_i, y_i) = \sum_{z=1}^{l} f_z(x_i, y_i) \times \Delta T_L(z) \quad \text{(S3)}$$

Where, l represents the total number of LCZ classes. $\Delta T_L(z)$ denotes the average temperature change trend of the *z-th* LCZ class during the period from t_1 to t_2.

1.4 Calculation of High-Resolution Temporal Prediction

By adding $\Delta T_L(z)$ to the high-resolution image at time t_1, the preliminary prediction at time t_2, denoted as T_{LPT}, is obtained:

$$T_{LPT}(x_{ij}, y_{ij}, t_2, d_2) = T_L(x_{ij}, y_{ij}, t_1, d_1) + \Delta T_L(z) \quad \text{(S4)}$$

T_{LPT} represents the high-resolution LST predicted based on the temporal component. (x_i, y_i) denotes the coordinates of the *j-th* high-resolution pixel within the *i-th* coarse-resolution pixel.

1.5 Calculation of Residual $R(x_i, y_i)$ at Coarse Resolution

Due to intra-LCZ variability, the residual between the observed change and the predicted change is calculated as:

$$R(x_i, y_i) = \Delta T_F(x_i, y_i) - \frac{1}{m}\left[\sum_{i=1}^{m} T_{LPT}(x_{ij}, y_{ij}, t_2, d_2) - \sum_{i=1}^{m} T_L(x_{ij}, y_{ij}, t_1, d_1)\right] \quad \text{(S5)}$$

$R(x_i, y_i)$ represents the prediction residual, defined as the difference between the observed true change at low resolution and the mean predicted change derived from high-resolution data.

1.6 Spatial Trend Estimation Based on Edge-Preserving Filtering (EPF)

Thin Plate Spline (TPS) interpolation is applied to downscale the coarse-resolution LST at time t_2 to obtain the spatial trend f_{TPS}. To preserve urban building edge details, edge-preserving filtering (guided by the LCZ map) is further applied to the interpolated results:

$$f_{TPS}(x, y) = a_0 + a_1 x + a_2 y + 0.5\sum_{i=1}^{N} b_i r_i^2 \ln r_i^2 \quad \text{(S6)}$$

Where, f_{TPS} denotes the thin plate spline function, which is used to capture the nonlinear spatial trend surface of temperature, and r_i represents the Euclidean distance between the interpolation point and the known points.

$$T_{LPS}(x_{ij}, y_{ij}, t_2, d_2) = \mathrm{EPF}(f_{TPS}(x_{ij}, y_{ij})) \quad \text{(S7)}$$

T_{LPS} is the spatial prediction after edge-preserving filtering. The EPF effectively reduces erroneous smoothing effects at LCZ boundaries (e.g., building edges), where by preserving sharp thermal gradients.

1.7 Residual Weighting Function Based on LCZ Homogeneity Index

A weighting function, CV, is designed by introducing the LCZ-based homogeneity index (HI). Within a moving window, if the LCZ class of the *k-th* pixel is the same as that of the central pixel, then I_k=1:

$$CV(x_{ij}, y_{ij}) = E_{ho}(x_{ij}, y_{ij}) \times HI(x_{ij}, y_{ij}) + R(x_{ij}, y_{ij}) \times (1 - HI(x_{ij}, y_{ij})) \quad \text{(S8)}$$

Where, CV is the composite weighting factor that determines the distribution proportion of the residual *R* at the sub-pixel level. E_{ho} represents the homogeneity error term. *HI* is the homogeneity index, which ranges from 0 to 1 and reflects the degree of consistency of LCZ classes within the local window. When the surrounding pixels all belong to the same LCZ class, HI approaches 1.

1.8 Residual Normalization and Allocation

The weights are normalized, and the residual *r* assigned to each high-resolution pixel is calculated as follows:

$$C(x_{ij}, y_{ij}) = CV(x_{ij}, y_{ij}) / \sum_{i=1}^{m} CV(x_{ij}, y_{ij}) \quad \text{(S9)}$$

Where, *C* represents the normalized weight (computed from Equation S9). *r* denotes the residual component allocated to the *k-th* high-resolution pixel, and

$$r(x_{ijk}, y_{ijk}) = m \times R(x_i, y_i) \times C(x_{ij}, y_{ij}) \quad \text{(S10)}$$

Where, *r* denotes the residual component allocated to the *k-th* high-resolution pixel.

1.9 Calculation of Total High-Resolution Change ΔT_L

The allocated residual is added to the LCZ class-based change to obtain:

$$\Delta T_L(x_{ij}, y_{ij}) = r(x_{ij}, y_{ij}) + \Delta T_L(z) \quad \text{(S11)}$$

ΔT_L represents the high-resolution temperature change prediction that integrates both the “class-level trend” and the “local residual details.”

1.10 Final Prediction $T_{L'}$ Based on Gaussian-Weighted Blending

To eliminate boundary effects caused by block-wise processing, a Gaussian distance weight W_K is introduced. Within a local neighborhood, n similar pixels of the same class are selected for weighted aggregation:

$$D_k = 1 + \sqrt{(x_k - x_{ij})^2 + (y_k - y_{ij})^2} \,/\, (w/2) \tag{S12}$$

where D_k represents the distance weight factor corresponding to pixel k, which determines its contribution during the patch fusion process. (x_k, y_k) represents the spatial coordinates of the current target pixel k. (x_{ij}, y_{ij}) represents the geometric center coordinates of the current calculation patch or block.

$$w_k = \frac{\exp(-D_k^2 / 2\sigma^2)}{\sum \exp(-D_k^2 / 2\sigma^2)} \tag{S13}$$

w_k represents the Gaussian spatial weight, which addresses seam artifacts in block mosaicking. The closer a pixel is to the target pixel, the exponentially higher its weight. Σ denotes the standard deviation of the Gaussian kernel, controlling the spatial spread of the weights.

$$T_{L'}(t_2, d_2) = T_L(t_1, d_1) + \sum_{k=1}^{n} w_k \times \Delta T_L(x_k, y_k) \tag{S14}$$

$T_{L'}$ is the first final forward prediction component generated by the model.

Step 2: Prediction of $T_{L''}(t_2, d_2)$

Similarly, if an auxiliary time t_3 exists (e.g., a time after t_2), the value at time t_2 can be backward-predicted from t_3 by following the same procedures (Steps 1.1–1.10). Where, t_3 is a reference date after the prediction time t_2. Through backward temperature difference estimation, information from the "future" is utilized to constrain and refine the results at t_2.

$$T_{M'}(t_2, d_2) = T_M(t_3, d_2) + \sum_{k=1}^{n_1} w_k \times \Delta T_M(x_k, y_k) \tag{S15}$$

where $T_{M'}(t_2, d_2)$ represents the predicted coarse-resolution Land Surface Temperature (LST) at the target date (t_2). $T_M(t_3, d_2)$represents the observed coarse-resolution LST (e.g., from MODIS or FY-2F) at the base date (t_3). $\sum_{k=1}^{n_1}$ represents the summation

operator for all similar pixels. k within a sliding window, where n_l represents the total number of similar pixels w_k represents the weight assigned to the k-th similar pixel, typically determined by spatial distance and spectral similarity to the central pixel. $\Delta T_M(x_k,y_k)$ represents the LST change for the coarse-resolution pixel at coordinates (x_k, y_k) between the base date (t_3) and the target date (t_2).

$$T_{L''}(t_2,d_2) = T_L(t_3,d_1) + \sum_{k=1}^{n} w_k \times \Delta T_L(x_k,y_k) \quad \text{(S16)}$$

where $T_{L''}(t_2,d_2)$represents the final predicted fine-resolution LST map (e.g., at 30 m resolution) for the target date (t_2).$T_L(t_3,d_1)$ represents the fine-resolution reference image (e.g., Landsat 8 actual observation) acquired at the base date (t_3). $\sum_{k=1}^{n_l}$ represents the summation term for weighted similar high-resolution pixels within the neighborhood at the fine scale. w_k represents the spatial weighting coefficient at the fine-resolution scale, used to accurately distribute temporal changes to specific land cover types. $\Delta T_L(x_k,y_k)$ represents the predicted fine-resolution pixel change, derived by decomposing the coarse-resolution temperature change (ΔT_M) to the sub-pixel scale using Local Climate Zone (LCZ) constraints and spectral unmixing.

Step 3: Final Fusion Prediction of $T_L(t_2,d_2)$

A one-step gated sum is introduced. Based on the temporal distances between t_1, t_3, and the prediction time t_2, the weights u_1 and u_2 are calculated as:

$$T_L(t_2,d_2) = \mu_1 \times T_{L'}(t_2,d_2) + \mu_2 \times T_{L''}(t_2,d_2) \quad \text{(S17)}$$

Where, $u_1+u_2=1$, and the weights are inversely proportional to the temporal distance. $T_L(t_2,d_2)$ represents the final output of the model, while u_1 and u_2 are the fusion weights.

A typical formulation is:

$$\mu_1 = \frac{1/|t_2-t_1|}{1/|t_2-t_1|+1/|t_3-t_2|}, \quad \mu_2 = \frac{1/|t_3-t_2|}{1/|t_2-t_1|+1/|t_3-t_2|} \quad \text{(S18)}$$

If t_2 is closer to t_1, then the weight u_1 will be larger, meaning the model places greater trust in the forward prediction result.

Supplementary S2. LST retrieval.

Although the native resolution of Landsat 8 TIRS bands is 100 m, the Level-1 product provided by USGS resamples these bands to 30 m to match the OLI bands. In this study, the resampled 30 m thermal band data were used for LST retrieval. In this study, LST was retrieved using the radiative transfer equation (RTE) method adapted to the spectral characteristics of Landsat 8 TIRS bands (Sekertekin 2019), implemented on the Google Earth Engine (GEE) platform. Atmospheric parameters, including atmospheric transmittance (τ), upwelling radiance (L↑), and downwelling radiance (L↓), were calculated following the method of Sekertekin (2019) using NASA's Atmospheric Correction Parameter Calculator (https://atmcorr.gsfc.nasa.gov/). Land surface emissivity (ε) was estimated using an NDVI-based approach, with NDVI derived from Landsat 8 OLI bands and fractional vegetation cover (FVC) obtained by thresholding. Emissivity for mixed pixels was calculated as an FVC-weighted combination of vegetation and soil values, while fixed values were assigned to fully vegetated and non-vegetated surfaces. The fundamental structure of the radiative transfer equation is as follows:

$$L_\lambda = [\varepsilon B(T_s) + (1-\varepsilon)L_\downarrow]\tau + L_\uparrow \quad \text{(S19)}$$

$$T_S = \frac{K_2}{\ln\left(\frac{K_1}{B(T_S)}\right)} - 273 \quad \text{(S20)}$$

where ε is the land surface emissivity; T_S is the actual surface temperature (◦C); $B(T_s)$ is the blackbody thermal radiation brightness; τ is the atmospheric transmittance in the thermal infrared band; $L_\uparrow$ and $L_\downarrow$ are the atmospheric upward and downward thermal radiation. For Landsat 8 thermal infrared band 10, K_1 = 774.89 W/(m^2·μm·sr) and K_2 = 1321.08 K. This part has been revised in Section 2.2.1 of the manuscript.