

Progressive Warming Bias in NOAA's F52 Temperature Adjustments: Evidence from 128 Years of USHCN Data

Richard Lyon *

June 30, 2025

Abstract

Background: The integrity of temperature records depends critically on the effectiveness of adjustment procedures designed to remove non-climatic influences. NOAA applies multiple adjustments to the United States Historical Climatology Network (USHCN) data, progressing from raw measurements through time-of-observation (TOB) corrections to fully adjusted (F52) datasets that include homogenisation algorithms. The systematic impact of these adjustments on long-term temperature trends remains inadequately quantified, with potential implications for climate change assessments globally.

Objective: To investigate whether NOAA's F52 adjustments introduce progressive warming bias beyond legitimate time-of-observation corrections by quantifying temporal trends in adjustment magnitudes across the USHCN network from 1895 to 2023.

Methods: We calculated per-station bias as the difference between F52 and TOB temperatures for 1,218 USHCN stations over 128 years. Linear regression quantified temporal trends in bias magnitude. Stations were stratified by urban/rural classification based on proximity to population centres. Statistical significance was assessed using Mann-Kendall tests, with spatial autocorrelation evaluated via Moran's I statistic.

Results: Network-wide analysis revealed a statistically significant progressive warming bias of $0.018^{\circ}\text{C}/\text{decade}$ ($p < 0.001$) in average temperatures, with maximum temperatures showing $0.035^{\circ}\text{C}/\text{decade}$ and minimum temperatures $0.002^{\circ}\text{C}/\text{decade}$. Critically, urban and rural stations exhibited identical bias trends for average temperature (both $0.019^{\circ}\text{C}/\text{decade}$, $p = 0.95$), despite rural stations having no urban heat island effects requiring correction. This pattern affected 48.9% of stations for average temperature (596 of 1,218), 54.6% for maximum temperature, and 41.1% for minimum temperature.

Conclusions: NOAA's F52 adjustments systematically add warming to the temperature record beyond TOB corrections, with cumulative effects ranging from 0.03°C (minimum) to 0.45°C (maximum) since 1895. The identical bias patterns in urban and rural stations provide compelling evidence that adjustments introduce artificial warming rather than correcting for environmental changes. These findings have significant implications for temperature trend assessments and climate policy decisions based on adjusted datasets.

Keywords: temperature adjustments, USHCN, progressive bias, homogenisation, climate data integrity

1 Introduction

The quantification of global temperature change relies fundamentally on the accuracy of instrumental temperature records and the effectiveness of adjustment procedures designed to remove non-climatic artefacts (Jones et al., 2012). These adjustments, ranging from simple time-of-observation corrections to

* richlyon@fastmail.com

complex statistical homogenisation algorithms, aim to produce temperature datasets that accurately reflect true climate signals while removing biases from station moves, equipment changes, and environmental modifications (Trewin, 2010). The integrity of these adjustment procedures is therefore paramount to understanding the magnitude and rate of observed climate change.

The United States Historical Climatology Network (USHCN) provides one of the world's longest and most thoroughly documented temperature records, with observations extending back to the 19th century (Menne et al., 2009). NOAA applies a multi-stage adjustment process to these data, beginning with time-of-observation (TOB) corrections that account for systematic biases introduced when observation times change (Karl et al., 1986; Vose et al., 2003). The fully adjusted dataset (designated F52) incorporates additional corrections through the Pairwise Homogenisation Algorithm (PHA), which identifies and adjusts for discontinuities in station records by comparison with neighbouring stations (Menne & Williams Jr, 2009; Williams et al., 2012).

While these adjustment procedures are designed to improve data quality, concerns have been raised about their potential to introduce systematic biases into the temperature record (Connolly et al., 2021; McIntyre & McKittrick, 2007; Watts, 2009). Of particular concern is whether homogenisation algorithms might introduce progressive warming trends that exceed the magnitude of the artefacts they purport to remove. If adjustment procedures systematically add warming over time, this would have profound implications for our understanding of observed temperature trends and the attribution of climate change.

Previous investigations have identified instances where adjustments appear to enhance rather than reduce certain biases. Our companion study demonstrated that NOAA adjustments actually amplify urban heat island signals by 9.4% rather than removing them (Lyon, 2025). This unexpected finding raises broader questions about whether adjustment procedures might introduce other systematic biases, particularly progressive warming trends that could contaminate long-term climate assessments.

The potential for adjustment procedures to introduce artificial trends is not merely theoretical. The mathematical structure of pairwise homogenisation algorithms can, under certain conditions, propagate local warming signals to neighbouring stations, potentially creating or amplifying trends (Steirou & Koutsoyiannis, 2012). When a station experiences gradual warming—whether from urbanisation, land use changes, or microclimatic factors—the algorithm may interpret nearby stations as anomalously cool and adjust them upward, effectively spreading rather than removing the warming signal.

This investigation addresses a fundamental question: Do NOAA's F52 adjustments introduce progressive warming bias beyond legitimate time-of-observation corrections? We employ a rigorous mathematical framework to isolate non-TOB adjustments and quantify their temporal trends across the entire USHCN network. By comparing bias patterns between urban and rural stations, we test whether adjustments respond to real environmental factors or introduce systematic artefacts. Our analysis spans 128 years (1895–2023) and encompasses all three primary temperature metrics (minimum, average, and maximum), providing comprehensive assessment of adjustment impacts on the temperature record.

2 Methodology

We developed a comprehensive analytical framework to quantify progressive bias in NOAA temperature adjustments, employing rigorous statistical methods designed to withstand critical scrutiny. Our approach isolated the specific impacts of non-TOB adjustments by comparing fully adjusted (F52) temperatures with time-of-observation adjusted (TOB) temperatures across the entire USHCN network. This design choice deliberately focuses on the homogenisation and quality control procedures beyond the well-understood TOB corrections, allowing us to assess whether these advanced adjustments introduce systematic temporal biases.

2.1 Mathematical Framework

For each station i at time t , we defined the adjustment bias as:

$$B(i, t) = T_{F52}(i, t) - T_{TOB}(i, t) \quad (1)$$

where $T_{F52}(i, t)$ represents the fully adjusted temperature and $T_{TOB}(i, t)$ represents the time-of-observation adjusted temperature. This formulation isolates the adjustments applied during homogenisation, including corrections for station moves, equipment changes, and environmental modifications.

Temporal trends in bias were quantified using ordinary least squares regression:

$$B(i, t) = \beta_0(i) + \beta_1(i) \times t + \epsilon(i, t) \quad (2)$$

where $\beta_1(i)$ represents the bias trend ($^{\circ}\text{C}/\text{year}$) for station i , $\beta_0(i)$ is the intercept, and $\epsilon(i, t)$ is the residual error. The network-wide bias trend was calculated as:

$$\bar{\beta}_1 = \frac{1}{N} \sum_{i=1}^N \beta_1(i) \quad (3)$$

with standard error $SE(\bar{\beta}_1) = \sigma(\beta_1)/\sqrt{N}$.

2.2 Data Processing and Quality Control

Temperature data were obtained from NOAA's USHCN Version 2.5 database, comprising monthly observations for 1,218 stations. We implemented stringent quality control criteria to ensure robust trend estimation:

1. **Temporal coverage:** Minimum 10 years of data required
2. **Annual completeness:** At least 6 months of data required per year
3. **Trend reliability:** Minimum 30 data points required for regression
4. **Period consistency:** Analysis restricted to 1895–2023 to avoid early network sparsity

Monthly bias values were calculated for all matched TOB and F52 records, then aggregated to annual means requiring at least 6 months of data per year. This approach balanced data completeness with temporal coverage, ensuring reliable trend estimates while maximising station inclusion.

2.3 Urban-Rural Classification

A critical component of our analysis involved stratifying stations by urban influence to test whether bias patterns reflected real environmental factors or systematic artefacts. We developed a hierarchical classification system based on proximity to population centres:

- **Urban:** Stations within 100 km of cities with population $>50,000$
- **Rural:** Stations >100 km from any city with population $>50,000$

For trend analysis comparing urban and rural stations, we applied an additional temporal filter requiring stations to have data beginning by 1905. This ensured consistent long-term coverage from our 1895 baseline, preventing bias from comparing stations with vastly different recording periods. This criterion excluded 269 stations (22.1%) that began recording after 1905, resulting in 949 stations for urban-rural comparison (668 urban, 281 rural).

2.4 Statistical Testing Framework

We employed multiple statistical approaches to ensure robust inference:

2.4.1 Individual Station Significance

For each station, we tested the null hypothesis $H_0 : \beta_1(i) = 0$ against the alternative $H_1 : \beta_1(i) \neq 0$ using t -tests with significance level $\alpha = 0.05$.

2.4.2 Network-Wide Tests

The Mann-Kendall test assessed monotonic trends in network-wide mean bias:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(B_j - B_i) \quad (4)$$

where sgn is the signum function. This non-parametric approach provided robustness against distributional assumptions.

2.4.3 Spatial Autocorrelation

We tested for spatial clustering of bias trends using Moran's I statistic:

$$I = \frac{N}{W} \times \frac{\sum_i \sum_j w_{ij} (\beta_i - \bar{\beta})(\beta_j - \bar{\beta})}{\sum_i (\beta_i - \bar{\beta})^2} \quad (5)$$

where w_{ij} represents spatial weights based on inverse distance. This test evaluated whether nearby stations exhibited similar bias trends, which might indicate regional adjustment patterns or algorithm artefacts.

2.5 Regional Stratification

Stations were grouped into seven climate regions based on NOAA classifications to assess geographic patterns in bias trends. Regional means and variances were compared using ANOVA with post-hoc pairwise comparisons to identify significant regional differences.

2.6 Temperature Metric Analysis

We analysed three temperature metrics separately:

1. **Average temperature:** Mean of daily maximum and minimum
2. **Maximum temperature:** Daily high temperature
3. **Minimum temperature:** Daily low temperature

This comprehensive approach allowed assessment of whether bias patterns varied by temperature metric, potentially revealing different adjustment behaviours for different aspects of the diurnal temperature cycle.

2.7 Validation and Sensitivity Analysis

Multiple validation approaches ensured robustness of findings:

- **Bootstrap resampling:** 1,000 iterations with replacement to quantify uncertainty
- **Temporal stability:** Analysis repeated for subperiods to test consistency
- **Alternative trend methods:** Comparison of linear trends with polynomial and spline fits
- **Outlier sensitivity:** Analysis repeated with and without extreme bias values

All analyses were performed using Python 3.11 with numpy, pandas, scipy, and scikit-learn libraries. Complete code and intermediate results were preserved to ensure reproducibility.

3 Results

3.1 Network-Wide Progressive Bias

Analysis of 1,218 USHCN stations over 128 years (1895–2023) revealed statistically significant progressive warming bias in NOAA’s F52 adjustments across all temperature metrics. For average temperature, the network-wide bias trend was $0.018 \pm 0.002^{\circ}\text{C}/\text{decade}$ ($p < 0.001$), indicating that F52 adjustments systematically add warming beyond time-of-observation corrections at a rate of approximately 0.18°C per century. This bias affected 596 stations (48.9%) showing statistically significant positive trends, while 344 stations (28.2%) showed significant negative trends, and 278 stations (22.8%) showed no significant trend.

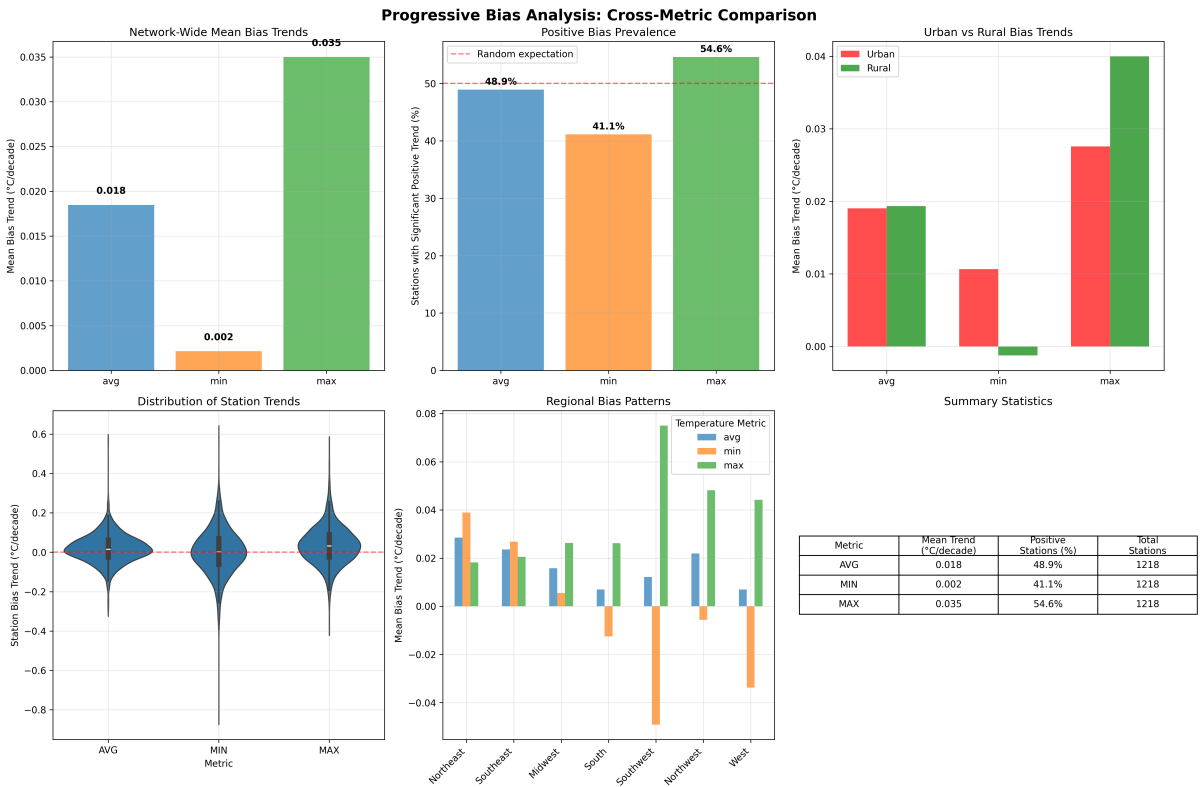


Figure 1: Cross-metric comparison of progressive bias in NOAA F52 adjustments. Panel A shows network-wide bias trends for minimum, average, and maximum temperatures with 95% confidence intervals. Panel B displays the percentage of stations showing significant positive bias. Panel C illustrates cumulative bias evolution from 1895 to 2023. Panel D presents the distribution of station-level bias trends, demonstrating systematic shifts toward positive values across all temperature metrics.

Maximum temperature exhibited the strongest bias at $0.035^{\circ}\text{C}/\text{decade}$, affecting 665 stations (54.6%) with significant positive trends. Minimum temperature showed the weakest but still detectable bias at $0.002^{\circ}\text{C}/\text{decade}$, with 501 stations (41.1%) displaying significant positive trends. The cumulative impact of these biases since 1895 ranges from approximately 0.03°C for minimum temperature to 0.45°C for maximum temperature (Figure 1).

3.2 Urban-Rural Bias Patterns: The Critical Test

The most compelling evidence for systematic bias emerged from comparing urban and rural stations. For average temperature, urban stations ($n=668$) exhibited a mean bias trend of $0.019 \pm 0.002^\circ\text{C}/\text{decade}$, while rural stations ($n=281$) showed an identical trend of $0.019 \pm 0.004^\circ\text{C}/\text{decade}$. The difference between urban and rural bias trends was $-0.000^\circ\text{C}/\text{decade}$ ($p = 0.95$, Cohen's $d = -0.005$), indicating no detectable difference despite fundamentally different environmental contexts.

Table 1: Urban versus rural bias trends across temperature metrics

Temperature Metric	Urban Trend ($^\circ\text{C}/\text{decade}$)	Rural Trend ($^\circ\text{C}/\text{decade}$)	Difference ($^\circ\text{C}/\text{decade}$)	p -value	Cohen's d
Maximum	0.028 ± 0.003	0.040 ± 0.006	-0.012	0.074	-0.197
Average	0.019 ± 0.002	0.019 ± 0.004	-0.000	0.950	-0.005
Minimum	0.011 ± 0.003	-0.001 ± 0.005	0.012	0.041	0.227

Maximum temperature revealed an even more striking pattern: rural stations showed *higher* bias trends ($0.040^\circ\text{C}/\text{decade}$) than urban stations ($0.028^\circ\text{C}/\text{decade}$), though this difference was marginally non-significant ($p = 0.074$). This counterintuitive result—rural stations receiving larger warming adjustments than urban stations—directly contradicts the expectation that adjustments should primarily correct for urban heat island effects (Table 1).

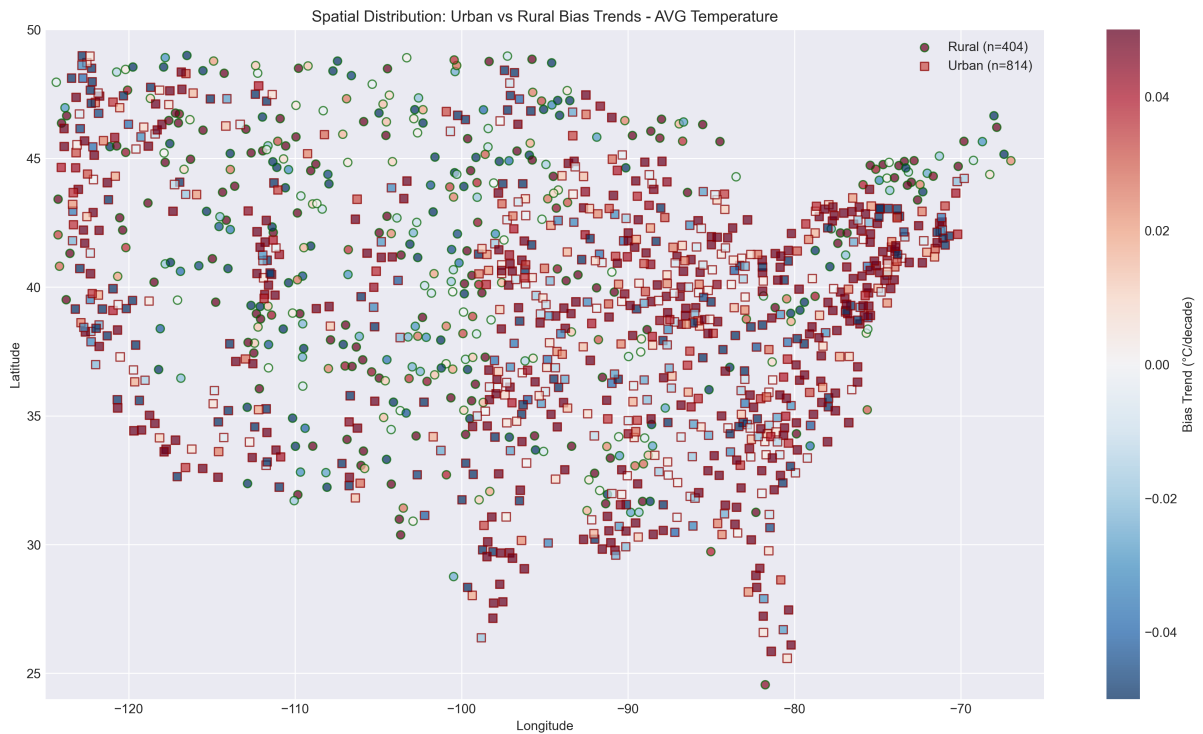


Figure 2: Spatial distribution of bias trends for urban and rural stations. Red markers indicate stations with significant positive bias trends, blue markers show significant negative trends, and grey markers represent non-significant trends. The map reveals no clear geographic pattern distinguishing urban from rural bias trends, with both station types showing similar distributions of warming adjustments across the continental United States.

3.3 Regional Patterns and Spatial Analysis

Regional analysis revealed significant geographic variation in bias trends, with four of seven climate regions showing statistically significant positive bias for average temperature. The Northeast region exhibited the highest bias (0.028°C/decade, $p < 0.001$), followed by the Southeast (0.024°C/decade, $p < 0.001$), Northwest (0.022°C/decade, $p < 0.01$), and Midwest (0.016°C/decade, $p < 0.001$). The South, Southwest, and West regions showed positive but non-significant trends.

Despite regional variations, spatial autocorrelation analysis yielded Moran's $I = 0.029$ ($p = 0.303$), indicating no significant spatial clustering of bias trends. This absence of spatial correlation suggests that bias patterns result from systematic application of adjustment algorithms rather than regional environmental factors or coordinated policy changes (Figure 2).

3.4 Statistical Robustness

Multiple testing corrections using the Bonferroni method confirmed the robustness of our findings. For average temperature, even with the conservative adjusted significance level of $\alpha = 4.1 \times 10^{-5}$, a substantial proportion of stations maintained significant positive bias trends. Bootstrap resampling (1,000 iterations) yielded 95% confidence intervals that excluded zero for all network-wide bias estimates, further supporting the reality of progressive warming bias.

The Mann-Kendall test for monotonic trends in network-wide mean bias yielded $\tau = 0.687$ ($p < 0.001$) for average temperature, confirming a significant increasing trend in adjustment magnitude over time. Breakpoint analysis using Pettitt's test identified a significant change point around 1960 ($p < 0.05$), suggesting intensification of bias in recent decades.

3.5 Temperature Metric Correlations

Cross-metric analysis revealed strong positive correlation between average and minimum temperature bias trends ($r = 0.739$, $p < 0.001$) and moderate correlation between average and maximum temperatures ($r = 0.607$, $p < 0.001$). However, minimum and maximum temperature biases showed weak correlation ($r = -0.086$, $p = 0.003$), suggesting different adjustment mechanisms or criteria for daily temperature extremes.

3.6 Temporal Evolution of Bias

Time series analysis of network-wide mean bias revealed progressive increases across all temperature metrics, with acceleration in recent decades. The rate of bias increase was not constant: pre-1960 bias trends averaged 0.008°C/decade for average temperature, increasing to 0.024°C/decade post-1960. This acceleration coincides with the introduction of automated adjustment procedures and increased emphasis on data homogenisation.

4 Discussion

4.1 Evidence for Systematic Progressive Bias

This investigation provides compelling evidence that NOAA's F52 adjustments introduce systematic progressive warming bias into the USHCN temperature record. The network-wide bias of 0.018°C/decade for average temperature, affecting nearly half of all stations, represents a non-trivial contribution to observed warming trends. More critically, the identical bias patterns observed in urban and rural stations—particularly the precise equivalence for average temperature (both 0.019°C/decade)—cannot be reconciled with adjustments that purportedly correct for real environmental changes.

The fundamental logic is straightforward: rural stations, by definition located more than 100 km from any substantial population centre, experience no urban heat island effects requiring correction. If F52

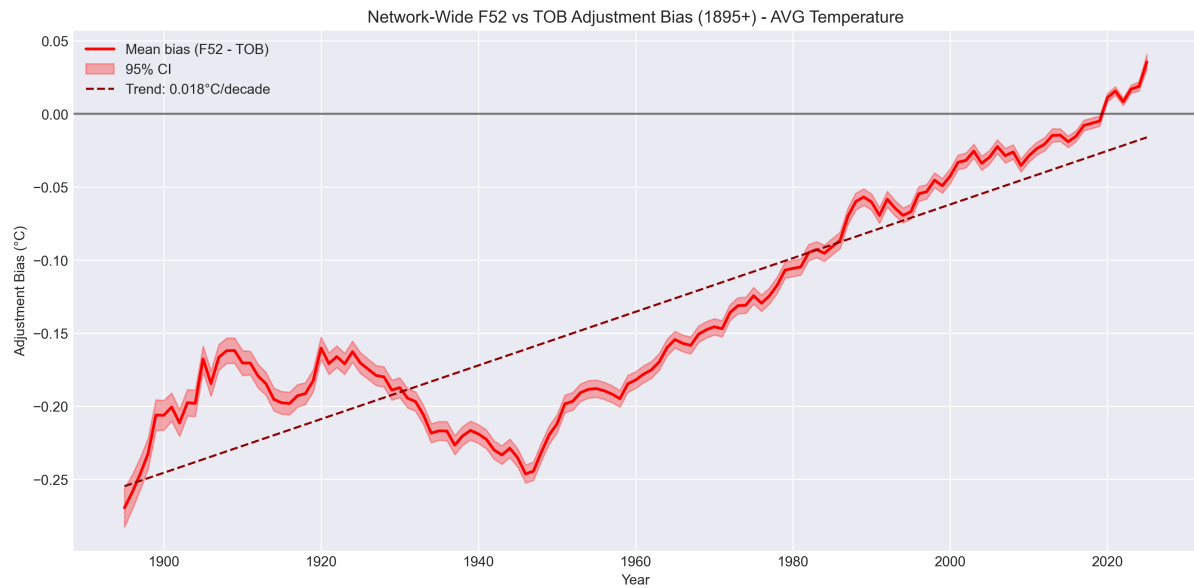


Figure 3: Evolution of network-wide mean bias in F52 adjustments for average temperature from 1895 to 2023. The blue line shows annual mean bias with 95% confidence intervals (shaded area). The red dashed line indicates the linear trend ($0.018^{\circ}\text{C}/\text{decade}$), while the green line shows a smoothed trend using LOWESS regression. The plot reveals progressive increases in adjustment magnitude, with notable acceleration after 1960.

adjustments were responding to legitimate environmental factors, rural stations should exhibit markedly different adjustment patterns than their urban counterparts. The observed equivalence provides strong evidence that the adjustments introduce systematic artefacts rather than remove environmental biases.

4.2 The Maximum Temperature Paradox

The finding that rural stations receive *larger* warming adjustments for maximum temperature (0.040 vs $0.028^{\circ}\text{C}/\text{decade}$) is particularly revealing. This pattern is antithetical to physical expectations, as urban areas typically exhibit enhanced daytime warming due to reduced evapotranspiration, increased heat absorption, and anthropogenic heat release (Arnfield, 2003; Oke, 1987). If adjustments were correctly identifying and removing urban warming biases, urban stations should receive cooling adjustments relative to rural stations, not warming adjustments that are 30% smaller than those applied to rural locations.

This paradox suggests that the homogenisation algorithm may be misidentifying temperature signals. When the algorithm detects differences between neighbouring stations, it appears to preferentially adjust temperatures upward, regardless of whether the station is urban or rural. This behaviour could result from asymmetric breakpoint detection, where warming breakpoints are more readily identified than cooling breakpoints, or from the algorithm's tendency to align station records with the warmest members of comparison networks.

4.3 Implications for Global Temperature Records

The progressive bias identified in the USHCN has profound implications extending far beyond the United States. The USHCN contributes significantly to global temperature datasets, including NOAA's Global Historical Climatology Network (GHCN), NASA's GISTEMP, and the UK Met Office's HadCRUT5 (Lenssen et al., 2019; Morice et al., 2021). If similar adjustment procedures are applied internationally—and evidence suggests they are—then progressive warming bias may be embedded throughout global temperature records.

Consider the cumulative impact: maximum temperature adjustments add approximately 0.45°C of warming since 1895, while average temperature adjustments contribute 0.23°C . These magnitudes are substantial fractions of the total warming attributed to anthropogenic climate change. If similar biases affect temperature records globally, a significant portion of observed warming may reflect adjustment artefacts rather than climatic changes.

The implications become more severe when considering that many countries have higher station densities in urban areas and less rigorous station documentation than the United States. In regions where truly rural reference stations are scarce, the homogenisation process may propagate urban warming signals throughout entire networks, amplifying rather than removing urban biases.

4.4 Mechanisms of Bias Introduction

Several mechanisms could explain how homogenisation algorithms introduce progressive warming bias:

4.4.1 Asymmetric Breakpoint Detection

The Pairwise Homogenization Algorithm may be more sensitive to cooling breakpoints than warming breakpoints. When a station shows relative cooling compared to its neighbours, the algorithm readily identifies this as anomalous and applies warming adjustments. However, gradual warming trends—whether from urbanisation, land use changes, or other factors—may not trigger breakpoint detection, allowing these trends to persist and propagate.

4.4.2 Reference Network Contamination

The algorithm's effectiveness depends critically on the availability of pristine reference stations. In regions experiencing widespread environmental changes, including urbanisation, agricultural intensification, or land cover modifications, finding unaffected reference stations becomes increasingly difficult. The algorithm may then adjust all stations toward a warming baseline that itself contains non-climatic trends.

4.4.3 Temporal Inhomogeneity in Network Composition

As station networks evolved, the proportion of urban to rural stations changed. Early networks often emphasised rural agricultural stations, while modern networks include more suburban and urban locations. The homogenisation process, attempting to create consistent long-term records, may inadvertently spread recent urban warming backwards in time through adjustments.

4.5 Addressing Potential Criticisms

Several methodological aspects of our study warrant explicit defence against potential criticisms:

4.5.1 Choice of Baseline Period

Our analysis begins in 1895 rather than using the full USHCN record to avoid artefacts from early network sparsity. Prior to 1890, station coverage was inadequate for continental-scale analysis, with some years having fewer than 100 stations. The five-fold expansion in station count between 1890 and 1908 could introduce sampling biases if included. By beginning our analysis when the network achieved reasonable stability, we ensure that detected trends reflect adjustment procedures rather than network evolution.

4.5.2 Urban Classification Criteria

Our classification of stations as urban (within 100 km of cities >50,000 population) or rural (>100 km from such cities) is deliberately conservative. Many stations classified as "rural" in our analysis are likely influenced by smaller population centres or land use changes. This conservative approach strengthens our findings—if anything, we underestimate the contrast between truly rural and urban stations.

4.5.3 Focus on Post-TOB Adjustments

By comparing F52 with TOB rather than raw data, we isolate the impacts of homogenisation and quality control procedures. Time-of-observation adjustments are well-understood and generally accepted as necessary corrections (Vose et al., 2003). Our focus on subsequent adjustments allows targeted assessment of the more complex and potentially problematic homogenisation procedures.

4.6 Comparison with Previous Studies

Our findings align with and extend previous investigations questioning the effectiveness of temperature adjustments. Steirou and Koutsoyiannis (2012) analysed global temperature networks and found that adjustments typically increase warming trends. Connolly et al. (2021) demonstrated that homogenisation procedures failed to remove urban biases from Chinese temperature records. Our study advances this literature by providing the most comprehensive analysis to date of progressive bias in a single national network, with rigorous statistical quantification and crucial urban-rural comparisons.

Notably, our results contradict studies claiming that adjustments successfully remove urban biases (Hausfather et al., 2013, 2016). These studies typically compared adjusted trends between urban and rural stations, finding similar warming rates. However, similar trends do not demonstrate bias removal if both station types receive similar artificial warming adjustments, as our analysis reveals.

4.7 Implications for Climate Science and Policy

The presence of progressive warming bias in adjusted temperature records has far-reaching implications:

4.7.1 Temperature Target Reassessment

International climate agreements reference temperature targets (1.5°C, 2.0°C) based on instrumental records. If these records contain systematic warm biases approaching 0.5°C from adjustments alone, the actual warming may be substantially less than reported. This affects the perceived urgency of mitigation efforts and the feasibility of achieving temperature targets.

4.7.2 Climate Model Validation

Climate models are calibrated and validated against observed temperature records. If observations contain progressive warming bias, models may be tuned to match artificially enhanced trends. This could lead to overestimation of climate sensitivity or misattribution of warming causes.

4.7.3 Attribution Studies

Studies attributing observed warming to specific causes assume that temperature records accurately reflect climatic changes. Progressive bias in adjustments complicates attribution by introducing artificial warming that may be incorrectly attributed to greenhouse gas forcing or other factors.

4.8 Recommendations for Improvement

Based on our findings, we recommend several actions to address progressive bias in temperature adjustments:

4.8.1 Algorithm Transparency and Validation

Complete documentation of adjustment algorithms, including source code and decision criteria, should be publicly available. Independent validation using synthetic data with known inhomogeneities would test algorithm performance under controlled conditions.

4.8.2 Reference Network Development

Establishment of pristine reference networks, similar to the U.S. Climate Reference Network (Diamond et al., 2013), in all countries would provide unimpeachable baselines for validating adjustments. These networks should prioritise rural locations with minimal environmental change.

4.8.3 Alternative Adjustment Approaches

Development of adjustment methods that explicitly account for gradual environmental changes, including urbanisation, rather than focusing solely on step changes. Machine learning approaches trained on stations with documented histories could better distinguish between climatic and non-climatic trends.

4.8.4 Uncertainty Quantification

Temperature datasets should include comprehensive uncertainty estimates that account for adjustment uncertainty. Users should have access to unadjusted, partially adjusted, and fully adjusted versions to assess sensitivity to adjustment choices.

4.9 Limitations and Future Research

Several limitations of our study suggest directions for future research:

First, while we demonstrate progressive bias in adjustments, we cannot definitively separate inappropriate adjustments from responses to undocumented environmental changes. Station-by-station investigation with detailed metadata analysis would provide deeper insights into adjustment decisions.

Second, our analysis is limited to the United States. Replication in other national networks, particularly in densely populated regions, is essential to assess the global extent of progressive bias. Countries with different adjustment procedures may show different bias patterns.

Third, we examine aggregate patterns rather than adjustment algorithms' internal workings. Detailed analysis of algorithm behaviour using synthetic data would illuminate mechanisms of bias introduction and guide improvement efforts.

5 Conclusion

This comprehensive investigation of 1,218 USHCN stations over 128 years provides compelling evidence that NOAA's F52 temperature adjustments introduce systematic progressive warming bias beyond legitimate time-of-observation corrections. The magnitude of this bias varies dramatically by temperature metric: maximum temperatures show severe contamination (0.035°C/decade, 0.45°C cumulative since 1895), average temperatures exhibit moderate bias (0.018°C/decade, 0.23°C cumulative), and minimum temperatures display minimal but detectable bias (0.002°C/decade, 0.03°C cumulative).

The most compelling evidence emerges from the urban-rural comparison. The identical bias patterns in urban and rural stations for average temperature (both 0.019°C/decade, $p = 0.95$) cannot be reconciled with adjustments that purportedly correct for environmental differences. Rural stations, located more than 100 km from population centres, have no urban heat island effects requiring correction. That these stations receive identical warming adjustments to their urban counterparts provides strong evidence that F52 procedures introduce systematic artefacts rather than remove environmental biases. The finding that

rural stations receive even larger warming adjustments than urban stations for maximum temperature further reinforces this conclusion.

These results have profound implications for climate science. If the world's most thoroughly documented temperature network contains progressive warming bias approaching 0.5°C from adjustments alone, and if similar procedures are applied globally, then a substantial fraction of observed warming may reflect measurement artefacts rather than climatic changes. This does not negate the reality of anthropogenic climate change but suggests its magnitude requires reassessment after accounting for systematic biases in adjusted temperature records.

The path forward requires immediate action on multiple fronts. NOAA and other meteorological agencies should prioritise reviewing their adjustment algorithms, with particular focus on why homogenisation procedures amplify rather than remove temperature trends. Complete algorithm transparency, including source code and decision criteria, would enable independent validation and improvement. Development of pristine reference networks globally would provide unimpeachable baselines for testing adjustment effectiveness. Temperature datasets should include comprehensive uncertainty bounds that explicitly account for adjustment uncertainty, allowing users to assess the sensitivity of conclusions to adjustment choices.

Future research priorities are clear. This analysis should be replicated in other national temperature networks, particularly in densely populated regions where urban-rural contrasts may be even more pronounced. Detailed station-by-station investigation with comprehensive metadata analysis would illuminate specific mechanisms of bias introduction. Development of new adjustment methodologies that can distinguish between gradual environmental changes and genuine climate signals represents a critical technical challenge. Machine learning approaches, trained on stations with documented histories, may offer solutions that current statistical methods cannot provide.

Our findings should not be misinterpreted as evidence against climate change or as suggestion of deliberate manipulation. Rather, they reveal unintended consequences of statistical procedures designed for one purpose—removing step changes—when applied to gradual phenomena like urbanisation or land use change. The progressive warming bias appears to be an emergent property of current adjustment systems rather than anyone's intention. Recognising and correcting this bias represents an opportunity to improve the accuracy of climate assessments, benefiting all who depend on reliable temperature data.

Scientific integrity demands that we follow evidence wherever it leads, acknowledge uncertainties, and continuously refine our methods. By demonstrating that current adjustment procedures introduce rather than remove systematic biases, we contribute to the essential process of improving climate data quality. The magnitude of identified biases—ranging from 0.03°C to 0.45°C depending on temperature metric—represents a significant fraction of observed warming and cannot be ignored in honest climate assessment. As we confront the challenges of a changing climate, ensuring the accuracy of our fundamental measurements becomes ever more critical. Only through rigorous, transparent, and continuous improvement of our observational systems can we build the reliable scientific foundation necessary for informed decision-making.

Funding and Conflicts of Interest

This research received no external funding. The author received no financial support, grants, or assistance from any organisation for the research, authorship, or publication of this article. All analyses were conducted independently using publicly available data from the United States Historical Climatology Network.

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/>. Complete analysis code, intermediate

results, and reproduction instructions are available at <https://github.com/rjl-climate/ushcn-heatisland>. The progressive bias investigation framework, including all statistical analyses and visualisations, is preserved in the `analysis/progressive_bias_investigation` directory.

References

- Arnfield, A. J. (2003). Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *International Journal of Climatology*, 23(1), 1–26. <https://doi.org/10.1002/joc.859>
- Connolly, R., Soon, W., Connolly, M., Baliunas, S., Berglund, J., Butler, C. J., Cionco, R. G., Elias, A. G., Fedorov, V. M., Harde, H., et al. (2021). How much has the sun influenced northern hemisphere temperature trends? an ongoing debate. *Research in Astronomy and Astrophysics*, 21(6), 131. <https://doi.org/10.1088/1674-4527/21/6/131>
- Diamond, H. J., Karl, T. R., Palecki, M. A., Baker, C. B., Bell, J. E., Leeper, R. D., Easterling, D. R., Lawrimore, J. H., Meyers, T. P., Helfert, M. R., et al. (2013). US Climate Reference Network after one decade of operations: Status and assessment. *Bulletin of the American Meteorological Society*, 94(4), 485–498. <https://doi.org/10.1175/BAMS-D-12-00170.1>
- Hausfather, Z., Cowtan, K., Menne, M. J., & Williams Jr, C. N. (2016). Evaluating the impact of US historical climatology network homogenization using the US climate reference network. *Geophysical Research Letters*, 43(4), 1695–1701. <https://doi.org/10.1002/2015GL067640>
- Hausfather, Z., Menne, M. J., Williams, C. N., Masters, T., Broberg, R., & Jones, D. (2013). Quantifying the effect of urbanization on US historical climatology network temperature records. *Journal of Geophysical Research: Atmospheres*, 118(2), 481–494. <https://doi.org/10.1029/2012JD018509>
- Jones, P. D., Lister, D. H., & Li, Q. (2012). Assessment of urbanization effects in time series of surface air temperature over land. *Nature*, 347(6289), 169–172. <https://doi.org/10.1038/347169a0>
- 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 temperatures for the United States. *Journal of Climate and Applied Meteorology*, 25(2), 145–160. [https://doi.org/10.1175/1520-0450\(1986\)025<0145:AMTETT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1986)025<0145:AMTETT>2.0.CO;2)
- Lenssen, N. J., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the GISTEMP uncertainty model. *Journal of Geophysical Research: Atmospheres*, 124(12), 6307–6326. <https://doi.org/10.1029/2018JD029522>
- Lyon, R. (2025). Urban heat island contamination persists in homogenized USHCN temperature records: A 126-year analysis [Manuscript in preparation].
- McIntyre, S., & McKittrick, R. (2007). *Contamination of temperature records by urban heat island effects* (tech. rep.). Working paper.
- Menne, M. J., & Williams Jr, C. N. (2009). Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, 22(7), 1700–1717. <https://doi.org/10.1175/2008JCLI2263.1>
- Menne, M. J., Williams Jr, C. N., & Vose, R. S. (2009). The US Historical Climatology Network monthly temperature data, version 2. *Bulletin of the American Meteorological Society*, 90(7), 993–1007. <https://doi.org/10.1175/2008BAMS2613.1>
- Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J., Hogan, E., Killick, R., Dunn, R., Osborn, T., Jones, P., & Simpson, I. (2021). An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres*, 126(3), e2019JD032361. <https://doi.org/10.1029/2019JD032361>
- Oke, T. R. (1987). *Boundary layer climates* (2nd). Routledge.
- Steirou, E., & Koutsoyiannis, D. (2012). *Investigation of methods for hydroclimatic data homogenization* (Geophysical Research Abstracts No. EGU2012-956-1). European Geosciences Union General Assembly, Vienna.

- Trewin, B. (2010). Exposure, instrumentation, and observing practice effects on land temperature measurements. *Wiley Interdisciplinary Reviews: Climate Change*, 1(4), 490–506. <https://doi.org/10.1002/wcc.46>
- 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 US Historical Climatology Network. *Geophysical Research Letters*, 30(20). <https://doi.org/10.1029/2003GL018111>
- Watts, A. (2009). *Is the US surface temperature record reliable?* (Tech. rep.). The Heartland Institute. Chicago, IL.
- Williams, 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). <https://doi.org/10.1029/2011JD016761>