Supplementary Materials

Overview of materials in this document:

- Description of the data used in this research (**Figure S1**)
- Contributions of land cover predictor variables to LST anomaly using our full mode (Eq 1, main text; **Figure S2**)
- Additional results from global GAMMs to show the apparent, high cooling ability of tree canopy on air and surface temperature based on the common practice of omitting likely covariates (i.e., univariate modeling; **Figure S3**)
- How distance from water modifies TCE from air temperature and LST at each time of day based on a 6000 m proximity threshold (**Figure S4)**
- TCE for LST in each city at 10% and 50% cover to understand whether the slope of the cooling curve is steeper where there is low or high canopy cover (**Figure S5**)
- Further exploration of the contribution of inadequate, moderate resolution tree canopy data to implausible LST TCE relationships (**Figure S6**, **Table S1**)
- Further detail on how modeling choices such as variable inclusion or inclusion/exclusion of local models using GWR as an example can lead to implausible LST TCE relationships (**Figure S7**, **Figure S8**).
- Supporting methods pertaining to the accuracy and provenance of our high-resolution urban tree canopy data (**Table S2**) and how a 6000 m water distance threshold was selected (**Figure S9**).

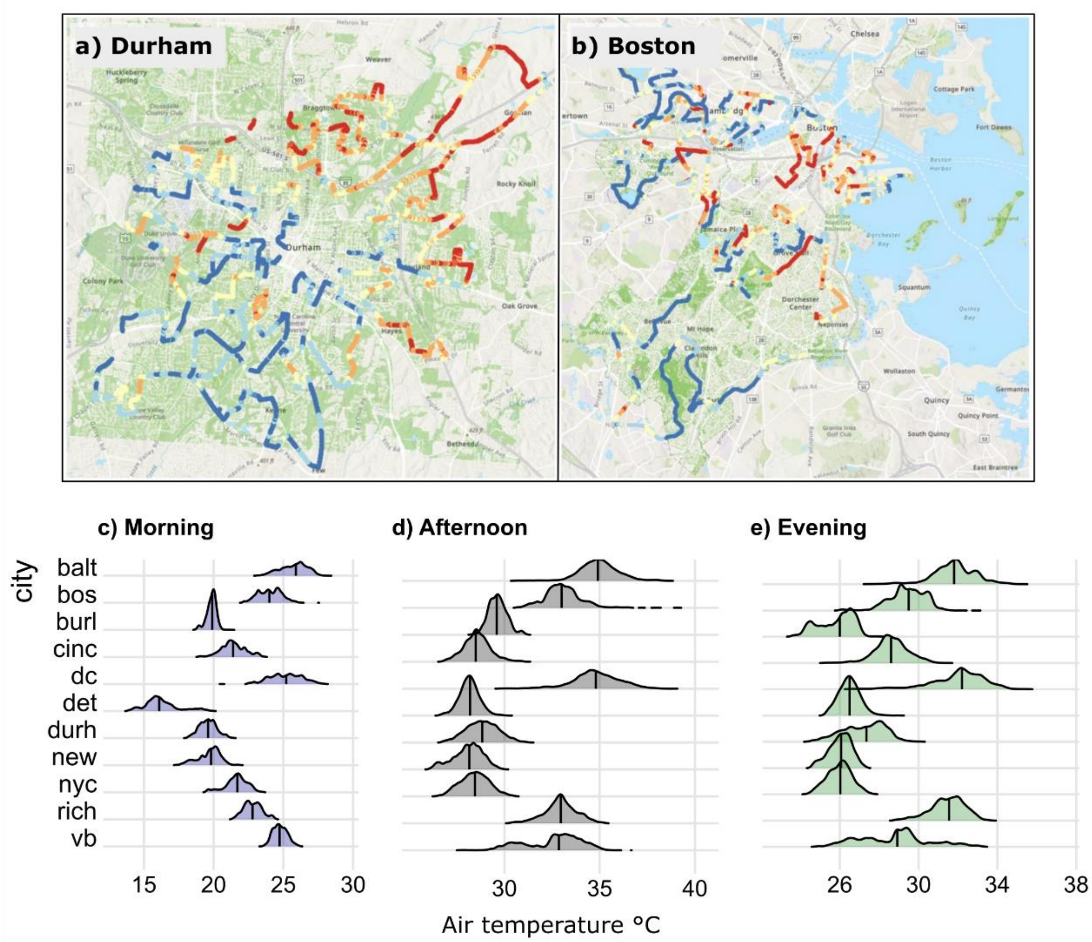


Figure S1: Data description: a) - b): Example mobile temperature point datasets from Durham (evening, n=35,419 samples) and Boston (afternoon, n=17,024 samples) where blue dots are in cooler and red dots in hotter areas. Green areas have greater canopy coverage. c) - e): Mobile air temperature distributions for our 11 cities in the morning: 6-7h, “afternoon”: 15-16h, “evening”:19-20h.

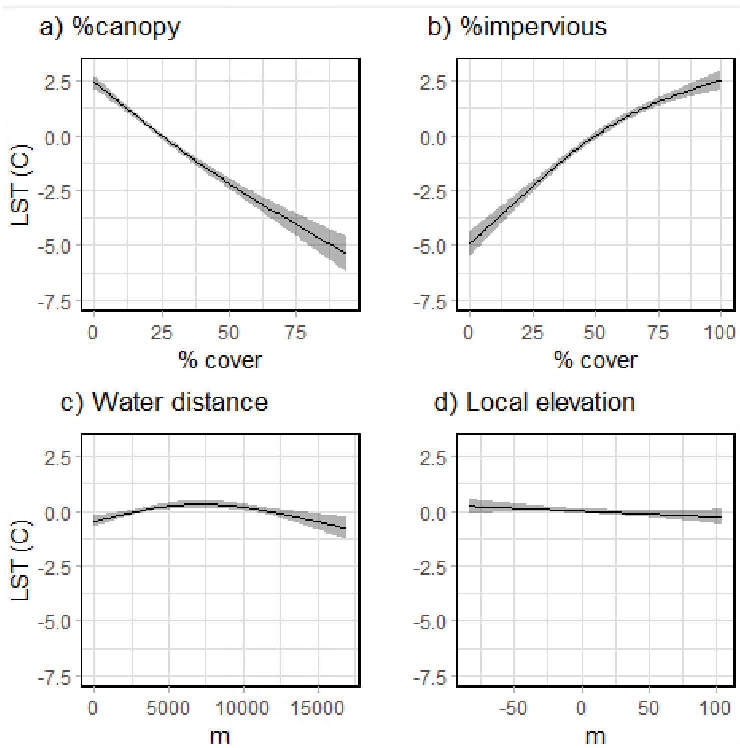


Figure S2: The multi-city, full-model GAMM but with LST as the response. See main text Figure 2 for the air temperature anomaly version of this plot.

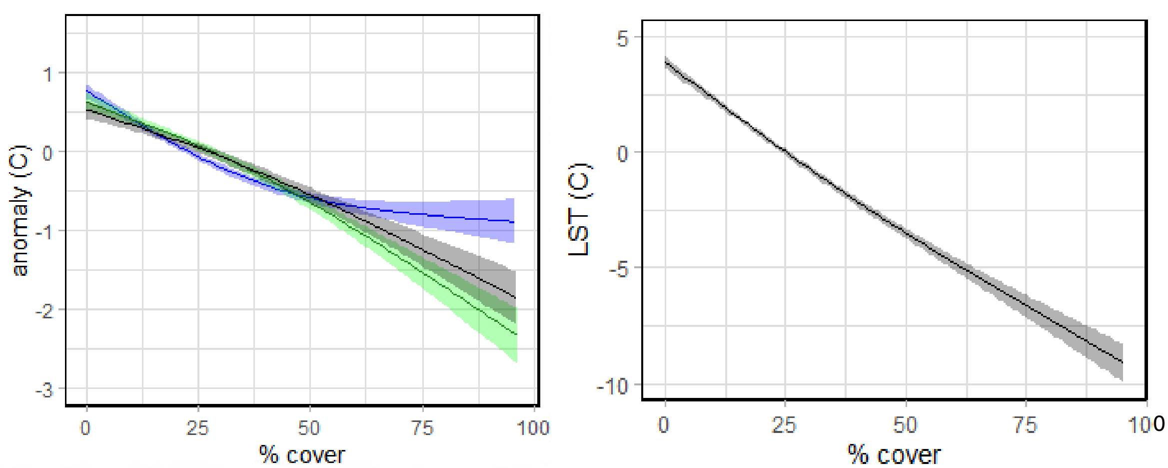


Figure S3: Univariate GAMM outputs (%canopy only) for: **a)** air temperature anomaly by time of day and **b)** LST at Landsat satellite overpass time. Mean model R^2 (across 50 model runs) for air R^2 = 0.54, 0.70, 0.69 for morning (am), afternoon (af), and evening (pm) respectively. Mean LST R^2= 0.64) for 50 runs at 0.02 sampling (not including spatial coordinates in model for comparability)

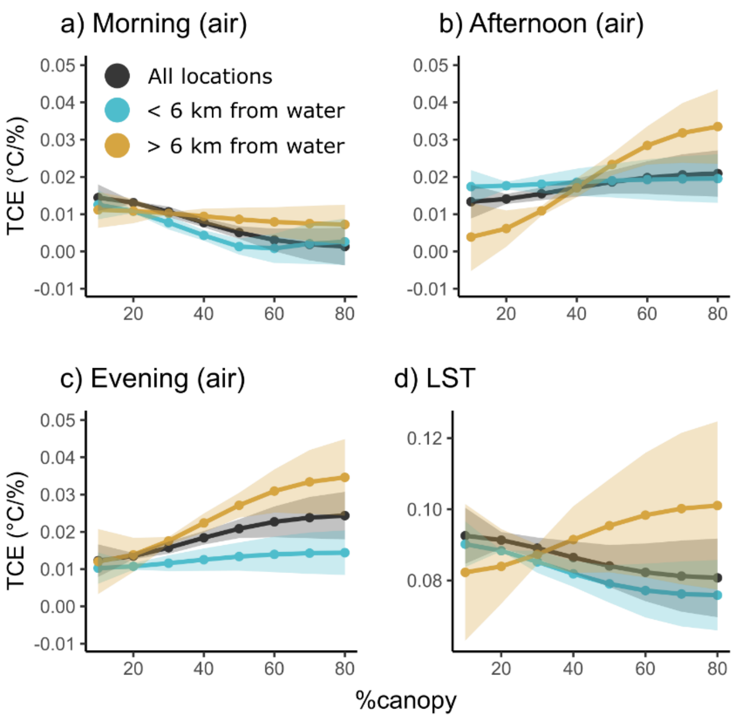


Figure S4: How distance from water modifies TCE from air temperature and LST at each time of day based on a 6000 m proximity threshold

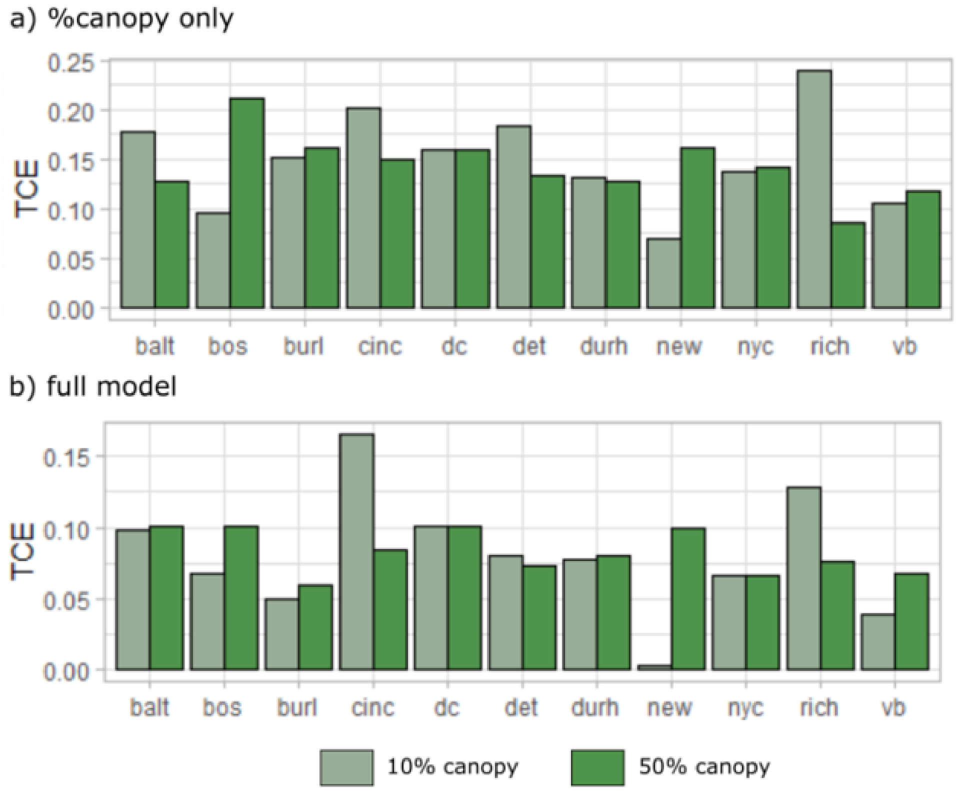
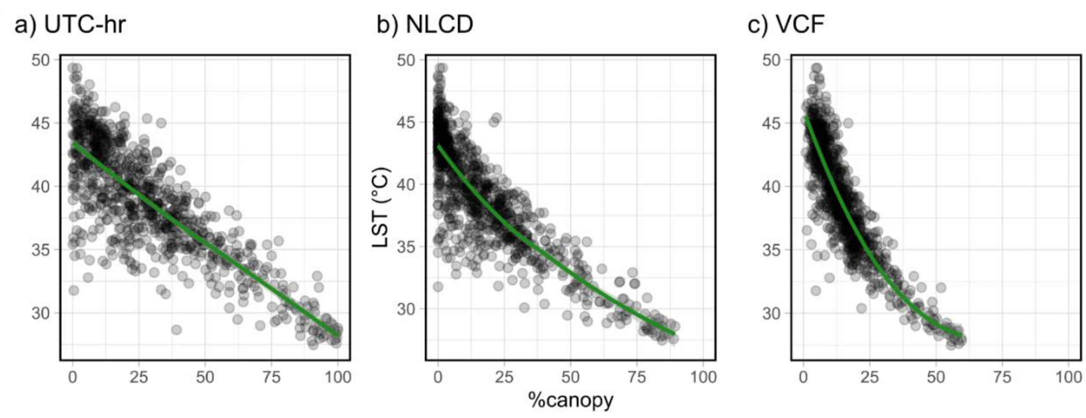
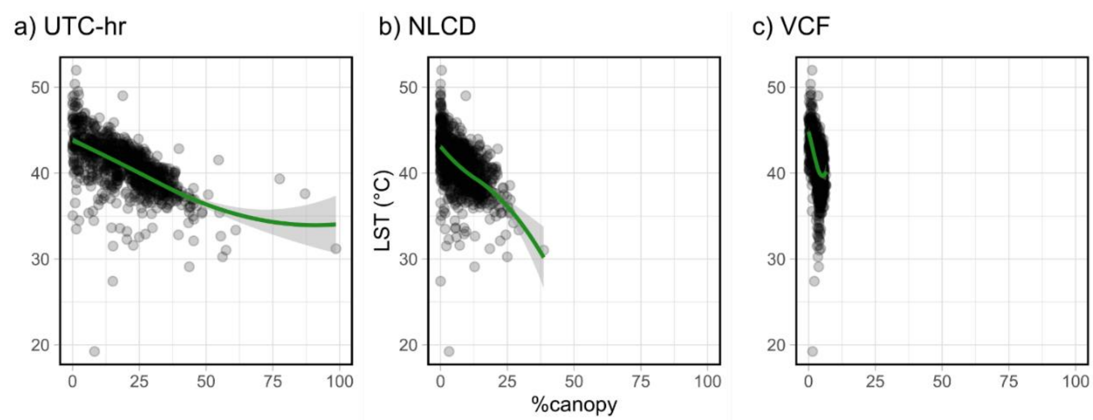


Figure S5: Tree cooling efficiency (TCE in °C/%cover) for LST at 10% and 50% cover based on **a)** univariate, canopy-only model and **b)** the full model.

Top row: Baltimore



Middle row: Boise



Bottom row: Detroit

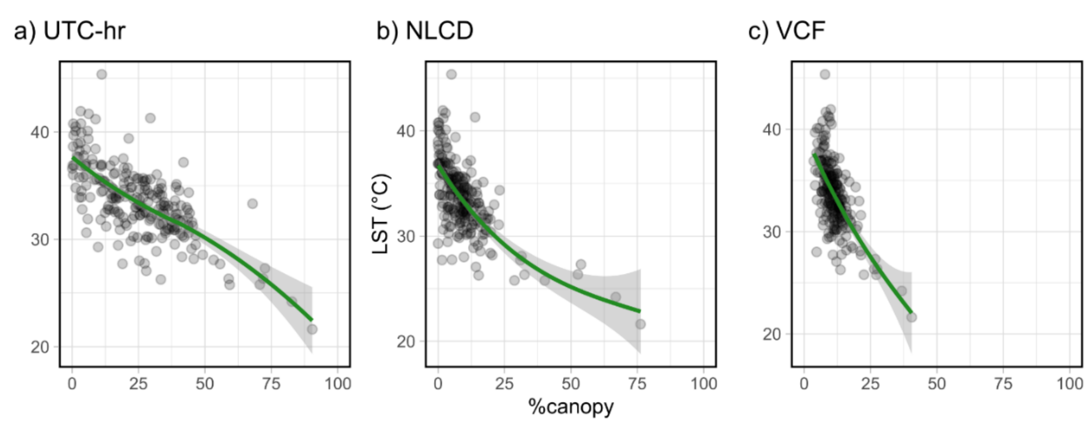


Figure S6: Canopy cover-LST relationships for each city and dataset sampled randomly within the bounding box of the UTC-hr dataset. Top row is Baltimore, middle is Boise, and bottom is Detroit. Datasets are: **a)** Urban Tree Canopy High Resolution (UTC-hr); **b)** National Land Cover Dataset (NLCD, 30m); **c)** Vegetation Continuous Fields (VCF, 30 m).

Table S1: Comparing NLCD 2021 and VCF against high-res %canopy in three cities. Slope, Intercept, and R2 are from simple linear regression where high resolution canopy cover is predicted using each moderate resolution dataset. *Underest*. is the mean percentage by which the metric underestimated canopy in densely built areas (areas with less than or equal to 20% canopy).

City	Metric	Slope	Intercept	R^2	Underest.
Baltimore	NLCD 21	1.1	7.8	0.88	69%
Boise	NLCD 21	1.6	7.2	0.61	72%
Detroit	NLCD 21	1.4	11.3	0.72	66%
Baltimore	VCF	1.9	-1.1	0.85	11%
Boise	VCF	5.7	-0.99	0.49	14%
Detroit	VCF	2.8	-7.5	0.73	73%

Details on GWR analysis and examination of issues related to scale

Beyond canopy data issues, we found that GWR estimates of LST TCE at low canopy cover were sensitive to the spatial unit of analysis as well as variable and local model inclusion decision making. To arrive at these conclusions, we conducted experiments by 1) adjusting GWR grid sizes from a minimum of 120 m (the size of a Landsat thermal band pixel) to a maximum of 700 m, 2) excluding or retaining insignificant local models, 3) comparing results from univariate, canopy-only models to multivariate models that also account for %impervious.

Overall estimates of LST TCE increased with grid size in accordance with Tobler's First Law of Geography (i.e., things that are closer are more similar than things that are farther away; **Figure S7**). With smaller grid sizes, shorter distance bands were selected via minimization of Akaike's Information Criterion (AIC), leading to smaller neighborhoods for local model evaluation. At the 120 m grid size (highlighted in main text), the search distance for neighbors was 1,025 m. This search distance then scaled roughly linearly with grid size: 2,012 m and 4,482 m search distances for 300 m and 700 m grids, respectively. Over larger distances, there is greater heterogeneity in both the urban fabric and environmental conditions enabling larger differences in TCE for a marginal change in canopy cover. At the 300 m grid size, the TCE magnitude estimated from multivariate GWR roughly aligned with the global estimate (**Figure S7b**). However, the global estimate suggests roughly linear cooling which was only achieved using GWR when all local models were retained (Figure S7b – dashed line), regardless of their statistical significance. We found that exclusion of "insignificant" models can lead to strong bias in the tree coverage of areas evaluated: The mean canopy cover in excluded areas of the 120 m model (**Figure S8a** hatched area), for example is 14% while the mean in included (significant) areas is 32%. Large portions of downtown Baltimore with rather homogeneously low canopy cover are excluded from the analysis, rendering conclusions regarding cooling at low %cover dubious (e.g., **Figure S7a,b**)

These edge effects that show up clearly in the %canopy-only models (**Figure S8a,b**) arise because something other than tree canopy is the primary driver of temperature change in that neighborhood. For example, the red arrow in **Figure S8a** points to an impervious-dominated, low canopy that is adjacent to a large park. The %canopy in the park is only slightly higher than just outside but LST is substantially lower, likely due to—rather than a change in canopy—the relative lack of impervious surface. To test this idea, we ran multivariate GWR models with *%canopy* and *%impervious* included as predictors. Those models showed us both that *%canopy* is not doing as much cooling as we may have thought based on univariate models (**Figure S7 -** black lines are lower than blue lines) and, perhaps more importantly, that the areas of very high TCE all but disappear (**Figure S8d-f)** upon introduction of a more locally relevant physical variable.

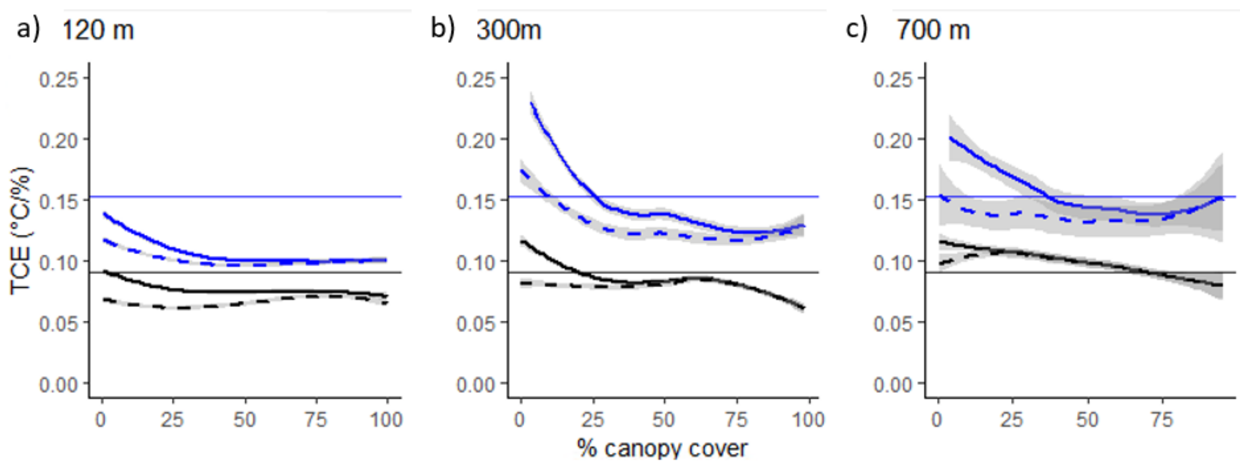


Figure S7: The impact of scale, variable inclusion, and insignificant model exclusion on LST TCE. Blue lines are univariate (canopy-only) models while black lines are multivariate (including impervious surface percent cover). The curves with solid lines are GWR estimates excluding insignificant local models while those with dashed lines retain all models. For reference, the thin lines are the univariate (blue) and multivariate (gray) LST TCE results from the global GAMM model. Neighborhood sizes tested were, a) 120 m (Landsat thermal pixel size); b) 300 m; c) 700 m.

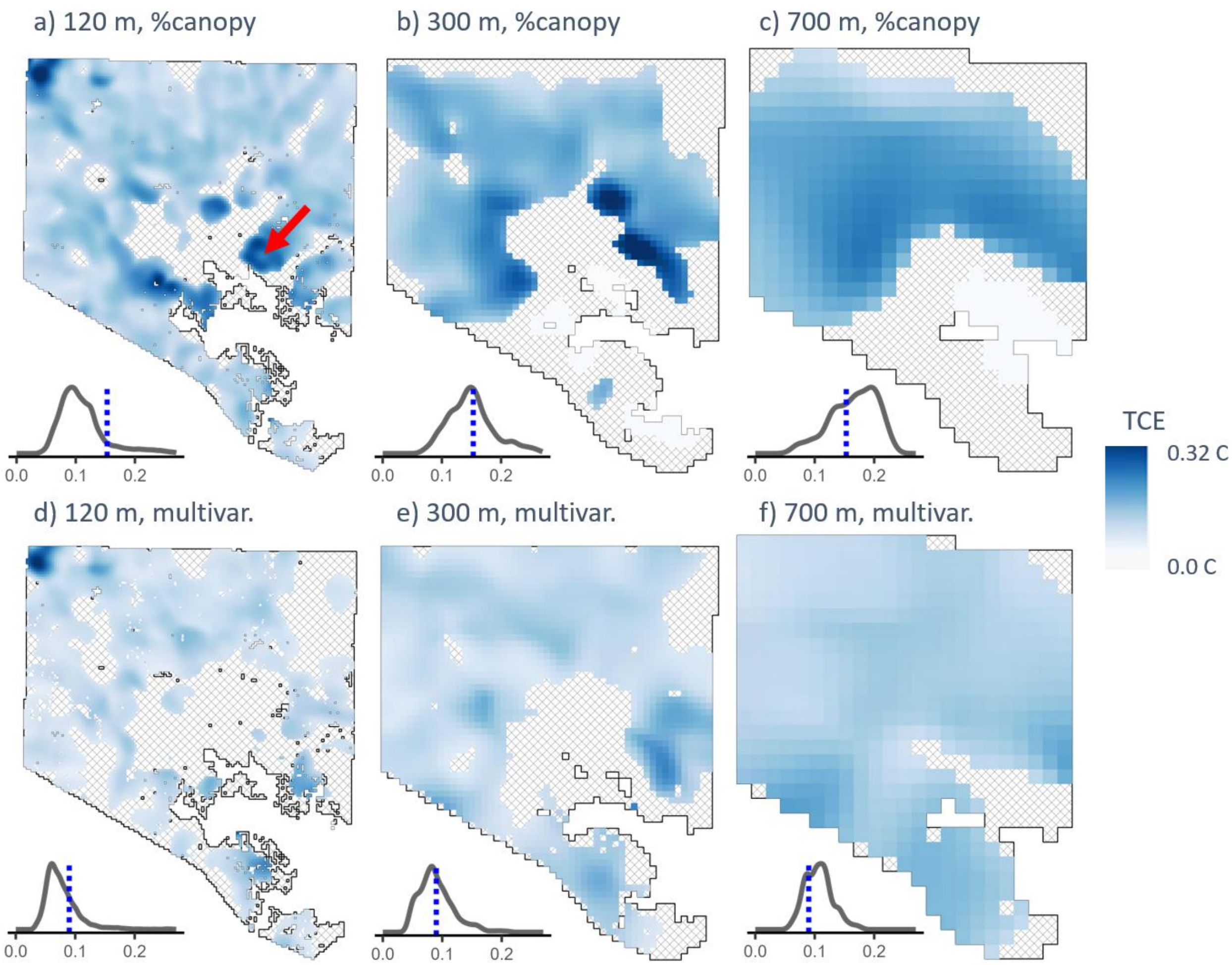


Figure S8: Tree cooling efficiency (TCE in °C/%cover) from GWR on LST at three grid sizes, only significant local models colored. Insignificant local models displayed with crosshatch. Density plots display the distribution of TCE values in each GWR map compared to the univariate (top row, 0.15°C%-1) or multivariate (bottom row, 0.09°C%-1) GAMM global model estimates. (a-c): TCE using models with %canopy as the only predictor variable; (d-f): TCE using multivariate GWR models including both %canopy and %impervious as predictors.

Table S2: Details on the production, QA/QC, and overall accuracy (OA) of the high-resolution urban tree canopy (UTC_{hr}) datasets. UVM-SAL is the University of Vermont's Spatial Analysis Lab. Accuracies are from comparison against 2014-2015 imagery in Google Earth except for New York City, whose accuracy assessment was conducted by UVM-SAL.

City	Map year	Provenance	Overall Acc.	Notes on QA/QC
Baltimore, MD	2018	UVM-SAL	92%	Manual QA/QC, 10,184 corrections made
Boise, ID	2021	UVM-SAL	90%	Not known
Boston, MA	2018	UVM-SAL		Full manual correction at 1:2000 scale
Burlington, VT	2016	UVM-SAL		Full manual correction at 1:3000 scale
Cincinnati, OH	2019	UVM-SAL		Full manual correction at 1:2500 scale
Detroit, MI	2014	UVM-SAL	87%	Full manual correction at 1:3000 scale
Durham, NC	2015	UVM-SAL		Full manual correction at 1:2500 scale
Newark, NJ	2010	UVM-SAL		Manual QA/QC, >5000 corrections made
New York City, NY	2016	UVM-SAL	98%	Accuracy assessment by SAL
Richmond, VA	2008	SAL-Vtech		Not known
Virginia Beach, VA	2008	UVM-SAL		Manual QA/QC, 18,194 corrections made
Washington, DC	2020	Alonzo 2021	91%	See OA
	Average OA		92%	

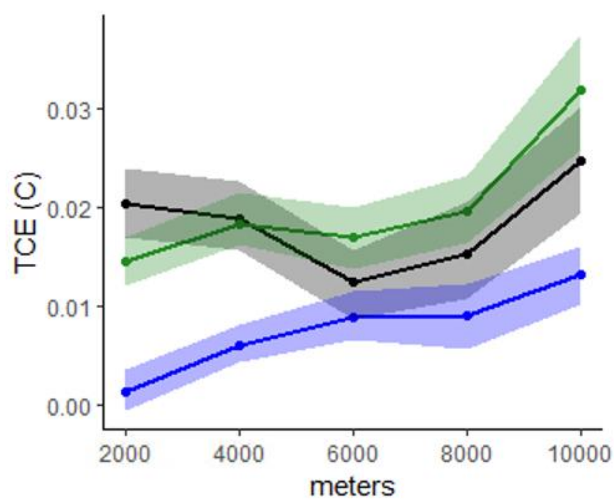


Figure S9: Air temperature TCE by distance from water at each time of day. 6000 m distance represents the best available inflection point. Cooling gets more efficient at all times of day after 6000 m.