Nature Communications
Supplementary Information for

Transfer learning reveals large discrepancies between air and land surface temperatures in cities

Authors: Yiwen Zhang[1], Lei Zhao[1,2,3,4*], TC Chakraborty[5], Priyam Mazumdar[6], Keer Zhang[7], Pierre Gentine[8*]

Affiliations:

[1]Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign; Urbana, IL, USA

[2]National Center for Supercomputing Applications, University of Illinois Urbana-Champaign; Urbana, IL, USA

[3]Institute for Sustainability, Energy, and Environment (iSEE), University of Illinois Urbana-Champaign; Urbana, IL, USA

[4]Department of Climate, Meteorology & Atmospheric Sciences, University of Illinois Urbana-Champaign; Urbana, IL, USA

[5]Pacific Northwest National Laboratory; Richland, WA, USA

[6]Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign; Urbana, IL, USA

[7]High Meadows Environmental Institute, Princeton University; Princeton, NJ, USA

[8]Department of Earth and Environmental Engineering, Columbia University; New York, NY, USA

*Corresponding author. Email: leizhao@illinois.edu, pg2328@columbia.edu

Contents of this file:

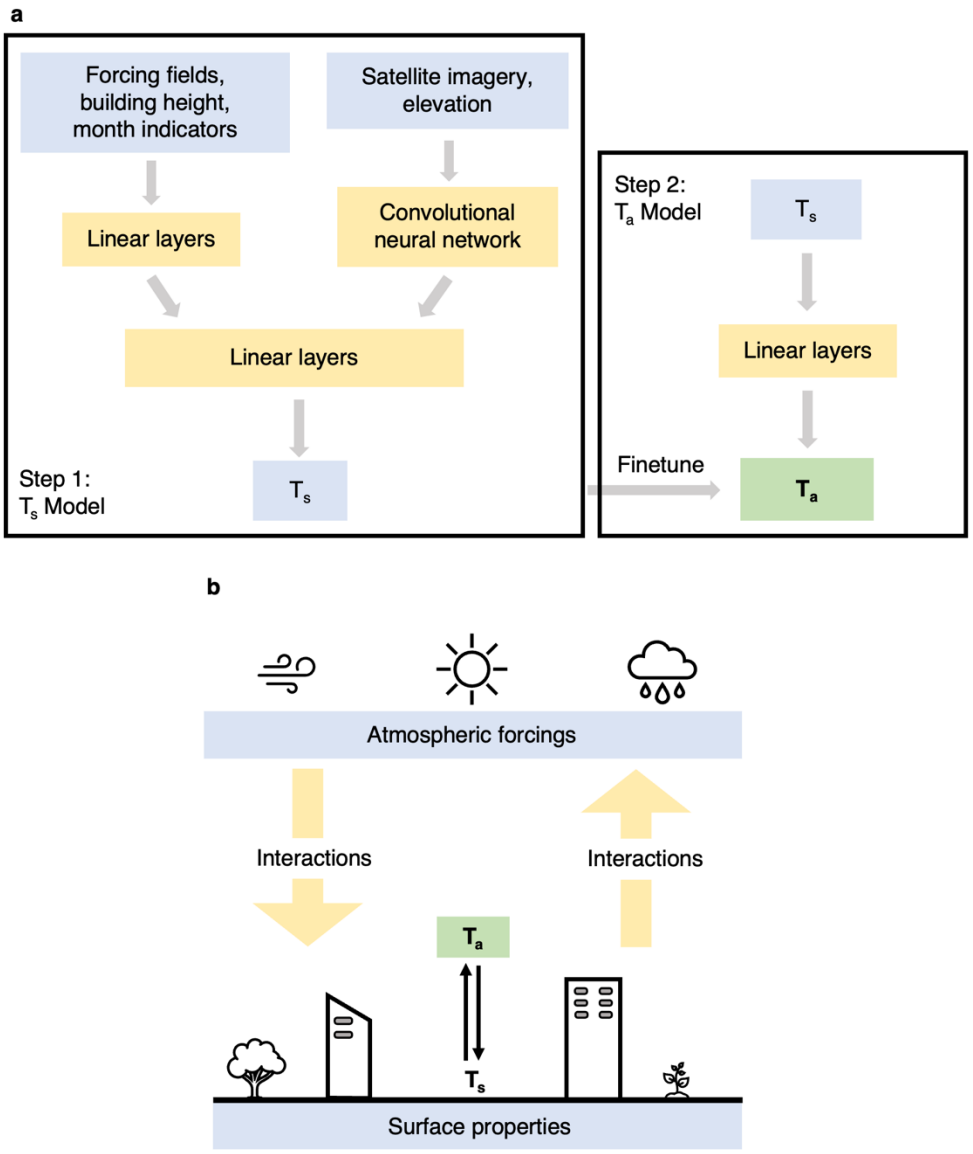

Fig. S1: The U-TL framework. a, Illustration of the U-TL framework consisting of two main steps: pretraining of the T_s (land surface temperature) model and fine-tuning of the T_a (near-surface air temperature) model. **b,** Land-atmosphere interactions that govern T_a, T_s and their relationship.

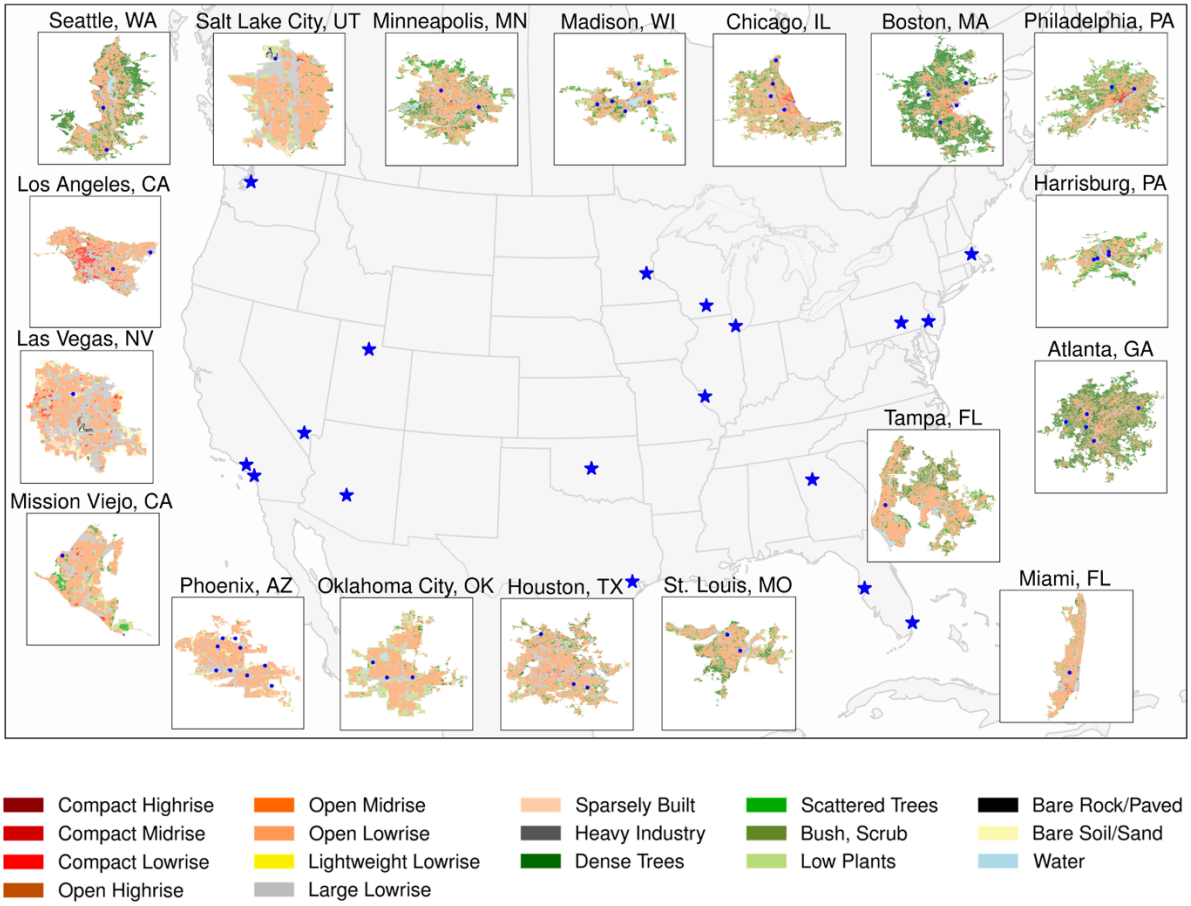

Fig. S2: Locations of the 52 meteorological stations from 18 cities selected for the fine-tuning step. Blue stars indicate cities with selected stations, and blue dots mark the specific station locations. The city backgrounds are based on local climate zone maps[1]. Basemap from Natural Earth (https://www.naturalearthdata.com/).

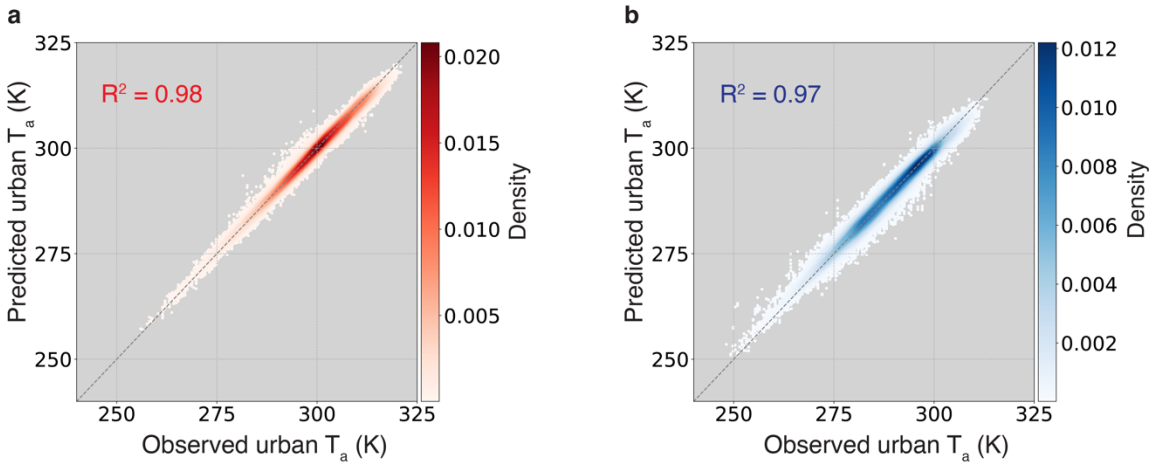

Fig. S3: Demonstration of the high accuracy of U-TL in predicting urban T_a (near-surface air temperature). U-TL's predicted urban T_a against unseen observations at daytime (**a**) and nighttime (**b**). A random 20% subset of testing samples is plotted for visualization, while the R^2 values represent the entire test dataset. Results are from the 9-fold cross-validation experiment.

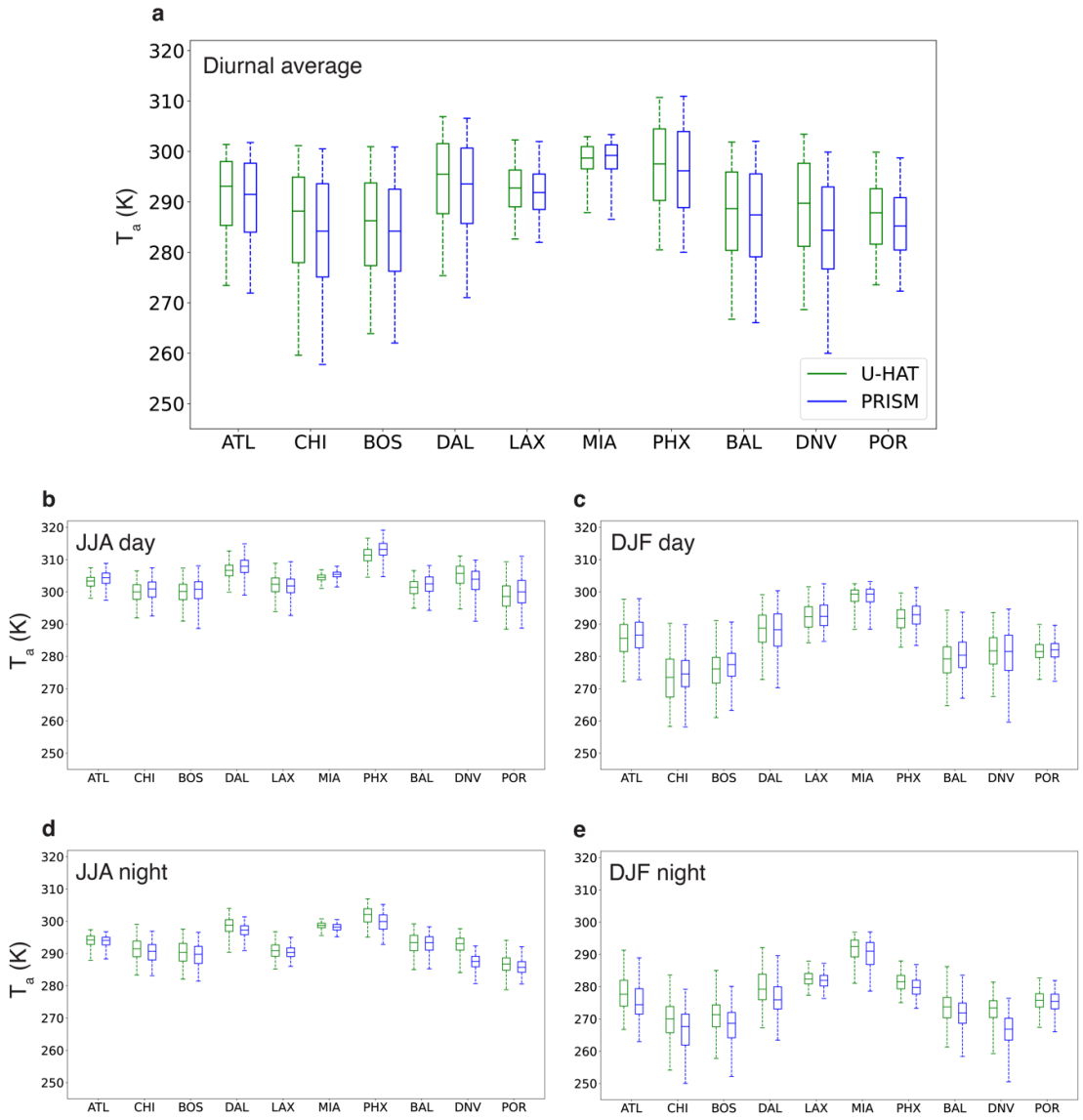

Fig. S4: Comparison of 2013–2023 distributions of T_a (near-surface air temperature) between U-HAT (green) and PRISM (blue). a, Distributions of diurnal average T_a. **b,c,** Distributions of daytime T_a during JJA (June–August) (**b**) and DJF (December–February) (**c**). **d,e,** Distributions of nighttime T_a during JJA (**d**) and DJF (**e**). Center bars represent the median, box edges the 25th and 75th percentiles, and error bars the 1st and 99th percentiles. ATL = Atlanta; CHI = Chicago; BOS = Boston; DAL = Dallas; LAX = Los Angeles; MIA = Miami; PHX = Phoenix; BAL = Baltimore; DNV = Denver; POR = Portland.

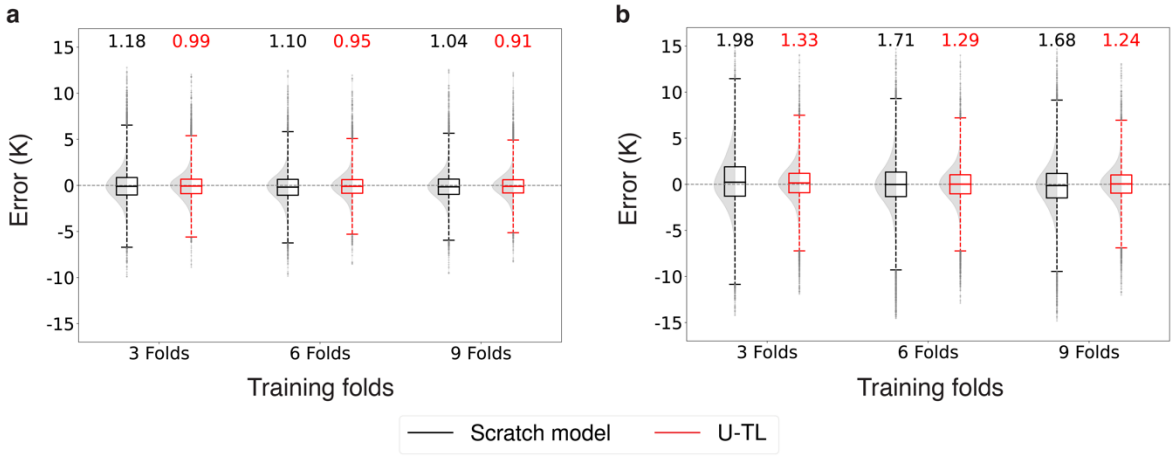

Fig. S5: Demonstration of the strong robustness of U-TL in predicting urban T_a (near-surface air temperature). Distributions of estimation errors for the scratch model (**black**) and U-TL **(red)** at daytime (**a**) and nighttime (**b**) using 3, 6 and 9 training folds. MAE for each configuration is denoted above the corresponding boxplot. Center bars represent the median, box edges the 25th and 75th percentiles, and error bars extend to 3×IQR (interquartile range) from the quartiles. Shaded violin plots indicate the underlying distribution.

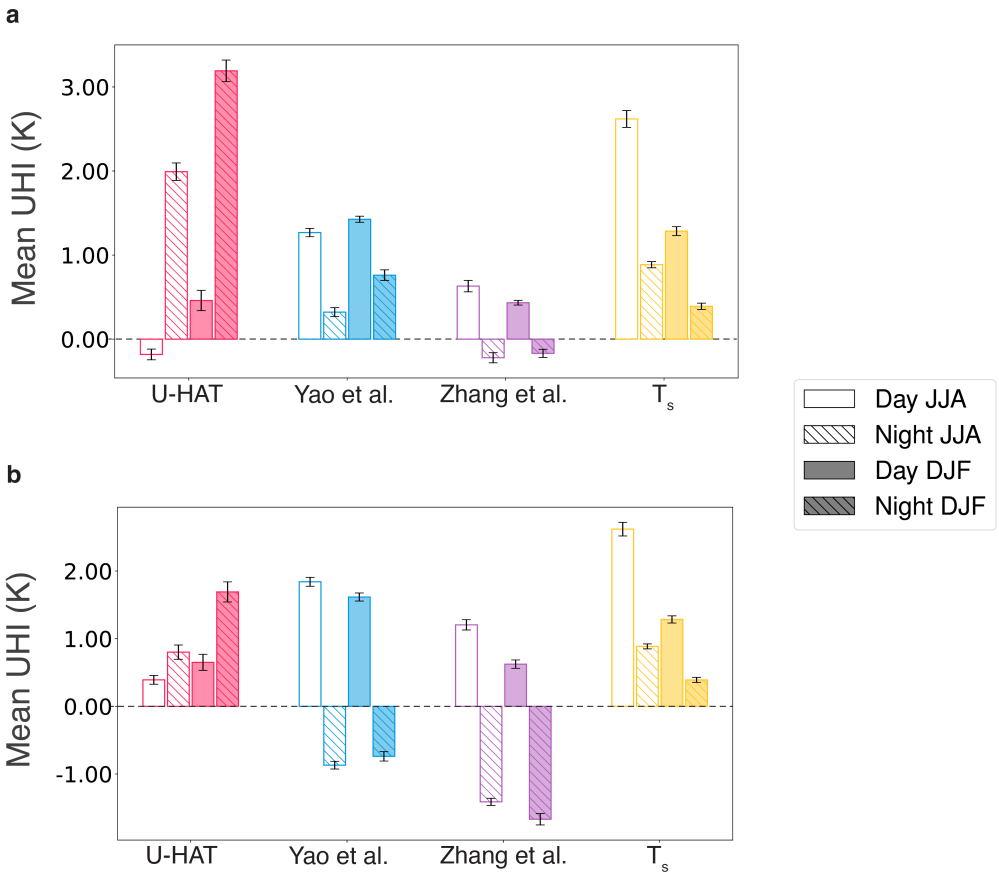

Fig. S6: Comparison of mean UHI$_a$ (canopy urban heat island) intensities between U-HAT and other two T_a (near-surface air temperature) datasets. Same as Fig. 2a but using rural reference temperatures from PRISM (**a**) and ERA5 (**b**). Error bars indicate the standard error of the mean.

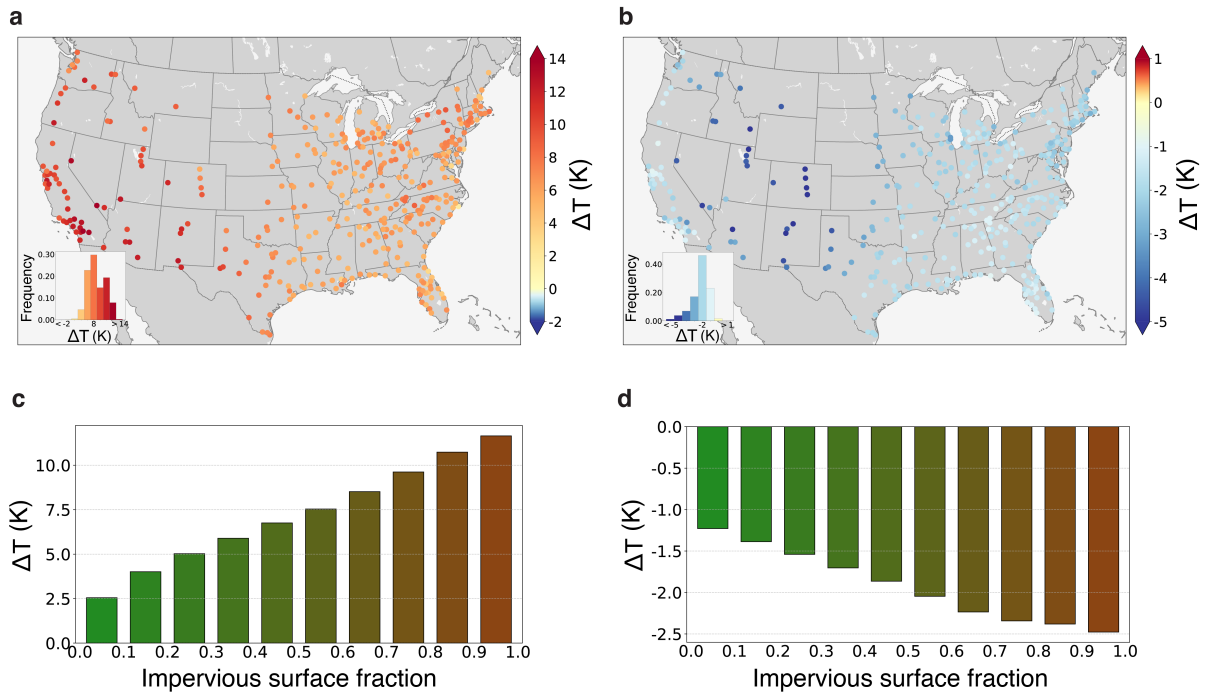

Fig. S7: Impacts of impervious surface fraction (ISF) on the differences between T_s (land surface temperature) and T_a (near-surface air temperature). a,b, Same as Fig. 3a,b for daytime (**a**) and nighttime (**b**) but calculated using only pixels with ISF higher than 0.5. **c,d,** Mean difference between 9-year mean T_s and T_a (ΔT) binned by ISF of the corresponding pixel at daytime (**c**) and nighttime (**d**). ISF is calculated as the fraction of ESA WorldCover pixels classified as "built-up" within each U-HAT pixel. Basemap from Natural Earth (https://www.naturalearthdata.com/).

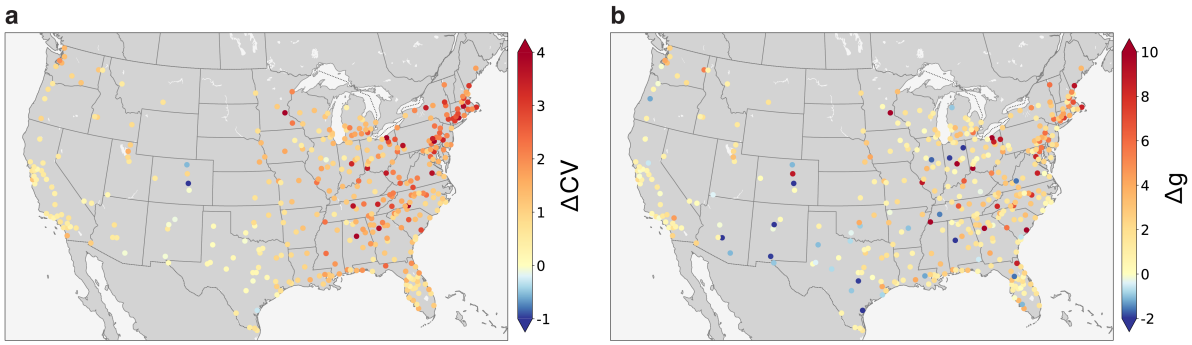

Fig. S8: Comparisons between heat exposure measured by T_s (HE_s) and T_a (HE_a). a, Map of the difference between the city-wise coefficient of variation of HE_s and HE_a (ΔCV). CV is calculated as the standard deviation divided by the mean of heat exposure within each city. **b,** Map of the difference between the city-wise skewness of HE_s and HE_a (Δg). The skewness is computed as the Fisher-Pearson coefficient of skewness. Basemap from Natural Earth (https://www.naturalearthdata.com/).

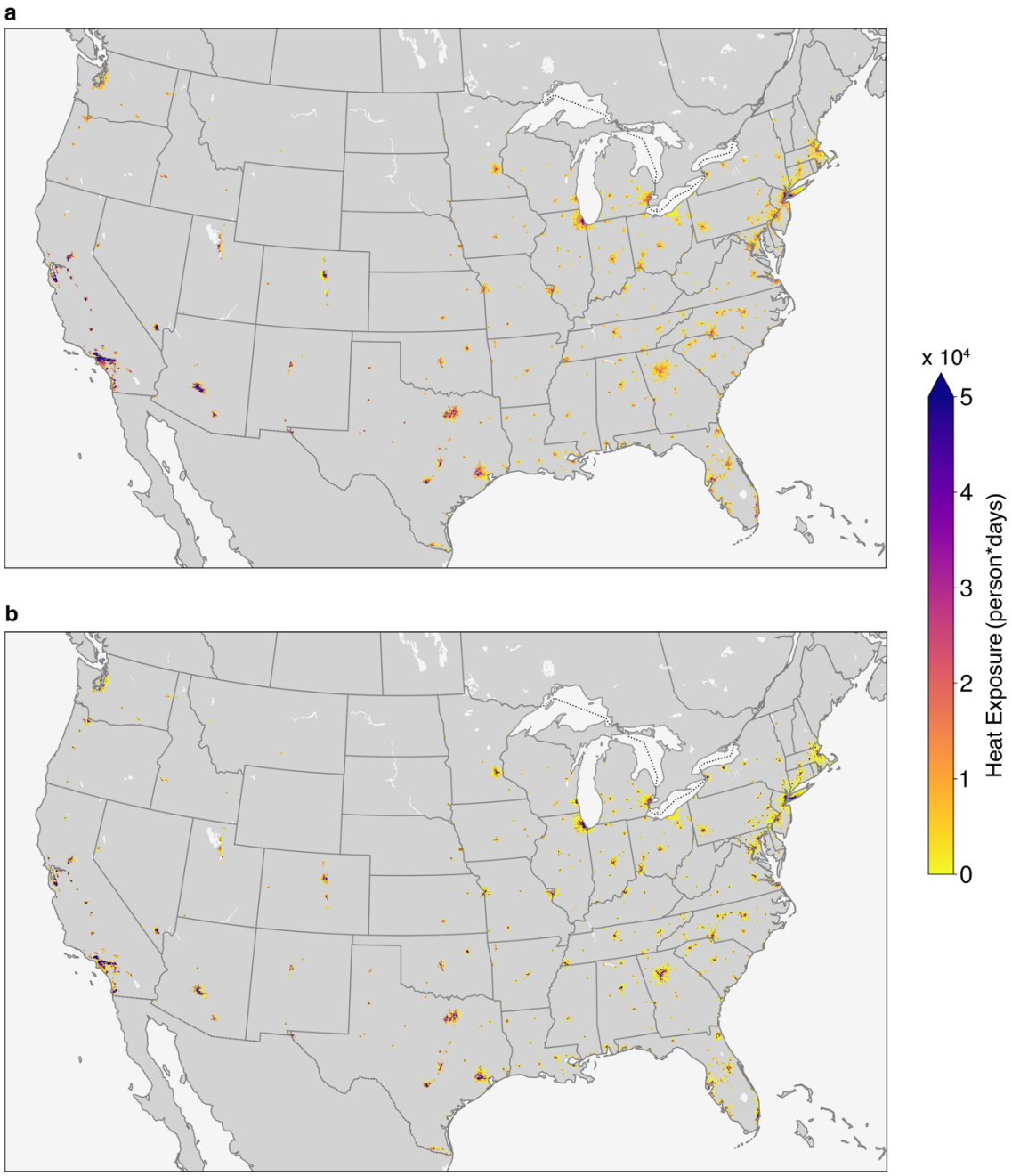

Fig. S9: Spatial distribution of heat exposure measured by T_s (HE_s) and T_a (HE_a). Results are shown for HE_a (**a**) and HE_s (**b**). Basemap from Natural Earth (https://www.naturalearthdata.com/).

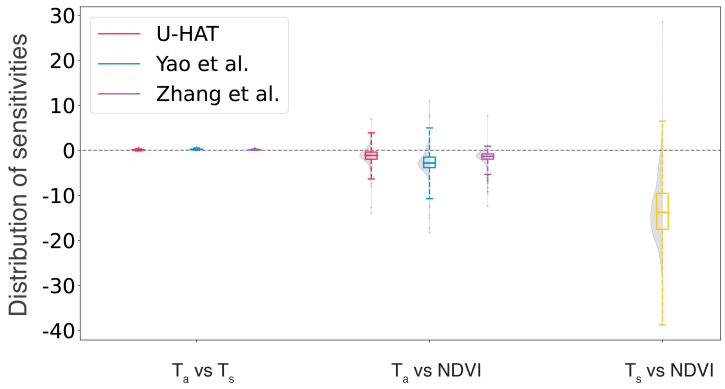

Fig. S10: Distributions of slopes from linear regressions between T_a (near-surface air temperature), T_s (land surface temperature) and NDVI (Normalized Difference Vegetation Index) across cities in the CONUS (contiguous United States). T_a estimates are from U-HAT, Yao et al.[2] and Zhang et al.[3]. Each data point represents the sensitivity of one variable to unit changes in the other, computed using JJA (June–August) mean values of variables from 2013–2020, based on all available pixels within one cluster. Center bars represent the median, box edges the 25th and 75th percentiles, and error bars extend to 3×IQR (interquartile range) from the quartiles. Shaded violin plots indicate the underlying distribution.

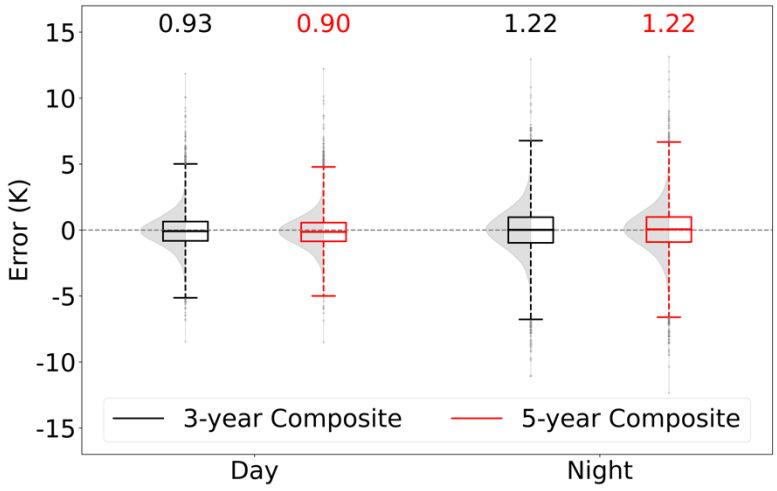

Fig. S11: Comparison of model performance using 3-year versus 5-year monthly median composite surface imagery. Distributions of estimation errors for models trained with 3-year (black) and 5-year (red) composites using 9 training folds at daytime and nighttime. MAE (mean absolute error) for each configuration is denoted above the corresponding boxplot. Center bars represent the median, box edges the 25th and 75th percentiles, and error bars extend to 3×IQR (interquartile range) from the quartiles. Shaded violin plots indicate the underlying distribution.

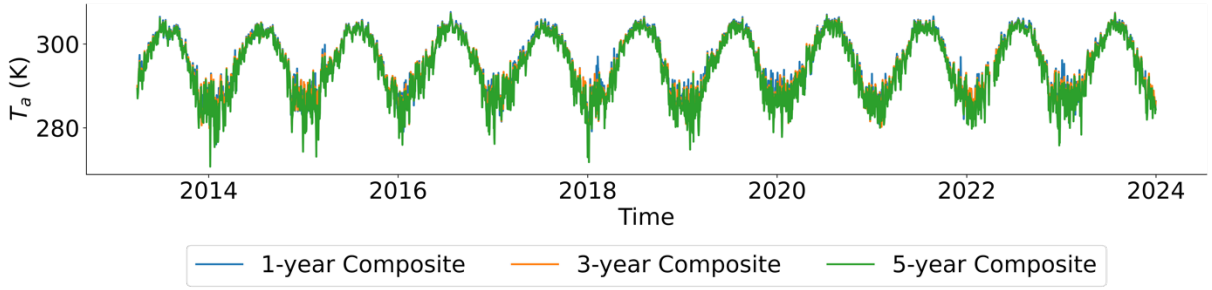

Fig. S12: Comparison of temporal patterns in daily mean urban air temperature estimated using different Landsat compositing windows. The blue, orange, and green lines represent results derived from 1-year, 3-year, and 5-year monthly median composites, respectively.

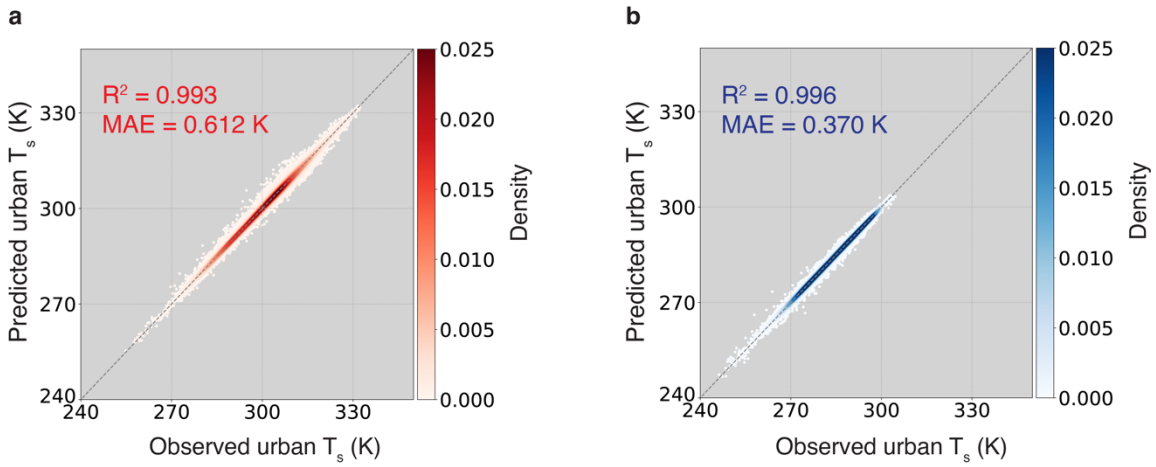

Fig. S13: Comparison between predicted and observed urban T_s (land surface temperature). Results are shown for daytime (**a**) and nighttime (**b**). A random 1% subset of testing samples is plotted for visualization, while the R^2 and MAE (mean absolute error) values represent the entire test dataset.

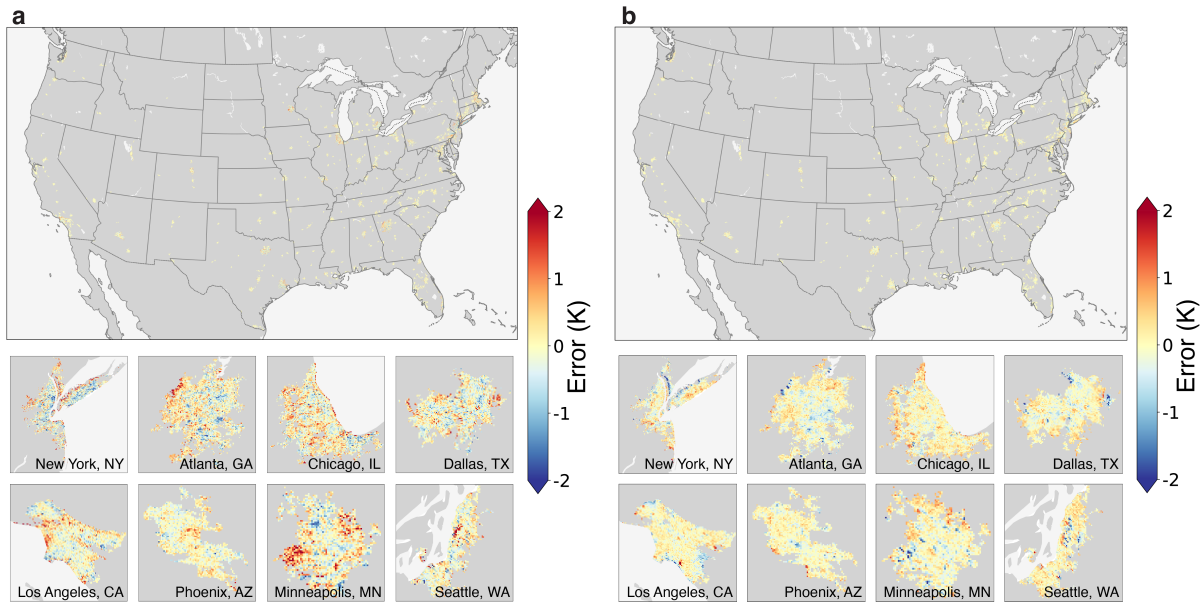

Fig. S14: Map of mean T_s (land surface temperature) estimation errors at each pixel. Results are shown for daytime **(a)** and nighttime **(b)**. Basemap from Natural Earth (https://www.naturalearthdata.com/).

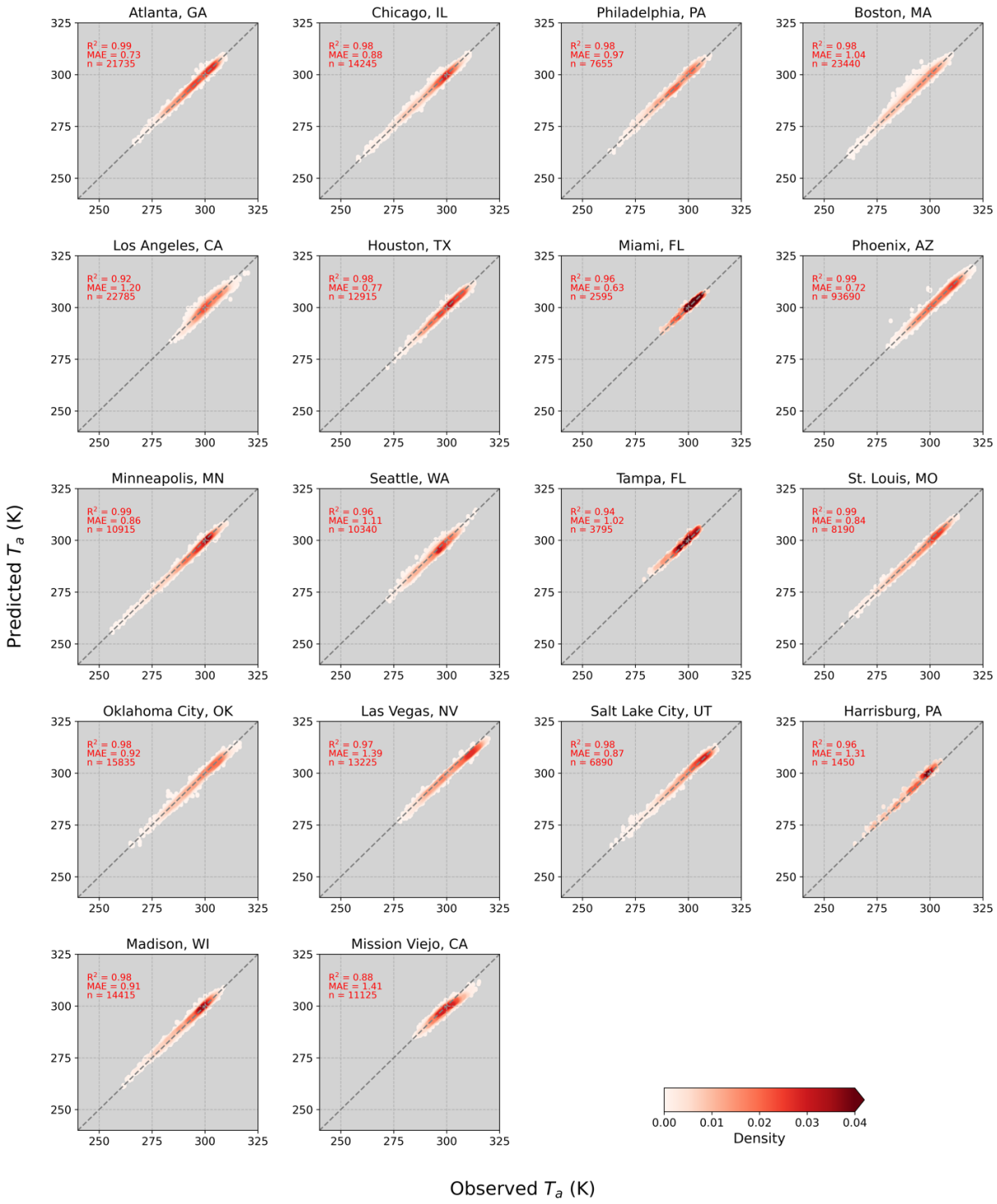

Fig. S15: U-TL's predicted urban T_a (near-surface air temperature) against unseen observations for each city at daytime. Results are from the 9-fold cross-validation experiment. R^2 and MAE (mean absolute error) are reported for each city.

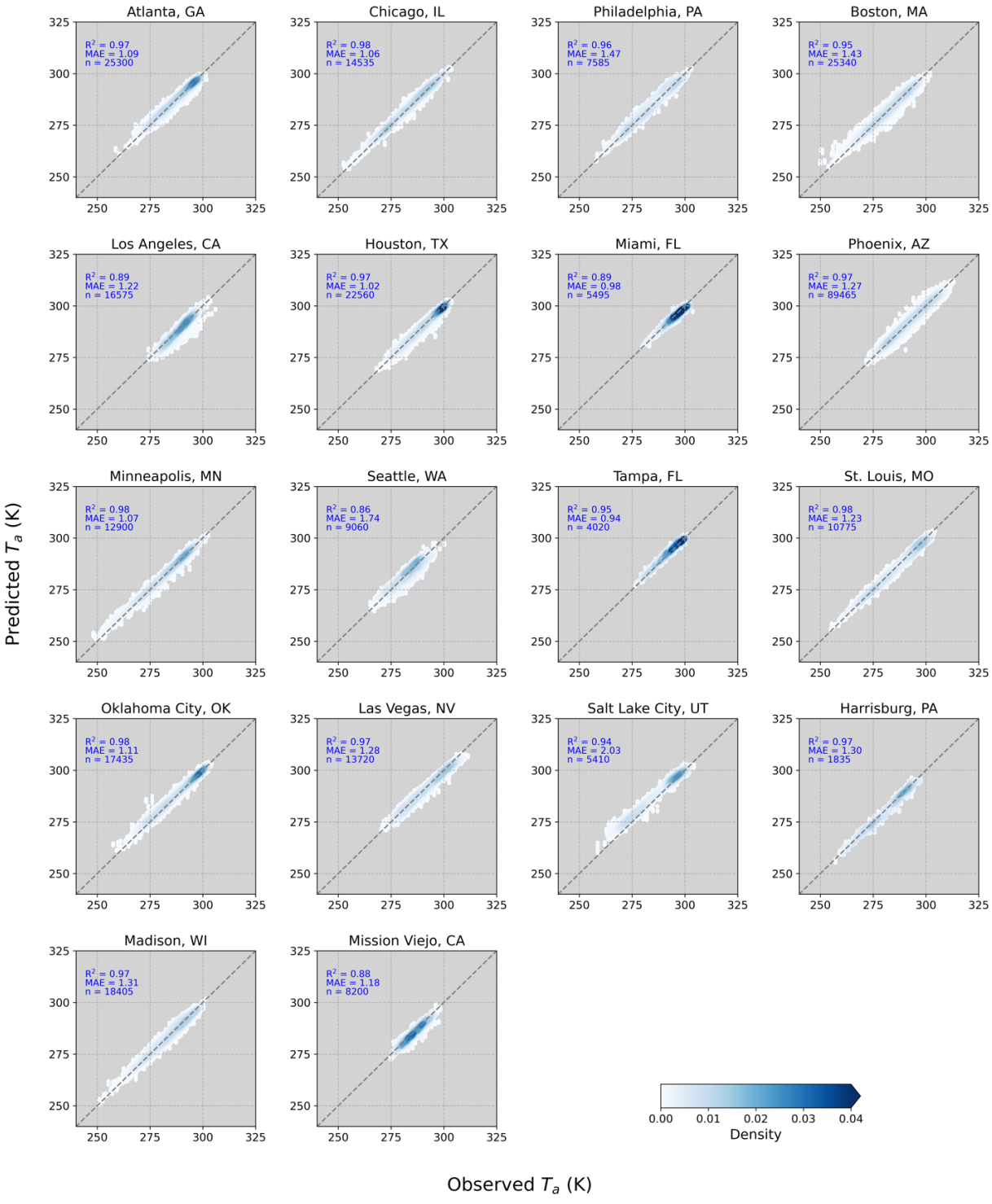

Fig. S16: U-TL's predicted urban T_a (near-surface air temperature) against unseen observations for each city at nighttime. Results are from the 9-fold cross-validation experiment. R^2 and MAE (mean absolute error) are reported for each city.

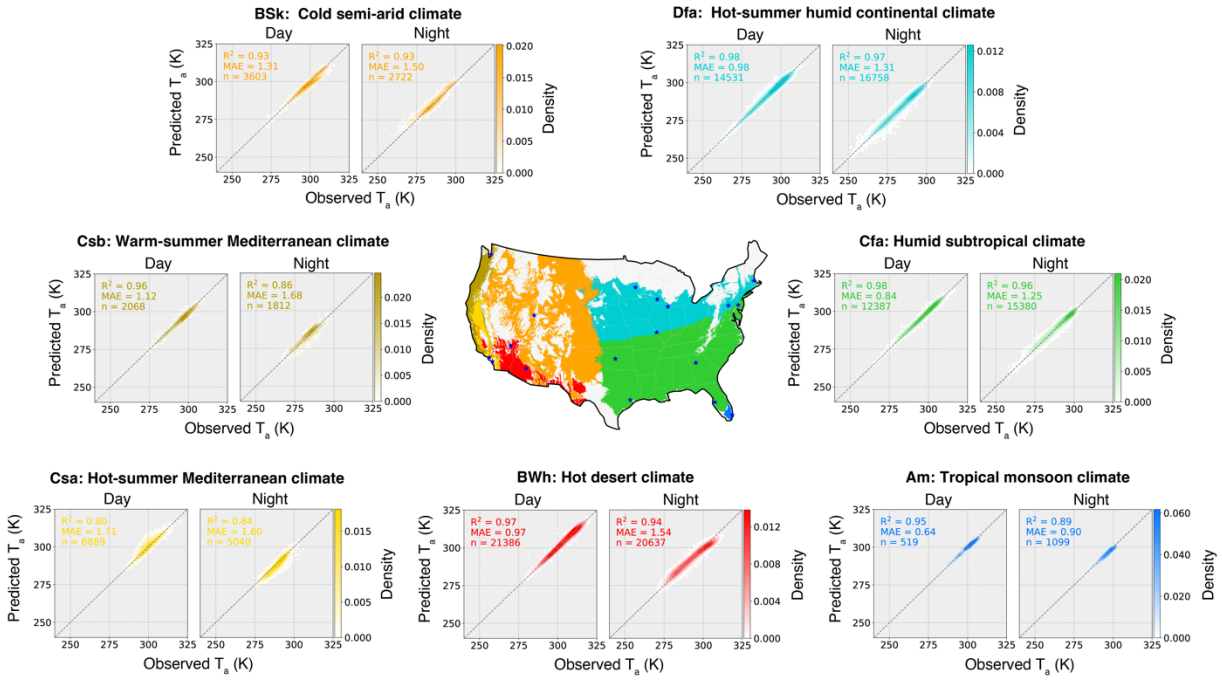

Fig. S17: U-TL's predicted versus observed urban T_a (near-surface air temperature) across Köppen-Geiger climate zones during daytime and nighttime. Each panel shows performance for a held-out climate zone from the leave-climate-zone-out cross-validation, with R^2, MAE (mean absolute error), and sample size (n) reported. Basemap from Natural Earth (https://www.naturalearthdata.com/).

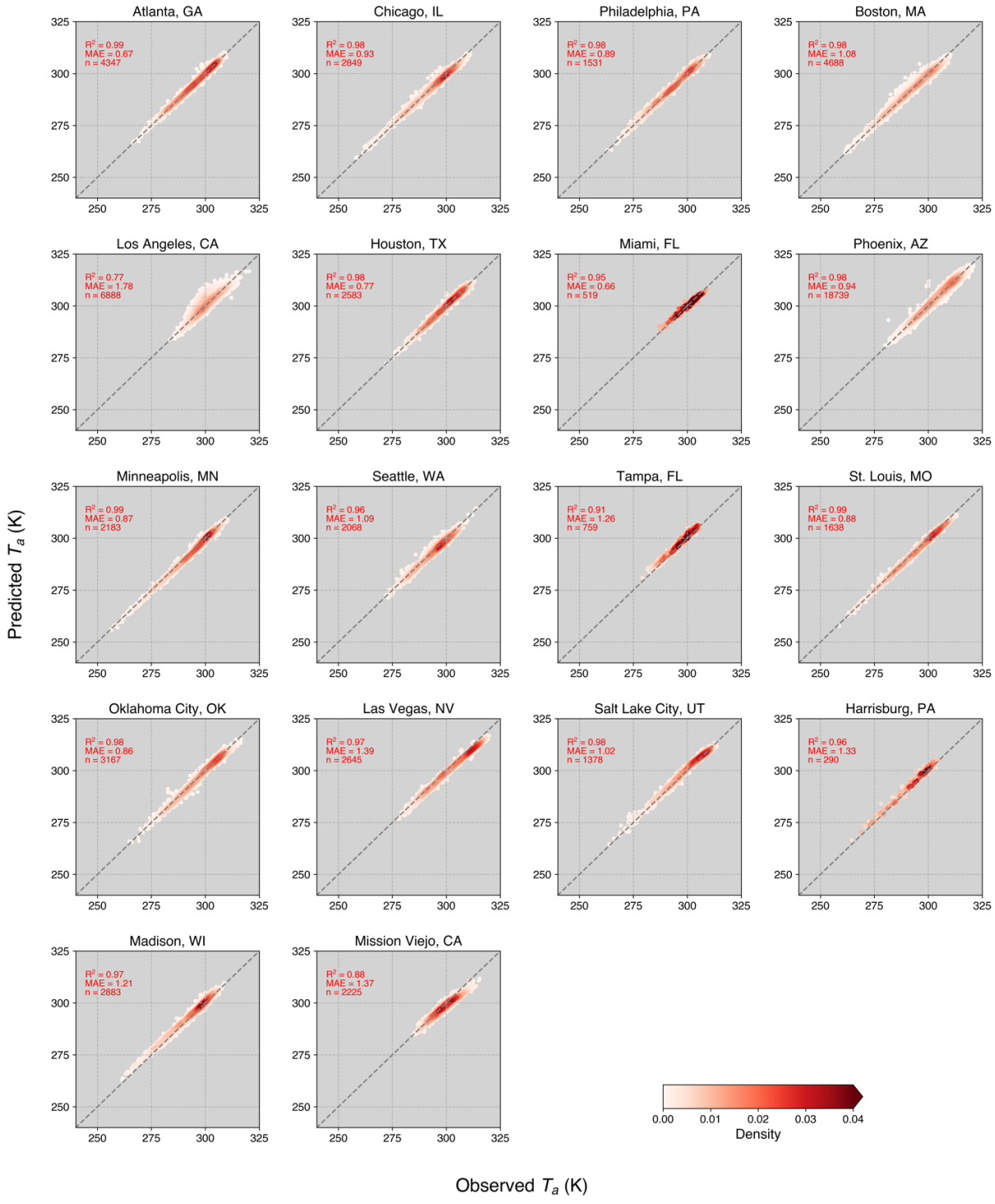

Fig. S18: U-TL's predicted urban T_a (near-surface air temperature) against unseen observations for each city at daytime. Results are from the leave-city-out cross-validation experiment. R^2 and MAE (mean absolute error) are reported for each city.

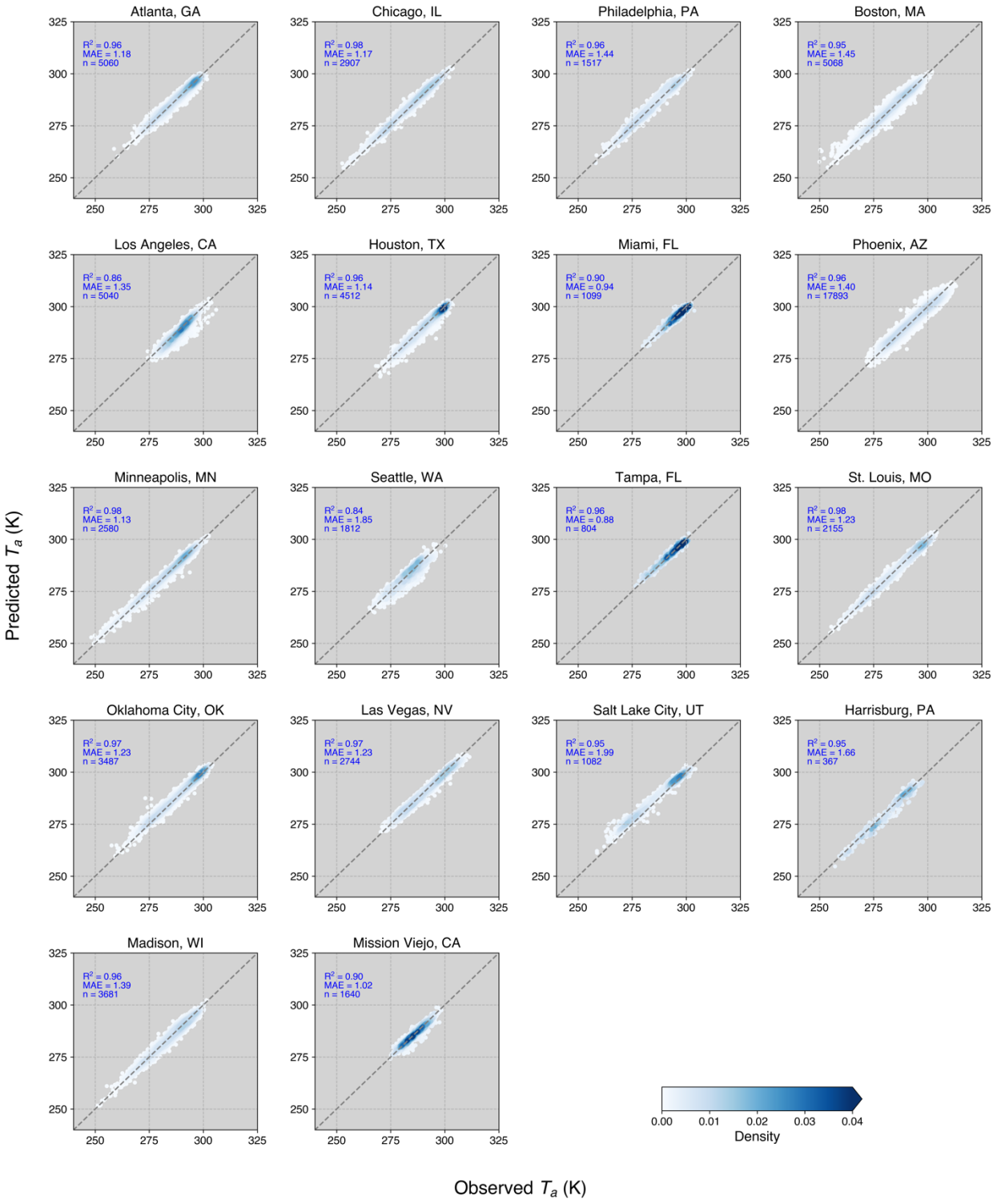

Fig. S19: U-TL's predicted urban T_a (near-surface air temperature) against unseen observations for each city at nighttime. Results are from the leave-city-out cross-validation experiment. R^2 and MAE (mean absolute error) are reported for each city.

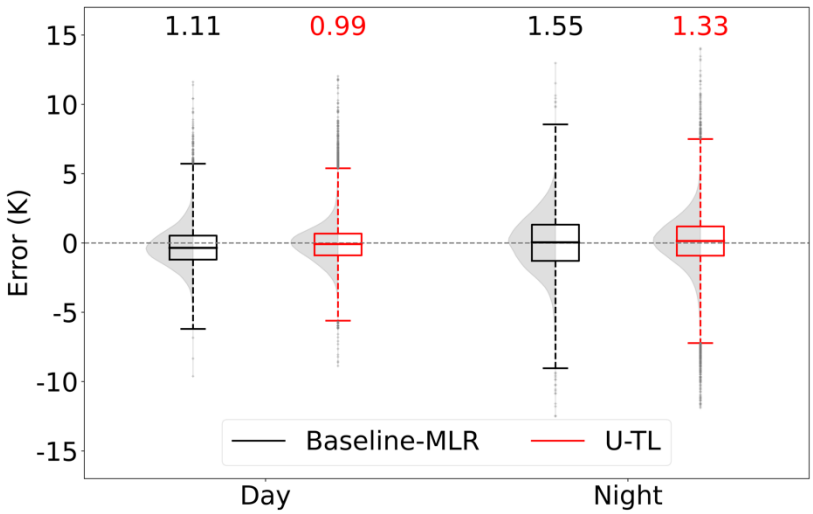

Fig. S20: Comparison between the performance of linear regression model and U-TL. Distributions of estimation errors for the linear baseline model (**black**) and U-TL **(red)** using 3 training folds at daytime and nighttime. MAE (mean absolute error) for each configuration is denoted above the corresponding boxplot. Center bars represent the median, box edges the 25th and 75th percentiles, and error bars extend to 3×IQR (interquartile range) from the quartiles. Shaded violin plots indicate the underlying distribution.

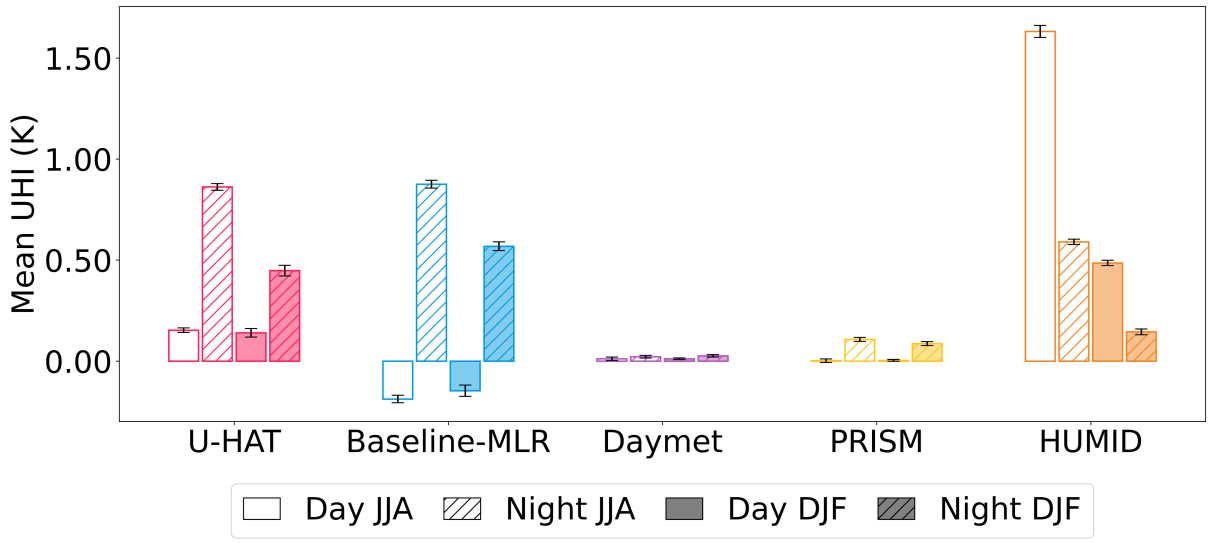

Fig. S21: Comparison of mean UHI_a (canopy urban heat island) intensities from U-HAT, linear regression (Baseline-MLR), Daymet, PRISM, and HUMID for all available cities in the CONUS (contiguous United States) during daytime and nighttime in JJA (June–August) and DJF (December–February), 2013–2020. Results are calculated using the Simplified Urban Extent (SUE) method. HUMID covers 2013–2018 due to data availability. Error bars show the standard error of the mean.

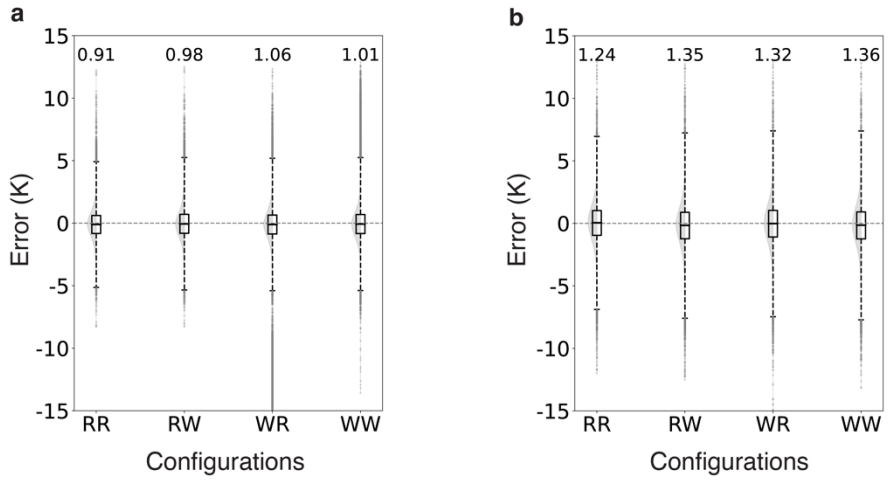

Fig. S22: Distribution of estimation errors for U-TL with randomly shuffled images. Results are shown for daytime (**a**) and nighttime (**b**) models trained on 9 folds of stations. RR = correct images in both pretraining and fine-tuning steps**,** RW = correct images in the pretraining step but mismatched images in the fine-tuning step, WR = mismatched images in the pretraining step but correct images in the fine-tuning step, WW = mismatched images in both pretraining and fine-tuning steps. MAE (mean absolute error) for each configuration is denoted above the corresponding boxplot. Center bars represent the median, box edges the 25th and 75th percentiles, and error bars extend to 3×IQR (interquartile range) from the quartiles. Shaded violin plots indicate the underlying distribution.

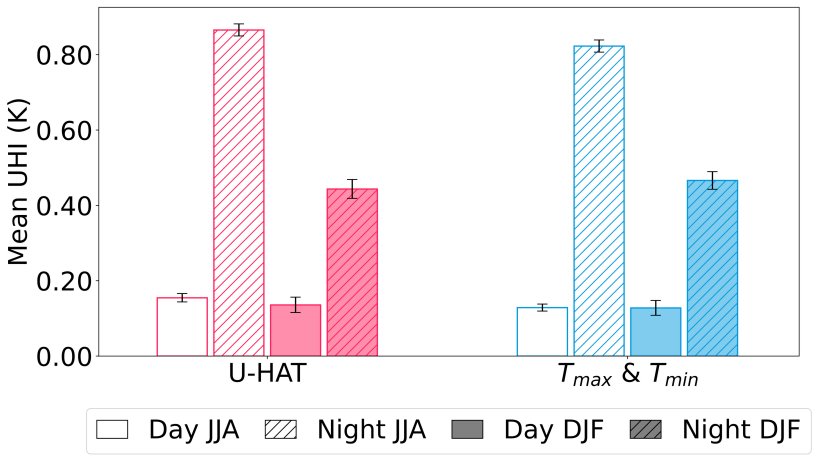

Fig. S23: Comparison of mean UHI_a (canopy urban heat island) intensities at 1:30 AM/PM and times of daily T_a extremes for all available cities in the CONUS (contiguous United States) during daytime and nighttime in JJA (June–August) and DJF (December–February), 2013–2020. Error bars show the standard error of the mean.

Table S1: Auxiliary data used in model training and U-HAT production.

Data	Source	Native resolution (spatial/temporal)	Bands or variables	Date range
Satellite imagery	Landsat 8 Surface Reflectance (Level 2, Collection 2, Tier 1)[4]	30 m/16 day	SR_B2, SR_B3, SR_B4, SR_B5, SR_B6	2013/04/01–2024/01/01
Satellite LST	Aqua MODIS (MYD11A1)	1 km/1 day	LST_Day_1km, LST_Night_1km	2013/04/01–2024/01/01
Forcing fields	ERA5[5,6]	0.25°/1 hour	ssrd, strd, sp, tp, u, v, t, r	2013/04/01–2024/01/01
Building height	Li, M et al.[7]	1 km/—	Building height	—
Elevation	NASA SRTM Digital Elevation[8]	30 m/—	Elevation	—

Table S2: Mean absolute error (MAE) grouped by dominant National Land Cover Dataset (NLCD) class at each station.

Class Number	Class name	Number of stations	Daytime MAE (K)	Nighttime MAE (K)
21	Developed open space	10	0.83	1.17
22	Developed low intensity	10	1.00	1.26
23	Developed medium intensity	14	0.87	1.15
24	Developed high intensity	14	0.96	1.22
52	Shrub/scrub	4	0.91	1.56

Table S3: Mean absolute error (MAE) grouped by dominant Local Climate Zone (LCZ) class at each station.

Class Number	Class name	Number of stations	Daytime MAE (K)	Nighttime MAE (K)
5	Open midrise	2	1.16	1.34
6	Open lowrise	7	0.98	1.07
8	Large lowrise	10	1.02	1.20
9	Sparsely built	1	0.72	1.05
12	Bare rock or paved	1	1.19	1.78
14	Low plants with scattered trees	25	0.83	1.24
15	Bare soil or sand	2	1.05	1.16
16	Low plants	4	0.84	1.38

Table S4: Summary of gridded air temperature datasets covering the contiguous United States. U-HAT is denoted in green, monthly-resolution datasets are denoted in yellow; dataset not available for download is denoted in gray.

Literature	Name of dataset	Summary of method	Spatial & temporal extent	Spatial & temporal resolution	Variables	Overall Accuracy reported (default: MAE)	Urban Accuracy against our selected stations	Urban Accuracy against selected stations in local networks
This study	U-HAT	U-TL	CONUS; 2013–2023	1 km; daily	T_a at 1:30 PM and 1:30 AM	Same as urban accuracy	0.99 K (1:30 PM, daily) 1.33 K (1:30 AM, daily) 0.64 K (1:30 PM, monthly mean) 0.86 K (1:30 AM, monthly mean)	0.68 K (1:30 PM, daily) 1.13 K (1:30 AM, daily) 0.59 K (1:30 PM, monthly mean) 0.73 K (1:30 AM, monthly mean)
Daly et al.[9]	PRISM	Regression-based spatial interpolation	CONUS; 1895–present	800 m; daily	T_{max}, T_{min}	0.36–0.77 ºC (T_{max}) 0.56–1.35 ºC (T_{min})	0.46 K (T_{max}) 1.42 K(T_{min})	0.62 K (T_{max}) 1.71 K (T_{min})
Oyler et al.[10]	TopoWx	Regression-based spatial interpolation	CONUS; 1948–2016	30-arcsec (~800m); daily	T_{max}, T_{min}	1.03 ºC (T_{max}) 1.06 ºC (T_{min})	0.93 K (T_{max}) 1.46 K (T_{min})	1.06 K (T_{max}) 1.71 K (T_{min})
MacDonald et al.[11]	–	Spatial interpolation	North America; 1901–2016	60-arcsec (~2 km); monthly	T_{max}, T_{min}	0.71 ºC (T_{max}) 1.02 ºC (T_{min})	0.54 K (T_{max}) 1.26 K (T_{min})	0.63 K (T_{max}) 1.57 K (T_{min})
Thornton et al.[12]	Daymet	Regression-based spatial interpolation	North America, Hawaii, and Puerto Rico;	1 km; daily	T_{max}, T_{min}	1.52 ºC (T_{max}) 1.78 ºC (T_{min})	0.76 K (T_{max}) 1.42 K (T_{min})	0.82 K (T_{max}) 1.69 K (T_{min})

			1950–present calendar year					
Li et al.[13]	–	Geographically weighted regression	CONUS, 2003–2016	1 km; daily	T_{max}, T_{min}	1.30 ºC (T_{max}) 1.50 ºC (T_{min})	Data not available	
Hooker et al.[14]	–	Geographically and climate space weighted regressions	Global; 2003–2016	0.05 º; monthly	T_{mean}	1.14–1.55 ºC	0.96 K (T_{mean})	1.01 K (T_{mean})
Verdin et al.[15]	CHIRTS -daily	Regression-based data fusion	60°S-70°N, 1983–2016	0.05° × 0.05°; daily	T_{max}, T_{min}	1.6 ºC (T_{max}) 1.4 ºC (T_{min}) for North America	1.44 K (T_{max}) 1.85 K (T_{min})	1.56 K (T_{max}) 1.63 K (T_{min})
Zhang et al.[3]	–	Regression with spatially varying coefficient	Global; 2003–2020	1 km; daily	T_{max}, T_{min}	1.58 ºC (T_{max}) 1.41 ºC (T_{min}) for impervious surfaces	2.13 K (T_{max}) 2.03 K (T_{min})	2.31 K (T_{max}) 1.99 K (T_{min})
Yao et al.[2]	GSHTD	Cubist model	Global, 2001–2020	1 km; monthly	T_{mean}, T_{max}, T_{min}	0.80 ºC (T_{mean}) 1.00 ºC (T_{max}) 1.06 ºC (T_{min})	1.33 K (T_{max}) 0.78 K (T_{min})	1.56 K (T_{max}) 0.68 K (T_{max})
Newman et al.[16]	HUMID	Physics-based model + bias correction	CONUS, 1981–2018	1 km; daily	T_{mean}, T_{max}, T_{min}	RMSE of 2.0 ºC (T_{max}) and 1.4 ºC (T_{min}) against Daymet	2.42 K (T_{max}) 1.47 K (T_{min})	3.00 K (T_{max}) 1.29 K (T_{min})

Table S5: Definitions of abbreviations used in this study.

Abbreviation	Full name
CONUS	Contiguous United States
CWS	Citizen weather station
DNN	Deep neural network
HE_a	Heat exposure measured by air temperature
HE_s	Heat exposure measured by skin temperature
ISF	Impervious surface fraction
MAE	Mean absolute error
NDVI	Normalized Difference Vegetation Index
SUE	Simplified Urban Extent
T_a	Near-surface air temperature
T_s	Land surface temperature
U-HAT	Urban high-resolution air temperature
U-TL	Urban transfer learning framework
UHI_a	Canopy urban heat island
UHI_s	Surface urban heat island
WMO	World Meteorological Organization

Supplementary References

1. Stewart, I. D. & Oke, T. R. Local Climate Zones for Urban Temperature Studies. *Bull. Am. Meteorol. Soc.* **93**, 1879–1900 (2012).
2. Yao, R. *et al.* Global seamless and high-resolution temperature dataset (GSHTD), 2001–2020. *Remote Sens. Environ.* **286**, 113422 (2023).
3. Zhang, T. *et al.* A global dataset of daily maximum and minimum near-surface air temperature at 1 km resolution over land (2003–2020). *Earth Syst. Sci. Data* **14**, 5637–5649 (2022).
4. Earth Resources Observation and Science (EROS) Center. Landsat 8-9 Operational Land Imager / Thermal Infrared Sensor Level-2, Collection 2. U.S. Geological Survey https://doi.org/10.5066/P9OGBGM6 (2020).
5. Copernicus Climate Change Service, Climate Data Store. ERA5 hourly data on pressure levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) https://doi.org/10.24381/CDS.BD0915C6 (2023).
6. Copernicus Climate Change Service, Climate Data Store. ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) https://doi.org/10.24381/CDS.ADBB2D47 (2023).
7. Li, M., Wang, Y., Rosier, J. F., Verburg, P. H. & van Vliet, J. Global maps of 3D built-up patterns for urban morphological analysis. *Int. J. Appl. Earth Obs. Geoinf.* **114**, 103048 (2022).
8. Farr, T. G. *et al.* The Shuttle Radar Topography Mission. *Rev. Geophys.* **45**, (2007).
9. Daly, C., Taylor, G. & Gibson, W. The Prism Approach to Mapping Precipitation and Temperature.

10. Oyler, J. W., Ballantyne, A., Jencso, K., Sweet, M. & Running, S. W. Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. *Int. J. Climatol.* **35**, 2258–2279 (2015).

11. MacDonald, H. *et al.* North American historical monthly spatial climate dataset, 1901–2016. *Sci. Data* **7**, 411 (2020).

12. Thornton, P. E. *et al.* Gridded daily weather data for North America with comprehensive uncertainty quantification. *Sci. Data* **8**, 190 (2021).

13. Li, X., Zhou, Y., Asrar, G. R. & Zhu, Z. Developing a 1 km resolution daily air temperature dataset for urban and surrounding areas in the conterminous United States. *Remote Sens. Environ.* **215**, 74–84 (2018).

14. Hooker, J., Duveiller, G. & Cescatti, A. A global dataset of air temperature derived from satellite remote sensing and weather stations. *Sci. Data* **5**, 180246 (2018).

15. Verdin, A. *et al.* Development and validation of the CHIRTS-daily quasi-global high-resolution daily temperature data set. *Sci. Data* **7**, 303 (2020).

16. Newman, A. J. *et al.* The High-resolution Urban Meteorology for Impacts Dataset (HUMID) daily for the Conterminous United States. *Sci. Data* **11**, 1321 (2024).