

Bigger cities have larger heat islands, denser ones do not, and growing ones get hotter

In Chang Hwang

ichwang@si.re.kr

The Seoul Institute

Article

Keywords:

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10965873/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Cities are warmer than their surroundings, and how much warmer shapes how heat waves are felt, how adaptation is planned and how urban heat is priced. Damage assessments describe the heat island with one rule: it grows with city population. Using 8,637 raw weather-station records for 910 cities over five decades, we show that it answers two questions with one number. Across cities, size matters: night-time heat islands rise + 0.66°C per tenfold population, in every region sampled, and at a given size denser cities carry no larger heat island. Within a city, what matters is the population added on the land it already occupies, + 0.28°C per log unit, found in North America only. A joint test rejects one coefficient for both. Both effects act at night, when heat is most dangerous, and in North America in winter as well as summer, so not through summer cooling alone.

Introduction

Cities are warmer than the countryside around them. That difference, the urban heat island (UHI), has been measured since Howard¹ recorded it in London in 1833. It was catalogued through the field-survey era by Landsberg² and synthesized by Oke³, then sharpened by methodological critique and classification frameworks^{4,5,6}. It is no longer only a question for urban climatology. The heat island adds to the warming that city residents face under climate change, so how large it is, and what makes it larger, bears on three questions at once: which cities need adaptation most, whether compact development carries a heat penalty, and how much urban heat costs when climate damages are priced. Yet the basic *structure* of the phenomenon has not been characterized from raw, multi-decade, multi-region air-temperature observations: how UHI scales with city characteristics, whether those relationships act between cities or within a city over time, and how they differ between day and night.

The heat island is the shadow of urban growth. Global warming reaches every city alike; the heat island is the additional burden a city creates for itself as it grows, and it is the part of the burden that a city's own decisions can reach. The relationship need not run one way. Heat raises energy demand, illness and mortality⁷ and lowers labour productivity, and where it makes a city less liveable it can slow the growth that produced it. Whether the heat island is treated as a side effect of growth or as a constraint on it, a city's ability to manage it depends on knowing which feature of growth produces it: the size a city reaches, the density it builds at, or the people it adds to the land it already occupies. Those are different levers, and a heat island that follows one of them is not managed by pulling another. Knowing the structure is therefore what turns a general concern about urban heat into a specific instrument for urban sustainability.

Two literatures have measured it, and each answers only part of the question. Field studies have established local UHI repeatedly but never generalized it into a global scaling relationship⁸. The classical reference coefficients were obtained under the specific conditions that maximize UHI: calm, clear nights, in one or two regions at a time³. Satellite studies have since established global scaling^{9,10}, mapping the surface UHI across US biomes¹¹ and global and Chinese cities^{12,13}. But satellites measure the

temperature of the surface, not of the air people live in, and the two diverge substantially¹⁴. Satellite products are also typically built on inverted, atmospherically corrected land-surface-temperature products rather than raw observations¹⁵. A systematic account of the regularities that actually govern the air-temperature UHI, built from raw station records at global scale, is missing.

This gap has a direct practical consequence. Global economic assessments of urban heat already rely on a scaling relationship for exactly this purpose: a single population coefficient, introduced by Oke¹⁶ and calibrated for economic use on Karl et al.'s¹⁷ mid-century US study¹⁸. It recently pushed the urban share of the global social cost of carbon to 78–93% in a follow-on assessment using the same parameterization¹⁹. That one number, a single population coefficient, is asked to do two jobs. It must say how heat exposure differs across cities of different sizes, which is what a planner deciding where to invest in adaptation needs. And it must say how a city's heat exposure will change as the city grows, which is what a damage assessment needs whenever, as is common, it is applied along a climate-change scenario that projects the century. The coefficient stands in for a relationship the classical literature measured only locally, at its calm-clear maximum. Whether it behaves as a single, time-invariant, universal number has never been tested directly, because no study has analyzed data of the necessary scale.

Here we assemble such a record and use it to analyze the structure of the urban heat island. The record consists of observations from 8,637 stations of the Global Historical Climatology Network-Daily (GHCN-Daily)²⁰, quality-controlled but deliberately not homogenized so as to measure the urban–rural divergence itself (see Data), and it spans 506 to 910 cities and up to five decades. Our contribution is not the scaling law itself, long known from single cities and, recently, globally in surface data. It is the law's *structure*, and a modern coefficient tested across regions. Specifically, the structure we address lets us determine whether the law acts between cities or within them, whether it holds across regions, whether it holds by day as well as by night, and whether its slope depends on how urbanized a city already is. The question we address here is not how much cities have warmed, nor what part of that warming is urban. Our question is what governs the size of the heat island at a point in time, what governs its change, and whether those are the same thing. The answer resolves into a small number of simple, testable rules. Each applies directly to an assessment or decision in urban heat management: which cities to prioritize for heat adaptation under climate change, whether the compact city is efficient for heat management, which coefficient a damage assessment should carry, and which hours of the day a policy target should aim at.

Results

Level versus trend

A single population coefficient, as used in current damage functions, is implicitly asked to do two different jobs. It must describe how heat exposure differs across cities of different sizes in different

settings, and how a city's heat exposure changes as the city grows. Whether one number can do both depends on whether "how cities differ" and "how a city changes" are governed by the same variable.

Between cities, city size sets how large the heat island is. Nocturnal UHI rises $+0.66^{\circ}\text{C}$ for every tenfold increase in population, the population size law of Oke¹⁶, hereafter the size law (95% confidence interval, hereafter 95% CI, $+0.32$ to $+0.99$, $n = 506$ cities), holding across the full size range from under 300,000 to over 5 million residents (Figs. 1a and 2). Within a city, tracked over time, it is not city size that governs how its heat island intensifies but the population the city adds inside the footprint it already occupies. Because the source database, the Global Human Settlement Layer Urban Centre Database (GHS-UCDB), holds each city's extent fixed across epochs (Methods), the within-city term means exactly that, population added onto land the city already occupies, which on that fixed footprint is its population density. The two-way fixed-effects panel of Eq. (2) gives an increase of $+0.28^{\circ}\text{C}$ per log unit of that density, hereafter per log unit, over 2000–2020, hereafter the primary window (95% CI $+0.15$ to $+0.42$, $n = 885$ cities). Extended back through 1975–2020, the full record, the term is $+0.22^{\circ}\text{C}$ ($+0.14$ to $+0.29$, $n = 910$ cities), stable across every start year tested (Supplementary Table S1). Both relationships also steepen while a city is actively urbanizing rather than mature, where mature is proxied by World Bank high-income status, which in this sample means cities whose population and footprint have largely stopped growing, and urbanizing by the lower income groups. The proxy does not distinguish a city that has reached the top of its growth path from one that has stopped growing for other reasons while room remains, a limitation taken up in the Discussion. For the size law the evidence is tentative. Splitting the daily-mean cross-section by World Bank income group, the slope is $+0.17^{\circ}\text{C}$ per tenfold population in high-income cities (95% CI $+0.08$ to $+0.25$) and about $+0.62^{\circ}\text{C}$ in upper-middle-income cities ($+0.34$ to $+0.91$). The latter group is small and concentrated in a few regions, so the contrast is confounded with region and climate (Supplementary Note 1, S1.4). For the within-city trend the same pattern is clearer. On the full record the term is $+0.17^{\circ}\text{C}$ per log unit in the 750 high-income cities (95% CI $+0.09$ to $+0.24$) and $+0.90^{\circ}\text{C}$ in the 53 cities that are not high-income and were still urbanizing over the period ($+0.44$ to $+1.35$) (Table 1).

Table 1

The structure of the air-temperature heat island: central estimates. Between-city (level) coefficients are from the 2001–2020 cross-section with country-clustered standard errors; within-city (trend) coefficients are from the epoch panel with city and year effects and city-clustered standard errors, with city-bootstrap intervals where noted. Size-law slopes are in °C per tenfold population, density and within-city terms in °C per log unit, the income premium in °C; 95% confidence intervals in parentheses, number of cities in *italics*. The full robustness suite is Supplementary Table S1.

quantity	estimate (95% CI)
Between cities: the level	
size law, night	+ 0.66 (+ 0.32 to + 0.99) <i>506</i>
size law, daily mean	+ 0.23 (+ 0.08 to + 0.39) <i>685</i>
size law, day	–0.16 (– 0.34 to + 0.01) <i>506</i>
population density at a given size, night	–0.63 (– 1.08 to – 0.18) <i>506</i>
per standard deviation, night: size / density	+ 0.43 (+ 0.29 to + 0.57) / –0.31 (– 0.53 to – 0.09) <i>506</i>
income premium of the richest 1% of cities, night / daily mean	+ 0.47 (+ 0.09 to + 0.88) <i>554</i> / +0.57 (+ 0.31 to + 0.87) <i>782</i>
Within cities: the trend	
population added on the fixed footprint, 2000–2020 (bootstrap CI)	+ 0.28 (+ 0.15 to + 0.42) <i>885</i>
the same, 1975–2020 (bootstrap CI)	+ 0.22 (+ 0.14 to + 0.29) <i>910</i>
between-city term of the joint model, night	–0.30 (– 0.57 to – 0.04) <i>611</i>
test that one coefficient fits both levels, 2000–2020 / 1975–2020	rejected, $p = 1.4 \times 10^{-3}$ <i>885</i> / $p = 4.1 \times 10^{-3}$ <i>910</i>
population on the moving GHS-WUP-MTUC boundary, 2000–2020	+ 0.19 (+ 0.08 to + 0.30) <i>733</i>
with building volume added: population / volume	+ 0.26 (+ 0.10 to + 0.41) / +0.02 (– 0.02 to + 0.06) <i>884</i>
By region	
size law, night: North America / Europe	+ 1.00 (+ 0.59 to + 1.40) <i>308</i> / +0.43 (+ 0.03 to + 0.83) <i>181</i>
size law, daily mean: Asia	+ 0.29 (+ 0.13 to + 0.45) <i>108</i>
within-city term, 2000–2020: North America / Europe / outside North America	+ 0.32 (– 0.02 to + 0.65) <i>331</i> / –0.03 (– 0.31 to + 0.25) <i>342</i> / +0.05 (– 0.07 to + 0.18) <i>554</i>
United States with climate-zone-by-year effects	+ 0.41 (+ 0.04 to + 0.77) <i>279</i>
By development phase	

quantity	estimate (95% CI)
size law, daily mean: upper-middle-income / high-income cities	+ 0.62 (+ 0.34 to + 0.91) 114 / +0.17 (+ 0.08 to + 0.25) 545
within-city term, 1975–2020: not high-income / high-income	+ 0.90 (+ 0.44 to + 1.35) 53 / +0.17 (+ 0.09 to + 0.24) 750

Population density measured on the fixed boundary behaves oppositely depending on which question is asked. Between cities, at a given population, the denser city carries the *smaller* heat island. The association is weak on its own and strengthens once city size is held, to -0.63°C per log unit of density (95% CI -1.08 to -0.18). Within a city over time, the same variable runs the other way: population density added inside the fixed footprint is followed by a larger heat island, at the $+0.28^{\circ}\text{C}$ per log unit reported above (95% CI $+0.15$ to $+0.42$). The contrast is not an artifact of estimating the two relationships separately. A joint (Mundlak) model²¹, Eq. (3), tests this directly. In this test two terms enter for every city at once: the city’s average density over the whole period, which captures how it compares with other cities, and its deviation from that average in each year, which captures how it changes over time. Estimated together, the two terms let us test directly whether their coefficients differ, and the estimates reject equality on both windows (Fig. 1; Table 1). The two coefficients therefore estimate different properties, belong to separate levels of variation, and cannot be represented by one number. This dissolves a long-standing paradox, in which urban-form studies have reported density as both UHI-lowering and UHI-raising depending on which comparison a given study happened to make^{22,23}. The practical consequence is this. A model calibrated on a cross-section of mature cities, whose population and footprint have largely stopped growing, inherits the negative between-city sign of density. Applied to a city that is still adding population to its footprint, that model predicts that densification cools the city, whereas the within-city evidence shows that it warms it. We use the word paradox deliberately rather than ambiguity or puzzle: density has precisely estimated, opposite signs depending on the comparison made, between cities or within them, which is the structure of a paradox rather than of an unresolved result.

That the within-city term does not depend on the boundary convention is shown under Robustness, and what the convention does and does not let the term mean is set out in the Discussion. The practical warning that follows is the paper’s central one: a coefficient estimated across cities and a coefficient estimated within them are not interchangeable inputs, because they are different responses of the same variable.

Both halves are nocturnal

The two halves of the separation, the between-city level and the within-city trend, differ in what drives them, but they share when they appear. Both are concentrated at night. In the level sample the North American heat island at night is several times its daytime value: $+0.77^{\circ}\text{C}$ (95% CI $+0.60$ to $+0.94$) against $+0.18^{\circ}\text{C}$ by day (-0.01 to $+0.36$). The size law itself changes with element (Fig. 2b). At night the

heat island grows with city size, at the $+ 0.66^{\circ}\text{C}$ per tenfold population reported above; in the daily mean the slope is about a third of that; and by day it is flat to slightly negative, at $- 0.16^{\circ}\text{C}$ per tenfold population ($- 0.34$ to $+ 0.01$). The daytime slope is a gradient across the whole size range rather than a threshold, where larger means larger by a given factor and not above any particular size. For every tenfold increase in population the daytime heat island is slightly smaller, though the interval reaches zero, so a larger city runs no warmer, and if anything slightly cooler, in daytime air temperature. The daytime sign is not tested directly here, given the paper's scope, but it is consistent with three daytime-suppressing mechanisms. Taller buildings cast more shade on streets and sensors, cutting the direct sunlight that would otherwise heat the air, and the same urban canyon geometry slows daytime heat storage³. Urban aerosol pollution scatters and absorbs incoming sunlight before it reaches the surface, a documented daytime dimming effect over cities²⁴. And a thermal-inertia effect stores daytime heat in building mass rather than releasing it into the air immediately, then lets it out again at night³. The daily-mean slope sits between the night and day values because the two pull in opposite directions, not because the mean channel is constructed by averaging them; each is computed independently from the station record. Urbanization is also proceeding globally, and most of the resulting population growth is concentrated in already-large cities²⁵. Because this study ties heat-island size directly to city size, the exposure an average resident faces is larger than the average over cities suggests: weighting cities by population more than doubles the nocturnal figure, from a city median of about half a degree to $+ 1.14^{\circ}\text{C}$ (bootstrap 95% CI $+ 0.86$ to $+ 1.38$).

The within-city trend shows the same asymmetry, and this is what makes the level/trend separation legible in the first place. Splitting the within-city panel by element, the between-city term of the joint model is negative at night ($- 0.30^{\circ}\text{C}$ per log unit, 95% CI $- 0.57$ to $- 0.04$) while the within-city term is positive, whereas by day both terms are positive. So the sign reversal that defines the separation is a nocturnal phenomenon specifically. By day, both terms move in the same direction, and no such paradox arises.

Regional generality

Both relationships were estimated separately by continent, across North America, Europe and Asia (Fig. 3). The size law's sign is positive in every region observed. It is steepest in North America ($+ 1.00^{\circ}\text{C}$ per tenfold population, 95% CI $+ 0.59$ to $+ 1.40$, $n = 308$) and about 2.3× shallower in Europe ($+ 0.43^{\circ}\text{C}$, $+ 0.03$ to $+ 0.83$, $n = 181$). In Asia only the daily-mean series is dense enough to estimate it, and there too it is positive ($+ 0.29^{\circ}\text{C}$ per tenfold population, $+ 0.13$ to $+ 0.45$, $n = 108$). It survives excluding any single region, the nine-step covariate ladder and the 384-specification extreme-bounds search described in Methods²⁶, remaining significant in 100% of the 256 specifications applicable to it. Two further checks ask whether the *slope* depends on development phase: a cross-sectional comparison of cities stratified by income group, and a longitudinal check on a fixed panel of the 364 cities with data at every epoch. Both point the same way: the slope is steeper, in point estimate, in still-urbanizing cities than in mature ones, and on the fixed panel it falls from $+ 0.27^{\circ}\text{C}$ per tenfold population in 1975 (95% CI $+ 0.04$ to $+ 0.50$)

to $+0.12^{\circ}\text{C}$ in 2020 (-0.01 to $+0.25$). The phase dependence is nonetheless suggestive rather than established, because the fixed panel consists almost entirely of mature cities and the cross-sectional contrast is confounded with region and climate (Supplementary Note 1, S1.4).

The within-city trend does not generalize in the same way. It survives a country-level jackknife, dropping any single country other than the United States without changing the result. But it collapses on excluding North America as a whole. Outside North America, adding population to a city's footprint no longer raises its heat island measurably: the within-city coefficient on the remaining 554 cities is $+0.05^{\circ}\text{C}$ per log unit (95% CI -0.07 to $+0.18$), against $+0.28^{\circ}\text{C}$ when North America is included. This does not look like a loss of statistical power. Dropping a third of the sample should widen the confidence interval around the estimate; instead the interval barely widens while the point estimate falls to near zero, which indicates a genuinely smaller effect estimated with comparable precision, not a result we simply failed to detect. Nor could we attribute it to which cities grew: restricting each region to the cities whose population on the fixed footprint grew at least as fast as the North American median leaves the gap intact (Supplementary Table S2). In our station panel the trend is a North American regularity rather than a global one. Nor could we attribute the contrast to how much the regressor moves. European cities did grow more slowly on their footprints over the primary window, so a European coefficient is estimated over a narrower range of change. Extending the window back to 1975 doubles that range, and the European coefficient stays at zero while the North American one stays positive (Supplementary Table S2). A longer window with twice the identifying variation returns the same regional difference, so the difference is in the response rather than in the range over which it is estimated. North America also supplies well over half of the study's stations and the longest unbroken records, which is a separate reason the trend is most clearly *measurable* there. Why the regions differ is narrowed here rather than settled. We could not attribute the difference to statistical power, growth pace, urban form, building volume, station siting, boundary definition, regional background trend or a rural reference crowded by neighboring cities: each was tested, and none accounts for it on our panel (Supplementary Note 1, S1.1). In particular the North American coefficient on population density survives country-by-year and climate-zone-by-year effects, which absorb any regional trend in the rural background. For the United States alone, once each climate zone carries its own trend, the heat island still rises by $+0.41^{\circ}\text{C}$ per log unit of density added on the fixed footprint (95% CI $+0.04$ to $+0.77$). Within North America the term is largest in the temperate, hot-summer zone, while Europe's temperate cities show none (Supplementary Table S2). East Asia shows none either, but on cities that barely grew over the period, so it cannot discriminate among explanations in the way the contrast between North America and Europe can. What the exclusions leave is a term that follows people rather than buildings, appears at night, and is concentrated where summers are hot and growth was fast. Three channels could produce such a term, anthropogenic heat per added resident^{27,28}, land-cover change inside the footprint^{3,23} and the materials of the building stock³; the Discussion sets them out and reports what our panel can test of each. On the regional question, the first channel is the one that differs between North America and Europe by far more than urban form does. More than 90% of US households run air conditioning against a small minority in Europe²⁹, and modeled waste heat from air conditioning alone raises night-time air temperature by about

1°C across a US metropolis³⁰. For Paris the same effect is still a scenario rather than a present condition³¹. Separating the channels requires data our station panel does not carry, such as city-epoch series of energy use, cooling load, vegetation change and building materials, and a night-time-lights test attempted as an activity proxy is inconclusive because the sensors saturate over the cores of developed-world cities (Supplementary Note 1, S1.1). The season is a partial discriminator that our panel does carry, since cooling waste heat and the evaporative cooling of vegetation are summer quantities whereas heating, traffic, industry and stored heat act in every season. On the full 1975–2020 record the North American night-time term is present in winter as well as summer, at similar size, and Europe shows none in either season (Table 2; Supplementary Note 1, S1.7), which speaks against cooling waste heat as the sole channel and against vegetation loss as the main one.

Table 2

The within-city population term by season, element and region, 1975–2020. Coefficient on population added on the fixed footprint, °C per log unit, city and year effects, 95% confidence intervals from city-clustered standard errors in parentheses, number of cities in italics. Summer is June to August and winter December to February of the same climatological year. The composition-free rural reference is refitted for each series, so the six panels differ in which cities and epochs qualify (Supplementary Note 1, S1.7).

element, season	pooled	North America	Europe	outside North America
night (TMIN), summer	+ 0.31 (+ 0.15 to + 0.47) 859	+ 0.44 (+ 0.20 to + 0.67) 330	+ 0.10 (– 0.23 to + 0.43) 321	+ 0.04 (– 0.17 to + 0.25) 441
night (TMIN), winter	+ 0.33 (+ 0.18 to + 0.49) 725	+ 0.33 (+ 0.11 to + 0.56) 330	– 0.09 (– 0.44 to + 0.27) 251	– 0.08 (– 0.29 to + 0.13) 359
daily mean (TAVG), summer	+ 0.15 (+ 0.07 to + 0.23) 921	+ 0.20 (+ 0.07 to + 0.34) 331	+ 0.02 (– 0.15 to + 0.19) 342	– 0.01 (– 0.11 to + 0.10) 470
daily mean (TAVG), winter	+ 0.23 (+ 0.14 to + 0.31) 910	+ 0.15 (+ 0.01 to + 0.30) 331	– 0.07 (– 0.26 to + 0.11) 342	– 0.02 (– 0.12 to + 0.08) 470
day (TMAX), summer	+ 0.09 (– 0.01 to + 0.19) 852	– 0.02 (– 0.17 to + 0.14) 330	– 0.16 (– 0.38 to + 0.06) 332	– 0.01 (– 0.14 to + 0.12) 457
day (TMAX), winter	+ 0.11 (– 0.01 to + 0.23) 811	– 0.04 (– 0.23 to + 0.14) 330	– 0.02 (– 0.22 to + 0.18) 305	+ 0.01 (– 0.12 to + 0.14) 428
night, summer minus winter	– 0.02 (– 0.14 to + 0.10) 723	+ 0.10 (– 0.07 to + 0.28) 330	+ 0.10 (– 0.16 to + 0.36) 250	
night, North America by zone: arid / temperate / continental, summer		+ 0.69 (+ 0.33 to + 1.05) 69 / +0.20 (– 0.02 to + 0.43) 112 / +0.23 (– 0.35 to + 0.81) 137		
night, North America by zone: arid / temperate / continental, winter		+ 0.68 (+ 0.03 to + 1.33) 69 / +0.22 (+ 0.05 to + 0.39) 112 / +0.22 (– 0.29 to + 0.73) 137		

The income premium

GDP per capita adds a non-linear driver on top of size. It is flat across most of the income range but rises sharply at the top, worth + 0.57°C of daily-mean UHI for the richest 1% of cities relative to the mean (95% CI + 0.31 to + 0.87) and + 0.47°C at night (+ 0.09 to + 0.88), and it is confined, like size, to the between-

city level. The within-city effect is null, so our panel gives no support to income as a direct cause of UHI change. A plausible but untested reading is that income instead marks which cities grew, with the within-city population term doing the work already established above. Full model and time-trend detail are in Supplementary Note 1 (S1.3).

Robustness

Both central relationships pass a full adversarial diagnostic suite. Estimates are identical under pooled and cross-sectional specification, neither sign nor significance changes across an extreme-bounds search of 384 covariate combinations, a permutation test that reshuffles the predictor 500 times returns no signal, and both reproduce closely on an independent, raw archive (GHCN-M v4³² unadjusted). The within-city trend is also unchanged under Driscoll–Kraay standard errors³³, shows no residual cross-sectional dependence on the Pesaran test³⁴, and is recovered on two satellite-based products, the YCEO surface heat island^{35,36} and the canopy-layer heat-island dataset of Yang et al.³⁷ (Supplementary Table S1). None of the remaining candidate covariates, built-up fraction, vegetation contrast, wind, coastal distance, or latitude, survives the same bar. Nor does building volume, added to the within-city panel from the GHS built-up volume layer (GHS-BUILT-V) on the same fixed footprint: the population term is unchanged and volume carries no coefficient of its own (Table 1; Supplementary Note 1, S1.6). The within-city term also survives the boundary convention. Re-measured on the multi-temporal GHS-WUP-MTUC delineation, which moves with the city³⁸, it is + 0.19°C per log unit (95% CI + 0.08 to + 0.30) on the 733 matched cities of the primary window, indistinguishable from the + 0.28°C on the fixed footprint. Where density is separately identifiable on that record, its own coefficient is zero (Supplementary Note 1, S1.5). The within-city response is also recovered by an independent design on the same station archive, a single long difference with population growth measured over a matching window, at + 0.22°C per log unit (95% CI + 0.12 to + 0.32, 493 cities) against + 0.28°C here³⁹. Nor does it depend on the rural-contamination screen. On a reference rebuilt from the station record with the exclusion radius swept from 0 to 30 km, the within-city term is significant at every radius, and over 2000–2020 it moves only from + 0.25°C per log unit with no screen to + 0.29°C at 25 km (from + 0.19 to + 0.22°C on the full record). Leaving the contamination in attenuates the term slightly toward zero, as expected, but does not remove it (Supplementary Table S1). The complete set of checks is in Supplementary Table S1 and Supplementary Notes 1–2.

Discussion

The air-temperature UHI resolves into a compact, testable structure. City size sets the level between cities. Population added inside the existing footprint sets the trend within them. Both halves of this separation, the level between cities and the trend within them, are concentrated at night, and a non-linear income premium sits on top of the size effect without displacing it. This resolves the paradox over whether density raises or lowers UHI: it depends on which of two questions is being asked. The between-city level generalizes across every region sampled here. The within-city trend is presently a North

American regularity. That contrast is itself a structural finding: it tells us which parts of this system are general properties of cities, and which are regionally contingent, at least on the record analyzed here.

Two readings of compactness

This structure also bears on urban planning and adaptation, not only on damage-model calibration. Compactness is not straightforwardly a heat penalty, and the longitudinal reading is not about compactness at all. Read cross-sectionally, comparing many cities at one point in time, the results support compact development. At a given population, that is, with the same number of people held in a smaller extent, denser cities carry a smaller nocturnal heat island (-0.31°C per standard deviation of $\ln$ density, 95% CI -0.53 to -0.09), consistent with a more compact built-up area storing and re-radiating less heat^{22,23}. Read longitudinally, tracking the same city over time, what raises the heat island is the population a city adds, at $+0.28^{\circ}\text{C}$ per log unit (95% CI $+0.15$ to $+0.42$) inside the footprint it already occupies. Density itself did not, on the moving-boundary record where the two can be separated (Robustness).

These two readings are not contradictory; they follow from different comparisons. Between cities, holding population fixed, a higher density means the same population is packed into a smaller urban extent. A city of smaller extent sits closer, on average, to the cooling rural boundary, so cooler rural air mixes more readily into the urban core, and it generates less contiguous built-up area for heat to accumulate over in the first place. A more sprawling city at the same population does the opposite, keeping its core farther from that cooling boundary and creating more contiguous area for heat to build up. Density here is best read as a proxy for how far a city's population is spread out^{22,23}.

The between-city reading above comes with a ready physical account, since a more compact extent sits closer to the cooling countryside. The within-city reading does not, and it needs one: the coefficient shows only that the night-time heat island grows as people are added to a fixed footprint, not how. The open question is therefore one of mechanism: through what channel does population added on a fixed footprint raise the night-time heat island? This is a different question from why that channel operates in North America and not elsewhere, which the regional analysis above addresses. Three hypotheses about the channel have a basis in the literature. The first is land cover: where more people have been added to land a city already occupies, a larger share of the surface is impervious, and impervious surface replaces vegetation, so rain runs off instead of evaporating and cooling the surface^{3,23}. The second is building mass: housing the added population adds building mass, which raises heat capacity, so more of the heat absorbed by day is stored and released at night³. The third is anthropogenic heat: the traffic, cooling, heating and industry of the added residents release heat directly, much of it at night^{3,27,28}. Our panel tests each of them in part, the first with land cover, the second with building volume and the third with season. Two land-cover series are available on the fixed footprint, the built-up share of the polygon and the vegetation of its core measured by satellite. Neither carries the night-time term. On the night channel, adding the built-up share takes about a tenth off the population coefficient in the pooled panel and

almost nothing in North America, where the term lives. Neither the built-up share nor the vegetation of the core or its contrast with the surroundings carries a night-time coefficient of its own (Supplementary Note 1, S1.6). Land cover is therefore at most a minor part of the night-time channel. Adding the building volume of each city's fixed footprint (GHS-BUILT-V, 2000–2020) to the panel takes none of the population term, which stays at $+0.26^{\circ}\text{C}$ per log unit (95% CI $+0.10$ to $+0.41$) with and without it. Volume carries no coefficient of its own at night, by day, in any region or in any climate zone (Table 1; Supplementary Note 1, S1.6). As far as building volume measures building mass, then, our panel gives no support to the second hypothesis: the within-city term does not behave as a proxy for building mass but as a measure of how many people, and how much of their activity, an already-built extent carries. What volume cannot capture is the material of the stock, its heat capacity per unit of volume, so heat storage remains possible only through materials rather than through the amount built. The third hypothesis is tested in part by season. The waste heat of air conditioning, like the evaporative cooling that vegetation supplies, is a summer quantity, whereas traffic, industry and heating act in every season. The seasonal split finds the North American term as large in winter as in summer, which speaks against a cooling-only version of the third hypothesis, and once more against the first, and leaves year-round anthropogenic heat, together with heat storage through the materials of the stock, as the readings our panel supports (Supplementary Note 1, S1.7). Taken together, the three tests give a provisional answer to the question posed above. Added residents warm the night mainly through what they do, the energy they use in every season, and not through the ground paved for them or the volume built to house them. Whether the materials of that stock add a storage term to this answer is the part a follow-up with building-stock and city-energy data can settle.

A planning argument that cites only one of the two readings is using half the evidence. The two readings pull in different directions for the same decision. The between-city reading says that, for cities of a given population, the more compact one carries the smaller night-time heat island, which supports building compactly. The within-city reading says that a city adding population to its footprint gets a warmer night as it does so, whether or not it becomes more compact in the process, which is a cost of growth rather than a verdict on form. The tension is confined to the night: by day both readings point the same way, so it is night-time minima, not daytime or annual-mean readings, on which a planner has to hold the two together. Read jointly, they say that a growing city should expect an additional warming cost while it grows, and that this cost is not evidence that a compact city is worse off once the transition is complete. As set out above, the between-city half of this argument holds in every region sampled, whereas the within-city half, the warming cost of growth, is established only for North America, and whether it carries to cities elsewhere is a question for later work.

Size, night-time burden and growth stage

Size predicts which cities carry the largest heat islands. What it predicts is not which cities have the highest absolute temperature, which climate sets, but which cities add most to their own regional temperature. The between-city level follows total population (at night, $+0.66^{\circ}\text{C}$ per tenfold population,

95% CI + 0.32 to + 0.99). So the cities carrying the largest heat islands are simply the largest ones, regardless of how compact they are. For triage, deciding where heat-health infrastructure is most needed, population is therefore the more useful screening variable. Population density defined on the fixed boundary is close to uninformative.

The burden is nocturnal and peaks in summer. Every part of the structure reported here is a night-time result: the size law holds at night and fades by day, the within-city term is a night-time term in every season, and the night-time level of the heat island is itself larger in summer than in winter in every region (Supplementary Note 1, S1.7). That timing matters for health because it is the failure of overnight cooling, not the daytime maximum, that removes the body's recovery period during a heat wave, and mortality studies report hot nights as a driver of warm-season deaths in their own right⁷. The heat island therefore adds most at the hours and in the season when the physiological margin is smallest, which carries two consequences for practice. An intervention judged on daytime or annual-mean temperature will understate what it delivers, because the quantity it acts on is largest at night; night-time minima are the metric on which cooling measures should be evaluated. And heat-warning and adaptation planning that use daytime thresholds alone miss the part of the exposure that the city itself contributes. The practical point is therefore one of target and instrument: a policy goal for urban heat should be set on night-time minima, and the measures chosen should be those that act on the heat island at night, whether they address the level a city's size has set or the increment its growth is adding.

What do these results imply for cities at different stages of growth? The structure was estimated on a sample in which most cities have largely finished growing, and the question is whether its coefficients carry to cities that have not. The evidence says that the coefficients depend on stage, in a direction that matters. In point estimate the daily-mean size-law slope is three- to fourfold steeper among the cities still building out than among the mature ones, and the within-city response is several times larger in the small group of cities that were still urbanizing over the period than in the high-income majority. That contrast is real in the data but is not yet an established effect, for a specific reason. The still-urbanizing cities are few, and they are concentrated in a few regions and in climates the sample otherwise lacks, so a steeper slope among them could reflect where they are and what climate they sit in rather than the stage they have reached. Stage, region and climate cannot be separated on this sample, and two further limits apply to the phase split. Mature here means the high-income cities, a proxy that cannot tell a city that has reached the top of its growth path from one that has stopped growing for other reasons while room remains. A first attempt to separate the two splits the high-income cities by the built-up share of their footprint, as a measure of room to grow. It finds the within-city response largest where room remains and where cities in fact grew, in the direction the stage reading implies, but the split is not stable across cells and is left to follow-up work (Supplementary Note 1, S1.4). The implication for adaptation planning follows regardless of which way that follow-up falls. If the stage effect is real, a rapidly growing city that borrows a slope estimated on European or North American cities, most of them mature, will understate how fast its own heat island is likely to grow; if it is not, the borrowed slope is merely the best available. Either way the safer practice for a growing city is to treat coefficients from mature cities as a lower bound rather than as a forecast, until an estimate from cities at its own stage exists.

How should the four planning points above, on compactness, on size, on the night-time burden and on growth stage, be used? One caution governs all of them. The coefficients behind them are structural relationships estimated across many cities: they identify which lever acts in which direction, and roughly how strongly, on average over the sample. They are not per-city forecasts. A city's own heat island departs from the sample relationship by several tenths of a degree for reasons the covariates do not capture, so the coefficients are appropriate for setting priors and for screening priorities across cities, and not for predicting what a specific intervention will do in a specific city. Between those two uses lies the aggregate. Over the cities of a country or a continent the idiosyncratic part averages out, so the coefficients can be used to project the expected effect of growth, or of a measure applied across many cities, on a regional portfolio. The region's own coefficient should be used where it differs from the pooled one, as the within-city term does. For exposure and damage assessment specifically, the results point to a refinement of current approaches rather than a replacement. The refinement is not to fix the sign of either term in advance. In our panel the night-time level is negative in about a third of cities and the daytime size-law slope is flat to slightly negative in the aggregate, so a parameterization that forces both terms to be positive overstates the daytime term across this sample and misstates the night-time term in a minority of its cities. Whether those negative values carry beyond this sample is not established: they are features of developed-world cities observed over 2001–2020, and both the daytime sign and the share of cities with a negative night-time level could differ in other periods or at other stages of growth. The recommendation is therefore a conditional one. An assessment should let the sign be determined by data from its own domain and period, and should treat a positive-only parameterization as an assumption to be tested rather than as a default.

What damage assessments should carry

Set against the coefficients that damage assessments currently carry, the structure also says how those coefficients should be read. Recalibrating them to a common functional form on matched coverage (Fig. 4) shows that both exceed this study's estimates. Karl et al.'s¹⁷ mid-century US slope runs about twice our contemporary US estimate, and Oke's¹⁶ calm-clear nocturnal maxima run three to five times above our all-weather estimates for the same regions. Neither gap reflects disagreement about how cities behave. Both reflect a difference in what was measured, an actively-urbanizing mid-century sample in Karl's case and peak calm-clear conditions rather than an all-weather climatological mean in Oke's.

What follows for an impact assessment that currently carries one of those coefficients is a set of substitutions rather than a single replacement. For the *level* of the heat island, the all-weather estimates here can stand in for the classical ones directly, and the substitution runs in one direction. A heat island projected from population with the nocturnal maxima of Oke, or with Karl's annual slope, is two to five times the level an all-weather station record supports. The share of urban damages that rests on that level therefore shrinks by the same factor. Which of our channels to use depends on what the damage function was calibrated to. A health function driven by hot nights should take the night-time slope; a function calibrated on daily-mean temperature should take the daily-mean slope, which is about a third of

the night-time one; and a daytime function should carry a slope near zero, or slightly negative, rather than a positive one. For the *change* in the heat island as a city grows, the between-city level is the wrong object altogether, and the within-city coefficient is the one to carry. Its size is instructive: Karl's mid-century United States slope, once expressed per log unit of population, lies close to the within-city response measured here, which is consistent with that slope having been estimated on cities that were actively adding people to their footprints. An assessment that has been using a Karl-type coefficient to project growth in North American cities has therefore been closer to the mark than one using it to set the present level.

Whether to apply global or regional values depends on the term. The level generalizes in sign across every region sampled but not in slope, so an assessment resolved by region should use the regional slope where one exists and the global one only for an aggregate. The within-city term does not generalize at all on this evidence. Applying the North American response to European cities would attribute growth-driven warming that the European record does not show, and applying it to the urbanizing world would extrapolate into conditions this panel does not observe. What the panel offers there is a bracket rather than a value: a lower bound near zero from mature European cities, a central estimate from North America, and, from the small non-high-income subsample whose response is several times larger, an indication that the increment in still-urbanizing cities may be well above the central estimate. The direction of the resulting change in an assessment is therefore not uniform. Damages that rest on the present heat-island level of developed-world cities are overstated by the classical coefficients; damages that rest on the growth increment in rapidly urbanizing cities are, on this evidence, under-constrained and possibly understated, and an assessment weighted toward the latter cities could move up rather than down. Carrying the bracket explicitly, and reporting the two terms separately rather than folding them into one population coefficient, is the practical recommendation this section supports.

The heat island as the shadow of growth

Read in the frame of the heat island as the shadow of growth set out at the start, the structure has a simple shape. The heat island a city carries today is set mostly by how large it has become, not by how densely it built, and it grows as the city adds people to the land it already occupies. Growth therefore pays for its own heat twice. It pays once in the level that comes with size, and again in the increment that comes with each round of infill, at night, when the body needs to cool. The first is close to a fixed cost of being a large city, and the measures that act on it are the cooling measures a city already knows, vegetation, surface materials and ventilation³, applied to a level that size sets. The second is a running cost that growth management can reach, and in our station panel it has been paid in North America and not, so far, outside it. Our panel narrows that difference without settling it. We could not attribute it to how the city is built, to the region's background trend or to the settled landscape around it. And because the North American term is present at night in winter as well as summer, we could not attribute it to cooling waste heat alone or to the vegetation that growth removes. What our panel leaves open is heat released in every season, the energy each added resident uses for heating, cooling and travel, and the

heat the building stock stores by day and gives back at night, which a city-energy follow-up can separate. Whether the heat island in turn slows the growth that produced it is not tested here, though one loop inside it is documented: heat raises cooling demand, whose waste heat raises night-time heat again^{30,40}. The panel does show the ingredients for a wider feedback, because the burden is largest in the largest cities and concentrates in the hours that matter most for health⁷. For a city asking what it can do, the result is a division of labour rather than a verdict against growth or against density. Compactness is not the penalty; size sets the level; growth on a fixed footprint sets the increment; and each is reached by a different instrument. A sustainability strategy that names which of the three it is acting on is, on this evidence, the one that can expect to see the heat island abate.

Limits of the estimates

Three limits bound how far these estimates reach. The first limit is what the within-city term can mean. Both terms are population counted on a footprint that GHS-UCDB holds fixed across epochs. Between cities that footprint differs, so density is a distinct quantity and enters as its own covariate. Within a city the footprint cannot move, so the column labeled density changes only when population changes, and the within-city density term is population added onto the extent the city already had, which is a consequence of the source data's delineation convention and holds in the data to six decimals (Supplementary Note 1, S1.5). As the moving-boundary check under Robustness shows, the results hold when a moving delineation from a different source is applied instead. In consequence, the within-city term this paper reports measures how many people an existing extent carries, not how compactly they are housed, so a claim about compactness has to rest on the between-city term alone. The second limit is predictive power, and it concerns how much of any one city's heat island the structure accounts for. The relationships reported here are averages across many cities, and they are strong as averages: the size law survives every test in Supplementary Table S1. But the statistical model explains only a modest share of how heat islands differ from one city to the next: the between-city variance it accounts for is about 5% with size alone and 23% with the full covariate set (Supplementary Table S1). The remainder is city-specific: the exact siting of the station, the local terrain, nearby water, the materials and layout of the district the station sits in, and the weather of the years measured, none of which the city-level covariates carry. That remainder amounts to roughly $\pm 0.3^{\circ}\text{C}$ of intrinsic per-city noise in station UHI, of the same order as the effects being estimated, and two consequences follow from noise of that size. The relationships describe what happens on average as size or population changes, so they are the right object for a prior, a screening rule or a regional aggregate, and the wrong object for a forecast of a named city. And an assessment that needs a city-specific number should treat the coefficients here as the systematic part of that number and add the local part from local observation, rather than reading the coefficients as if they were the whole. The clearest limit, though, is coverage. Our panel is drawn from a developed-world, extratropical station network; the humid tropics and the developing world, where current and future urban heat exposure concentrates, remain essentially unobserved. Only about 7% of the sample lies in the tropics. Even those are mostly atypical tropical cities, dry or at high elevation, and the humid-tropical megacities in which most of the tropical urban population lives are essentially absent²⁵. The few tropical cities the sample does hold already behave differently. Their mean night-time

heat island is slightly negative, -0.08°C (95% CI -0.69 to $+0.53$) against $+0.64^{\circ}\text{C}$ ($+0.51$ to $+0.76$) across the arid, temperate and continental zones together, so the typical tropical city in this sample is a mild heat sink rather than an island. The size-law slope among them is the steepest of any zone (Fig. 2c; Supplementary Table S3). Both features rest on 38 cities and should not be pressed, but they show that the tropics are not simply the temperate world at a higher baseline. There is also a general reason for caution. The size law's sign is not fixed even within the sample: it is positive at night and absent by day, so the law already changes across one dimension we can observe. A relationship that changes across an observed dimension gives no warrant for assuming that it stays fixed across an unobserved one, and humid tropical versus temperate climate is such a dimension here. A test of whether the size-law slope differs by climate zone, an interaction of city size with zone in the level regression, does not reject a common slope ($p = 0.50$; Supplementary Table S3). That result should not be read as showing that the tropical slope equals the temperate one. With 38 tropical cities the test has little power: the tropical slope is estimated at $+1.14^{\circ}\text{C}$ per tenfold population with a 95% CI from $+0.27$ to $+2.00$, wide enough to contain both the temperate slope of $+0.51^{\circ}\text{C}$ ($+0.15$ to $+0.88$) and a slope four times as steep. The data therefore cannot distinguish a tropical size law like the temperate one from a much steeper one, which is an absence of evidence for a difference rather than evidence against one. Extrapolating the law's form with confidence into unsampled conditions is therefore not something our panel can support. We see this analysis as establishing a testable structure, level versus trend and their nocturnal and regional dependence, for a broader observational effort to extend to a wider set of regions, including low-income and tropical countries and the developing world more generally.

Methods

Data

We build an air-temperature UHI record from GHCN-Daily²⁰, NOAA's global compilation of daily maximum and minimum temperature from land surface stations. We use it rather than a satellite or reanalysis product for two reasons. It is the largest available archive of raw, station-level near-surface air temperature, which is the quantity that governs human heat exposure. It also offers the multi-decade, multi-region scale that a test of the size law requires.

The record is built city by city. For each city in the GHS Urban Centre Database, the station nearest its centroid is the urban station. Where two urban centers resolve to the same nearest station, only the center whose centroid is nearer keeps it, so that each physical station enters once; the cities dropped by that rule are listed in the reproduction bundle and with the data descriptor cited below. A matched set of rural reference stations is selected separately, and the UHI is the difference between the two. Of the 40,800 stations that report temperature, 8,637 pass quality and completeness screening and enter this analysis.

The data are quality-controlled but deliberately not homogenized. Homogenization removes the gradual warming of a city relative to its rural neighbors, and that trend is what this analysis is designed to

measure. The risk that homogenization removes it is real, and the literature is split on this exact point for homogenized US station products^{41–55}. At one end, the urban–rural signal has been shown to be sensitive to the homogenization procedure itself⁴¹. At the other, a homogenized product reports no detectable urban–rural difference at all⁴².

UHI is defined as

$$\text{UHI} = T_{\text{urban}} - \text{median}(T_{\text{rural}}) + \Gamma \cdot (z_{\text{urban}} - \text{median}(z_{\text{rural}})), \quad (1)$$

Here T_{urban} is the annual (or nocturnal, or daytime) mean at the station nearest the urban centroid within 25 km. The rural reference median(T_{rural}) is the median over at least 3 clean stations in a 40–150 km annulus around that centroid. z is station elevation, and $\Gamma = 6.5^{\circ}\text{C km}^{-1}$ is a fixed environmental lapse rate. The lapse term corrects for rural stations that sit at a different altitude from the city center. Rural stations lying within 12 km of any other urban center are excluded from the reference, so that the reference is not itself warmed by a neighboring city.

All series are aggregated by us from raw daily maximum and minimum temperature. Only values whose GHCN-Daily quality flag is blank are retained, and each month requires ≥ 20 valid days and each year ≥ 11 valid months. The rural reference for the within-city trend panel is estimated once rather than re-selected at each period (see Rural reference for the trend panel). This prevents a station entering or leaving the reference set from being misread as a change in the city itself.

Seasonal series are built from the same daily record by the same rules. For each station and year, the summer mean is the mean of June, July and August, and the winter mean is the mean of the preceding December with January and February. Each month requires ≥ 20 valid days and each season all three months. The series are built separately for TMIN, TMAX and the daily mean. 7,792 of the study's stations yield 479,335 station-seasons. The composition-free reference is then refitted for each season and element, so the six seasonal panels differ slightly in which cities and epochs qualify (Table 2). Full technical validation is in the companion data descriptor⁵⁶. It covers the selection criteria, city-station matching, sensitivity to the rural-reference radius and lapse rate, and independent-archive replication.

Candidate covariates are drawn from GHS-UCDB R2024A⁵⁷ and station metadata: total population, population density, GDP per capita, built-up fraction, urban–rural contrast in the normalized difference vegetation index (NDVI), wind, coastal distance, latitude, and Köppen zone. Elevation is not used as a covariate, since it is already corrected for in the UHI construction. The usable sample varies with each analysis's requirements and is reported per result. The cross-city analyses use 719 cities, 542 of them with a nocturnal series (TMIN, the daily minimum). The within-city trend analysis is centered on five epochs from 2000–2020 (885 cities, 3,787 city-epochs). For robustness it is extended back through 1975 as a ten-epoch, 910-city (6,736 city-epoch) panel.

Population density is the GHS-UCDB column. Its denominator is each city's delineated area, held fixed across epochs. That fixed denominator gives the column two meanings. Between cities the denominator

varies, so the column is a density. Within a city it does not vary, so the column is population on a fixed footprint, and the within-city coefficient should be read accordingly. Results estimated on the moving GHS-WUP-MTUC delineation³⁸, where the two are separable, are reported in Supplementary Note 1, S1.5.

Benchmarked against all 11,161 GHS urban centers, the analyzed cities are a developed-world slice. About 15% of world urban population is represented. Roughly 95% of that population is in high-income cities, and the climates are dominantly temperate or continental. Coverage concentrates in North America, Europe, and East Asia. The tropics, Africa, South America, and the rest of the developing world are essentially absent. Supplementary Fig. S1 breaks the sample's UHI level down by continent, income group, climate zone and country. The night UHI exceeds the annual mean in 15 of the 18 groups where both can be shown, and the intervals widen sharply wherever a group is small, which is why the tropical and upper-middle-income groups cannot bear weight on their own. City size, by contrast, is sampled evenly, at 6–14% of cities in every size class from under 300,000 to over 5 million population. The population-size law is therefore estimated across the full size axis rather than a restricted slice. The geographic and climatic covariates (climate zone, latitude, income) are less evenly sampled. They enter, however, as additive offsets that shift each city's baseline level up or down as a whole, and they do not interact with the size or density slopes themselves.

Statistical estimation

UHI level is analyzed as a cross-section on city-mean UHI. To check whether a driver's effect depends on which other variables happen to be controlled for, we run an extreme-bounds search. The search refits the model under 384 different combinations of the candidate covariates. Every combination is constrained to include population, density, or both. A covariate is classified as robust if it is significant in $\geq 90\%$ of the specifications that include it, with a stable sign and confidence bounds excluding zero. A covariate that is significant only in the one specification we happened to choose is not robust. A nine-step covariate ladder complements the search. Controls are added to the cross-city size relationship one at a time on a constant sample, in the order climate zone, absolute latitude, income group, GDP per capita, population density, built-up fraction, urban-station elevation, coastal distance and station siting. A size effect that was a proxy for any of them would be seen to decay as that control enters. GDP's functional form is chosen via nested joint tests and the Akaike information criterion (AIC). AIC is a standard statistical criterion for comparing how well competing curve shapes fit the data.

UHI trend is analyzed on an epoch panel. The primary panel is the five 2000–2020 GHS epochs. As a robustness check it is extended back through 1975 (ten epochs). Standard errors are city-clustered throughout, and the trend is estimated with two approaches. The first is a two-way (city, year) fixed-effects regression. The second is a correlated-random-effects (Mundlak) model, which enters each city's mean density and its deviation from that mean in a single model. Both absorb time-invariant confounds and identify effects from within-city change. The fixed-effects form is

$$\text{UHI}_{ct} = \beta \ln D_{ct} + \gamma' X_{ct} + \mu c + \lambda t + \varepsilon_{ct}, \quad (2)$$

where D_{ct} is the GHS-UCDB population density of city c at epoch t . X_{ct} holds the covariates carried in a given specification (built-up fraction and the income spline). μ_c and λ_t are city and year effects, and β is the within-city coefficient reported in Results. The Mundlak form is

$$UHI_{ct} = \beta_w (\ln D_{ct} - \ln \bar{D}_c) + \beta_b \ln \bar{D}_c + \gamma' X_{ct} + \lambda_t + \varepsilon_{ct}, \quad (3)$$

where $\ln \bar{D}_c$ is the city's mean over the epochs it contributes. β_w equals β of Eq. (2), and β_b is the between-city coefficient. Whether the two are equal is a Wald test on their difference. This is the form's advantage: it lets the between- and within-city coefficients be tested directly against each other. One technical adjustment is needed. In an unbalanced panel, a deviation taken from raw values is not orthogonal to the year effects, so the year means are removed before the city mean and deviation are formed.

The trend panel's rural reference is not the re-selected median used for the level but a fitted one. It is described in the next section.

All level analyses are run for both the daily mean (TAVG) and the daily minimum (TMIN). This lets us test directly whether the level/trend separation is itself a nocturnal result, rather than an assumption carried over from the level alone.

Rural reference for the trend panel

The reference used for the level is the median over whichever nearby clean rural stations report in the period measured. That is harmless for a single climatological average, because more reporting stations only sample the background better. For a difference between two periods it is not harmless. A station entering or leaving the reference set moves the median, and the estimator reads that movement as a change in the city.

The trend panel therefore uses a fitted reference. A station-level regression, $T(s,t) = \alpha_s + \tau(c,t)$, is fit across all of a city's clean rural stations. The station term α_s absorbs each station's own level. The period term $\tau(c,t)$ is identified from the overlaps between stations rather than from whichever stations happen to report. This gives the composition-free reference $R(c,t) = \text{median}_s(\alpha_s) + \tau(c,t)$. No station-year is discarded, and unlike the re-selected median the construction does not become less reliable as the window lengthens, so the full 1975–2020 record remains a valid check on the trend even though it is not the primary window.

Two alternative constructions were carried for comparison. The first restricts the reference to stations valid at every epoch a city contributes, which holds composition constant by discarding stations. The second is the re-selected median itself. The first agrees with the fitted reference. Both disagree with the re-selected median on the long window. That disagreement is what identifies the composition drift as real rather than as an artifact of either method.

Two rules apply to the reference stations themselves. First, rural stations within 12 km of any other urban center are excluded, so that the reference is not itself contaminated by a neighboring city. This contamination is correlated with urbanization, so leaving it in attenuates the within-city signal toward zero. Second, the series are not homogenized (see Data), because a gradual urban trend relative to rural neighbors is precisely the local inhomogeneity that pairwise adjustment algorithms are designed to remove. Neither the radius of the first rule nor its presence changes the result. Without any screen the within-city term is attenuated slightly, from + 0.28 to + 0.25°C per log unit over 2000–2020, and at every radius from 3 to 30 km it lies between + 0.25 and + 0.29 (Robustness; Supplementary Table S1).

Statistics

All hypothesis tests are two-tailed and evaluated at a 5% significance level. Every estimate quoted in the text carries a 95% confidence interval, and the exact P value of each robustness check is given in Supplementary Table S1. Regression coefficients take theirs from the clustered standard error under the normal approximation. The central within-city terms take theirs from the city-cluster bootstrap. The population-weighted mean and the median take theirs from a city bootstrap. Standard errors are city-clustered unless stated otherwise, with country-clustered and bootstrap alternatives reported as robustness checks. No statistical methods were used to predetermine sample size. The analysis uses the full set of cities meeting the station-quality and matching criteria described above.

The primary window for the within-city trend is 2000–2020. The full 1975–2020 record is carried alongside as confirmation rather than as the primary estimate. The reason lies in the covariate, not in the reference. Pre-2000 GHS-UCDB population is reconstructed into the modern urban footprint rather than observed within the footprint of the day. That reconstruction is also why the covariate's denominator is frozen, so that within a city the column is population rather than density. From 2000, population, like GDP per capita, is drawn from inputs closer to the observation date, and the GHS-UCDB series becomes city-resolved. The panel is therefore best identified from 2000 onward. The full record then serves as a check under twice the identifying variation. The covariate's within-city standard deviation is 0.20 over 1975–2020 against 0.09 over 2000–2020, and the coefficient is the same.

Use of large language models

During the preparation of this work, the author used large language model AI assistants for language editing, and for data analysis and figure generation through author-reviewed code. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Data availability

Every input the analysis reads, as it was read, is archived in a single reproduction bundle on Zenodo under CC-BY-4.0 (10.5281/zenodo.22647642): the city-level files of the GHCN-Daily air-temperature urban heat-island record documented in the companion data descriptor⁵⁶ (the station-year panel, the city–station matching with duplicate urban-station matches removed so that each physical station

enters once, and the city UHI panels), and, from its companion city panels, the GHS-UCDB predictors, the co-located satellite panels, the building-volume, vegetation and night-time-light extractions, the moving-delineation statistics and city crosswalk behind the comparison reported in Results, and the seasonal station series and city-epoch panels behind Table 2. The bundle states, file by file, which inputs are copies of the public source records and which are filtered, and lists the source licences. All raw inputs are public: GHCN-Daily (NOAA), GHCN-M v4 (NOAA, used in the independent-archive robustness check), GHS-UCDB R2024A (EU Joint Research Centre), YCEO v4 (used in the cross-instrument robustness check), and the Yang et al. (2024) UHII dataset.

Code availability

The scripts that regenerate every number and figure from the bundled inputs are archived in the same Zenodo reproduction bundle (10.5281/zenodo.22647642), under an MIT license, together with the logs of a complete run and the figures it produced.

Declarations

Competing interests

The author declares no financial or non-financial competing interests.

Author Contribution

I.C.H. designed the study and the methods, wrote the analysis code, performed the statistical analysis, curated the data, prepared the figures, and wrote, reviewed and edited the manuscript.

Acknowledgement

This work was supported by the Korea Environmental Industry & Technology Institute (KEITI) through the Climate Change R&D Project for New Climate Regime, funded by the Ministry of Climate, Energy and Environment (formerly the Ministry of Environment) of the Republic of Korea (RS-2023-00218794), and by The Seoul Institute (Project 2026-BR-17, Sectoral Climate-Change Damage Costs and Policy Benefits for Seoul). The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Data Availability

Every input the analysis reads, as it was read, is archived in a single reproduction bundle on Zenodo under CC-BY-4.0 (<https://doi.org/10.5281/zenodo.22647642>): the city-level files of the GHCN-Daily air-temperature urban heat-island record documented in the companion data descriptor (Hwang, submitted

to Earth System Science Data; ref. 56 in the manuscript), namely the station-year panel, the city–station matching with duplicate urban-station matches removed so that each physical station enters once, and the city UHI panels; and, from its companion city panels, the GHS-UCDB predictors, the co-located satellite panels, the building-volume, vegetation and night-time-light extractions, the moving-delineation statistics and city crosswalk behind the comparison reported in Results, and the seasonal station series and city-epoch panels behind Table 2. The bundle states, file by file, which inputs are copies of the public source records and which are filtered, and lists the source licences. All raw inputs are public: GHCN-Daily (NOAA), GHCN-M v4 (NOAA, used in the independent-archive robustness check), GHS-UCDB R2024A (EU Joint Research Centre), YCEO v4 (used in the cross-instrument robustness check), and the Yang et al. (2024) UHII dataset.

References

1. Howard, L. The Climate of London, Deduced from Meteorological Observations (Harvey and Darton, 1833).
2. Landsberg, H.E. The Urban Climate (Academic Press, 1981).
3. Oke, T.R. The energetic basis of the urban heat island. *Q. J. R. Meteorol. Soc.* 108, 1–24 (1982).
4. Arnfield, A.J. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.* 23, 1–26 (2003).
5. Stewart, I.D., Oke, T.R. Local Climate Zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* 93, 1879–1900 (2012).
6. Chapman, S., Watson, J.E.M., Salazar, A., Thatcher, M., McAlpine, C.A. The impact of urbanization and climate change on urban temperatures: a systematic review. *Landsc. Ecol.* 32, 1921–1935 (2017).
7. Kim, S.E. et al. Mortality risk of hot nights: a nationwide population-based retrospective study in Japan. *Environ. Health Perspect.* 131, 057005 (2023).
8. Stewart, I.D. A systematic review and scientific critique of methodology in modern urban heat island literature. *Int. J. Climatol.* 31, 200–217 (2011).
9. Peng, S. et al. Surface urban heat island across 419 global big cities. *Environ. Sci. Technol.* 46, 696–703 (2012).
10. Manoli, G. et al. Magnitude of urban heat islands largely explained by climate and population. *Nature* 573, 55–60 (2019).
11. Imhoff, M.L., Zhang, P., Wolfe, R.E., Bounoua, L. Remote sensing of the urban heat island effect across biomes in the continental USA. *Remote Sens. Environ.* 114, 504–513 (2010).
12. Clinton, N., Gong, P. MODIS detected surface urban heat islands and sinks: global locations and controls. *Remote Sens. Environ.* 134, 294–304 (2013).
13. Zhou, D., Zhao, S., Liu, S., Zhang, L., Zhu, C. Surface urban heat island in China's 32 major cities: spatial patterns and drivers. *Remote Sens. Environ.* 152, 51–61 (2014).

14. Venter, Z.S., Chakraborty, T., Lee, X. Crowdsourced air temperatures contrast satellite measures of the urban heat island and its mechanisms. *Sci. Adv.* 7, eabb9569 (2021).
15. Zhou, D. et al. Satellite remote sensing of surface urban heat islands: progress, challenges, and perspectives. *Remote Sens.* 11, 48 (2019).
16. Oke, T.R. City size and the urban heat island. *Atmos. Environ.* 7, 769–779 (1973).
17. Karl, T.R., Diaz, H.F., Kukla, G. Urbanization: its detection and effect in the United States climate record. *J. Clim.* 1, 1099–1123 (1988).
18. Estrada, F., Botzen, W.J.W., Tol, R.S.J. A global economic assessment of city policies to reduce climate change impacts. *Nat. Clim. Chang.* 7, 403–406 (2017).
19. Estrada, F., Lupi, V., Botzen, W.J.W., Tol, R.S.J. Urban and non-urban contributions to the social cost of carbon. *Nat. Commun.* 16, 4193 (2025).
20. Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E., Houston, T.G. An overview of the Global Historical Climatology Network-Daily database. *J. Atmos. Ocean. Technol.* 29, 897–910 (2012).
21. Mundlak, Y. On the pooling of time series and cross section data. *Econometrica* 46, 69–85 (1978).
22. Zhou, B., Rybski, D., Kropp, J.P. The role of city size and urban form in the surface urban heat island. *Sci. Rep.* 7, 4791 (2017).
23. Zhao, L., Lee, X., Smith, R.B., Oleson, K. Strong contributions of local background climate to urban heat islands. *Nature* 511, 216–219 (2014).
24. Jin, M., Shepherd, J.M., Zheng, W. Urban surface temperature reduction via the urban aerosol direct effect: a remote sensing and WRF model sensitivity study. *Adv. Meteorol.* 2010, 681587 (2010).
25. United Nations, Department of Economic and Social Affairs, Population Division. World Urbanization Prospects: The 2018 Revision (United Nations, 2019).
26. Leamer, E.E. Let's take the con out of econometrics. *Am. Econ. Rev.* 73, 31–43 (1983).
27. Allen, L., Lindberg, F. & Grimmond, C. S. B. Global to city scale urban anthropogenic heat flux: model and variability. *Int. J. Climatol.* 31, 1990–2005 (2011).
28. Varquez, A. C. G., Kiyomoto, S., Khanh, D. N. & Kanda, M. Global 1-km present and future hourly anthropogenic heat flux. *Sci. Data* 8, 64 (2021).
29. International Energy Agency. The Future of Cooling: Opportunities for Energy-Efficient Air Conditioning (IEA, 2018).
30. Salamanca, F., Georgescu, M., Mahalov, A., Moustauoui, M. & Wang, M. Anthropogenic heating of the urban environment due to air conditioning. *J. Geophys. Res. Atmos.* 119, 5949–5965 (2014).
31. de Munck, C. et al. How much can air conditioning increase air temperatures for a city like Paris, France? *Int. J. Climatol.* 33, 210–227 (2013).
32. Menne, M.J., Williams, C.N., Gleason, B.E., Rennie, J.J., Lawrimore, J.H. The Global Historical Climatology Network Monthly Temperature Dataset, Version 4. *J. Clim.* 31, 9835–9854 (2018).
33. Driscoll, J.C., Kraay, A.C. Consistent covariance matrix estimation with spatially dependent panel data. *Rev. Econ. Stat.* 80, 549–560 (1998).

34. Pesaran, M.H. General diagnostic tests for cross-sectional dependence in panels. *Empir. Econ.* 60, 13–50 (2021).
35. Chakraborty, T., Lee, X. A simplified urban-extent algorithm to characterize surface urban heat islands on a global scale and examine vegetation control on their spatiotemporal variability. *Int. J. Appl. Earth Obs. Geoinf.* 74, 269–280 (2019).
36. Chakraborty, T., Lee, X. Yale Center for Earth Observation (YCEO) Surface Urban Heat Islands, Version 4, 2003–2018. NASA Socioeconomic Data and Applications Center (SEDAC) [data set] <https://doi.org/10.7927/S5M5-ZK14> (2023).
37. Yang, Q. et al. A global urban heat island intensity dataset: generation, comparison, and analysis. *Remote Sens. Environ.* 312, 114343 (2024); dataset [data set] <https://doi.org/10.6084/m9.figshare.24821538>.
38. Schiavina, M., Alessandrini, A., Melchiorri, M. & Dijkstra, L. GHS-WUP-MTUC R2025A: GHS-WUP multitemporal urban centres, obtained from the Degree of Urbanisation grids and linked across epochs, multitemporal 1950–2100. European Commission, Joint Research Centre [data set] <https://doi.org/10.2905/1ea967e5-bedc-4cf3-a0b0-3851742ee7e2> (2026).
39. Hwang, I. C. The urban heat island's share of observed city warming: a three-term decomposition of raw station records for 553 cities in 37 countries. Preprint at <https://ssrn.com/abstract=7430049> (2026).
40. Takane, Y., Kikegawa, Y., Hara, M. & Grimmond, C. S. B. Urban warming and future air-conditioning use in an Asian megacity: importance of positive feedback. *npj Clim. Atmos. Sci.* 2, 39 (2019).
41. Hausfather, Z., Menne, M.J., Williams, C.N., Masters, T., Broberg, R., Jones, D. Quantifying the effect of urbanization on U.S. Historical Climatology Network temperature records. *J. Geophys. Res. Atmos.* 118, 481–494 (2013).
42. Peterson, T.C. Assessment of urban versus rural in situ surface temperatures in the contiguous United States: no difference found. *J. Clim.* 16, 2941–2959 (2003).
43. Jones, P.D., Groisman, P.Y., Coughlan, M., Plummer, N., Wang, W.-C., Karl, T.R. Assessment of urbanization effects in time series of surface air temperature over land. *Nature* 347, 169–172 (1990).
44. Jones, P.D., Lister, D.H., Li, Q. Urbanization effects in large-scale temperature records, with an emphasis on China. *J. Geophys. Res.* 113, D16122 (2008).
45. Kalnay, E., Cai, M. Impact of urbanization and land-use change on climate. *Nature* 423, 528–531 (2003).
46. Parker, D.E. Large-scale warming is not urban. *Nature* 432, 290 (2004).
47. Parker, D.E. A demonstration that large-scale warming is not urban. *J. Clim.* 19, 2882–2895 (2006).
48. Sun, Y., Zhang, X., Ren, G., Zwiers, F.W., Hu, T. Contribution of urbanization to warming in China. *Nat. Clim. Chang.* 6, 706–709 (2016).

49. Zhou, L. et al. Evidence for a significant urbanization effect on climate in China. *Proc. Natl Acad. Sci. USA* 101, 9540–9544 (2004).
50. Menne, M.J., Williams, C.N. Homogenization of temperature series via pairwise comparisons. *J. Clim.* 22, 1700–1717 (2009).
51. Muller, R.A. et al. A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinform. Geostat. Overv.* 1:1 (2013).
52. Wickham, C. et al. Influence of urban heating on the global temperature land average using rural sites identified from MODIS classifications. *Geoinform. Geostat. Overv.* 1:2 (2013).
53. Rohde, R.A., Hausfather, Z. The Berkeley Earth Land/Ocean Temperature Record. *Earth Syst. Sci. Data* 12, 3469–3479 (2020).
54. Hansen, J. et al. A closer look at United States and global surface temperature change. *J. Geophys. Res.* 106, 23947–23963 (2001).
55. Hansen, J., Ruedy, R., Sato, M., Lo, K. Global surface temperature change. *Rev. Geophys.* 48, RG4004 (2010).
56. Hwang, I. C. A raw, homogenization-free air-temperature urban heat-island dataset for 1,117 cities from GHCN-Daily, 1941–2020. (submitted).
57. Mari Rivero, I. et al. GHS-UCDB R2024A; GHS Urban Centre Database 2025. European Commission, Joint Research Centre [data set] <https://doi.org/10.2905/1a338be6-7eaf-480c-9664-3a8ade88cbcd> (2024).

Figures

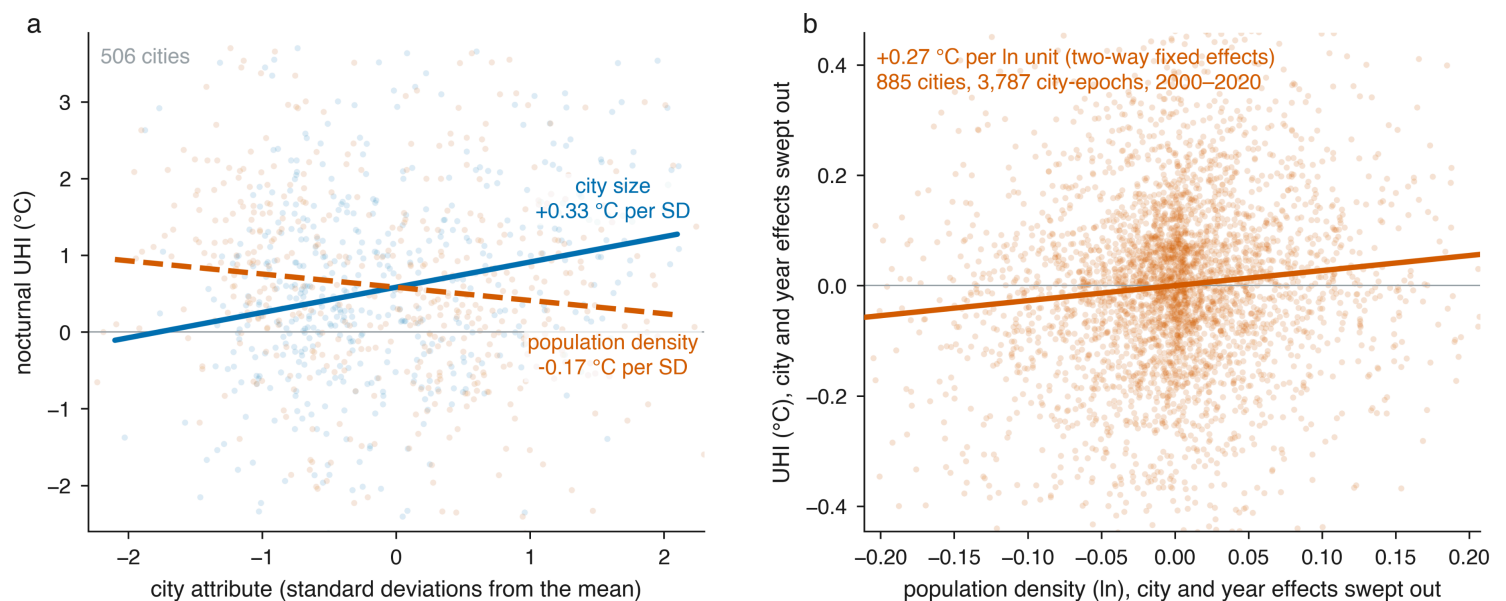


Figure 1

Between-city and within-city coefficients of night-time UHI on population and density. **a**, Between cities: partial coefficients of nocturnal UHI on city size and on population density, °C per standard deviation of the log variable, with 95% CIs from country-clustered standard errors (506 cities). Density is shown alone (−0.17 °C, 95% CI −0.40 to +0.06) and with size held (−0.31 °C, −0.53 to −0.09); size is +0.43 °C (+0.29 to +0.57). **b**, Within cities: residual UHI against residual log density after city and year effects, 2000–2020 (885 cities, 3,787 city-epochs), with the fitted slope of +0.28 °C per log unit (+0.22 °C on 1975–2020). Within a city the density column is population on a fixed footprint. Both panels are residualized on the same fixed effects and correspond to the two terms of the joint (Mundlak) model.

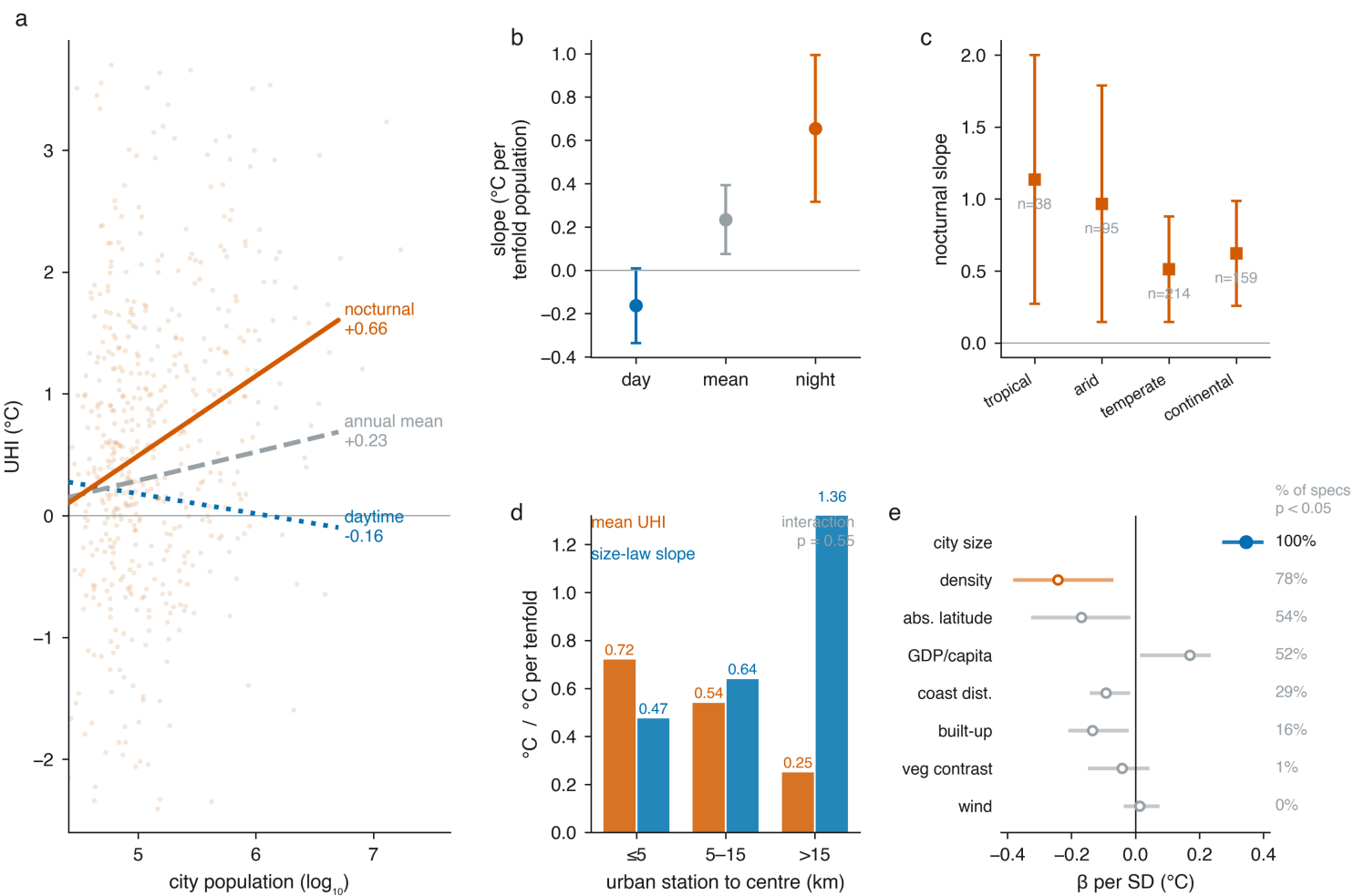


Figure 2

The population-size law by element, climate zone and station siting. **a**, Nocturnal UHI against $\log_{10}$ city population for the 542 cities with a nocturnal series, with the day, daily-mean and night size laws fitted on the level sample. **b**, Size-law slope by element, °C per tenfold population with 95% CIs: day −0.16 (−0.34 to +0.01), daily mean +0.23 (+0.08 to +0.39), night +0.66 (+0.32 to +0.99). **c**, Night slope by Köppen climate zone with 95% CIs (tropical n = 38). **d**, Mean night UHI and night size-law slope by urban-station distance from the city centroid (up to 5 km, 5 to 15 km and beyond 15 km). **e**, Extreme-bounds search over 384 covariate combinations: the share of specifications in which each candidate driver of nocturnal UHI is significant.

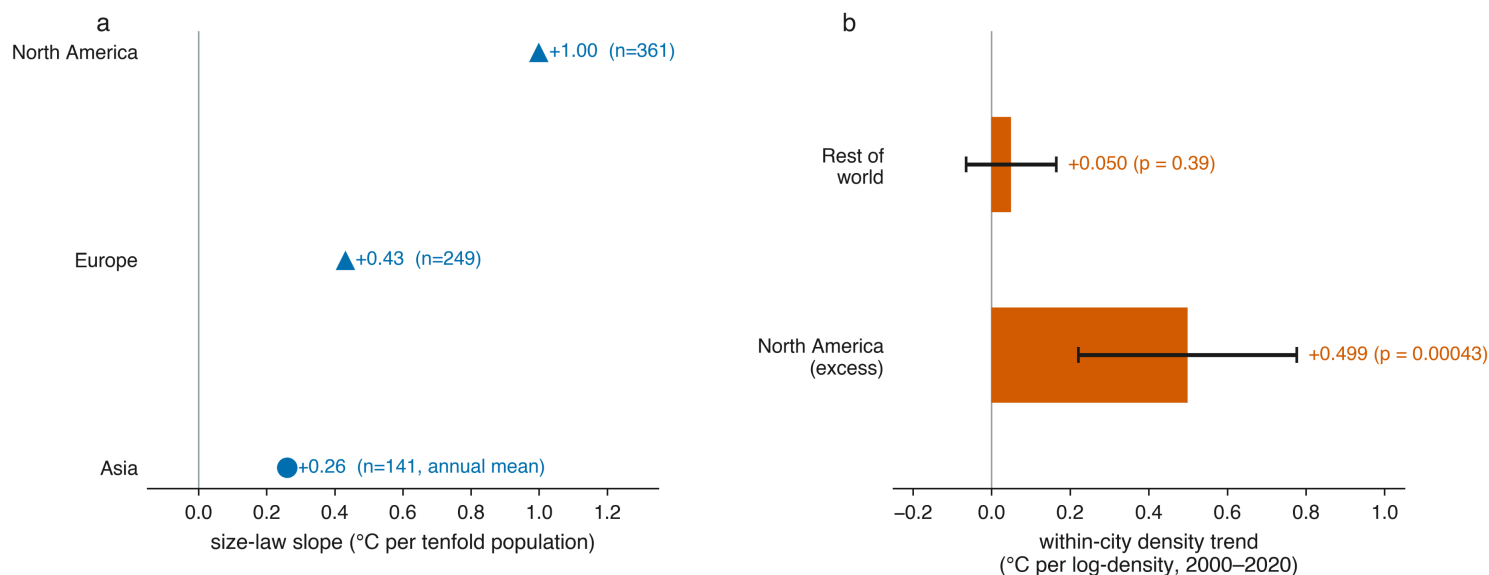


Figure 3

The size law and the within-city term by region. **a**, Size-law slope by region, °C per tenfold population with 95% CIs: night element for North America (n = 308) and Europe (n = 181), daily mean for Asia (n = 108), whose nocturnal sample is too small to fit. **b**, The within-city term in the regional-interaction model, 2000–2020: the rest-of-world coefficient (+0.050 °C per log unit, 95% CI –0.06 to +0.16) and the North American excess (+0.499 °C, +0.22 to +0.78).

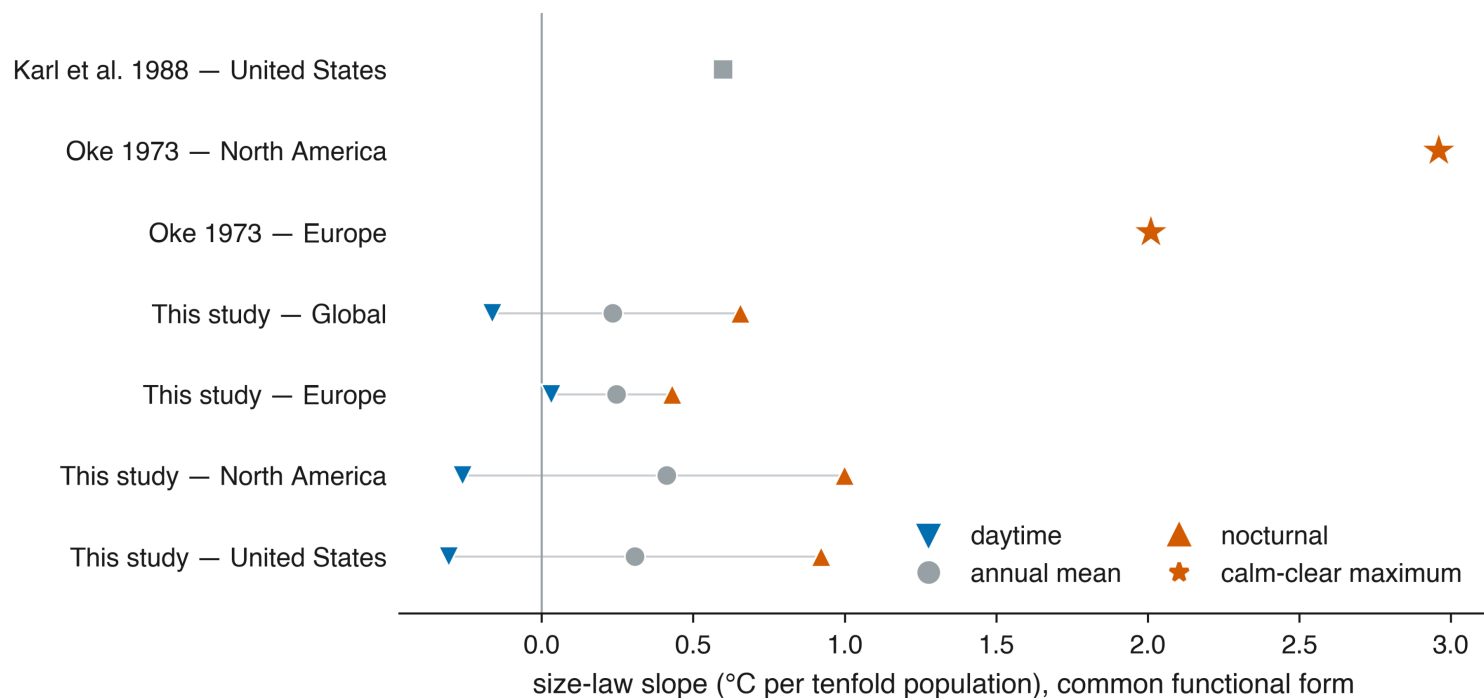


Figure 4

Published size-law coefficients and this study's estimates on a common functional form. Slopes in °C per tenfold population on matched coverage: Oke's¹⁶ calm-clear nocturnal maxima for North America

and Europe, Karl et al.'s¹⁷ annual-average United States slope, and this study's all-weather night, daily-mean and day estimates for the same regions (night +0.66 global and +1.00 North America; daily mean +0.23 global and +0.31 United States; day -0.16).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [uhinpjUSSl.pdf](#)