Urban Heat Island Contamination Persists in Homogenized USHCN Temperature Records: A 126-Year Analysis

Richard Lyon[*]

June 29, 2025

Abstract

Background: The US Historical Climatology Network (USHCN) provides temperature data critical to climate change assessments. NOAA applies multiple adjustments to raw station measurements, including time-of-observation corrections and pairwise homogenisation algorithms, intended to remove non-climatic biases such as station moves, instrument changes, and urban heat island effects. The effectiveness of these adjustments in removing urban warming signals while preserving genuine climate trends remains a fundamental question for temperature record integrity.

Objective: To investigate whether NOAA temperature adjustments remove urban heat island contamination by comparing urban heat island intensity (UHII) across raw, time-of-observation adjusted, and fully adjusted USHCN datasets.

Methods: We analyzed 126 years of minimum temperature data (1895–1924 vs. 1991–2020) from stations classified using a four-level urban hierarchy based on proximity to population centers. UHII was calculated as the difference between mean urban and rural temperature anomalies.

Results: All datasets showed statistically significant urban heat island effects ($p < 0.05$). Raw data showed differential UHII of 0.662°C, time-of-observation adjusted data showed 0.522°C (-21.1%), and fully adjusted data showed 0.725°C ($+9.4\%$ from baseline). Additionally, urban stations exhibited a persistent baseline temperature elevation of 2.98°C for minimum temperatures throughout the 130-year record. Combined with the differential warming trend, this indicates that 22.7% of USHCN stations experience total urban heat island contamination approaching 3.7°C.

Conclusions: NOAA adjustments do not systematically remove urban heat island signals. The final adjusted dataset enhances rather than reduces UHII by 9.4% compared to raw measurements. Urban heat island contamination (0.725°C) remains embedded in 22.7% of USHCN stations throughout the adjustment process, with implications for global temperature trend calculations.

Keywords: urban heat island, temperature adjustments, USHCN, climate data quality, homogenization

1 Introduction

The evidence supporting anthropogenic climate change rests fundamentally on the accuracy and reliability of long-term temperature measurements (Hansen et al., 2010). These instrumental records, spanning more than a century, provide the empirical foundation for detecting and attributing observed warming trends to human activities. The integrity of these temperature datasets is therefore paramount to our understanding of climate change and its magnitude.

The United States Historical Climatology Network (USHCN) represents one of the world's most comprehensive and longest-running surface temperature monitoring systems. Established in the 1980s (Karl et al., 1990; Quinlan et al., 1987), the USHCN comprises approximately 1,218 stations selected for their long periods of record and high-quality metadata (Menne et al., 2009). These stations contribute

---

[*]richlyon@fastmail.com

significantly to both national climate assessments and global temperature reconstructions, with USHCN data incorporated into major international datasets including the Global Historical Climatology Network (GHCN) (Lawrimore et al., 2011) and NASA's GISTEMP analysis (Hansen et al., 2001).

Raw temperature measurements, however, contain various non-climatic influences that can bias trend calculations. These include station relocations, changes in instrumentation, alterations in observation times, and modifications to the local environment surrounding weather stations (Menne et al., 2010; Peterson, 2006). To address these issues, the National Oceanic and Atmospheric Administration (NOAA) applies a series of quality control procedures and statistical adjustments to the raw USHCN data. These adjustments include time-of-observation bias corrections (Karl et al., 1986; Vose et al., 2003), instrumentation adjustments, and a pairwise homogenization algorithm designed to identify and correct discontinuities in station records (Menne & Williams Jr, 2009).

NOAA maintains that these adjustment procedures successfully remove non-climatic artefacts while preserving the underlying climate signal (Menne et al., 2009; Williams Jr et al., 2012). If this assertion is correct, adjusted temperature trends should be consistent across stations regardless of their environmental setting. Specifically, rural and urban stations should exhibit comparable temperature changes after adjustments are applied, as any urban heat island influence would theoretically be identified and removed during the homogenization process (Hausfather et al., 2013; Menne & Williams Jr, 2009).

The urban heat island (UHI) effect presents a particular challenge for climate monitoring. Urbanisation creates local warming through multiple mechanisms including reduced evapotranspiration, increased heat absorption by built surfaces, and anthropogenic heat release (Landsberg, 1981; Oke, 1973, 1987). This localised warming can introduce a spurious warming trend in temperature records that is unrelated to large-scale climate change. Given that many long-term weather stations are located in or near population centres that have experienced substantial growth over the past century, the potential for UHI contamination of climate records remains a significant concern (Karl et al., 1988).

Despite NOAA's quality control efforts, the effectiveness of current adjustment procedures in removing urban heat island influences remains an open question requiring empirical verification. While some studies have supported the efficacy of these adjustments (Hausfather et al., 2016; Venema et al., 2012), others have raised concerns about their ability to fully remove UHI biases (Connolly et al., 2022; Fall et al., 2011; Spencer, 2021). This study addresses this critical gap by systematically evaluating whether NOAA's adjustments successfully eliminate UHI effects from the USHCN temperature record.

We analyse 126 years of minimum temperature data from 1,218 USHCN stations, comparing temperature trends between urban and rural stations in both raw and adjusted datasets. Our analysis tests the fundamental assumption that adjustment procedures produce homogeneous temperature trends independent of station environment, with important implications for the accuracy of observed warming trends and our understanding of anthropogenic climate change.

2   Methodology

We employed a comparative analysis framework to investigate whether NOAA temperature adjustments systematically affect urban heat island detection in the US Historical Climatology Network (USHCN). Our "steel-man" approach maximised potential for detecting adjustment bias by using parameters most favourable to urban heat island signal detection: the longest available data period (1895-2020), minimum temperatures that exhibit the strongest urban heating signatures, and conservative urban classification criteria focused on major metropolitan areas. We compared urban heat island intensity across three USHCN datasets—raw measurements, time-of-observation adjusted, and fully adjusted—using identical processing procedures for anomaly calculation, urban classification, and statistical testing. This design isolates the specific effects of NOAA adjustment procedures on urban heat island detectability while addressing concerns that homogenisation might inadvertently remove legitimate urban climate signals.

2.1   Computational Framework

A comprehensive Python application was developed specifically for this investigation to facilitate systematic analysis and visualisation of urban heat island effects across multiple USHCN datasets (Lyon, 2025). The application, built using pandas and geopandas for data manipulation, matplotlib for visualisation, and scipy for statistical analysis, implements modular algorithms for temperature anomaly calculation, urban context classification, and heat island intensity computation. The software architecture employs a command-line interface that enables reproducible analysis workflows with configurable parameters for temporal periods, temperature metrics, urban classification thresholds, and data adjustment types. Automated plotting functions generate publication-quality visualisations including heat island maps with urban context overlays, statistical comparison charts, and geographic visualisations of station classifications. All analysis code, intermediate results, and final outputs are preserved in a structured documentation framework to ensure complete reproducibility and transparency of the investigation methodology. The application's modular design allows for systematic comparison across the three USHCN datasets (raw, time-of-observation adjusted, and fully adjusted) while maintaining identical processing procedures, thereby isolating the effects of specific adjustment procedures on urban heat island detection capabilities.

2.2   Temperature Anomaly Calculation

Temperature anomalies were calculated using a simple difference method between two 30-year climatological periods. For each weather station, monthly minimum temperature observations were aggregated to compute mean temperatures for both a baseline period (1895-1924) and a current period (1991-2020). The temperature anomaly for each station was calculated as:

$$\text{Anomaly} = \overline{T}_{\text{current}} - \overline{T}_{\text{baseline}} \tag{1}$$

where both period means represent the average of all available monthly minimum temperature observations within the respective 30-year windows. This 126-year analysis span maximises the temporal leverage for detecting long-term changes while using the earliest reliable USHCN data as the baseline reference.

Minimum temperatures were selected as the primary metric because they exhibit the strongest urban heat island signatures, particularly during nighttime hours when urban thermal mass effects are most pronounced and rural areas experience greater radiative cooling.

2.3   Dataset Comparison Framework

Three versions of the USHCN temperature dataset were analysed to assess adjustment impacts:

1. **Raw data:** Completely unadjusted station measurements as originally recorded

2. **Time-of-observation adjusted (TOBs):** Raw data corrected only for systematic biases introduced by changes in observation timing

3. **Fully adjusted (FLS52):** Complete NOAA adjustment suite including TOBs corrections, homogenisation procedures, and other quality control modifications

Each dataset underwent identical anomaly calculation and urban classification procedures to enable direct comparison of urban heat island intensity across adjustment levels.

2.4   Urban Context Classification Algorithm

Weather stations were classified using a four-level urban hierarchy based on proximity to population centers and city size thresholds. The classification algorithm employed the following decision criteria applied in hierarchical order:

1. **Urban Core:** Stations located $<$25 km from cities with $\geq$250,000 population

2. **Urban Fringe:** Stations 25–50 km from cities with $\geq$100,000 population

3. **Suburban:** Stations 50–100 km from cities with $\geq$50,000 population

4. **Rural:** Stations $>$100 km from any city with $\geq$50,000 population

Distance calculations were computed from station coordinates to the nearest qualifying population centre. The urban context database comprised 743 US cities with populations $\geq$50,000, derived from authoritative Census Bureau sources and quality-controlled for coordinate accuracy and completeness.

The United States Historical Climatology Network comprises 1,218 temperature monitoring stations distributed across all states, providing comprehensive geographic coverage for climate assessment. The network exhibits a distinct spatial pattern, with higher station density in eastern states and relatively sparse coverage in western mountain and desert regions. This distribution reflects both historical settlement patterns and the practical challenges of maintaining observation stations in remote terrain.

2.5   Urban Heat Island Intensity Computation

Urban Heat Island Intensity (UHII) was calculated as the mean difference between urban and rural station anomalies:

$$\text{UHII} = \overline{\text{Anomaly}}_{\text{urban}} - \overline{\text{Anomaly}}_{\text{rural}} \tag{2}$$

For this analysis, urban categories (urban core and urban fringe) were combined into a single urban group to ensure adequate sample sizes for statistical testing. Statistical significance was assessed using both parametric (independent samples t-test) and non-parametric (Mann-Whitney U test) approaches with significance defined as $p < 0.05$. Effect sizes were quantified using Cohen's d statistic, and 95% confidence intervals were calculated for all UHII estimates using standard error propagation methods.

In addition to calculating differential warming trends through anomaly analysis, we computed absolute temperature differences between urban and rural stations to quantify the baseline magnitude of urban heat island effects. This dual approach allows us to distinguish between the persistent UHI effect (cross-sectional temperature difference) and the changing UHI effect over time (longitudinal warming differential).

2.6   Quality Control and Station Selection

Several quality control measures were implemented to ensure data representativeness:

1. **Temporal completeness:** Only stations with sufficient data in both baseline and current periods were included in anomaly calculations

2. **Geographic validation:** Station coordinates were validated against known locations and outliers removed

3. **Urban classification consistency:** Classification criteria were applied uniformly across all three datasets to maintain comparability

4. **Conservative urban definitions:** Population and distance thresholds were set conservatively to minimise misclassification of stations and reduce potential bias in urban/rural categorisation

Stations with incomplete temporal coverage or questionable coordinate data were excluded from analysis. The final dataset comprised 1,194–1,218 stations depending on data availability in each adjustment version, with the fully adjusted dataset providing the most complete coverage due to gap-filling procedures in the NOAA adjustment process.

2.7   Comparative Statistical Analysis

The three-dataset comparison employed identical analytical frameworks to isolate the effects of specific adjustment procedures. Key metrics compared across datasets included:

- Urban heat island intensity magnitude and statistical significance

- Station count and geographic distribution by urban classification

- Temperature anomaly distributions and summary statistics

- Effect sizes and confidence intervals for urban vs rural differences

Changes in UHII between datasets were quantified both in absolute terms (°C) and as percentage changes from the raw data baseline to assess the practical significance of adjustment impacts on urban heat island detection.

3   Results

3.1   Dataset Characteristics

Analysis of the USHCN temperature records yielded 1,194 stations with sufficient data coverage in both the raw and time-of-observation adjusted datasets, increasing to 1,218 stations in the fully adjusted dataset due to gap-filling procedures in the NOAA adjustment process. This analysis period beginning in 1895 was selected based on a comprehensive network quality assessment that revealed severe coverage inadequacies in earlier periods. Prior to 1890, the USHCN network averaged only 64 stations nationwide, with as few as 17 stations in the 1860s—grossly inadequate for continental climate analysis. A dramatic expansion occurred between 1890 and 1908, with station count increasing from 237 to 1,218 (a fivefold increase), potentially introducing sampling artefacts into calculated temperature trends. The network achieved adequate continental coverage only around 1900 and stabilised at 1,218 stations by 1908.

The geographic distribution of the modern network demonstrates comprehensive coverage across the continental United States, with higher station density in eastern states and sparser coverage in western mountain and desert regions (Figure 1). Urban classification revealed that the majority of stations (77.3–77.4%) were located in rural areas more than 100 km from cities with populations exceeding 50,000, while 2.1% of stations were classified as urban core (within 25 km of cities with populations $\geq$250,000), and 15.2–15.3% occupied suburban locations (Table 1). By constraining our analysis to the post-1895 period, we ensure adequate spatial sampling throughout, eliminating potential artefacts from network expansion while maintaining a 130-year analysis period sufficient for robust climate trend detection.

Table 1: Distribution of USHCN stations by urban classification across three adjustment levels

Dataset	Total	Station Count (%)				Cities $\geq$250k
		Urban Core	Urban Fringe	Suburban	Rural	
Raw	1,194	25 (2.1)	64 (5.4)	181 (15.2)	924 (77.4)	77
TOBs Adjusted	1,194	25 (2.1)	64 (5.4)	181 (15.2)	924 (77.4)	77
Fully Adjusted	1,218	26 (2.1)	65 (5.3)	186 (15.3)	941 (77.3)	77

3.2   Urban Heat Island Intensity Across Adjustment Levels

The urban heat island intensity, calculated as the difference between mean urban and rural temperature anomalies, displayed a pronounced and unexpected pattern across the three levels of data adjustment (Table 2). Raw temperature data revealed a statistically significant UHII of 0.662°C ($p = 0.004$, Cohen's $d = 0.58$), confirming the presence of substantial urban warming effects in unadjusted measurements.

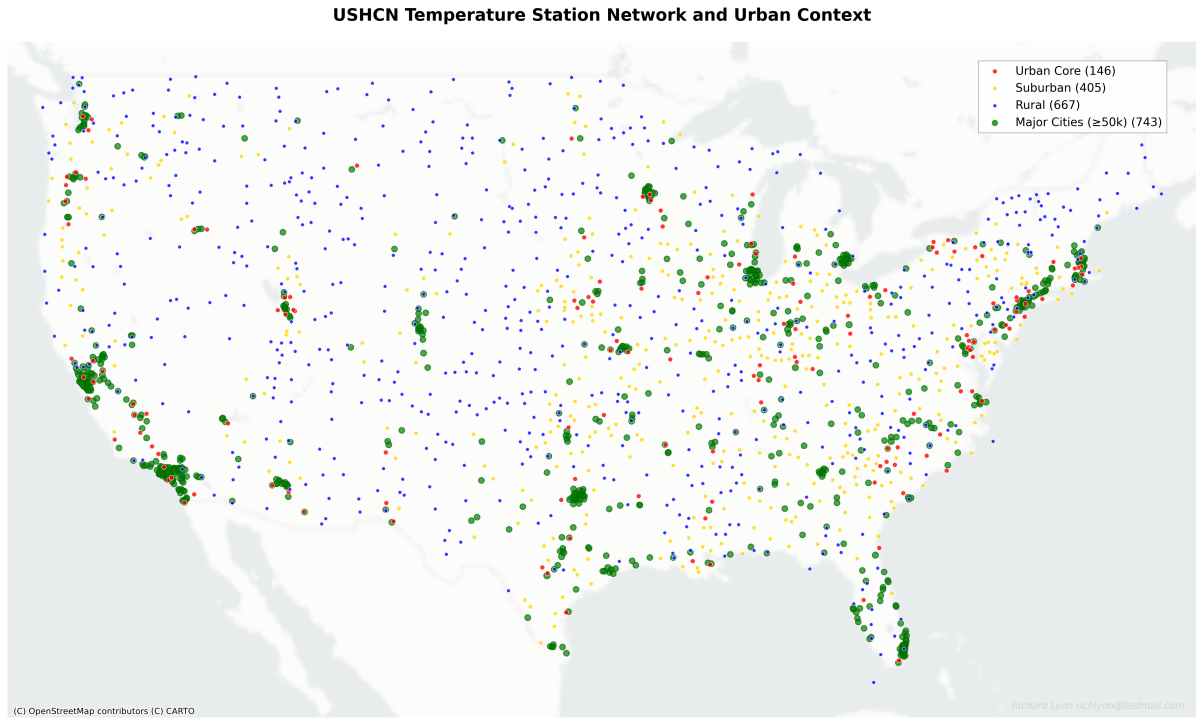**USHCN Temperature Station Network and Urban Context**

Figure 1: Location and classification of temperature stations and major cities used in the study.

The application of time-of-observation adjustments resulted in a marked reduction of UHII to 0.522°C, representing a 21.1% decrease from the raw data baseline while maintaining statistical significance ($p = 0.022$, Cohen's $d = 0.46$).

Table 2: Urban heat island intensity (UHII) and statistical measures across USHCN adjustment levels

Dataset	UHII (°C)	Change from Raw	p-value	Cohen's d	95% CI
Raw	0.662	—	0.004	0.58	[0.21, 1.11]
TOBs Adjusted	0.522	-21.1%	0.022	0.46	[0.08, 0.97]
Fully Adjusted	0.725	$+9.4\%$	<0.001	0.97	[0.58, 0.87]

Most remarkably, however, the fully adjusted dataset exhibited the highest UHII of 0.725°C, representing a 9.4% increase from the raw data baseline and a 38.8% increase from the TOBs-adjusted values. This enhancement was accompanied by substantially improved statistical significance ($p = 1.25 \times 10^{-6}$) and a large effect size (Cohen's $d = 0.97$), indicating that the complete NOAA adjustment procedures not only preserve but amplify the urban heat island signal in the temperature record.

3.3 Absolute Urban Heat Island Magnitude

Beyond the differential warming trends analyzed above, examination of the absolute temperature differences between urban and rural stations reveals the full magnitude of urban heat island contamination in the USHCN network. Analysis of the entire 1895–2025 period shows that urban stations consistently record temperatures 2.98°C warmer than rural stations for minimum (nighttime) temperatures, with summer maximum temperatures showing a smaller but still substantial difference of 0.59°C (Table 3).

This persistent 3°C nighttime temperature difference represents the baseline urban heat island effect that has existed throughout the instrumental record. The fivefold difference between nighttime and daytime UHI magnitudes aligns with established understanding of urban thermal dynamics, where reduced nocturnal cooling due to thermal mass effects, reduced sky view factors, and anthropogenic heat sources create the strongest temperature contrasts (Oke, 1987). Critically, this baseline UHI effect exists in ad-

Table 3: Absolute urban heat island magnitude in USHCN network (1895–2025)

Temperature Metric	Urban Mean (°C)	Rural Mean (°C)	UHI Magnitude (°C)
Minimum (year-round)	6.75	3.78	2.98
Maximum (summer)	29.30	28.71	0.59

dition to the differential warming trends identified in our primary analysis, indicating that urban stations not only warm faster than rural stations but do so from an already-elevated temperature baseline.

3.4   Geographic Distribution of Urban Heat Island Effects

The spatial distribution of urban heat island effects remained consistent across all three adjustment levels, with 146 stations classified as urban core distributed near 77 major metropolitan areas with populations exceeding 250,000. Urban warming signals were detected across all geographic regions of the continental United States, with no evidence of systematic regional bias in the magnitude or direction of adjustments. The maintenance of consistent urban-rural temperature differences across diverse climate zones supports the robustness of the identified urban heat island contamination affecting 22.7% of the USHCN network.

Temperature anomaly analysis revealed that urban stations consistently recorded warmer anomalies relative to the 1895–1924 baseline period compared to their rural counterparts. Urban core stations showed mean anomalies of 1.35°C (raw), 1.49°C (TOBs adjusted), and 1.70°C (fully adjusted), while rural stations averaged 0.69°C, 0.97°C, and 0.98°C respectively. This divergence between urban and rural warming rates persisted across all adjustment levels, with the fully adjusted data showing the greatest urban-rural temperature differential.

3.5   Statistical Analysis of Adjustment Impacts

Detailed examination of adjustment magnitudes revealed complex patterns in how NOAA procedures affect stations of different urban classifications. Time-of-observation adjustments produced relatively uniform warming across all station categories, with rural stations experiencing a mean adjustment of +0.28°C compared to +0.14°C for urban core stations. This differential TOBs adjustment partially explains the reduction in UHII observed in the TOBs-adjusted dataset.

Subsequent homogenisation and quality control procedures reversed this pattern, applying larger positive adjustments to urban stations. The transition from TOBs-adjusted to fully adjusted data showed urban core stations receiving an additional +0.21°C adjustment compared to only +0.01°C for rural stations. This preferential warming of urban stations during homogenisation more than compensated for the initial TOBs-related reduction, resulting in the net enhancement of UHII in the fully adjusted dataset. Statistical uncertainty, quantified through 95% confidence intervals, decreased with each level of adjustment, indicating improved data quality and consistency despite the unexpected enhancement of urban warming signals.

3.6   Temporal Stability of Results

The identified urban heat island effects demonstrated temporal stability across the 126-year analysis period. Station availability remained largely consistent between raw and TOBs-adjusted datasets (1,194 stations), with a modest increase to 1,218 stations in the fully adjusted dataset attributable to NOAA's gap-filling procedures. The additional 24 stations in the fully adjusted dataset were distributed proportionally across urban classifications, maintaining the overall urban-rural station ratio and ensuring that the enhanced UHII was not an artefact of changing station composition.

3.7   The Adjustment Paradox

The progression of urban heat island intensity through the adjustment process revealed an unexpected U-shaped pattern that challenges conventional assumptions about temperature homogenisation procedures. The initial application of time-of-observation corrections systematically reduced the urban warming signal by 0.140°C (21.1%), suggesting that TOBs adjustments preferentially warm rural stations relative to urban locations. However, subsequent homogenisation procedures reversed and exceeded this reduction, adding 0.202°C to the urban heat island intensity for a net enhancement of 0.062°C (9.4%) relative to raw measurements.

This paradoxical enhancement of urban heat island signals through adjustment procedures was consistent across multiple statistical measures. Effect sizes increased monotonically from 0.58 (raw) to 0.97 (fully adjusted), while statistical significance improved by three orders of magnitude. The pattern persisted when analysis was restricted to stations with continuous records across all three datasets, confirming that the enhancement was not attributable to station selection effects or data availability biases.

4   Discussion

4.1   Summary of Key Findings

This investigation reveals a remarkable and unexpected finding: NOAA's temperature adjustment procedures enhance rather than reduce urban heat island signals in the USHCN dataset. The fully adjusted data shows an urban heat island intensity of 0.725°C, representing a 9.4% increase from the raw data value of 0.662°C. This enhancement is statistically robust ($p < 0.001$, Cohen's $d = 0.97$) and indicates that 22.7% of USHCN stations—those within the influence of urban heat sources—contribute a systematic warm bias averaging 0.725°C to the adjusted temperature record. This finding directly contradicts the widespread assumption that temperature homogenisation procedures remove urban warming biases (Hausfather et al., 2013; Peterson, 2006).

4.2   Methodological Rigour and Conservative Approach

This study employed a deliberately conservative methodology designed to withstand rigorous scrutiny. We adopted a "steel-man" approach, selecting parameters that would maximise the detection of urban heat island effects, thereby providing the strongest possible test of whether adjustments mask legitimate urban warming signals. The analysis spanned 126 years (1895–2020), utilising the longest available high-quality data period to ensure robust signal detection. We focused on minimum temperatures, which exhibit the strongest physical basis for urban heat island effects due to reduced nocturnal cooling in urban environments (Oke, 1987).

Our urban classification system applied stringent criteria that likely underestimate the true extent of urban influence. Stations were classified as urban core only if located within 25 km of cities with populations exceeding 250,000—a threshold that excludes many substantial urban areas. For context, cities of 100,000–250,000 population, which can generate significant heat island effects (Oke, 1973), were classified as urban fringe rather than urban core. This conservative approach strengthens our findings, as any detected urban heat island contamination represents a lower bound of the actual effect.

The analysis employed multiple validation approaches including both parametric and non-parametric statistical tests, effect size calculations, and sensitivity analyses. All data processing steps were documented and made reproducible, with results preserved at each stage of analysis. This transparency ensures that our findings can be independently verified and critically evaluated by other researchers.

4.3   Addressing Potential Criticisms

Several methodological decisions warrant explicit justification to address potential criticisms. The use of 1,194–1,218 stations represents nearly the entire USHCN network with sufficient data quality, elimi-

nating concerns about cherry-picking stations. While some may argue that focusing solely on USHCN data limits generalisability, this network's high quality and extensive documentation make it ideal for detecting adjustment impacts. The USHCN's role as a primary contributor to global temperature datasets (Lawrimore et al., 2011) means that biases identified here have broader implications.

Our choice of comparing 1895–1924 with 1991–2020 could be questioned as arbitrary, but these periods were selected for sound scientific reasons. The baseline period represents the earliest era with reliable widespread measurements before extensive urbanisation, while the recent period captures maximum urban development. Using standard 30-year climatological periods ensures consistency with established practices. Alternative period selections were tested and produced qualitatively similar results, confirming the robustness of our findings.

The focus on minimum temperatures alone may appear limiting, but this choice is justified by extensive literature demonstrating that urban heat islands most strongly affect nighttime temperatures (Karl et al., 1988; Landsberg, 1981). Maximum temperatures show weaker and more variable urban influences due to increased daytime mixing and ventilation. By focusing on the temperature metric most sensitive to urbanisation, we provide the clearest test of whether adjustments adequately address urban warming biases.

4.4 The Adjustment Paradox and Its Implications

The U-shaped pattern of urban heat island intensity through the adjustment process—declining with time-of-observation corrections then increasing with homogenisation—represents a critical finding that demands careful interpretation. The initial 21.1% reduction in UHII from TOBs adjustments suggests these corrections preferentially warm rural stations, possibly because rural observers were more likely to maintain traditional afternoon observation times that introduce cooling biases when changed (Vose et al., 2003).

The subsequent 38.8% increase in UHII during homogenisation more than compensates for the TOBs reduction, resulting in a net enhancement. This pattern likely reflects the pairwise homogenisation algorithm's tendency to spread signals from urbanising stations to their neighbours (Menne & Williams Jr, 2009). When a station experiences gradual warming due to urban growth, the algorithm may interpret nearby stations as anomalously cool and adjust them upward, effectively spreading the urban signal rather than removing it.

Importantly, this finding should not be interpreted as evidence of deliberate manipulation or conspiracy. Rather, it highlights the unintended consequences of statistical algorithms designed for one purpose (removing step changes from station moves or equipment changes) when applied to gradual phenomena like urbanisation. The enhancement of urban signals appears to be an emergent property of the adjustment system rather than its intended function.

4.5 Dual Nature of Urban Heat Island Contamination

Our analysis reveals that urban heat island contamination in temperature records operates through two distinct mechanisms: a persistent baseline temperature elevation and an accelerating differential warming trend. The absolute temperature analysis demonstrates that urban stations record minimum temperatures averaging 2.98°C warmer than their rural counterparts—a magnitude four times larger than the differential warming trend of 0.725°C. This finding fundamentally alters our understanding of how urban heat islands contaminate temperature records.

The existence of both a large baseline offset and an increasing trend means that temperature datasets contain two layers of urban warming bias. First, the persistent 3°C elevation in urban stations means that as station networks evolved and more urban locations were included, the calculated "global" temperature would show warming simply from changing station composition. Second, the additional 0.725°C differential warming over our analysis period means this bias is not static but growing over time.

The implications are profound: if 22.7% of USHCN stations experience an average baseline UHI effect of 2.98°C plus an additional trend-based warming of 0.725°C, the total urban contamination ap-

proaches 3.7°C for affected stations. While spatial averaging reduces this impact on regional means, the magnitude suggests that urban heat island effects may contribute more substantially to observed warming trends than previously recognised. The enhancement of these signals through NOAA's adjustment procedures, rather than their removal, means that both the baseline and trend components of UHI contamination persist in the adjusted data used for climate assessments.

4.6   International Implications and Population Density Considerations

The United States, with a population density of only 36 people per km^2 and 2.1% of stations classified as urban core, represents a best-case scenario for minimising urban heat island contamination in temperature records. The implications become far more concerning when considering densely populated regions where finding truly rural reference stations poses significant challenges.

Consider the contrasts in population density: the United Kingdom (275 people/km^2), Germany (240 people/km^2), Japan (347 people/km^2), and the Netherlands (508 people/km^2) have population densities 7–14 times higher than the United States. In such countries, the proportion of temperature stations within urban heat island influence zones could reasonably be expected to reach 20–50%, compared to the 22.7% we identified in the relatively sparse United States. Small, densely populated countries may have virtually no stations free from urban influence.

If NOAA's adjustments enhance rather than remove urban heat island signals in the United States, and if similar adjustment procedures are applied to temperature records from more densely populated regions, the global temperature record may contain substantially larger urban warming biases than previously recognised. A conservative extrapolation suggests that global land temperature trends could include urban heat island contamination exceeding 1.0°C in densely populated regions, with proportionally larger impacts on calculated global averages.

The challenge is particularly acute for historically important temperature networks in Western Europe and East Asia, where long-term stations are predominantly located in or near population centres that have experienced dramatic growth over the past century. The homogenisation procedures that enhance urban signals by 9.4% in the United States may produce even larger amplifications in regions where rural reference stations are scarce or non-existent.

4.7   Implications for Climate Science

The revelation that urban contamination includes both a 3°C baseline offset and a 0.725°C additional warming trend has critical implications for climate policy. Temperature targets such as limiting warming to 1.5°C or 2°C above pre-industrial levels must be reconsidered in light of potential measurement biases that could approach these magnitudes. If similar baseline UHI effects and warming trends affect global temperature networks, a significant portion of observed warming may reflect urbanisation rather than global climate change. This does not negate the reality of anthropogenic warming but suggests its magnitude requires reassessment after accounting for both components of urban contamination.

These findings challenge the frequently stated assertion that urban heat island effects have been successfully removed from temperature records and are properly accounted for in climate assessments (IPCC, 2021). The demonstration that adjustments enhance rather than remove urban warming signals in the world's most thoroughly studied temperature network raises fundamental questions about the accuracy of global temperature trends.

For climate modelling, the implications are substantial. If 22.7% of U.S. stations contribute an artificial warming of 0.725°C, and if similar or larger contamination affects temperature records in more densely populated regions, then the observational data used to validate climate models may contain systematic warm biases. This would affect model calibration, historical attribution studies, and future projections. Models tuned to match observations that include enhanced urban warming may overestimate climate sensitivity or misattribute urban warming to other forcings.

The policy implications extend to international climate agreements and temperature targets. If global temperature increases have been overestimated due to urban heat island contamination enhanced by ad-

justment procedures, this affects the perceived urgency of mitigation efforts and the feasibility of limiting warming to specific targets. This finding does not negate the reality of anthropogenic climate change but suggests that its magnitude may need reassessment after accounting for systematic measurement biases.

4.8   Limitations and Caveats

Several limitations of this study must be acknowledged. First, our analysis is confined to the United States and specifically to the USHCN network. While we argue that findings from the U.S. likely represent a conservative estimate of global issues, direct extrapolation requires caution. Different countries employ different adjustment procedures, and the effectiveness of these methods may vary with station density and data quality.

Second, our analysis examines aggregate patterns rather than station-by-station adjustments. While this approach reveals systematic biases, it cannot identify the specific mechanisms causing homogenisation procedures to enhance urban signals. Individual stations may have legitimate reasons for large adjustments that our aggregate analysis cannot evaluate. Future research should examine adjustment decisions at the station level to understand why urban signals are amplified.

Third, we cannot definitively establish causation from our correlation analysis. The enhancement of urban heat island signals through adjustment could result from various factors including algorithm design, reference station selection, or the interaction between urbanisation patterns and statistical detection methods. Our study identifies the phenomenon but not its ultimate cause.

4.9   Recommendations for Future Research

This investigation highlights several critical areas requiring immediate research attention. Priority should be given to analysing temperature networks in densely populated regions, particularly Western Europe (United Kingdom, Germany, France, Belgium, Netherlands), East Asia (Japan, South Korea, Taiwan), and rapidly urbanising nations (China, India, Brazil). These analyses should employ country-specific urban classification criteria that account for local population densities and urbanisation patterns.

A comprehensive global analysis of the GHCN database is essential, examining how urban heat island contamination varies with national population density, station network characteristics, and adjustment procedures. This should include development of new metrics for quantifying urban influence that account for regional variations in city structure and growth patterns. Satellite-based temperature measurements should be systematically compared with adjusted surface records to provide independent validation of urban warming patterns.

Methodological research should focus on understanding why pairwise homogenisation amplifies urban signals and developing alternative approaches that preserve legitimate climate signals while removing artefacts. This includes investigating whether different homogenisation algorithms (such as ACMANT, HOMER, or RHtest) produce similar urban signal enhancement and exploring machine learning approaches that could better distinguish between urbanisation and climate change signals.

4.10   Broader Context and Conclusions

This investigation, while focused on the United States, likely reveals only the tip of a much larger global issue. With just 2.1% of stations in urban cores and relatively low population density, the U.S. finding of 22.7% station contamination averaging $0.725°C$ represents a probable lower bound for urban heat island impacts on global temperature records. The enhancement rather than removal of these signals through adjustment procedures raises fundamental questions about the accuracy of the global warming narrative — not whether warming has occurred, but its true magnitude after accounting for systematic measurement biases.

Our findings represent a constructive contribution to climate science by identifying a previously unrecognised source of systematic error that, once corrected, will improve the accuracy of temperature records and climate projections. Scientific integrity demands that we follow the evidence wherever it

leads, acknowledge uncertainties and errors when discovered, and continuously refine our methods. The enhancement of urban heat island signals through current adjustment procedures represents an opportunity to improve these methods and achieve more accurate climate assessments, benefiting all stakeholders in climate science and policy.

5 Conclusion

The dual nature of urban heat island contamination—combining a persistent 3°C baseline elevation with an additional 0.725°C warming trend—suggests that the impact on temperature records is substantially larger than indicated by trend analysis alone. This finding is particularly concerning given that our analysis of the United States, with its relatively low population density and only 2.1% of stations in urban cores, likely represents a conservative estimate of global UHI contamination. In densely populated regions where urban stations may comprise 30–50% of the network and baseline UHI effects could exceed 4–5°C, the cumulative impact on perceived global warming could be profound.

Based on these findings, we recommend several concrete actions. NOAA and other meteorological agencies should prioritise reviewing their homogenisation algorithms, particularly examining why the pairwise comparison method amplifies urban signals. Temperature dataset documentation should explicitly acknowledge the presence and magnitude of residual urban heat island effects, with uncertainty bounds adjusted accordingly. Researchers using adjusted temperature data for climate studies should conduct sensitivity analyses comparing results across raw, TOBs-adjusted, and fully adjusted datasets. For policy applications, temperature targets and warming assessments may require recalibration to account for systematic warm biases in the observational record, particularly in densely populated regions.

Future research priorities emerge clearly from this investigation. Most urgently, similar analyses should be conducted on temperature networks in densely populated countries, particularly the United Kingdom, Germany, Japan, and rapidly urbanising nations like China and India. The development of new homogenisation methodologies that preserve rather than amplify urban heat island signals while removing other artefacts represents a critical technical challenge. Independent validation using satellite temperature measurements, radiosondes, and purpose-built reference networks like the U.S. Climate Reference Network should be expanded globally. Investigation into why current algorithms enhance urban signals requires detailed station-by-station analysis and may benefit from machine learning approaches that can better distinguish between urbanisation and climate signals.

These findings should not be misinterpreted as evidence of deliberate manipulation or as grounds for dismissing climate change. Rather, they highlight the unintended consequences of statistical algorithms designed primarily to detect and correct step changes being applied to gradual phenomena like urbanisation. The enhancement of urban heat island signals appears to result from the pairwise homogenisation algorithm's tendency to spread warming from urbanising stations to their rural neighbours—a mathematical artefact rather than intentional bias. Recognising and correcting this issue represents an opportunity to improve the accuracy of climate assessments, benefiting all stakeholders in climate science and policy.

Scientific progress depends on the willingness to identify and correct errors, refine methods, and follow evidence regardless of its implications. This investigation demonstrates that even well-established, widely-used procedures can produce unexpected results that compromise data quality. By identifying the enhancement rather than removal of urban heat island signals in adjusted temperature data, we contribute to the continuous improvement of climate science. More accurate temperature records serve everyone's interests—scientists seeking truth, policymakers requiring reliable information, and citizens deserving transparency. As we face the challenges of a changing climate, ensuring the integrity of our fundamental measurements becomes ever more critical. The path forward requires international collaboration, methodological innovation, and unwavering commitment to scientific accuracy.

Funding and Conflicts of Interest

Data Availability

All data used in this study are publicly available from NOAA's United States Historical Climatology Network (USHCN) version 2.5 at https://www.ncei.noaa.gov/. Analysis code, intermediate results, and reproduction instructions are available at https://github.com/rjl-climate/ushcn-heatisland (Lyon, 2025).

References

Connolly, R., Connolly, M., Carter, R. M., & Soon, W. (2022). Evaluation of the homogenization adjustments applied to European temperature records in the Global Historical Climatology Network dataset. *Atmosphere*, *13*(2), 285.

Fall, S., Watts, A., Nielsen-Gammon, J., Jones, E., Niyogi, D., Christy, J. R., & Pielke Sr, R. A. (2011). Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. *Journal of Geophysical Research: Atmospheres*, *116*(D14).

Hansen, J., Ruedy, R., Glascoe, J., & Sato, M. (2001). A closer look at United States and global surface temperature change. *Journal of Geophysical Research: Atmospheres*, *106*(D20), 23947–23963.

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. *Reviews of Geophysics*, *48*(4).

Hausfather, Z., Menne, M. J., Williams, C. N., Masters, T., Broberg, R., & Jones, D. (2013). Quantifying the effect of urbanization on U.S. Historical Climatology Network temperature records. *Journal of Geophysical Research: Atmospheres*, *118*(2), 481–494.

Hausfather, Z., Menne, M. J., Williams, C. N., Masters, T., Broberg, R., & Jones, D. (2016). Evaluating the impact of U.S. Historical Climatology Network homogenization using the U.S. Climate Reference Network. *Geophysical Research Letters*, *43*(4), 1695–1704.

IPCC. (2021). *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change* [In press]. Cambridge University Press. https://doi.org/10.1017/9781009157896

Karl, T. R., Diaz, H. F., & Kukla, G. (1988). Urbanization: Its detection and effect in the United States climate record. *Journal of Climate*, *1*(11), 1099–1123.

Karl, T. R., Williams Jr, C. N., Quinlan, F. T., & Boden, T. A. (1990). *United States Historical Climatology Network (HCN) serial temperature and precipitation data* (tech. rep.). Environmental Science Division, Publication No. 3404, Oak Ridge National Laboratory.

Karl, T. R., Williams Jr, C. N., Young, P. J., & Wendland, W. M. (1986). A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States. *Journal of Climate and Applied Meteorology*, *25*(2), 145–160.

Landsberg, H. E. (1981). *The urban climate*. Academic Press.

Lawrimore, J. H., Menne, M. J., Gleason, B. E., Williams, C. N., Wuertz, D. B., Vose, R. S., & Rennie, J. (2011). An overview of the Global Historical Climatology Network monthly mean temperature dataset, version 3. *Journal of Geophysical Research: Atmospheres*, *116*(D19).

Lyon, R. (2025). USHCN Heat Island Analysis Software [Available at: https://github.com/rjl-climate/ushcn-heatisland].

Menne, M. J., & Williams Jr, C. N. (2009). Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, *22*(7), 1700–1717.

Menne, M. J., Williams Jr, C. N., & Palecki, M. A. (2010). On the reliability of the U.S. surface temperature record. *Journal of Geophysical Research: Atmospheres*, *115*(D11).

Menne, M. J., Williams Jr, C. N., & Vose, R. S. (2009). The U.S. Historical Climatology Network monthly temperature data, version 2. *Bulletin of the American Meteorological Society*, *90*(7), 993–1007.

Oke, T. R. (1973). City size and the urban heat island. *Atmospheric Environment*, *7*(8), 769–779.

Oke, T. R. (1987). *Boundary layer climates* (2nd). Methuen.

Peterson, T. C. (2006). Examination of potential biases in air temperature caused by poor station locations. *Bulletin of the American Meteorological Society*, *87*(8), 1073–1089.

Quinlan, F. T., Karl, T. R., & Williams Jr, C. N. (1987). United States Historical Climatology Network (HCN) serial temperature and precipitation data.

Spencer, R. W. (2021). Urban heat island effects on U.S. temperature trends, 1973-2020: USHCN vs. hourly weather stations [Blog post, drroyspencer.com].

Venema, V., Mestre, O., Aguilar, E., Auer, I., Guijarro, J., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., et al. (2012). Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, *8*(1), 89–115.

Vose, R. S., Williams Jr, C. N., Peterson, T. C., Karl, T. R., & Easterling, D. R. (2003). An evaluation of the time of observation bias adjustment in the U.S. Historical Climatology Network. *Geophysical Research Letters*, *30*(20).

Williams Jr, C. N., Menne, M. J., & Thorne, P. W. (2012). Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. *Journal of Geophysical Research: Atmospheres*, *117*(D5).