

Importance of spatially-continuous urban surface properties in urban-resolving Earth system modeling

Yifan Cheng¹, Lei Zhao^{1,2,3}, Keith Oleson⁴, TC Chakraborty⁵, Keer Zhang⁶, Cenlin He⁷, Joyce Yang¹

¹Department of Civil and Environmental Engineering, The Grainger College of Engineering, University of Illinois Urbana-Champaign, Urbana, IL, USA;

²Institute for Sustainability, Energy, and Environment (ISEE), University of Illinois Urbana-Champaign, Urbana, IL, USA;

³Department of Climate, Meteorology & Atmospheric Sciences, University of Illinois Urbana-Champaign, Urbana, IL, USA;

⁴Climate and Global Dynamics Laboratory, NSF National Center for Atmospheric Research, Boulder, Colorado, USA;

⁵Atmospheric, Climate, and Earth Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA;

⁶High Meadows Environmental Institute, Princeton University, Princeton, NJ, USA;

⁷Research Applications Laboratory, NSF National Center for Atmospheric Research, Boulder, Colorado, USA

* Correspondence to: Lei Zhao (leizhao@illinois.edu)

Contents of this file

Text S1 to S4
Figures S1 to S13
Table S1

Introduction

The supporting information contains detailed description of data preparation (Text S1, Figure S1, S2), simulation configuration (Text S2), additional information on validation dataset (Figure S3, Table S1) and evaluation (Text S3, S4), as well as complementary results, including the corresponding wintertime validation results (Figures S4-S7) and extended analyses (Figures S8-S13).

Text S1.

Generation of surface datasets

All surface datasets were generated using the mksurfdata_esmf tools available with CESM. The tool accepts urban input data at maximum resolution of $0.05^\circ \times 0.05^\circ$ due to computational and data storage constraints. This limitation poses the issue of double interpolation for higher resolutions ($1/32^\circ$ in our case), since we need to first aggregate the 1-km data to 0.05° , then the tool remaps and adjusts the urban percentage to ensure all land unit weights sum up to 100%. To preserve actual urban extent from U-Surf in USURF simulations, we manually replaced the PCT_URBAN variable in tool-generated surface datasets (hereafter referred to as DEF_urban) with aggregated values (TRUE_urban). In normal cases where DEF_urban is less than 100%, we adjusted all other land unit fractions (natural vegetation, crop, lake, wetland, glacier) proportionally using the scale factor $(1 - \text{TRUE_urban})/(1 - \text{DEF_urban})$. For edge cases where DEF_urban is equal to 100% without baseline reference for proportional adjustment, we assigned excess urban percentage to natural vegetation. The comparison between DEF_urban and TRUE_urban is shown in Figure S2. Note that although U-Surf no longer relies on mapping between urban density classes to parameters, the default code still reads in urban data by density class. Thus, we assigned the total urban percentage and surface parameters for each grid exclusively to the MD class in all USURF runs (i.e., urban percentage in TBD and HD were set as zero, ensuring they wouldn't invoke the urban scheme for extra calculation).

Text S2.

Calculation of urban land unit aerodynamic constants

In the point-scale simulations, we modified the code of calculating urban land unit aerodynamic constants by introducing a new variable called wall_to_plan_area_ratio (λ_w) to be consistent with the method adopted in the Urban-PLUMBER project. This variable is defined as: $\lambda_w = 2H/W(1 - WT_{roof})$, where WT_{roof} represents the roof fraction with respect to the urban surface (WTLUNIT_ROOF) and H/W is the canyon height-to-width ratio (CANYON_HWR).

The modification replaces the original calculation method (eqn. S1-S2) for plan area index (λ_p) and frontal area index (λ_f) with an alternative approach (eqn. S3-S4) based on relationships derived from Porson et al., (2009) and Masson et al., (2020):

$$\lambda_p = \frac{H/W}{H/W+1} = \frac{H}{H+W} \quad (S1)$$

$$\lambda_f = (1 - \lambda_p) \cdot H/W \quad (S2)$$

$$\lambda_p = WT_{roof} \quad (S3)$$

$$\lambda_f = \frac{\lambda_w}{\pi} = \frac{2 \cdot H/W}{(1 - \lambda_p) \cdot \pi} \quad (S4)$$

Both methods then use the calculated plan area index to derive the displacement height:

$$z_d = 1 + \alpha^{-\lambda_p} \cdot (\lambda_p - 1) \cdot H \quad (S5)$$

where α is a constant coefficient with the value of 4.43.

The displacement height and frontal area index subsequently determine the surface roughness length:

$$z_0 = H \cdot (1 - \frac{z_d}{H}) \cdot \exp[-(0.5 \cdot \frac{\beta \cdot C_d}{\kappa^2} \cdot (1 - \frac{z_d}{H}) \cdot \lambda_f)^{-0.5}] \quad (S6)$$

where β is a constant coefficient with the value of 1, C_d is the drag coefficient as used in Grimmond & Oke (1999) with the value of 1.2, and κ is the von Karman constant of 0.4.

Text S3.

Calculation of relative humidity in CLMU

In CLMU, the saturation vapor pressure e_s given the air temperature T_C is computed using an 8th-degree polynomial fit:

$$e_s(T_C) = 100 \times (a_0 + a_1 T_C + a_2 T_C^2 + \dots + a_8 T_C^8) \text{ if } T_C \geq 0^\circ C \quad (S7)$$

$$e_s(T_C) = 100 \times (c_0 + c_1 T_C + c_2 T_C^2 + \dots + c_8 T_C^8) \text{ if } -75^\circ C \leq T_C < 0^\circ C \quad (S8)$$

where $a_0, \dots, a_8$ and $c_0, \dots, c_8$ are empirical coefficients for water and ice, respectively.

Then the saturation specific humidity is calculated as:

$$q_s(T_C, p) = \frac{0.622 \cdot e_s(T_C)}{p - 0.378 \cdot e_s(T_C)} \quad (S9)$$

The relative humidity is computed as:

$$RH = \min(100\%, \frac{q}{q_s(T_C, p)} \times 100\%) \quad (S10)$$

The values of empirical coefficients are documented as below:

a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
6.112	0.444	0.143e-01	0.264e-03	0.306e-05	0.196e-07	0.892e-10	-0.373e-12	0.209e-15
c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
6.111	0.503	0.188e-01	0.421e-03	0.614e-05	0.603e-07	0.388e-09	0.149e-11	0.263e-14

Text S4.

Calculation of Variable Importance in Projection (VIP)

In the supervised Principal Component Analysis (sPCA), we calculated Variable Importance of Projection (VIP) to identify the top 10 most important drivers of information loss, as displayed in the biplot (Figure 11):

$$VIP_i = \sqrt{p \frac{\sum_{a=1}^A s_a (\frac{w_{ia}^2}{\|w_a\|^2})}{\sum_{a=1}^A s_a}}, s_a = \|t_a\|^2 q_a^2 \quad (S11)$$

where p is the number of predictors, A is the number of components. Subscripts a and i stands for individual component and predictor. t, w represents the score and weight matrix of input feature X , respectively. q is the loading matrix of Y . The top 10 dominant drivers underlying the spatial information loss of Bowen ratio and net radiation are listed below:

Bowen ratio	X	μ_{WTR}	σ_{WTR}	μ_{HT}	μ_{HWR}	σ_{ALBpr}	σ_{HWR}	μ_{ALBpr}	σ_{ALBr}	μ_{ALBir}	μ_{EMpr}
	VIP	1.758	1.648	1.645	1.337	1.098	1.030	1.001	0.847	0.845	0.842
Net radiation	X	μ_{WTP}	σ_{HT}	σ_{ALBpr}	σ_{WTP}	σ_{ALBr}	μ_{ALBw}	σ_{WTR}	σ_{HWR}	σ_{ALBir}	μ_{HWR}
	VIP	1.331	1.317	1.250	1.168	1.147	1.119	1.117	1.105	1.083	1.015

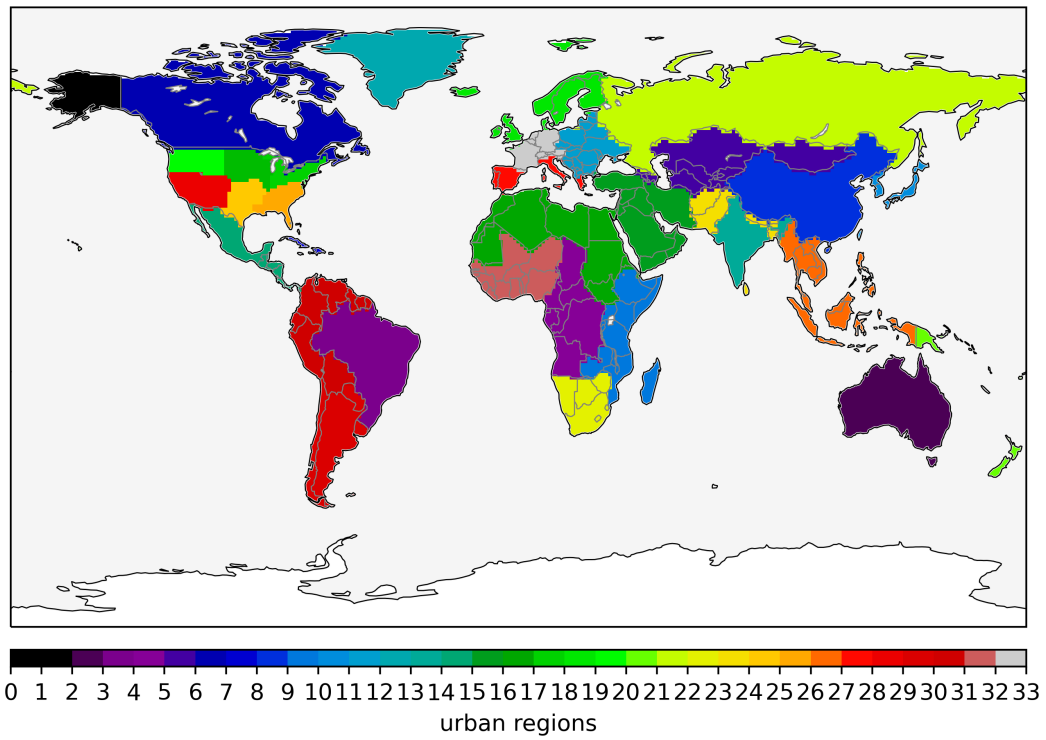


Figure S1. The thirty-three regions used to define the global variability of urban surface properties in Jackson et al., (2010) [adapted from Figure S1 in Cheng et al., 2025].

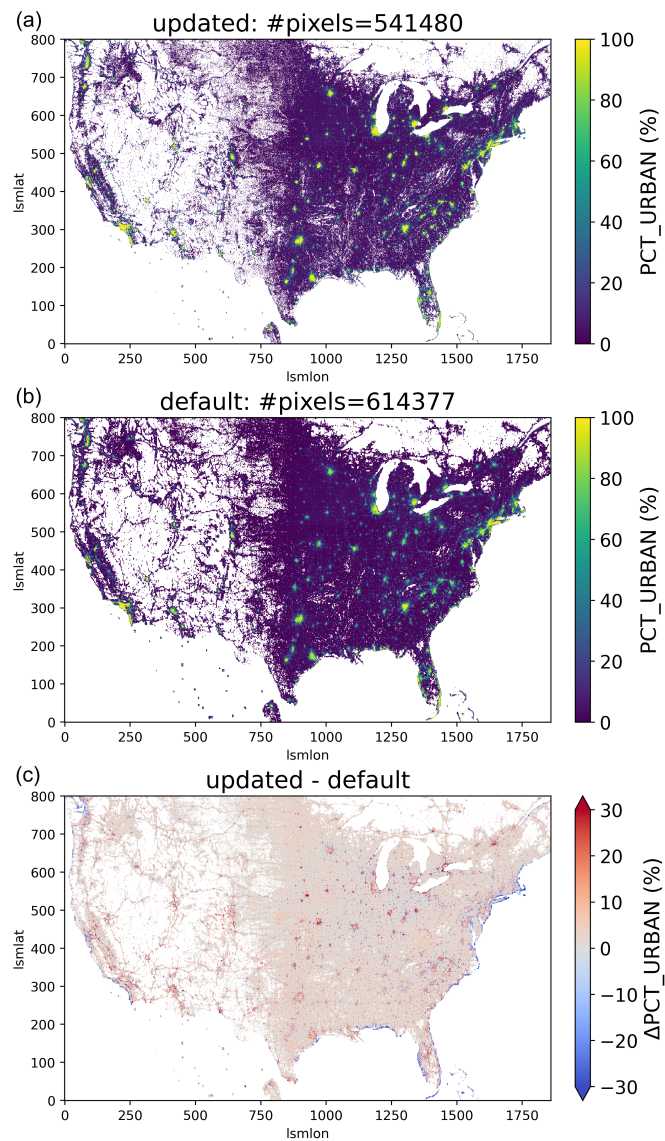


Figure S2. Urban percentage adjustments in high-resolution simulations (USURF_UHR run). (a) Updated urban percentage using U-Surf raw data (TRUE_urban), (b) default urban percentage from mksurfdata_esmf tool (DEF_urban), (c) difference between updated and default values.

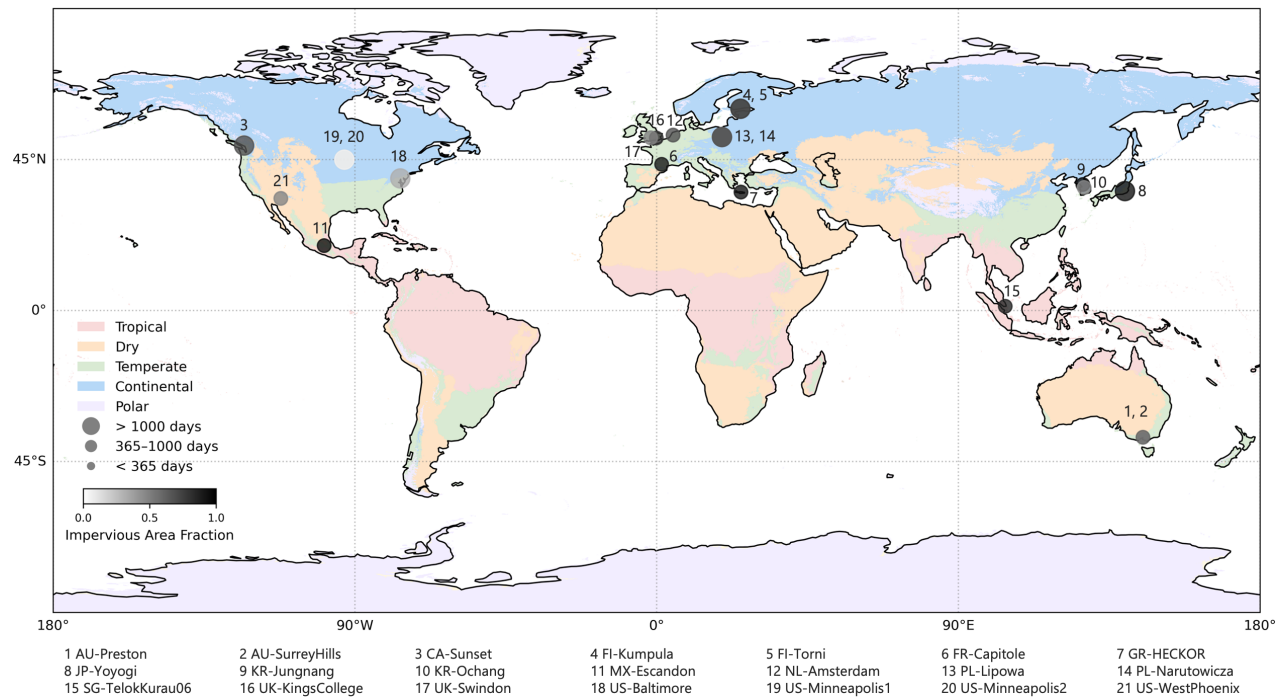


Figure S3. Locations of 21 Urban-PLUMBER stations. Circle size indicates the length of observation period available (>1000 days, 365-1000 days, <365 days), and circle color represents the impervious area fraction at each site. Observations at the Minneapolis KUOM tall tower are separated into residential (US-Minneapolis1) and recreational (US-Minneapolis2) sectors. Background climate zones are based on the 0.01° Köppen-Geiger map from Beck et al. (2018). [Adapted from Figure 1 in Lipson et al., (2022) and Figure A1 in Sun et al., (2025).]

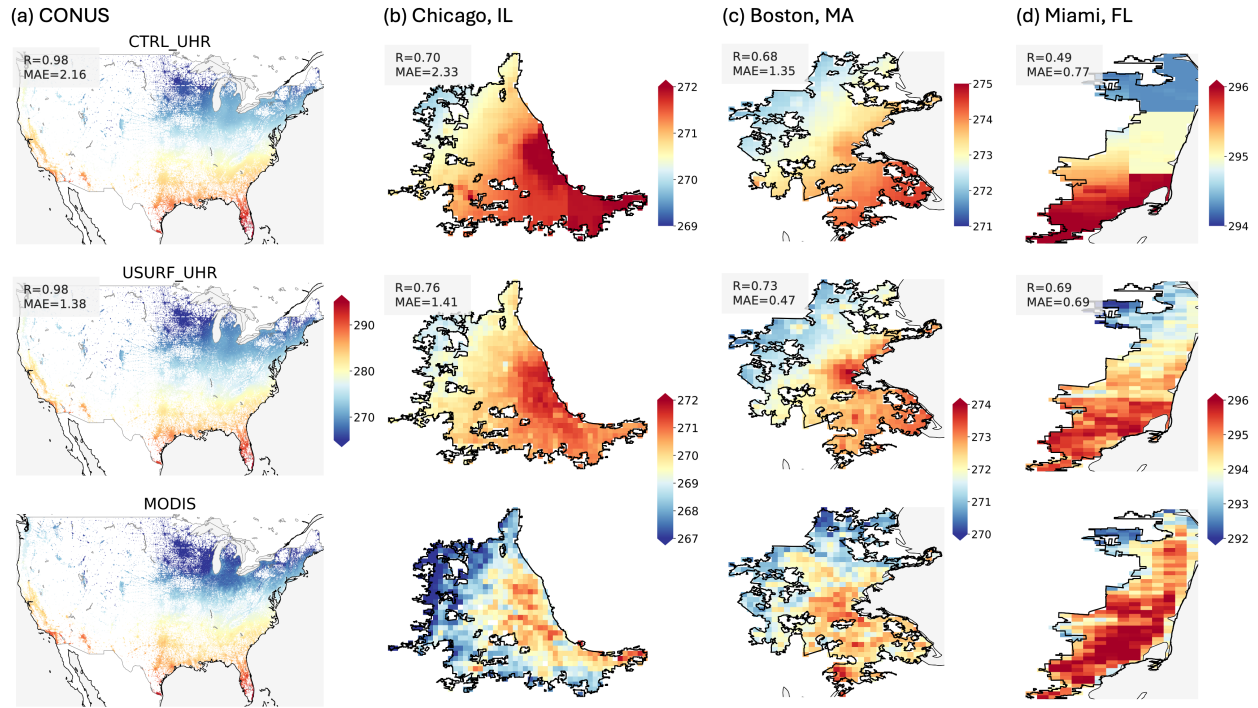


Figure S4. Five-year (2010-2014) wintertime average urban land surface temperature (LST) comparison between land-only CESM simulations and MODIS observations at $0.03125^\circ \times 0.03125^\circ$ resolution. The performance metrics correlation coefficient (R) and mean absolute error (MAE) [K] are demonstrated in (a) CONUS and selected U.S. cities: (b) Chicago, IL, (c) Boston, MA, (d) Miami, FL, only using common urban pixels. The city boundary shapefiles are from the U.S. Census Bureau (2024). Panel (a) shares a common colorbar. In panels (b)-(d), CTRL uses separate colorbars due to its narrower spatial distribution; while USURF and MODIS share the same colorbars for spatial pattern comparison.

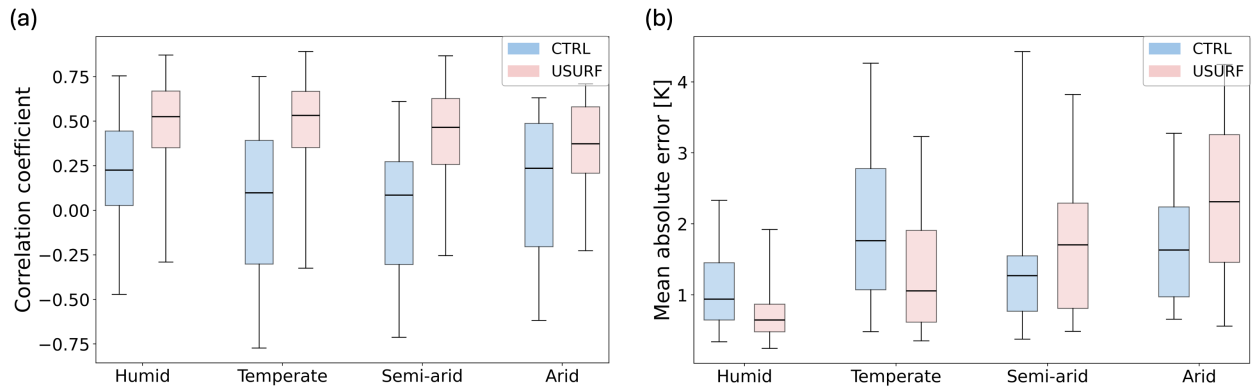


Figure S5. Five-year (2010-2014) wintertime average land surface temperature (LST) correlation coefficient and mean absolute error (unit: K) validated against MODIS LST over 453 CONUS cities as defined in the U.S. Census Bureau (2024). Box and whisker plots show the 25th percentile, the median, and the 75th percentile (the bottom, middle, and top horizontal bars) and extend to the 1st and 99th quantiles.

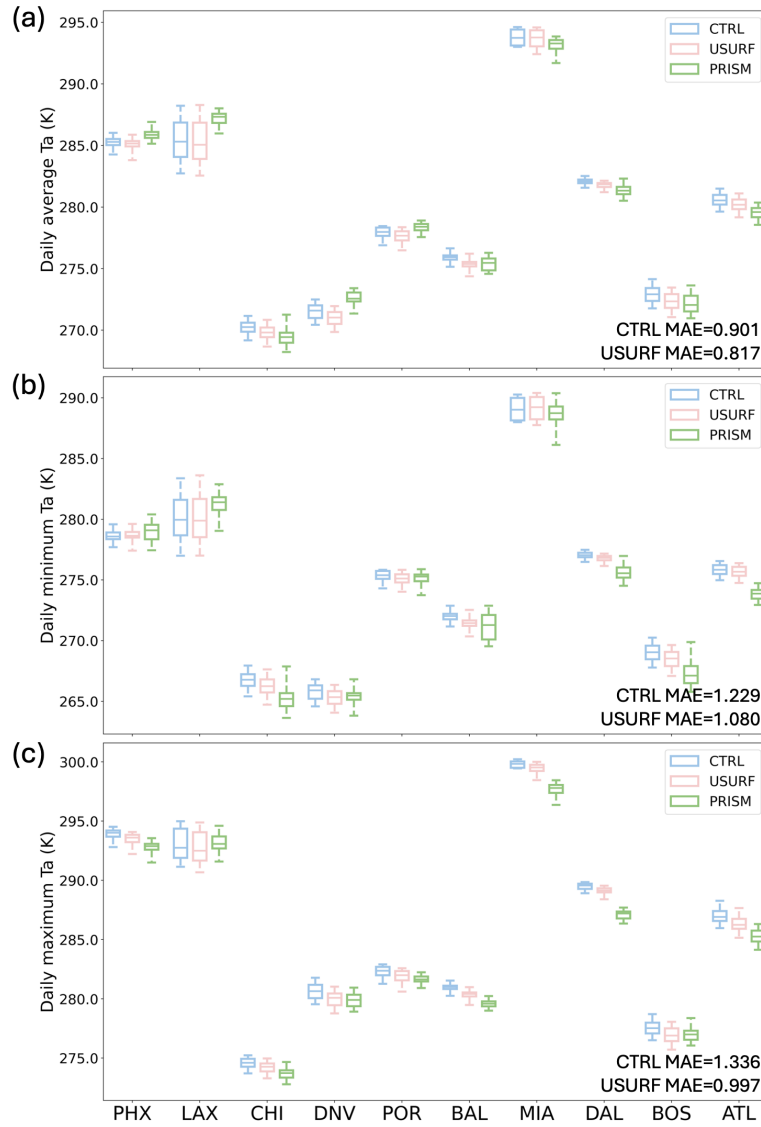


Figure S6. Distribution of five-year (2010-2014) wintertime (a) average daily average, (b) minimum and (c) maximum near-surface air temperature (T_a) for 10 major CONUS cities: Phoenix, AZ; Los Angeles, CA; Chicago, IL; Denver, CO; Portland, OR; Baltimore, MD; Miami, FL; Dallas, TX; Boston, MA; Atlanta, GA. Numbers on the bottom right corner stand for the city average mean absolute error (MAE) comparing USURF and CTRL simulations with PRISM dataset as reference, respectively. Box and whisker plots show the 25th percentile, the median, and the 75th percentile (the bottom, middle, and top horizontal bars) and extend to the 1st and 99th quantiles.

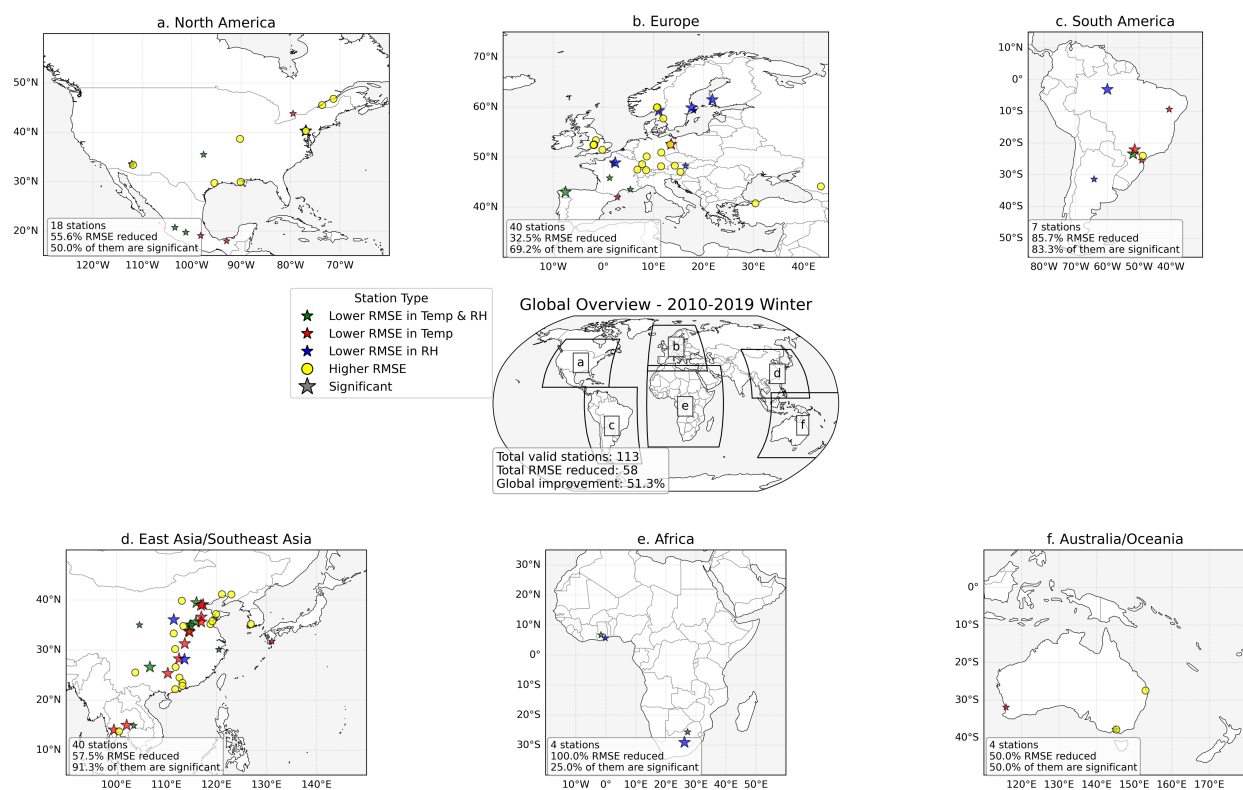


Figure S7. Improvement in wintertime model performance for daily average near-surface air temperature and relative humidity at 113 available urban stations during 2010–2019, shown across six continents: (a) North America, (b) Europe, (c) South America, (d) East/Southeast Asia, (e) Africa, and (f) Australia/Oceania. The modeled results are from $0.23 \times 0.31^\circ$ land-only CESM simulations as described in Table 2. The improvement is demonstrated by the decrease in the root mean squared error (RMSE) in this plot.

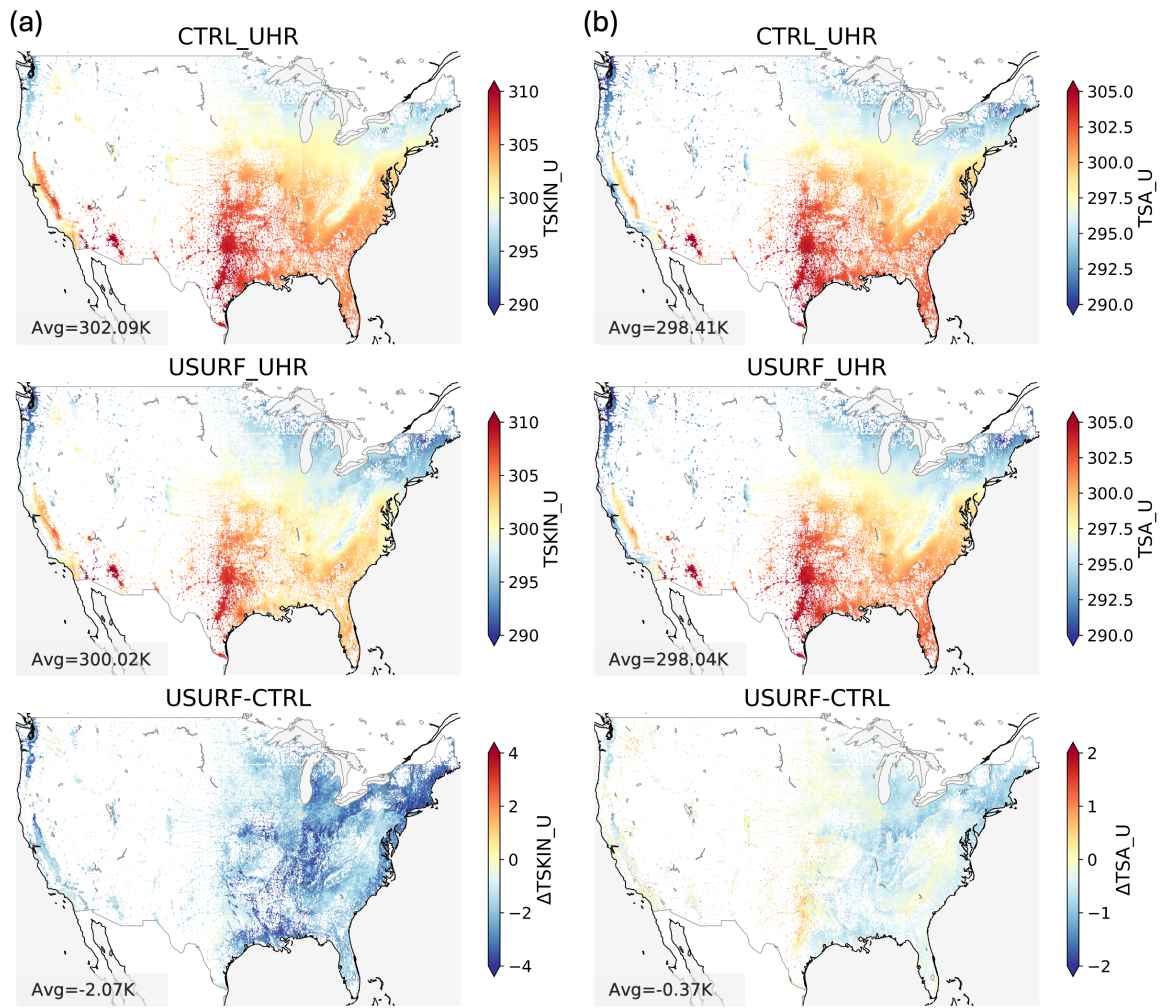


Figure S8. Summertime temperature differences between USURF and CTRL simulations for urban (a) land surface temperature (TSKIN_U) and (b) near-surface air temperature (TSA) (unit: K). Maps show 2010-2014 averages from CTRL_UHR and USURF_UHR runs described in Table 2.

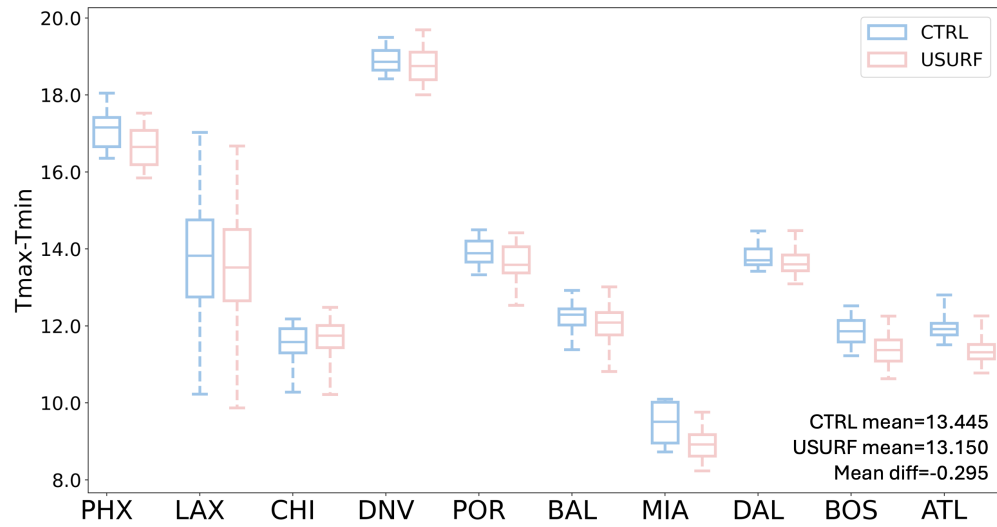


Figure S9. Distribution of five-year (2010-2014) summertime daily diurnal range of near-surface air temperature (T_a) for 10 major CONUS cities: Phoenix, AZ; Los Angeles, CA; Chicago, IL; Denver, CO; Portland, OR; Baltimore, MD; Miami, FL; Dallas, TX; Boston, MA; Atlanta, GA. Numbers on the bottom right corner stand for the city average mean values from CTRL and USURF simulations and the average difference, respectively. Box and whisker plots show the 25th percentile, the median, and the 75th percentile (the bottom, middle, and top horizontal bars) and extend to the 1st and 99th quantiles.

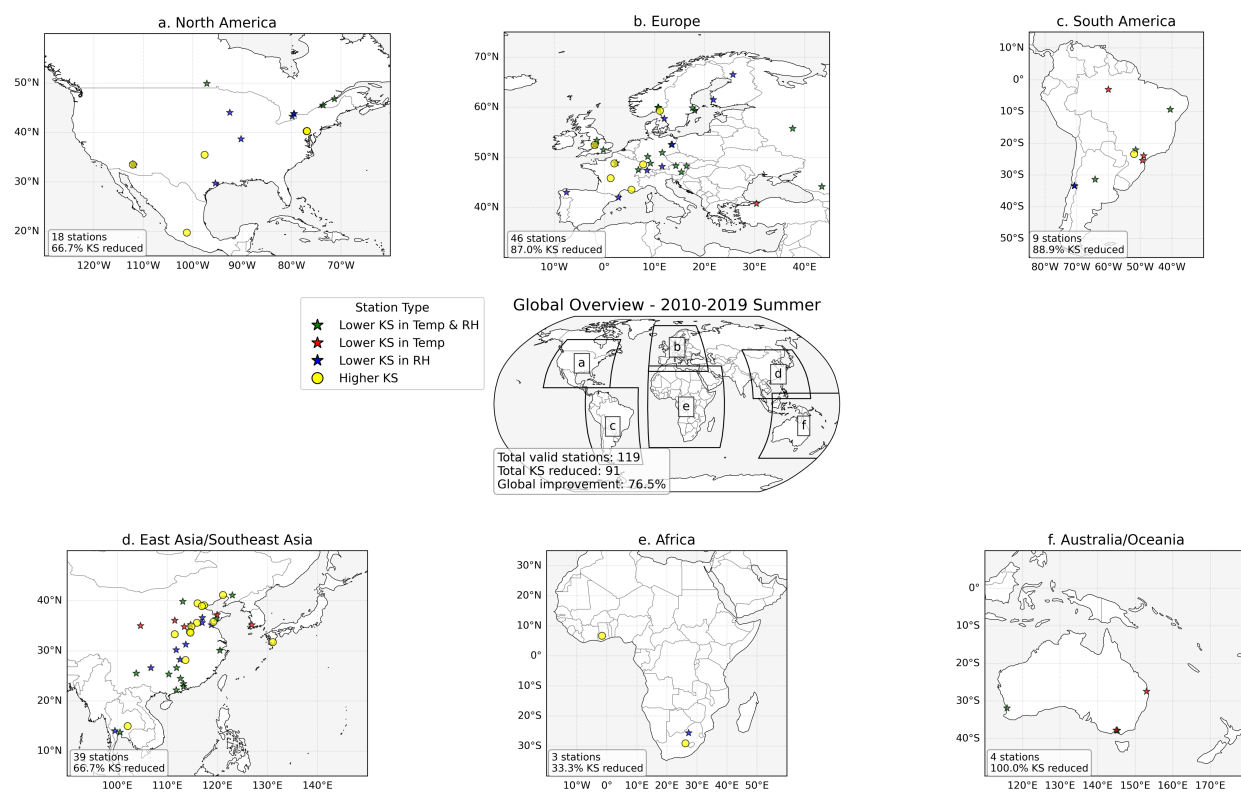


Figure S10. Same as Figure S7, but for summertime average with improvement determined by the decrease in the Kolmogorov–Smirnov statistic.

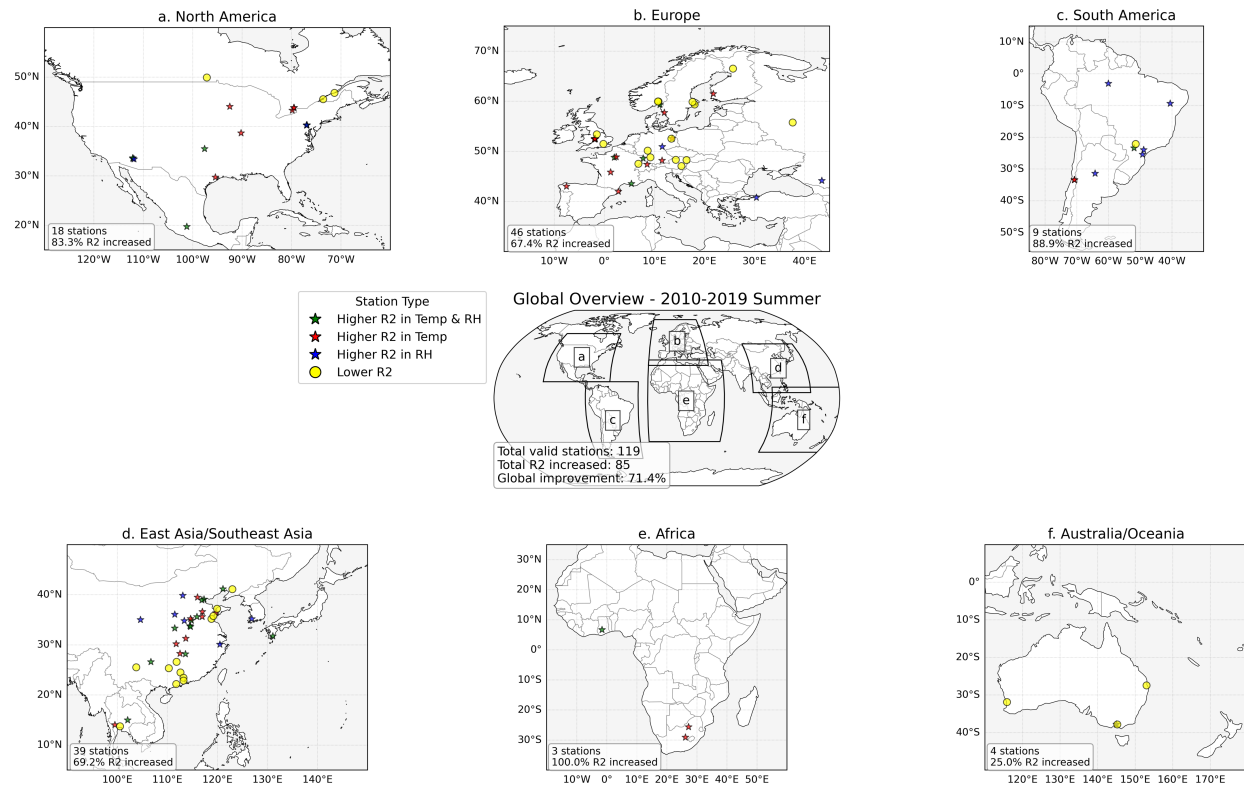


Figure S11. Same as S10, but with improvement determined by the increase in the coefficient of determination (R^2).

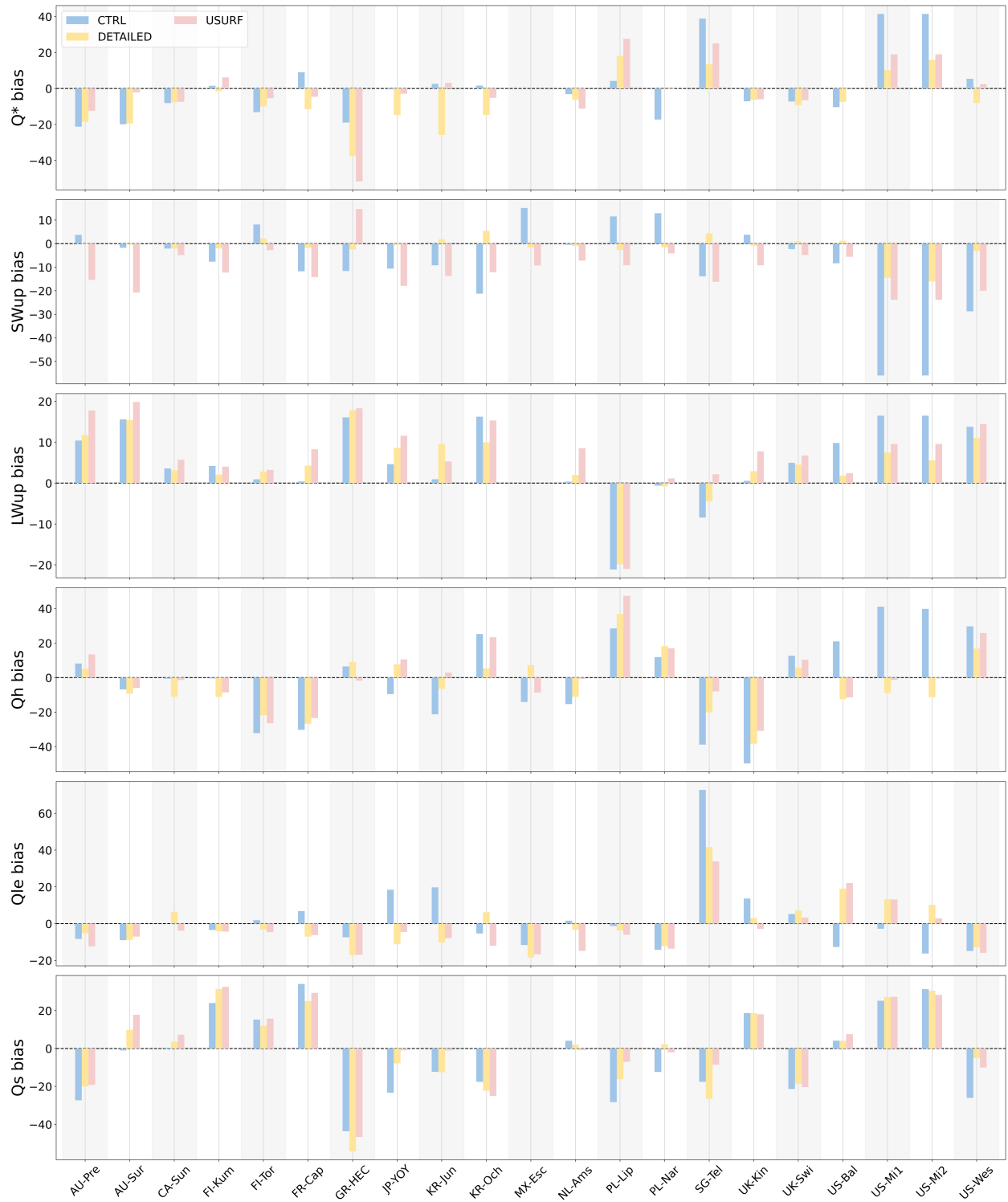


Figure S12. Comparison of long-term averaged bias of surface energy fluxes at 21 Urban-PLUMBER sites across CTRL, DETAILED and USURF simulations. Fluxes include upwelling shortwave radiation (SW_{up}), upwelling longwave radiation (LW_{up}), net radiation (Q^*), sensible heat flux (Q_h), latent heat flux (Q_{le}), and residual term (Q_s) representing the sum of storage heat flux (Q_{stor}) and anthropogenic heat flux (Q_{anth}). The unit of bias is W/m^2 .

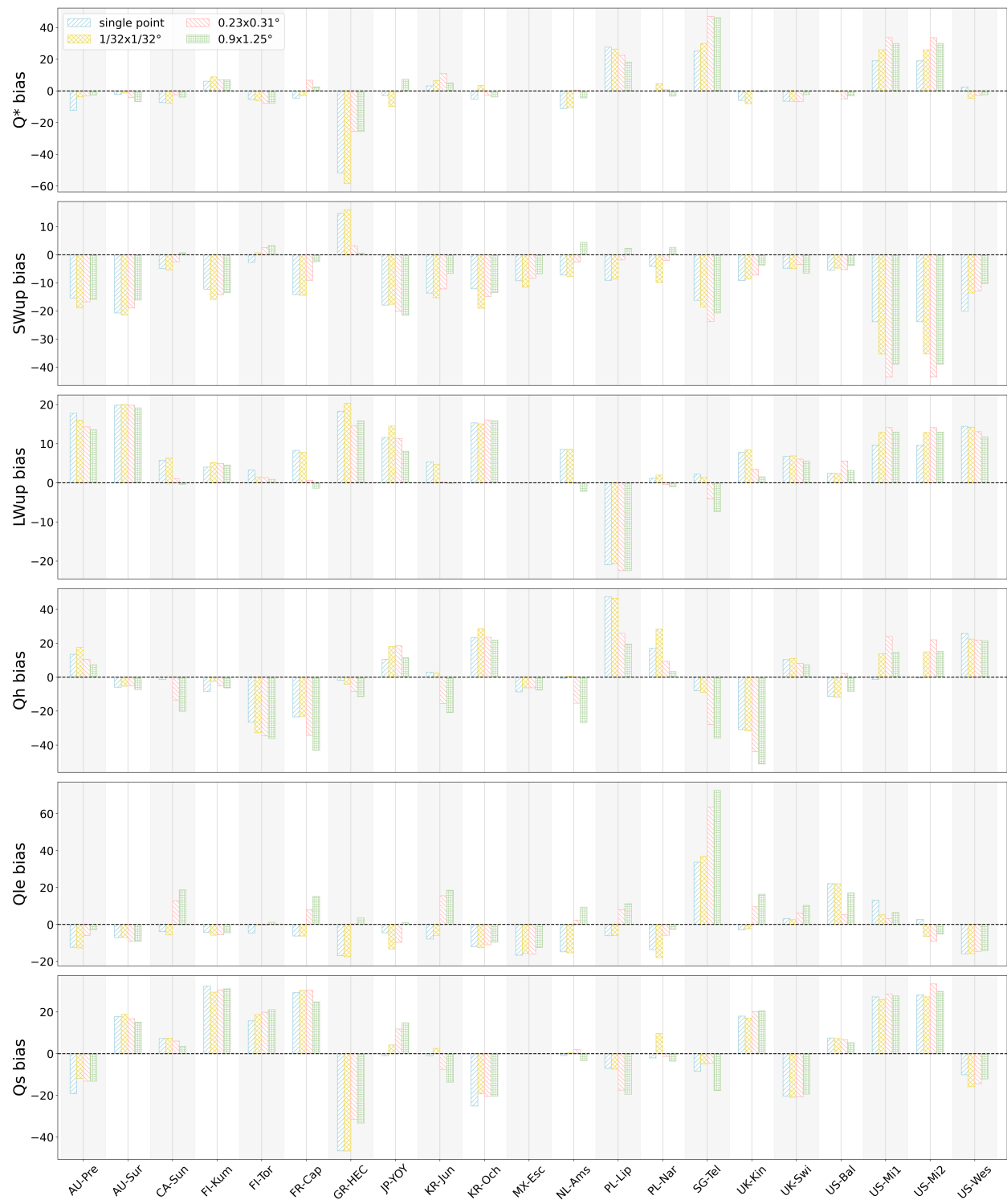


Figure S13. Same as Figure S12, but for comparison across different spatial resolutions of simulations using U-Surf parameters.

Table S1. Analysis period of 21 Urban-PLUMBER sites.

Site_name	Location	Longitude	Latitude	Time_analysis_start	Time_analysis_end
AU-Preston	Preston, Melbourne, Australia	145.0145	-37.7306	08/12/03 03:30	11/28/04 13:00
AU-SurreyHills	Surrey Hills, Melbourne, Australia	145.099	-37.8265	02/23/04 05:00	07/19/04 23:30
CA-Sunset	Sunset, Vancouver, Canada	-123.0784	49.2261	01/01/12 00:00	12/31/16 23:30
FI-Kumpula	Kumpula, Helsinki, Finland	24.9611	60.2028	12/31/10 22:30	12/31/13 22:00
FI-Torni	Hotel Torni, Helsinki, Finland	24.9387	60.1678	12/31/10 22:30	12/31/13 22:00
FR-Capitole	Capitole district, Toulouse, France	1.4454	43.6035	02/20/04 00:30	03/01/05 00:00
GR-HECKOR	Kornarou Sqr., Heraklion (Irakleio), Crete, Greece	25.1328	35.3361	06/30/19 22:00	06/30/20 22:00
JP-Yoyogi	Yoyogi, Tokyo, Japan	139.6845	35.6645	03/31/16 15:00	03/31/20 14:00
KR-Jungnang	Jungnang, Seoul, South Korea	127.0794	37.5907	01/24/17 16:00	04/29/19 06:30
KR-Ochang	Ochang, South Korea	127.4344	36.7197	06/07/15 15:00	07/26/17 00:00
MX-Escandon	Escandon, Mexico City, Mexico	-99.1761	19.4042	06/01/11 17:00	09/13/12 14:00
NL-Amsterdam	Amsterdam, The Netherlands	4.8929	52.3665	01/01/19 00:00	10/13/20 10:00
PL-Lipowa	Lipowa Street, Łódź, Poland	19.4453	51.7625	01/01/08 00:00	12/31/12 23:00
PL-Narutowicza	Narutowicza Street, Łódź, Poland	19.4811	51.7733	01/01/08 00:00	12/31/12 23:00
SG-TelokKurau06	Telok Kurau, Singapore	103.9112	1.3143	04/30/06 16:30	03/31/07 16:00
UK-KingsCollege	Kings College, London, United Kingdom	-0.1167	51.5118	04/04/12 00:00	01/01/14 00:00
UK-Swindon	Swindon, United Kingdom	-1.7981	51.5846	05/11/11 19:00	04/25/13 11:00
US-Baltimore	Cub Hill, Baltimore, United States	-76.5215	39.4128	01/01/02 05:00	01/01/07 04:00
US-Minneapolis1	KUOM Tall Tower, Minneapolis, Minnesota, United States	-93.18836	44.9984	06/01/06 18:00	05/29/09 13:00
US-Minneapolis2	KUOM Tall Tower, Minneapolis, Minnesota, United States	-93.18836	44.9984	06/01/06 18:00	05/29/09 13:00
US-WestPhoenix	West Phoenix, Arizona, United States	-112.1426	33.4839	12/16/11 18:30	01/01/13 06:30