Biases due to widespread use of low-cost sensors for urban heat stress assessments

Sarah Berk,[a], TC Chakraborty[b], Angel Hsu[a]

[a] *University of North Carolina at Chapel Hill, Chapel Hill, NC*

[b] *Pacific Northwest National Laboratory, Richland, WA*

Corresponding authors: TC Chakraborty, tc.chakraborty@pnnl.gov, Angel Hsu, angel.hsu@unc.edu

Early Online Release: This preliminary version has been accepted for publication in *Bulletin of the American Meteorological Society*, may be fully cited, and has been assigned DOI 10.1175/BAMS-D-25-0222.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT

Urban heat stress is an area of critical research interest due to its relevance for public health and policy. Given the lack of operational-grade weather stations within cities, different types of low-cost sensors have been used to assess urban heat stress. However, these sensors are traditionally not designed for capturing fine scale differences in temperature and humidity (i.e., the main components of moist heat), which are expected across a city, and their siting is often suboptimal due to logistical challenges. Here, through several exercises over lawn and rooftop settings using 41 Kestrel sensors, we demonstrate significant issues with the use of such low-cost sensors for urban heat stress assessments. Issues stem from use of sensors without radiation shields, and even with cheaper non-aspirated shields, exposure to confounding environmental factors, and local land cover influences. Unshielded sensors overestimated temperature by up to 0.7 °C relative to shielded counterparts. For sensors housed in unaspirated radiation shields, daytime measurement error was strongly correlated with diurnal temperature range (lawn $r = 0.86$; unshaded rooftop $r = 0.97$). Humidity cross-sensor variability exhibited a weaker correlation with its diurnal range. Heat index, derived from temperature and humidity, combined biases in the two measurements, magnifying resulting errors (14% higher than standard deviation for temperature). Finally, sensors near anthropogenic heat sources showed cross-sensor variability up to four (temperature) and five (humidity) times higher than interference free sites. Based on these potential mismeasurements in urban heat stress gradients seen using inexpensive sensors, we provide recommendations on their appropriate deployment in urban environments.

SIGNIFICANCE STATEMENT

Heat stress is a major concern in cities, where temperatures are often higher than in nearby rural areas. To identify which neighborhoods face the greatest risks, communities and researchers have increasingly turned to low-cost sensors to measure temperature and humidity. Our study shows that these low-cost sensors must be used with caution when comparing conditions across a city, especially in hot and humid conditions. Inaccurate readings risk muting or magnifying the true extent of heat exposure and misguiding resilience planning. We highlight the importance of proper shielding and placement of sensors, and underscore the need for stricter standards to ensure reliable location-based comparisons of temperature and humidity.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

CAPSULE (BAMS ONLY)

Low-cost sensors are widely used to monitor urban heat, but shielding, siting, and accounting for local interferences are critical for robust intra-city comparisons.

1. Introduction

Global climate change and rapid urbanization has made addressing extreme heat a priority in cities, which frequently experience elevated air temperatures (T_a) in relation to their rural references, known as the urban heat island (UHI) effect (Howard 1833; Mohajerani et al. 2017). Urbanization often also reduces relative humidity (RH), producing the urban dry island (UDI) effect. However, this pattern is not universal; in arid and semi-arid regions, irrigation and differences in land surface properties can instead make urban areas cooler and more humid than their surroundings, known as the "oasis effect" (Fan et al. 2017; Georgescu et al. 2011). Together these processes modulate moist heat, a function of both temperature and humidity, within cities (Chakraborty et al. 2022; Meili et al. 2022). Across many regions of the United States, humidity can amplify heat stress burden over what temperature alone would suggest (Narayanan et al. 2025), as high humidity limits evaporative cooling through sweat (McGregor and Vanos 2018).

Impervious surfaces reduce evapotranspiration, while the thermal properties of common urban materials such as concrete and asphalt result in increased heat storage and release (Oke 1982; Oke et al. 2017). Additional heat inputs from anthropogenic sources, including vehicles, buildings, and other human activities, can further increase sensible heat flux within urban environments (Allen et al. 2011; Oke et al. 2017). Urban surface geometry, including street canyons, can provide shading (Danniswari et al. 2022; Hamdi and Schayes 2008; Wang and Xu 2021) but also trap heat, restrict ventilation and slow nighttime cooling (Oke, 1981; Oke et al., 1991; WMO, 2023a). Together, these factors alter the local radiation balance, turbulence, humidity and heat storage within cities. This heat hazard is not uniform across a city, and varies significantly due to local differences in land cover, building morphology, anthropogenic heat and surface materials (WMO, 2023a). Consequently, given the high population densities of urban centers, the health risks associated with moist heat can have disproportionately higher impacts in cities and even across neighborhoods.

Accurately estimating intra-urban patterns of T_a and RH is therefore essential for comparing physiologically relevant heat hazard across different parts of a city (Chakraborty

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

et al. 2023; Hsu et al. 2021). Furthermore, assessing spatial variability in heat exposure at policy-relevant scales is needed for guiding targeted mitigation strategies (Chakraborty et al. 2022) and informing climate resilience (Martilli et al. 2020). Conventional meteorological datasets, such as reanalysis products (e.g., ERA5; Hersbach et al., 2020) are too coarse to capture these variations (Li et al. 2024; Nogueira et al. 2022), and typically do not resolve urban-scale signals (Chen et al. 2024). Satellite-sensed land surface temperature is available at higher resolution (70 m to 1 km) but is less relevant for human health and comfort than T_a and moist heat stress indices (Chakraborty et al. 2022; Venter et al. 2021).

While T_a and RH are common meteorological variables, standard weather stations are rarely located within the urban environment (Li et al. 2023; Muller et al. 2013). Weather stations used to represent urban conditions are frequently located at airports, providing measurements from a single site that may not reflect conditions within the urban core (Li et al. 2023) and across neighborhoods. These stations are also required to follow strict siting guidelines from the World Meteorological Organization (WMO), which recommends placing sensors over short grass on flat, unobstructed terrain, and positioned away from trees, buildings, or other structures that might shield sensors from full exposure to sunlight and wind (WMO, 2023b). While this setup ensures consistency in climate-scale measurements, these requirements are difficult to fulfill over urban environments, which are dominated by impervious surfaces and densely packed buildings, making it challenging to find unobstructed sites for standard sensor siting. More importantly, some of these siting requirements do not necessarily make sense for examining human-centric exposure and impacts at the local scale.

To address these limitations, researchers increasingly are turning to sensor networks to monitor urban microclimate (Muller et al. 2013). Since standard meteorological sensors can be expensive, and multiple sensors are needed to capture the heterogeneity of urban environments, low-cost, consumer grade sensors are often utilized. Various types of low-cost sensors have been used in mobile campaigns (via cars, bikes or on foot) (Shandas et al. 2019; Shi et al. 2021; Wang et al. 2023), stationary sensor networks (Chapman et al. 2015; Fenner et al. 2019; Shi et al. 2021), or in crowdsourced citizen weather station data (Chakraborty et al. 2022; Fenner et al. 2019; Venter et al. 2021; Zumwald et al. 2021).

To enable meaningful comparison both across and within urban environments, however, robust and established methodologies are essential (Stewart 2011). Outdoor deployments introduce a range of challenges not captured in standard testing environments, yet most T_a

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

sensor calibrations are conducted under controlled conditions (Abdinoor et al. 2025). To address this challenge, studies carry out evaluation of low-cost sensor accuracy by comparison to a research grade reference sensor, with a primary focus on T_a accuracy (Ahmed and Jänicke 2025; Bell et al. 2015; Fenner et al. 2021; Gubler et al. 2021; Meier et al. 2017). In such approaches, low-cost sensors are typically calibrated against the reference sensor, modelling the relationships between the low-cost and research grade sensors (Abdinoor et al. 2025; Chakraborty et al. 2017; Couzo et al. 2024; Nan et al. 2025).

Low-cost sensors are commonly deployed with naturally ventilated radiation shields to reduce radiative heating. This setup reduces costs and avoids the need for power at deployment sites. However, even when shielded, solar radiation - and to a lesser extent longwave radiation - can result in T_a measurement bias across sensors, particularly under conditions of strong solar radiation and low wind speeds (Abdinoor et al. 2025; Bell et al. 2015; Gubler et al. 2021; Jenkins 2014; Yamamoto et al. 2017; Young et al. 2014). Some calibration approaches incorporate radiation measurements as explanatory variables to correct these biases. However, this requires radiation measurements to be taken at the same location as the temperature observations (Mauder et al. 2008; Yamamoto et al. 2017). Because radiation sensors are comparatively expensive, such approaches are often impractical for large low-cost sensor networks.

Biases in RH measurement can also arise due to their dependence on T_a observations. RH sensors (hygrometers) measure the amount of water vapor in the air and derive RH using the observed T_a, meaning biases in T_a can propagate into RH calculations, in addition to other potential sources of error. Bell et al (2015) evaluated a number of low-cost weather stations and found bias varied by instrument, with some sensors exhibiting persistent wet or dry biases, while others showed wet bias at high RH and dry bias at low RH. Additionally, RH sensors may experience calibration drift over time (Bell et al. 2015; da Cunha 2015).

A further challenge in evaluating sensor performance is the presence of microclimate variability, which can produce differences in temperature observations even over short distances. In addition to radiative heating, natural convection due to buoyancy forces or forced convection due to windspeeds can be influenced by localized variation (Mauder et al. 2008). Mauder et al (2008) examined variability among 25 sensors in an agricultural setting over a two-day period and reported RMSE differences of 0.05-0.14 °C relative to a randomly

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

selected sensor within the group, despite relatively uniform environmental conditions and a small diurnal temperature range (9.5 to 16 °C).

However, comparable studies examining measurement variability within dense urban environments remain limited. In urban areas, land cover, shading, wind disruption, and anthropogenic heat amplify these effects, making localized variability a key concern for sensor networks that prioritize spatial coverage. Previous studies have shown that sensor biases can differ between sites with different environmental characteristics (Gubler et al. 2021; Nan et al. 2025; Yamamoto et al. 2017), but less attention has been given to understanding how much variability may occur among sensors deployed within the same site under seemingly similar environmental conditions.

Rather than focusing solely on the accuracy of an individual sensor to a reference, this study evaluates the variability among measurements from siting multiple co-located Kestrel sensors. Understanding this variability is important for interpreting spatial differences observed in urban sensor networks. If sensors located close together have substantially different measurements, it becomes difficult to distinguish true spatial variability from measurement uncertainty. Here, we investigate how sensor biases and inconsistent measurement techniques affect the use of low-cost sensors for assessing intra-urban heat variability. We focus on T_a and RH, as these variables can be measured reliably with widely available low-cost sensors.

2. Methods

a. Sensor Specification

We used 41 Kestrel Drop D2 sensors (Kestrel, 2025), measuring T_a and RH with a precision of 0.1 °C and 0.1%, and an accuracy ± 0.5 °C for T_a, ±2 % for RH (Nielsen-Kellerman. 2020). Kestrel Drop D2 sensors are factory calibrated at standards of the National Institute of Standards and Technology against traceable reference instruments (Ametek DTI-050 Digital Temperature Indicator for T_a and Edgetech HT120 Humidity Transmitter for RH) over a controlled measurement range (Nielsen-Kellerman., 2020; -10 to 55 °C for T_a and 10 to 90% for RH at 25°C). Calibration drift for T_a measurements is noted as negligible for the life of the product, and for RH is less than 0.25% per year.

Kestrel sensors are widely used in urban heat research due to their affordability (~$109 per unit as of August 2025) and ease of deployment in both mobile and stationary

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

configurations. They have been employed across a broad range of applications, including the evaluation of urban heat mitigation strategies, such as street tree canopies (Segura et al. 2022) and cool roofs (Elnabawi et al. 2023) as well as studies assessing heat exposure risks in vulnerable populations (Karanja et al. 2023; Mukhopadhyay and Weitz 2022). In a comparative validation study of commonly used T_a sensors (Bailey et al. 2020), Kestrel sensors exhibited the highest correlation with research grade weather station data.

Each of the sensors were housed in a low-cost passive radiation shield (~$21 per unit as of January 2026, La Crosse Technology, 2026) unless stated otherwise to prevent the sensors from heating up in direct sunlight. A benefit of the Kestrel Drop D2 sensor, by comparison to some other low-cost devices, is that the T_a probe is suspended using a coiled wire outside of the sensor casing, minimizing the influence of the casing on measurements. Humidity is measured through perforations in the sensor housing which contains a hygrometer (Nielsen-Kellerman. 2020).

b. Experimental Setup

To confirm that no significant systemic bias existed amongst the sensors, all sensors were deployed in a room inside a building, and the temperature raised to 32 °C via central heating to match the higher T_a values expected in the local area. Sensors were additionally placed in a plant growth chamber set to 30, 25, and 20 °C for 6 h each, within which RH was sequentially adjusted to 90, 65 and 40 % at 2 h intervals (Table A1, Fig. A1a-d). The selected T_a setpoints were chosen to represent typical warm-season conditions at which larger cross-sensor variability may be expected during the outdoor experiments, to confirm the absence of systematic bias under controlled conditions representative of those encountered in field deployments. RH setpoints were selected to span the range achievable by the plant growth chamber. To provide a comparison with research grade observations in field conditions, the low-cost sensors were also deployed adjacent to a research-grade (more accurately, operational) reference at an airport site, with the sensors positioned approximately 2-5 m from the reference (Fig A1e-i, State Climate Office of North Carolina, NC State University, 2026). Sensors were then tested outdoors under realistic conditions to determine how various factors can impact the measurements. Configurations are detailed in Fig. 1.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

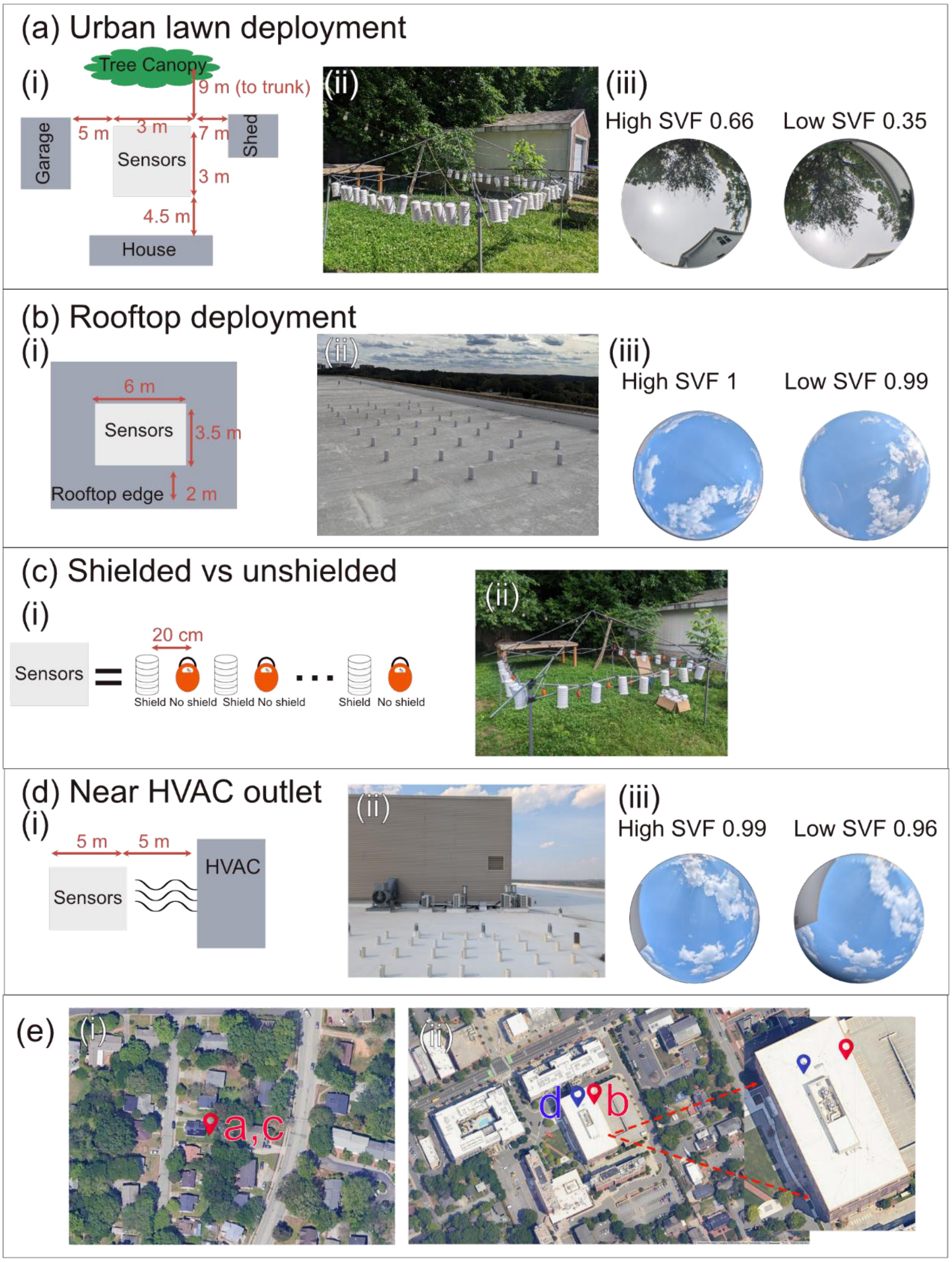

Fig. 1. Outline of the sensor configurations assessed, including (i) sensor placement and distances to surrounding obstacles, (ii) site photographs, and (iii) the highest and lowest sky view factors (SVFs) each site with corresponding fisheye images. Panels show: (a) 41 sensors housed in radiation shields on the urban lawn at 1.5 m height, 20 cm distance (b) 41 sensors housed in radiation shields on the urban rooftop at 7cm height (for safety reasons, elevated structures were not permitted on the rooftop), (c) alternating sensors housed inside and outside of radiation shields on the urban lawn (same SVFs as panel a), (d) sensors placed 5 - 10 m from HVAC units on the rooftop, and (e) satellite images showing locations of panels

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0222.1.

(a–d). Sensors on the lawn were spaced 20 cm apart (center-to-center between radiation shields), while rooftop sensors were spaced 0.5 m apart. Further details in Table A2. Satellite Imagery: Google Earth Engine, 2026 Airbus, Maxar Technologies. Distances in meters are reported to the nearest 0.5 m and in centimeters to the nearest 5 cm.

Sensors recorded data every ten minutes. We calculated hourly means within a ±30-minute window to capture hourly conditions and minimize the influence of short lived microclimate fluctuations, following WMO best practices (WMO, 2023). Hourly reporting intervals correspond with conventional weather station data, and aggregation across this period also reflects a practical balance between capturing diurnal variability and managing constraints on battery life and data storage, which are common in low-cost sensor deployments requiring manual data retrieval. Prior to averaging, data were quality controlled by filtering for outliers based on an exceedance of three standard deviations (Beele et al. 2022; Chapman et al. 2017) from the hourly mean of each sensor (i.e., within the six measurements per hour); however, no observations were excluded through this process.

Sensors were deployed at two sites in Chapel Hill, North Carolina, U.S.A., to assess the influence of solar exposure and land cover type: a semi-shaded urban lawn (35.905° N, 79.064° W, Fig A2a) and a fully exposed rooftop (35.912° N, 79.057° W). The lawn site consisted of flat terrain with short grass in a suburban residential area, with Local Climate Zone (LCZ) 6, open lowrise (Demuzere et al. 2022), and was semi-shaded. The rooftop was a flat, white painted surface on an eleven-story office building (average floor area of 2,800 m^2), typical of large institutional and office structures in the area (LCZ 5, open midrise), with multiple HVAC units present. Unless specified, sensors were positioned in open areas away from such equipment to minimize localized heating effects. The rooftop location is 1.3 km away from the lawn site and experienced no shading from adjacent buildings.

Meteorological observations were obtained from two stations in the North Carolina ECONet network. CHAP (35.934° N, 79.064° W) provided solar radiation, windspeed and rainfall to characterize conditions during deployments (Fig. A3), as well as reference T_a and RH for comparisons. KRDU (35.892° N, 78.782° W) provided cloud cover hours (State Climate Office of North Carolina, NC State University 2026).

Chapel Hill is characterized as humid subtropical (Koppen Geiger Cfa) (Peel et al. 2007), with precipitation distributed throughout the year. The lawn sensors (Fig. 1a) took measurements over a period of 16 full days (3-18 May 2025) with both clear sky conditions and days of heavy precipitation (Fig. A3a), with the overall mean measured T_a of 20.5 °C and

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

RH 79%. The roof top deployment (Fig. 1b) took place over a period of 6 days (16-20 October 2024) with clear sky conditions (Fig. A3b) with the period mean T_a 14.6 °C, RH 59%. The alternating shielded and unshielded sensors (Fig. 1c) were deployed for three clear sky days (20 – 22 May, mean T_a 22.5 °C, RH 66% from shielded sensors only, Fig A3c). The rooftop deployment near a localized heat source (Fig. 1d) occurred from 20–22 September 2024, with a mean T_a of 25.2 °C and mean RH of 72%. Precipitation occurred on the final day of the three-day period (Fig A3d). No control measurements were conducted away from the HVAC units during this deployment; therefore, measurements likely reflect localized warming associated with the nearby heat source rather than representative rooftop background conditions. However, this configuration was specifically designed to represent the influence of localized anthropogenic heat sources, which are commonly encountered in practical deployments near buildings. Deployments were scheduled outside of the winter months to focus on periods where extreme heat is a bigger public health issue.

c. Heat Stress Calculations and Statistical Analyses

The heat index (HI), representing the feels-like temperature in shaded conditions due to combined impact of T_a and RH, is calculated following standard U.S. National Weather Service (USNWS) practice based on Steadman (1979) and Rothfusz (NOAA 2022; 1990). When HI is below 80 °F (26.7 °C), equation (1) is used (with T_a in °F), consistent with USNWS guidance that the Rothfusz regression is not appropriate under conditions that yield HI values of below 80 °F.

$$HI = 0.5 \times (T_a + 61 + (T_a - 68) \times 1.2) + RH \times 0.094 \tag{1}$$

HI values are first calculated using Steadman's formula (equation 1), and if this produces values of HI greater than 80 °F, values are recalculated using the Rothfusz equation, based on a regression of Steadman's table (Rothfusz 1990; Steadman 1979).

$$\begin{aligned} HI = & -42.379 + 2.04901523\,T_a + 10.14333127\,RH - 0.22475541\,T_a\,RH \\ & - 6.83783 \times 10^{-3}{T_a}^2 - 5.481717 \times 10^{-2}RH^2 + 1.22874 \\ & \times 10^{-3}{T_a}^2RH + 8.5282 \times 10^{-4}T_aRH^2 - 1.99 \times 10^{-6}{T_a}^2RH^2 \end{aligned} \tag{2}$$

For RH less than 13% and T_a between 80 and 112 °F, adjustment 1 is subtracted from HI (NOAA 2022).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

$$Adjustment\ 1 = \frac{(13 - RH)}{4} \sqrt{\frac{17 - abs(T_a - 95)}{17}} \tag{3}$$

For RH greater than 85% and T_a between 80 and 87 °F, adjustment 2 is added to HI (NOAA 2022).

$$Adjustment\ 2 = \frac{(RH - 85)}{10} \times \frac{87 - T_a}{5} \tag{4}$$

Humidex, a unitless index for moist heat, is also calculated as per the Environment and Climate Change Canada (ECCC) (Masterton and Richardson 1981), in equation (5) (with T_a in °C).

$$Humidex = T_a + 0.5555 \times \left(6.11e^{5417.7530 \times \frac{1}{273.15 + T_d}} - 10\right) \tag{5}$$

Where T_d is dew point temperature (°C), directly measured by the sensors.

To evaluate the agreement of low-cost sensors, the standard deviation (SD, equation (6)) of sensor readings was computed to quantify variability across sensors.

$$SD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}} \tag{6}$$

Where x_i are the observations, $\bar{x}$ the mean, and N is the number of observations.

Pearson's correlation coefficient (Pearson 1895) was used to assess the relationship between sensor variability and environmental conditions, including diurnal range and cloud cover, with a significance level $p < 0.05$.

To isolate the influence of diurnal variability independent of the mean measurement conditions, an adjusted diurnal range was calculated using linear regression. The diurnal range was regressed against the measurement mean, and the residuals from this regression were used as the adjusted diurnal range in subsequent analyses.

$$D_i = \beta_0 + \beta_1 \bar{X}_i + \varepsilon_i \tag{7}$$

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

Where D_i is the diurnal range, defined as the difference between daily maximum and minimum values, $\bar{X}_i$ the daily mean measurement, and ε_i the residuals representing the adjusted diurnal range.

One-sided analysis of variance (ANOVA) tests was carried out on the groups of sensor hourly mean values to measure how much the means of any individual sensor differ relative to the variability within the group. Paired t-tests were carried out to examine differences between day and night SD (equation 6), and the sensor shielded unshielded pairs (Fig. 1c).

A least squares regression was performed to examine calibration against a median sensor during the 3-18 May lawn deployment (Fig. 1a). On 8 May, sensor positions were randomly permutated using a random number generator (Fig. A2b) to test calibration based on both sensor ID and position. The regression was fit calibrating sensors to the median sensor for days 3-8 May, based on equation 8.

$$T_{a_{sensor}} = b + aT_{a_{median\ sensor}} \tag{8}$$

Where a and b are the gradient and intercept.

Model performance was evaluated using R^2 and RMSE. The resulting calibration equations were then applied to the data from 10-18 May, based on the sensor ID (same sensor, different position), and sensor position (different sensor but same position).

Sky view factors (SVFs) are used to quantify the shade exposure within and across the site locations, calculated using Fisheye lens photographs and a modified Steyn method (Middel et al. 2018; Steyn 1980). Each fisheye is split into n = 36 rings to account for the distortion in the fisheyes and the contribution of each section summed.

$$SVF = \frac{\pi}{2n}\sum_{i=1}^{n}\left(\frac{\pi(2i-1)}{2n}\right)\left(\frac{p_i}{t_i}\right) \tag{9}$$

Where $\frac{p_i}{t_i}$ is the ratio of sky pixels to total pixels.

Data processing and analysis were performed in Python v3.13.2 with standard scientific libraries, including NumPy and pandas for numerical operations and data manipulation,

SciPy for statistical tests, Matplotlib and Seaborn for visualizations and OpenCV for image processing.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

3. Results

Here, we analyzed the standard deviation (SD) amongst daily hourly mean sensor readings at each deployment location and sensor arrangement.

a. Urban Lawn Deployment

For the lawn deployment, the SD amongst sensor readings in the small deployment areas shows a clear diurnal cycle, highlighting key times of day where spatial comparisons should be made with more caution. Of note, at night (defined as periods when incoming solar radiation was zero using ERA5-Land reanalysis data (Muñoz Sabater 2019)), SD amongst sensors is minimal (mean = 0.09 °C, range 0.04 – 0.14 °C), but greatly increases during daytime hours (mean = 0.37 °C, range 0.09 – 0.70 °C), when the atmosphere is less stable and radiation can impact sensor accuracy (Fig. 2). A paired t-test comparing daytime and nighttime T_a SD showed a significant difference ($t = 8.04$, $p < 0.05$), confirming greater values during the day than at night.

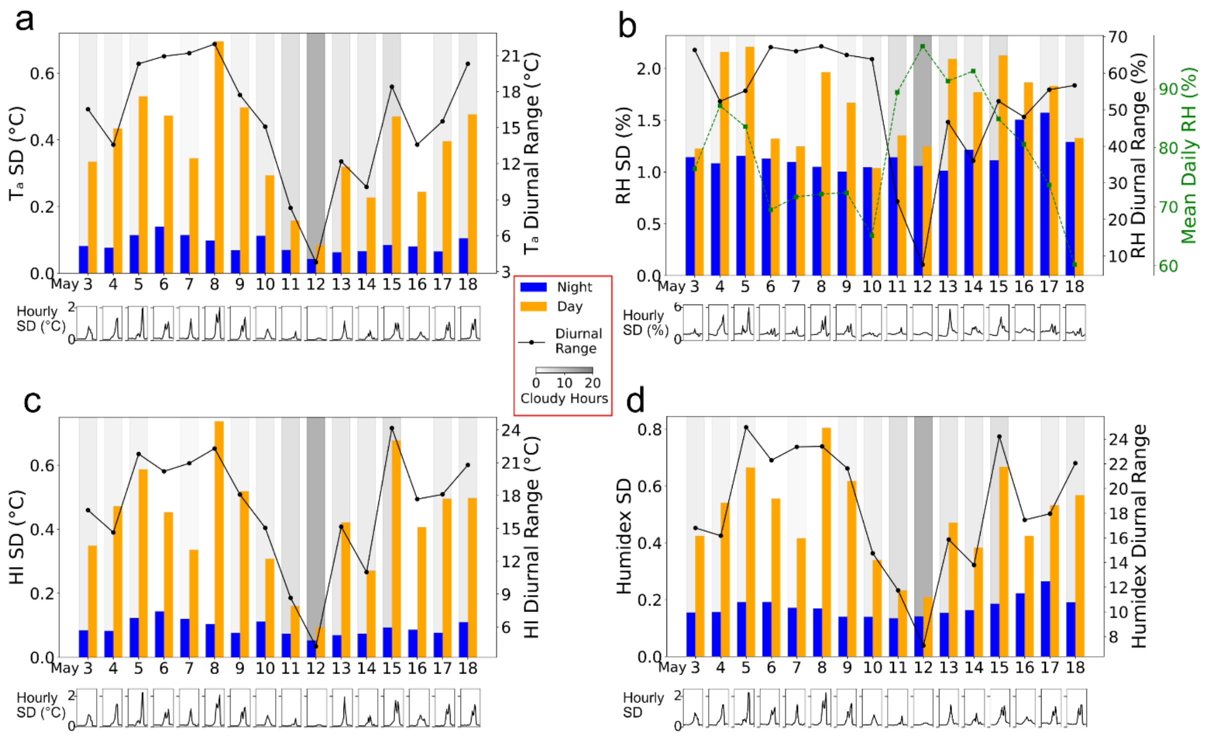

Fig. 2. Nighttime and daytime SD (bars, primary y axis) and diurnal range (lines, secondary y axis) for 41 sensors each day in an urban lawn: (a) T_a, (b) RH with a third y axis (green) showing mean daily RH, (c) HI, and (d) Humidex. Small hourly plots beneath bars show the shape of the diurnal cycle of SD for the day. The number of cloudy hours in a day (where the sky is 50%+ covered by cloud), obtained from ECONet KRDU airport weather station, is indicated by gray shading.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0222.1.

The magnitude daytime SD in T_a is highly correlated with the diurnal T_a range, with Pearsons's correlation coefficient (*r*) (Pearson 1895) of 0.86 (Fig A4a, Table A3) during the 16-day urban lawn deployment (Fig. 2a). Strong positive correlations are also observed at night (r = 0.78). Figures 2 further illustrates the diurnal pattern of SD for each day, with larger versions in Figure A5a. During daytime hours, SD is most pronounced in transitional periods when T_a is rising or falling rapidly, with many days exhibiting a bimodal diurnal pattern, with two peaks occurring, around 13:00 and 18:00. A negative correlation was observed between a higher number of cloudy hours and SD (r = -0.60), consistent with the relationship between low diurnal ranges and increased cloud cover.

As with T_a, SD in RH is larger during the day (mean = 1.65%, range 1.04 - 2.21%) than at night (mean = 1.16%, range 1.00 – 1.57%), with a significant paired t-test (t = 4.76, $p < 0.05$). However, unlike T_a, the relationship between RH SD and diurnal range (Fig. 2b), is only seen after controlling for the mean RH, with larger mean-adjusted diurnal range (ε_i, equation (7)) in RH positively correlated with daytime SD (r = 0.54, Fig. A4d, Table A3). During the night, even after controlling for the mean RH, no correlations were found between measurement SD and diurnal RH range. RH SD is also found not to have any significant relationship with cloud cover hours based on Pearsons correlation.

Because the equations used to calculate HI (equation 1-2) are functions of both Ta and RH, any uncertainty in these measurements propagates into HI. This is reflected in the SD of HI (Fig. 2c), which is amplified in relation to T_a, with the overall mean HI SD 14% higher than that of T_a (HI mean = 0.42 °C vs T_a= 0.37 °C, HI range 0.09 – 0.74 °C). Despite this amplification, the general pattern of variability mirrors that of T_a, with strong correlations between HI SD and diurnal cycle during that day (r = 0.87) and night (r = 0.69) (Fig. A4e, Table A3). HI nighttime SD is of similar magnitudes to T_a (HI mean = 0.09 °C, range 0.05 – 0.14 °C). Similarly, Humidex (equation 5), a function of T_a and dewpoint temperature, T_d (which is used to calculate RH) is influenced by SD in both measurements (daytime mean SD = 0.50, range 0.13 – 0.26, nighttime mean SD = 0.17, range 0.21 – 0.80), and both HI and Humidex have higher SD as a percentage of the mean (Coefficient of Variation, Fig. A6). Whilst the results presented here focus on SD, further statistical measures, including box plots of absolute sensor bias and RMSE are shown in Fig. A7.

To determine if the observed SD stemmed from systematic sensor or positional biases, we performed a regression-based calibration against the median sensor (equation 8), showing

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

good performance with RMSE ranging from 0.2 to 0.9 °C of (mean 0.5 °C) and R^2 from 0.98 to 1 (mean 0.99) (Fig A8). However, once applied a separate test period, neither sensor ID nor position consistently predicted bias magnitude or direction, and calibration adjustments often increased measurement spread (Fig. A9). Further evaluation was done using one-way ANOVA tests, performed for each variable (T_a, RH, HI and Humidex) to determine whether the mean readings of individual sensors differed significantly from the group mean. No significant differences were observed (all F-statistics not significant).

Standard meteorological practice involves taking measurements at sub-hourly intervals (typically 1-10 minutes), which are then aggregated to an average to represent hourly conditions (WMO, 2023). This can highlight outliers and gives a full representation of the conditions during the hour. Using hourly means reduced T_a SD compared to single-point measurements (Fig. A10a), while single measurements also produced a larger diurnal range. A similar increase in the range of measurements logged is seen for RH (Fig. A10b), which again are inherited in both HI and Humidex (Fig. A10cd).

b. Rooftop Deployment

For sensors deployed on the unshaded rooftop, daytime T_a SD was higher than night (daytime mean SD = 0.63 °C, range 0.25 – 0.9 °C, nighttime mean SD = 0.15 °C, range 0.04 – 0.27 °C), with differences between the two significant (paired t-test $t = 8.83$, $p<0.05$) (Fig. 3a). Across the six-day deployment, SD was strongly correlated with the T_a diurnal range (daytime $r = 0.97$, nighttime $r = 0.93$, Fig A11), reflecting the results of the lawn deployment. A bimodal diurnal pattern is also apparent (Fig. 3, Fig A5b). On the unshaded rooftop, the first peak in T_a SD tends to be earlier in the day than the semi-shaded lawn, at around 9:00, whereas the second peak occurs around 18:00, consistent with the lawn results.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

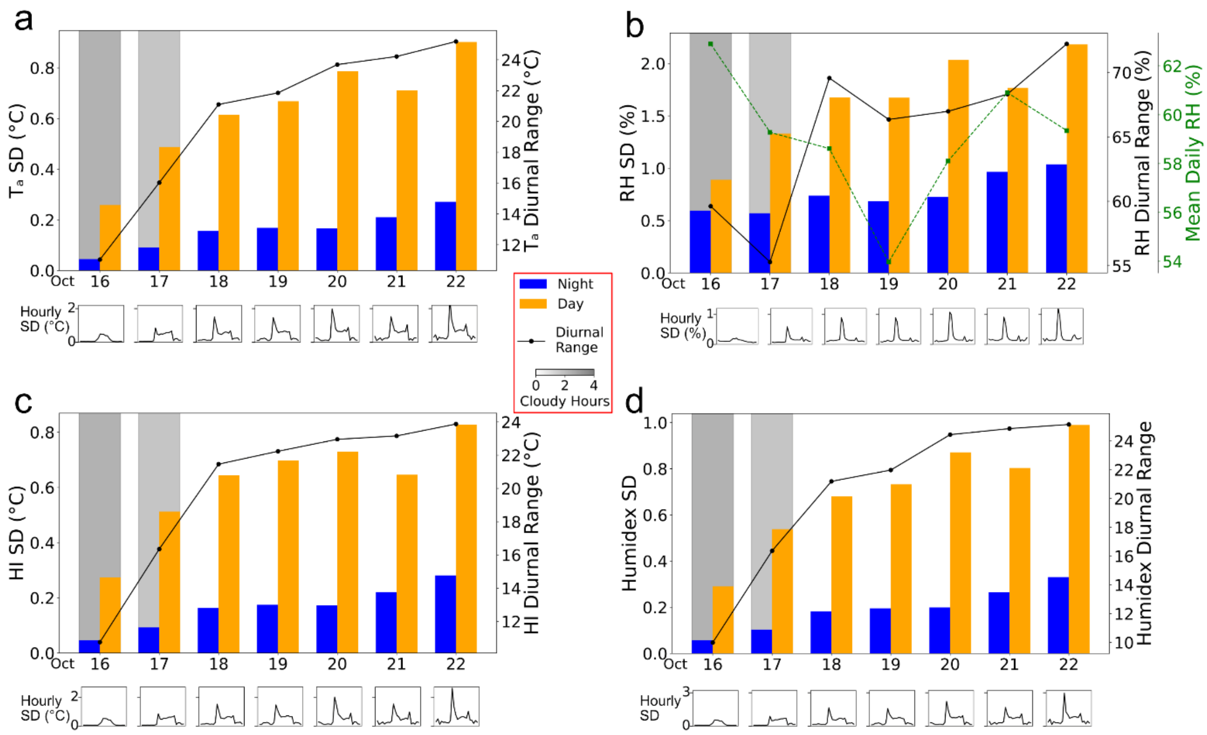

Fig. 3. Nighttime and daytime SD (bars, primary y axis) and diurnal range (lines, secondary y axis) for 41 sensors each day in an urban rooftop: (a) T_a, (b) RH with a third y axis (green) showing mean daily RH, (c) HI, and (d) humidex. Small hourly plots beneath bars show the shape of the diurnal cycle of SD for the day. The number of cloudy hours in a day (where the sky is 50%+ covered by cloud), obtained from ECONet KRDU airport weather station, is indicated by gray shading.

Daytime RH SD (mean = 1.65%, range 0.89 – 2.18%) was again found to be significantly larger than night (mean = 0.76%, range 0.57 – 1.04%) (paired t-test $t = 7.26$, $p<0.05$). In the rooftop setting, over an impervious surface, RH SD and diurnal RH range correlated ($r = 0.80$, Fig A11c), contrasting the lawn deployment, where a correlation between the two was only seen after controlling for the mean RH in the mean-adjusted value.

During the rooftop deployment, the daytime HI SD (mean = 0.62 °C, range = 0.27 – 0.83 °C) is comparable to that of T_a, and in this case is slightly smaller (by 2%). During this time, RH was lower (mean daily RH 59% for the rooftop compared to 79% during lawn deployment) and therefore contributed less to the HI (equations 1 - 4), as high humidity more strongly amplifies heat stress. Nighttime HI SD is similarly comparable to T_a with a mean of 0.16 °C (range = 0.05 – 0.28 °C). Humidex SD exceeds that of both Ta and HI, with a daytime mean of 0.70 °C (range = 0.29–0.99 °C) and a nighttime mean of 0.19 °C (range = 0.06–0.33 °C). Further statistical measures for the rooftop deployment are shown in Fig. A12.

c. Alternating shielded and unshielded

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

T_a measurements taken without radiation shielding exhibit a daytime positive bias compared to those from shielded sensors, with the deployment indoors (inside a heated room) added as a control (Fig. 4). This discrepancy is most pronounced during the warmest hours of the day (13:00-18:00, mean T_a difference = 0.25 °C, mean HI difference = 0.43 °C, mean Humidex difference = 0.47 °C), where SD is also higher in unshielded sensors (difference in SD for unshielded T_a = +0.21 °C, HI = +0.32 °C, Humidex = +0.31 °C, Fig. 4acd). This is consistent throughout the three days of deployment (Fig. A13). In contrast, during nighttime (between ~22:00 and ~6:00), when solar radiation is absent, the mean T_a recorded by both shielded and unshielded sensors converge.

Although shielding reduced some cross sensory variability, the passive low-cost shield was not completely effective. In a two-day comparison against an adjacent research-grade reference (Fig. A1), the shielded low-cost sensors reproduced the overall diurnal pattern of T_a but still recorded higher T_a observations during warmest hours of the day, particularly on the day with the greater T_a range.

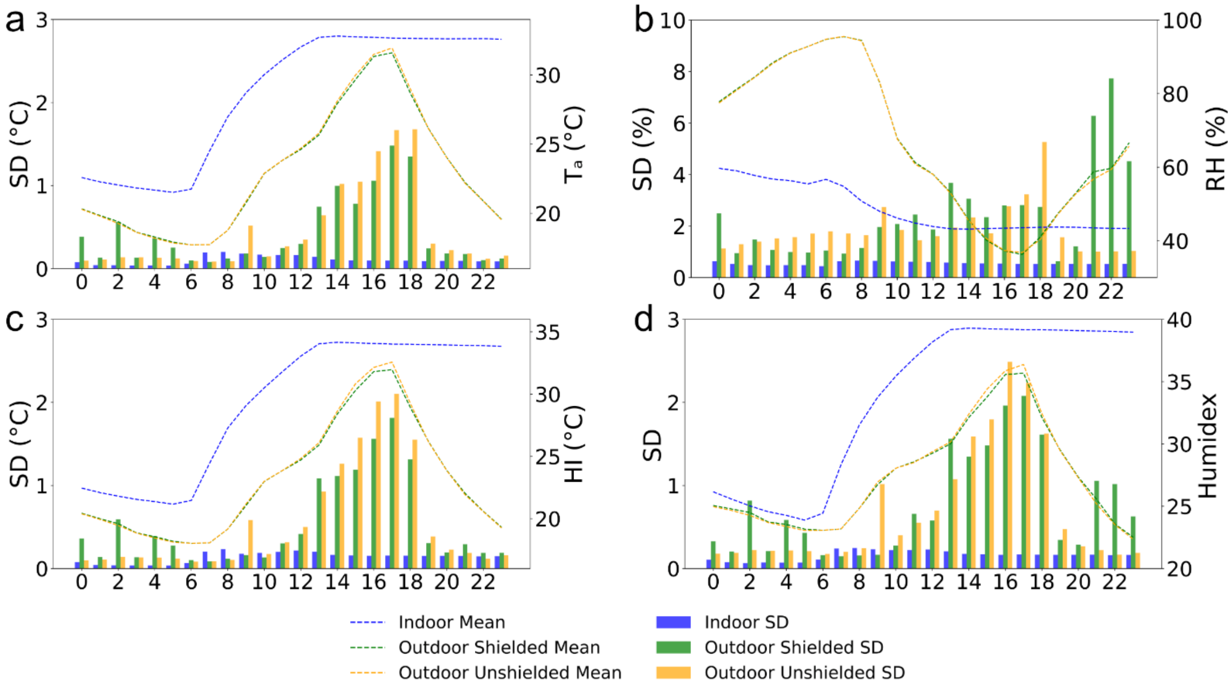

Fig. 4. Comparisons of SD across 37 sensors when shielded, unshielded and in absence of solar radiation inside a building. Hourly mean SD for (a) T_a, (b) RH (c) HI, (d) Humidex (bars, primary axis) and mean value (lines, secondary axis) from sensors placed indoors and in an urban lawn setup with and without radiation shields (Fig. 1c).

This shielding versus ventilation tradeoff with unaspirated, passive radiation shields means there can be a lag in the T_a readings during temperature rise due to the lack of forced

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0222.1.

ventilation. The impact of ventilation is potentially apparent on the May 20 result (Fig. A14a), where the SD for shielded briefly exceeds unshielded sensors.

Unlike T_a, RH measurements showed minimal differences between unshielded and shielded across the full period (mean shielded-unshielded difference = 0.2%) and during the warmer hours (13:00-18:00, mean shielded-unshielded difference = -0.2%). These differences are small relative to the overall variability of RH (RH SD unshielded = 2.8%, shielded = 2.5%), indicating shielding has a limited influence on RH measured values. While some hourly differences in RH SD between configurations are visible (Fig. 4b), these were not statistically significant (paired t-test across hourly values $t = 1.29$, $p = 0.20$). Of note, however, is that the deployment took place over days with no precipitation. The signal of RH SD appears more in the Humidex than HI observations, with evening (9pm and 10pm) high SDs in RH (Fig. 4). This is because Humidex is more sensitive to RH than HI (Chakraborty et al. 2022; Sherwood 2018).

d. Rooftop deployment near HVAC outlet

A sensor's immediate surroundings can have a profound effect on the accuracy and reliability of micrometeorological observations. In urban settings, proximity to anthropogenic heat sources such as air conditioning (AC) outlets, vents, or machinery can significantly bias measurements, leading to misleading spatial patterns if not properly accounted for. To quantify the impact of such localized heat sources, we conducted a comparison with sensors deployed on the rooftop, by deploying sensors in a location near HVAC units for a three-day period (Fig. A15). The daytime SD in T_a was 2.82 °C (range 2.46 - 3.25 °C) and RH 8.02% (range 6.29 – 8.96 %), over four (T_a) and approaching five (RH) times higher than the rooftop sensor deployment away from the outlets (Fig. 1b, 0.63 °C, 1.65%). A similar trend was observed at night, with a SD of 0.80 °C T_a (range 0.51 - 0.96 °C) and 9.5% RH (range 3.86-13.53%) near the HVAC unit compared to just 0.15 °C and 0.76% RH in the interference-free setting. It should be noted the interference free deployment (Fig 1b) and deployment near the HVAC unit (Fig. 1d) occurred at a different time, although the average diurnal T_a range during both deployments was nearly identical (20.6 °C near unit vs. 20.4 °C for the isolated sensor). RH was higher during the HVAC deployment (73% versus 59%), which may have additionally inflated RH SD, as rainfall occurred during the final day (Fig. A3d).

4. Discussion and Conclusion

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

There is a need to establish clear guidelines and quality control measures for making robust urban heat stress comparisons when using low-cost sensors. Low-cost sensors are not just limited by calibration accuracy, but by how they interact with microclimate at very small spatial scales, making siting, shielding, and environmental context critical for interpreting intra-urban variability. Here, we quantify how measurements from identical sensors may differ when placed in close proximity (i.e., meters apart), due to sensor interaction with localized microclimatic conditions rather than purely sensor-specific differences that can be calibrated for.

Our findings suggest environmental factors, such as shading or heat sources, and methodological choices, such as site selection or shielding, introduce substantial variability that can undermine the accuracy of both spatial comparisons and estimates of UHI and UDI magnitude if not carefully accounted for. We demonstrate that the SD of T_a and RH sensors can degrade significantly when employed in outdoors, consistent with other studies showing decreased sensor performance under field conditions (Chakraborty et al. 2022; Yamamoto et al. 2017; Yulizar et al. 2023). Although we did not observe calibration drift in RH measurements during our deployments, such drift has been reported as a potential issue in other studies (Bell et al. 2015; da Cunha 2015). These results highlight the importance of conducting sensor validation and calibration under real-world, field-based conditions to ensure reliability and accuracy of environmental monitoring, particularly in heterogeneous urban settings.

a. Radiation Shielding and Ventilation

Consistent with previous research (Gubler et al. 2021; Sun et al. 2015; Yamamoto et al. 2017) larger variability in daytime T_a, HI and Humidex SD (Fig. 2, Fig. 3) likely reflects the influence of solar heating. Unshielded sensors consistently recorded higher daytime temperatures during the hottest hours of the day (Fig. 4), artificially inflating daily maximums, potentially skewing the spatial variability of the UHI across variably shaded parts of the city. This finding establishes the necessity of using some form of radiation shielding for outdoor measurements. This requirement is standard for meteorological measurements but not always implemented, which presents a major issue for citizen weather stations (Alerskans et al. 2022; Bell et al. 2015; Cornes et al. 2020; Fenner et al. 2021; Jenkins 2014). Failure to shield sensors may also interfere with assessments of cooling measures such as due to tree canopy or shading structures, as comparison sensors exposed to direct sunlight will show

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

inflated T_a values. Additionally, adequate shielding is important for accurate RH measurements, which require the sensor to remain dry and, as RH is derived using T_a, artificially high T_a values will lead to underestimated RH (da Cunha 2015). Our results here also suggest that a focus on investing in higher quality shields may be preferable to budgeting more on the sensors themselves, as comparison against a reference sensor showed elevated daytime T_a across the low-cost sensors.

In addition to absolute radiation levels, rapid changes in solar radiation appear to exacerbate T_a measurement SD. For the unshaded rooftop deployment, peaks in SD occurred in the morning (around 9:00), and in the semi shaded urban lawn early afternoon (around 13:00), likely later due to the influence of shade delaying solar heating. SD is high again for both locations around 18:00. Yamamoto et al (2017) noted that abrupt shifts in solar intensity led to large fluctuations in sensor readings. These rapid changes are also associated with transitions in boundary layer processes, such as increased turbulence and convection, which may further contribute to measurement variability, due to microscale differences in sensor ventilation. Unaspirated shields can retain heat and minor differences in airflow across sensors can lead to variability in readings (Nakamura and Mahrt 2005), meaning the thermal response of each sensor becomes more susceptible to local conditions, contributing to increased measurement divergence during periods of strong solar forcing. Young et al (2014) similarly finds the largest errors in sensor readings occur around sunrise, potentially due to two mechanisms; enhanced radiation errors caused by the low solar angle, and inadequate ventilation in radiation shields leading to heat buildup and stagnant air. Notably, these errors varied by deployment, suggesting limited generalizability. This issue is reflected in our own findings, where SD was not consistent across days (Fig. 2, 3), and therefore difficult to correct for. Even with radiation shielding, sensors may respond inconsistently to radiative heating due to differences in passive ventilation. Low wind speeds during deployments (Fig. A3), resulted in the unaspirated shields receiving little natural ventilation (Harrison 2010).

b. Landcover and Surroundings

Other environmental factors, such as land cover, can further influence sensor performance. Nan et al (2025) demonstrated that T_a calibration equations optimized for specific land cover types (e.g., grass, forest, synthetic turf) yielded varying levels of accuracy, indicating that local surface characteristics can influence sensor behavior and that a one-size-fits-all calibration may not be effective across diverse urban landscapes. Our

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

findings also highlight that RH presents distinct challenges. Because sensor-measured RH is derived from T_a, it is indirectly affected by radiation, although the relationship is less clear than with T_a (Jenkins 2014), but its measurement variability is also closely tied to land cover types. In our deployments, grass cover dampened the correlation between diurnal range and RH error, whereas the impervious rooftop location, lacking surface moisture, exhibited stronger correlations (Fig. 3, Table A3). This contrast indicates that water retention and evapotranspiration processes in grass may alter localized RH variability. While the magnitude of RH error over both surfaces is similar, different processes appear to dominate depending on the surface type.

The need for metadata on sensor surroundings has long been considered an important consideration in urban micrometeorology (Stewart 2011). Sensors placed near localized heat sources (Fig. 1d) exhibited highly inflated SEs over seven times greater in T_a and almost ten times for RH than those in undisturbed locations (Fig. A15), although RH variability may be further increased due to rainfall. These elevated errors are likely influenced by strong local temperature gradients and turbulent mixing associated with anthropogenic heat sources, which can introduce rapid, small-scale variability that is difficult for sensors to consistently resolve. These findings highlight that proximity to anthropogenic heat sources can strongly influence measurements; whether this represents bias or meaningful signal depends on the research objective. For studies aiming to characterize broader ambient conditions, careful siting away from such localized influences is important, whereas studies focused on microscale heat exposure may intentionally include them. In all cases, recording sensor surroundings is critical. Simple metadata fields (WMO, 2023a) such as proximity to buildings or mechanical systems could help screen for potential problematic sensors and improve confidence in spatial comparisons. In crowdsourced networks, where detailed metadata is often lacking, this emphasizes the need for anomaly detection and quality control (Fenner et al. 2021; Napoly et al. 2018).

c. Measurement Frequency

This practical demonstration may help guide decisions on trade-offs between sampling frequency, battery life, and data storage, especially for studies where data retrieval requires physically visiting the sensors, as with the Kestrel devices used here. Comparisons between single point and hourly averaged measurements (Fig. A10), confirm that averaging reduces

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

SD, resulting in more stable spatial comparisons. Low sampling frequencies are more vulnerable to transient effects, such as passing clouds, which can distort results.

d. Managing sensor variability

The results of the regression-based calibration suggest that measurement uncertainty cannot be explained by intrinsic sensor characteristics alone but instead reflects interactions between sensor behavior and environmental conditions. In particular, factors such as diurnal temperature range (Fig. 2, 3), radiation exposure and ventilation are likely to modulate the sensor biases. Increased variability, especially during daytime, can complicate the interpretation of statistically significant T_a or RH differences, particularly in short-term or sparsely sampled deployments.

Measurement SD was lowest at night, when atmospheric conditions are more stable and there is no solar radiation. During the daytime, variability amongst sensor measurements increases, particularly during periods of rapid heating or cooling. This difference is reflected in the strong correlation of SD with the diurnal temperature range. As a result, spatial comparisons are more robust at night or for days when the diurnal temperature range is relatively small. However, given the importance of heat on human health and comfort both during day and night, knowing the extent of this discrepancy can help. Although radiation is found to be a strong predictor for measurement bias in previous studies (Mauder et al. 2008; Yamamoto et al. 2017), the use of low-cost sensors is often motivated by the need to minimize costs, and dedicated radiation measurements are typically not feasible due to expense. Diurnal temperature range, for example, may serve as a useful proxy for expected variability and could help establish estimates of expected sensor SD, which can be used to establish robustness in spatial comparisons of heat stress. Additionally, where resources allow, co-locating two or three sensors at a single site (Diamond et al. 2013) could help to provide an estimate of potential variability and improve confidence in the measurements.

Since Chapel Hill is a humid climate, both T_a and RH are important components of the moist heat stress burden. Here we see sensor SD in both measurements compound in heat stress indices (Fig. 2,3). In drier climates, T_a typically dominates heat stress variability; so errors in RH would have a smaller impact. Additionally, the relatively low wind speeds typical of Chapel Hill (Fig. A3) limit airflow around sensors, whereas windier locations, such as coastal cities, may benefit from enhanced natural ventilation. We should also note here that different moist heat stress indices have different sensitivity to T_a and RH, with relevance for

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0222.1.

cross-sensor variability depending on the index chosen (Chakraborty et al. 2022; Sherwood 2018).

As cities increasingly seek to map urban temperatures and develop policies to mitigate and adapt to extreme heat, ensuring measurement reliability is essential. For many public health and planning applications, identifying spatial disparities and relative differences in heat exposure across neighborhoods is critical for targeting interventions and reducing inequities, even when absolute temperature precision is limited (Hsu et al. 2021; Johnson 2022). When used appropriately, low-cost sensors can play a valuable role in identifying spatial disparities in urban heat and evaluating the effectiveness of targeted interventions. By offering practical guidance to minimize cross-sensor variability and localized measurement artifacts, this study supports the development of robust, standardized urban temperature monitoring.

Although our analysis focuses on a single sensor and shield type, we expect the qualitative findings to be transferable to other low-cost T_a and RH sensors. In particular, factors such as sensor shielding, siting and ventilation conditions, alongside proximity to localized heat sources are likely to influence performance across sensor types. However, the magnitude of inter-sensor variability will vary depending on sensor quality and specifications. Therefore, future work should evaluate different sensor and shielding configurations, a broader range of urban forms, and diverse meteorological conditions to test the robustness of these relationships. This study was conducted within a single city and climate; extending the analysis to multiple cities would help assess the transferability of results and the suitability of low-cost sensor networks for city-to-city comparisons. Finally, the analysis focuses on short-term deployments; longer-term studies with exposure to a wider range of meteorological features, such as wind speed, and solar angles and would help increase confidence in the use of low-cost sensor networks for monitoring urban heat stress.

Acknowledgments.

This work was supported in part by a grant/cooperative agreement from the National Aeronautics and Space Administration (NASA) under Award No 80NSSC24K0505;22-IDS22-0146. Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. We also acknowledge the North Carolina State Climate Office for assistance with comparisons against a research grade sensor from the ECOnet network, and the Phytotron at Duke University for support with plant growth chamber experiments.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

Data Availability Statement.

The dataset supporting this study, *"Biases due to widespread use of low-cost sensors for urban heat stress assessments (dataset)"*, is openly available on UNC Dataverse at: https://doi.org/10.15139/S3/LMALFL.

APPENDIX

Appendix A: Additional Figures

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

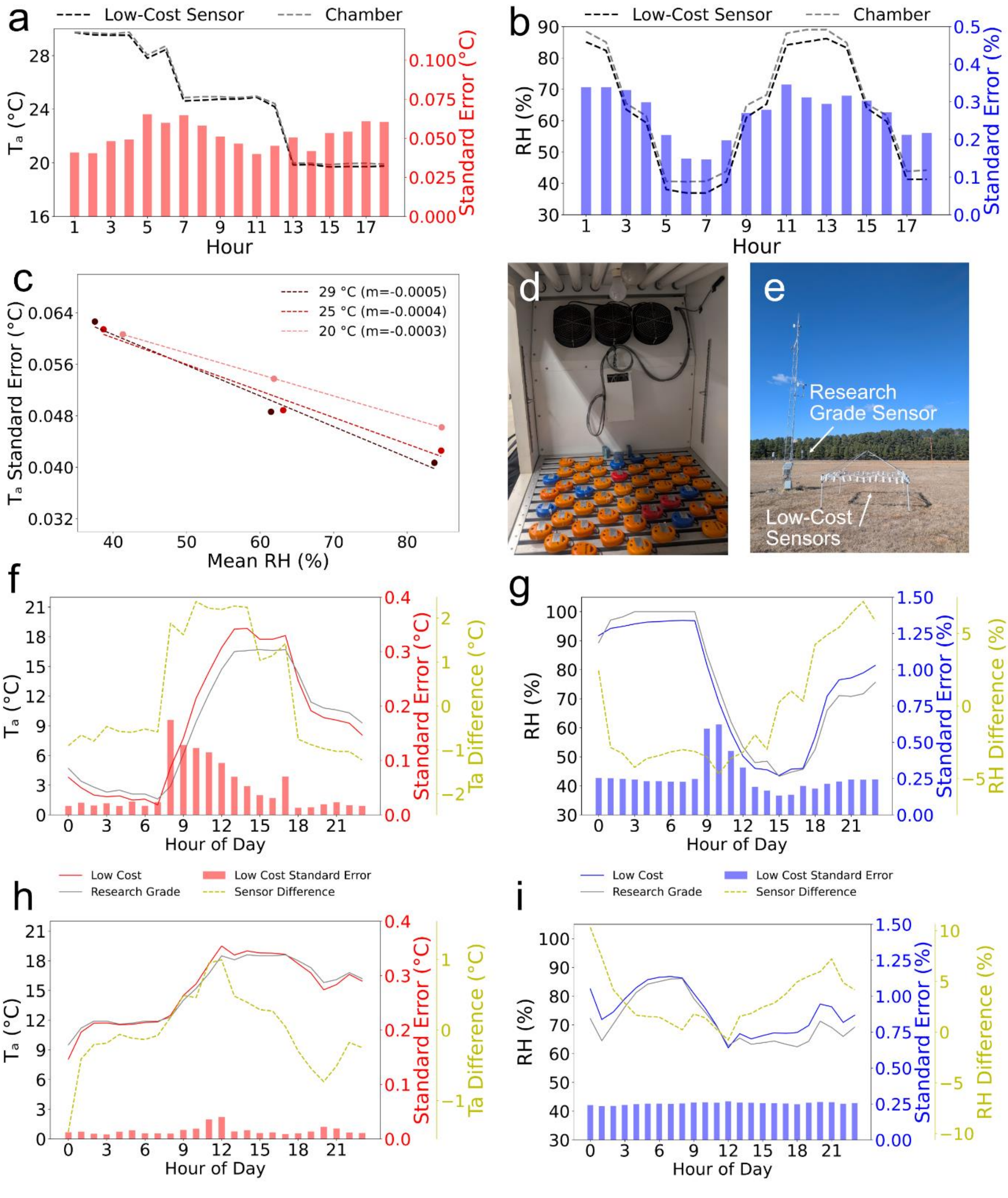

Fig. A1. Plant growth chamber and research-grade (more accurately, operational) sensor comparison for 41 low-cost sensors. Panels (a–b) show sensor mean values (black lines, primary y-axis) and standard error (bars, secondary y-axis) during a plant growth chamber experiment with T_a sequentially set to 30, 25, and 20 °C for 6 h each, while RH was stepped from 90 to 65 to 40 % for 2 h per step (18 h total), with actual chamber observations (gray lines). Panel (a) shows T_a and (b) RH. Panel (c) shows the relationship between RH mean and T_a standard error at each T_a step with a linear best-fit line. Photographs of the experimental setup are shown in (d) the plant growth chamber and (e) the research-grade sensor station (ECOnet CHAP (35.934° N, 79.064° W)), with the low-cost sensors positioned

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0222.1.

2 to 5 m from the station at a height of 1.5 m. Panels (f–i) show field comparisons between low-cost sensors and the research-grade station over two days. Lines (primary y-axis) show top of the hour measurements from low-cost sensors and the research-grade reference, bars (secondary y-axis) show the low-cost sensor standard error, and dashed lines (tertiary y axis) show the measurement difference (research-grade minus low-cost). Panels show (f) T_a (°C) 17th February 2026, g) RH (%) 17th February 2026, (h) T_a (°C) 18th February 2026, and (i) RH (%) 18th February 2026.

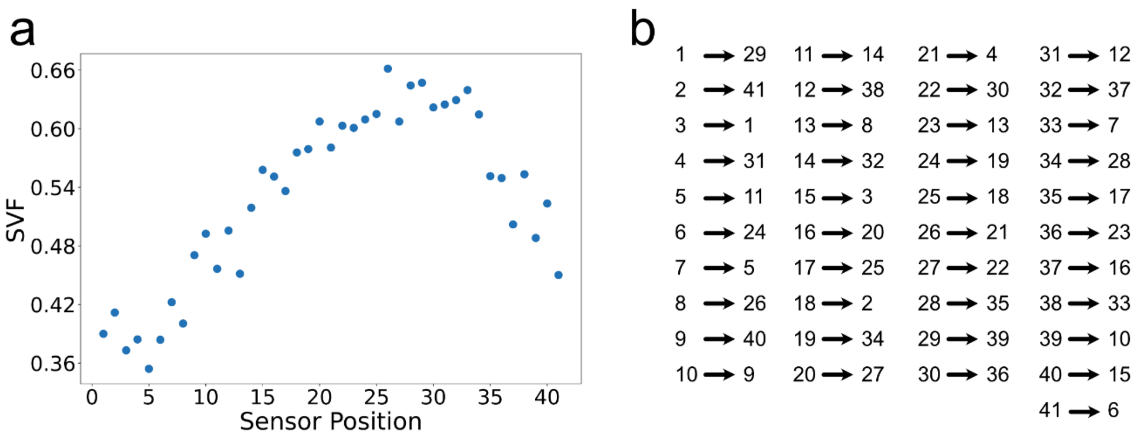

Fig. A2. Position related information for the urban lawn deployment. Panels show (a) Sky View Factors (SVFs) for sensors in the semi-shaded urban lawn setting. Positions of the sensors are labeled from 1 to 41 starting from the upper left hand side of the setup and proceeding anticlockwise. Low SVF indicates more shaded areas, while higher values indicate more exposed locations. (b) Permutations of the sensor positions carried out on 9 May. For example, the sensor originally at position 1 was moved to position 29, sensor in 2 to 41, etc.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0222.1.

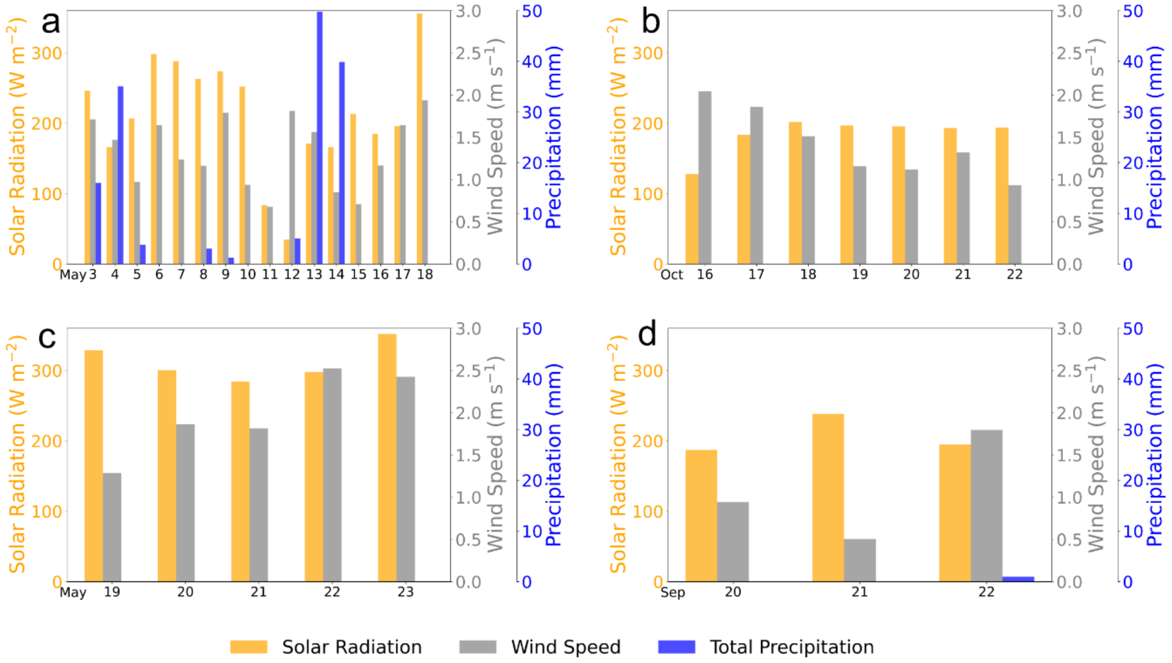

Fig. A3. Meteorological conditions during deployments (a–d), including hourly mean solar radiation, hourly mean wind speed, and total daily precipitation from the CHAP station of the North Carolina ECONet.

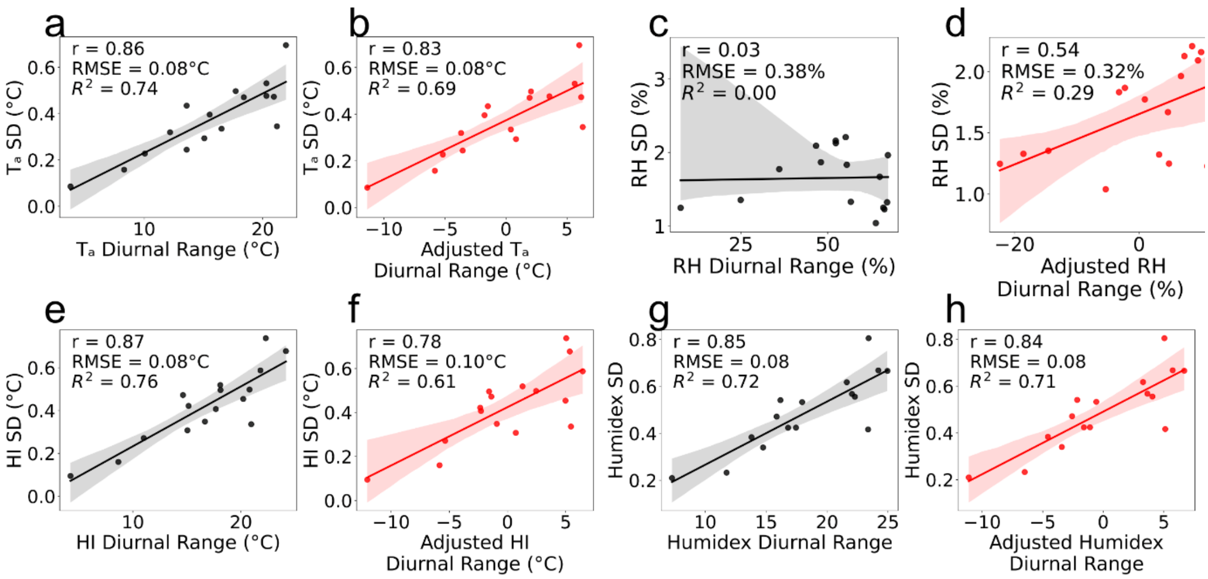

Fig. A4. Relationships between diurnal range and daytime measurement SD for T_a, RH, HI and Humidex during the lawn deployment. Panels show relationships using the raw diurnal range (a, c, e, g) and the adjusted diurnal range (b, d, f, h, equation (5)). Least-squares regression lines with 95% confidence intervals are shown (gray for unadjusted relationships and red for adjusted relationships). Values shown within each panel include Pearson's correlation coefficient (r) and the fitted regression root mean square error (RMSE) and coefficient of determination (R^2). Panels correspond to (a-b) Ta, (c-d) RH, (e-f) HX, and (g-

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0222.1.

h) HI. Pearson correlation coefficients for relationships between diurnal range and nighttime SD are provided in Table A3.

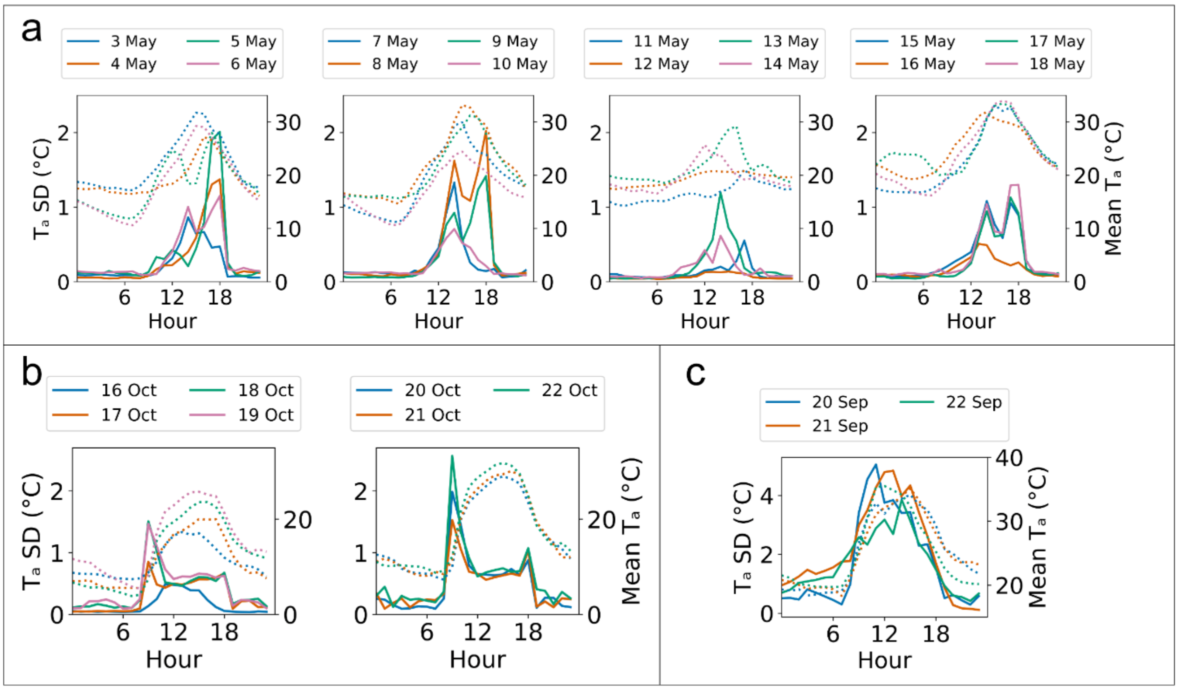

Fig. A5. Diurnal cycles of SD (solid line) and mean (dashed line) for T_a across three deployments. Panels show (a) Lawn site: hourly SD shown for individual days, grouped into four-day subsets (four plots total). (b) Rooftop calibration site: hourly SD for individual days, grouped into four-day subsets, with the final plot containing three days. (c) HVAC influence experiment: hourly SD for all days shown together on a single plot.

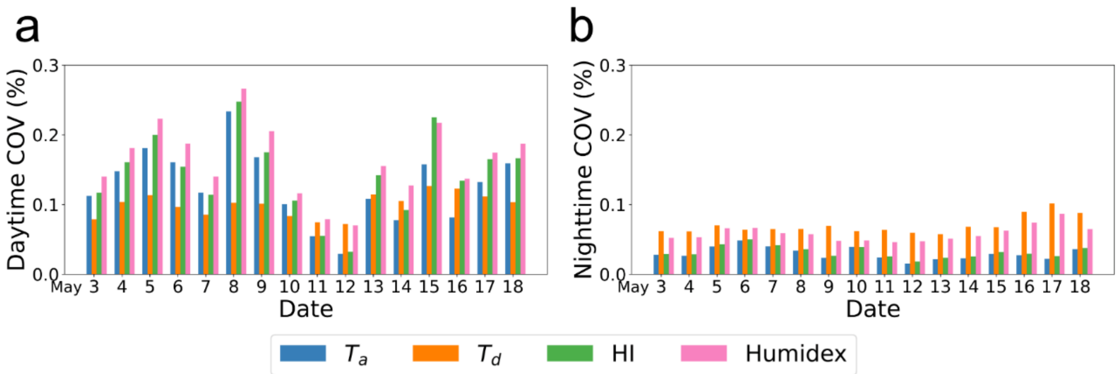

Fig. A6. Coefficient of variation (COV) for the heat stress measurements, including dewpoint temperature (T_d). Panels show (a) daytime measurements and (b) nighttime measurements. The COV is defined as the measurement SD expressed as a percentage of the mean (in Kelvin).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

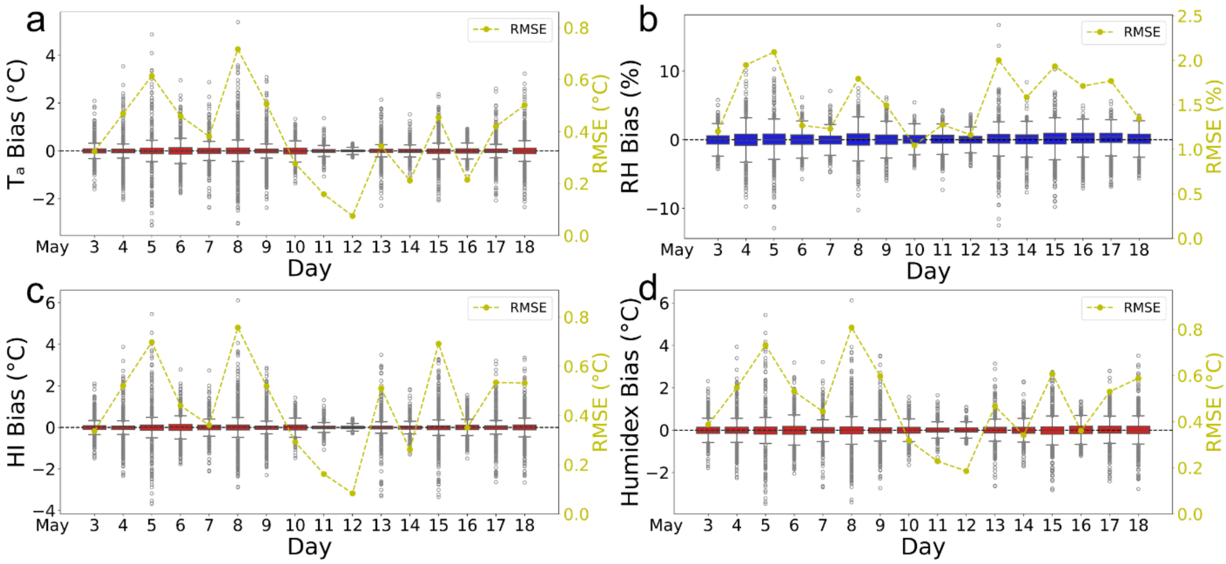

Fig. A7. Absolute bias (box plots, primary y-axis) and root mean square error (RMSE; lines, secondary y-axis) for sensors deployed in the urban lawn setting. Box plots show the median, interquartile range (IQR), whiskers (1.5× IQR), and outliers. Results are shown for (a) T_a (°C), (b) RH (%), (c) HI (°C), and (d) Humidex.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0222.1.

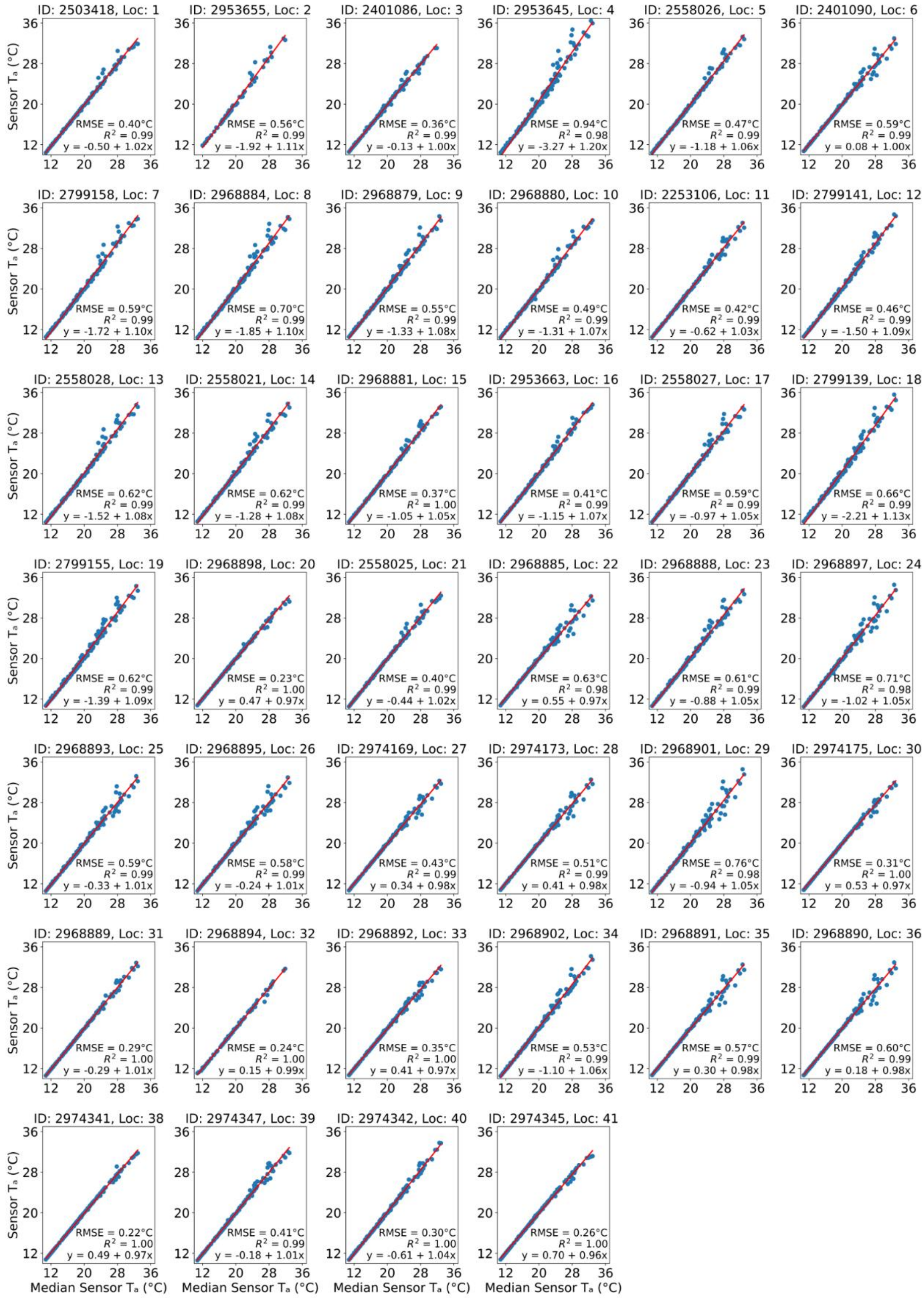

Fig. A8. Least squares regression fits of T_a for each individual sensor against the median sensor. Each subplot shows the sensor ID and its location number. Regression statistics, including the root mean square error RMSE, R^2, and the fitted linear equation, are displayed on each plot.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

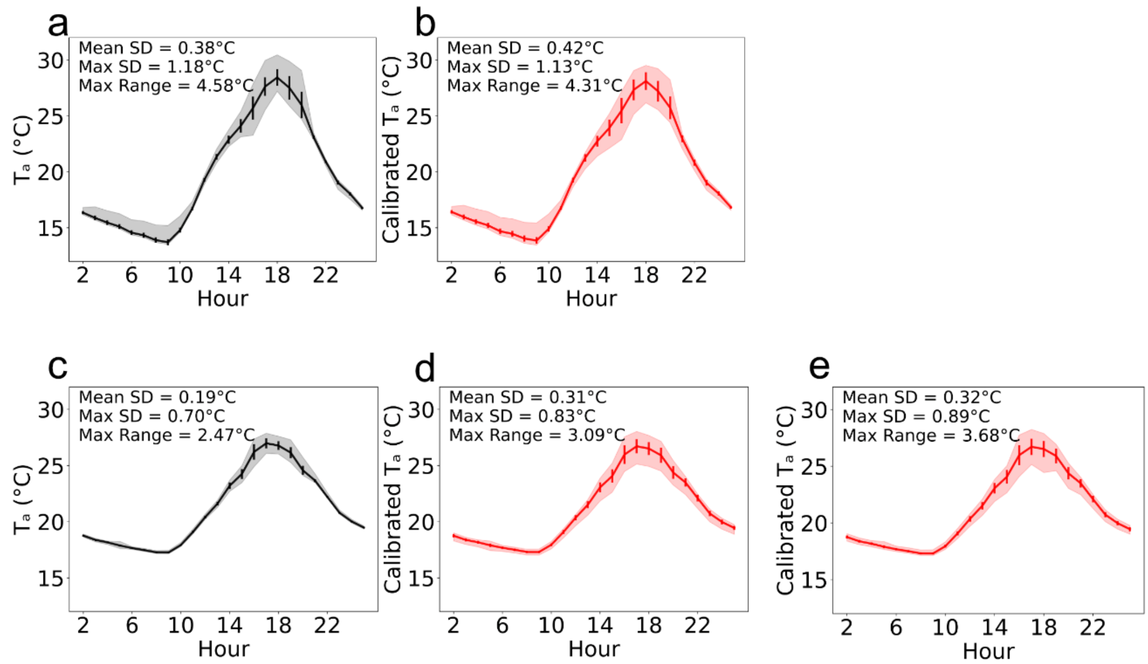

Fig. A9. Mean T_a before (black) and after (red) sensor calibration based on urban lawn from 3-8 May and tested on 10-18 May. Error bars indicate SD across sensors, and shaded areas show the full sensor range for (a) 3-8 May, uncalibrated, (b) 3-8 May calibration applied, (c) 10-18 May uncalibrated, (d) 10-18 May, with 3-8 May calibration applied based on sensor ID, and (e) 10-18 May calibration with 3-8 May applied based on sensor location/position. SVFs for the sensor positions shown in Fig. A2.

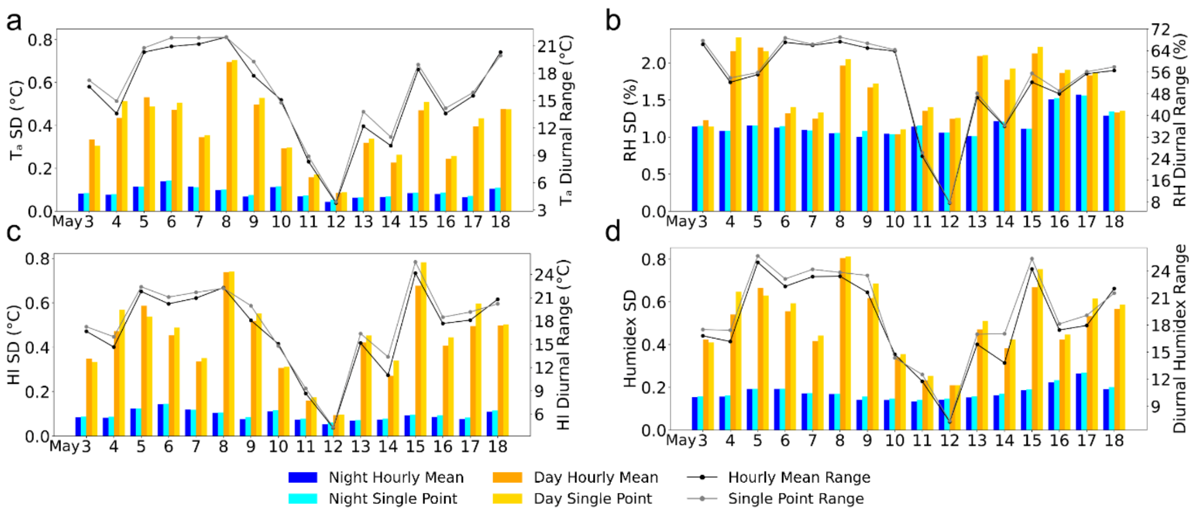

Fig. A10. Comparison of measurements derived from six 10-min observations (hourly mean) and a single top-of-the-hour observation across 41 sensors. Bars show the SD under daytime and nighttime conditions, while lines indicate the diurnal range of each variable. Results are shown for (a) T_a, (b) HI, (c) Humidex, and (d) RH.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0222.1.

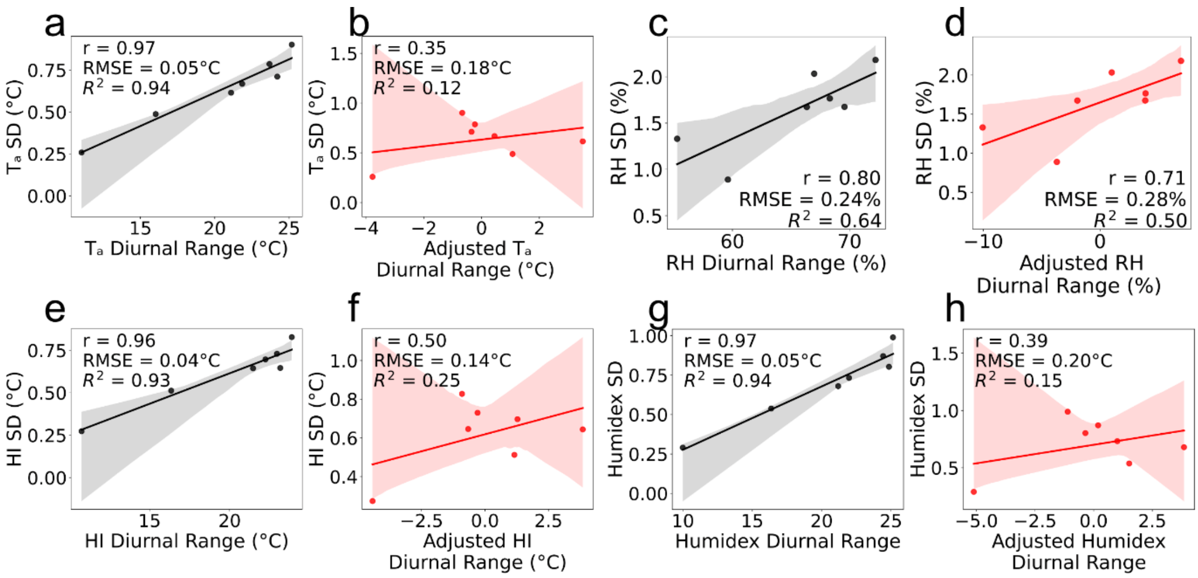

Fig. A11. Relationships between diurnal range and daytime measurement SD for T_a, RH, HI and Humidex during the rooftop deployment. Panels show relationships using the raw diurnal range (a, c, e, g) and the adjusted diurnal range (b, d, f, h, equation (5)). Least-squares regression lines with 95% confidence intervals are shown (gray for unadjusted relationships and red for adjusted relationships). Values shown within each panel include Pearson's correlation coefficient (r) and the fitted regression root mean square error (RMSE) and coefficient of determination (R^2). Panels correspond to (a-b) Ta, (c-d) RH, (e-f) HX, and (g-h) HI. Pearson correlation coefficients for relationships between diurnal range and nighttime SD are provided in Table A3.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0222.1.

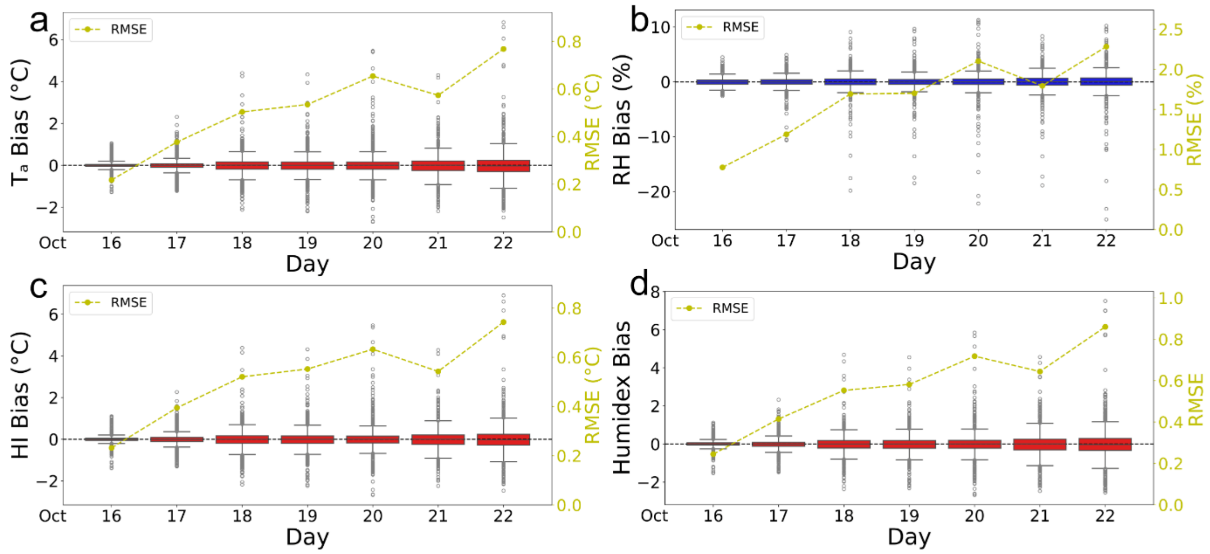

Fig. A12. Absolute bias (box plots, primary y-axis) and root mean square error (RMSE; lines, secondary y-axis) for sensors deployed in the rooftop setting. Box plots show the median, interquartile range (IQR), whiskers (1.5× IQR), and outliers. Results are shown for (a) T_a (°C), (b) RH (%), (c) HI (°C), and (d) Humidex.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

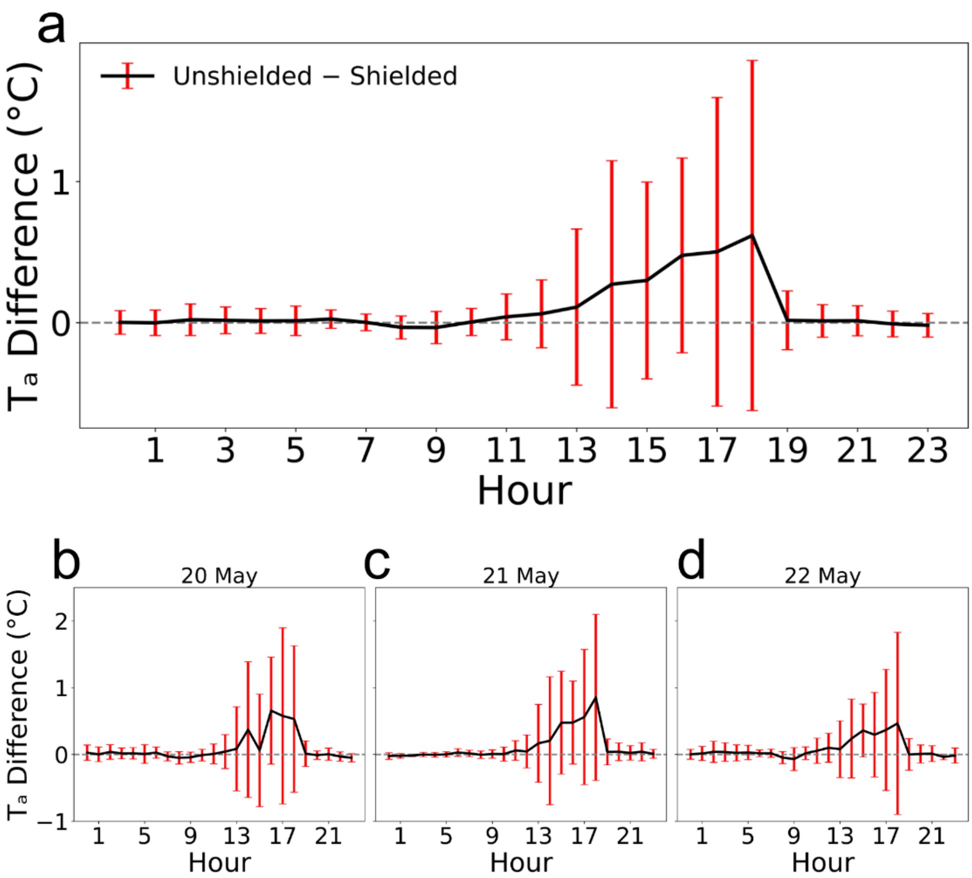

Fig. A13. Mean T_a difference between shielded and unshielded sensors with 95% confidence intervals, calculated using differences between adjacent sensor pairs. (a) average differences over the 3-day period. Daily differences for (b) 20 May, c) 21 May, and d) 22 May.

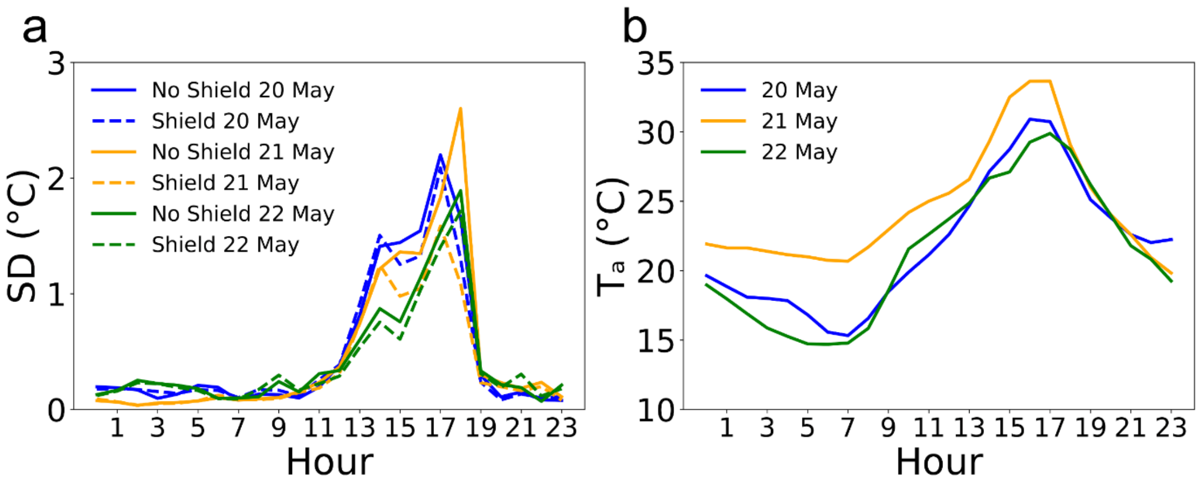

Fig. A14. (a) SD of temperature for each day sensors were deployed in alternating shielded and unshielded setups. (b) Mean T_a diurnal cycle for the corresponding days.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0222.1.

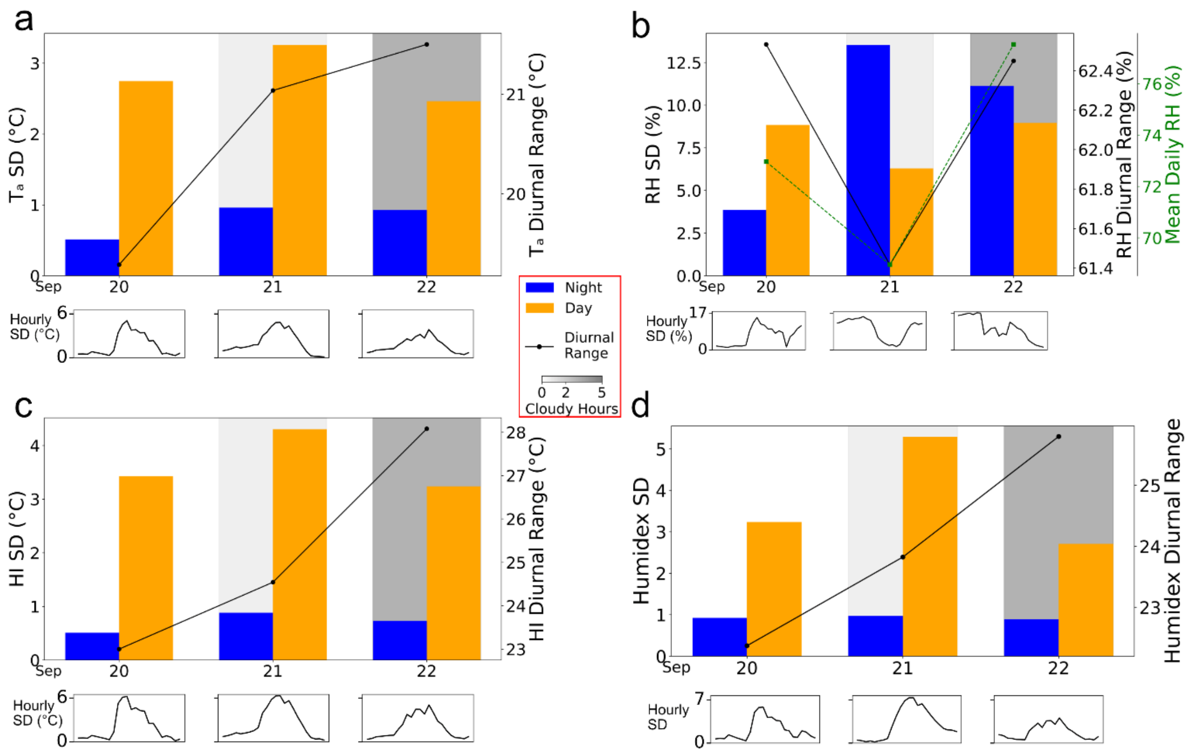

Fig. A15. Nighttime and daytime SD (bars, primary y axis) and diurnal range (lines, secondary y axis) for 41 sensors each day in the rooftop location near HVAC: (a) T_a, (b) RH with a third y axis (green) showing mean daily RH, (c) HI, and (d) Humidex. Small hourly plots beneath bars show the shape of the diurnal cycle of SD for the day. The number of cloudy hours in a day (where the sky is 50%+ covered by cloud), obtained from ECONet KRDU airport weather station, is indicated by gray shading.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

	Chamber Setpoint		Chamber Actual		Sensor Measurement	
Hours	T_a (°C)	RH (%)	T_a (°C)	RH (%)	T_a (°C)	RH (%)
1 – 2	30	90	29.7	86.5	29.6	83.5
3 – 4	30	65	29.7	63.5	29.5	61.4
5 – 6	30	40	28.4	40.6	28.1	37.5
7 – 8	25	40	24.9	42.3	24.6	38.7
9 – 10	25	65	24.9	66.5	24.8	63.2
11 – 12	25	90	24.7	88.4	24.5	84.7
13 – 14	20	90	20.0	86.9	19.9	84.6
15 – 16	20	65	19.9	63.5	19.7	61.8
17 – 18	20	40	19.9	44.0	19.7	41.3

Table A1. Plant growth chamber T_a and RH setpoints, corresponding chamber actual values, alongside mean sensor observations recorded in the setpoint period. Actual chamber conditions differed from setpoints because RH was maintained through water misting in the chamber, which caused fluctuations in both T_a and RH. Data during the initial transition to the first setpoint was excluded, and all subsequent measurements were retained.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

Deployment	Location	Sensor shielding	Measurement period	Sensor height, spacing	Number of sensors	Additional details
(a)	Urban Lawn	Shielded	3 – 18 May 2025	1.5 m, 20 cm	41	On 9th May sensor order permutated SVFs varied from 0.35 to 0.66
(b)	Rooftop	Shielded	16 – 22 October 2024	7 cm, 50 cm	41	Very unshaded (SVFs 0.99+)
(c)	Urban Lawn	Alternate shielded/ unshielded	19 – 23 May 2025	1.5 m, 20 cm	37	Same SVF distribution as (a). Four sensors had missing data, meaning 37 used in analysis.
(d)	Rooftop	Shielded	20 – 22 September 2024	7 cm, 50 cm	41	Near HVAC unit Very unshaded (SVFs 0.96+)
Indoor	Inside building	Unshielde d	20 – 21 November 2024	n/a, 8 cm	41	Inside heated room in building
(A1d)	Plant Growth Chamber	Unshielde d	11 – 12 February 2026	n/a, 8 cm	41	R2 plant growth chamber
(A1e)	Horace Williams Airport	Shielded	17 – 18 February 2026	1.5 m, 20 cm	41	Research grade comparison

Table A2. Summary of sensor deployment configurations used in the experiments outlined in Fig. 1.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

	Measurement	Unadjusted correlation		Mean adjusted correlation	
		Lawn	Rooftop	Lawn	Rooftop
Day	AT	**0.86**	**0.97**	**0.83**	0.35
	HI	**0.87**	**0.96**	**0.78**	0.50
	Humidex	**0.85**	**0.97**	**0.84**	0.39
	RH	0.03	**0.80**	**0.54**	0.71
Night	AT	**0.78**	**0.93**	**0.85**	0.26
	HI	**0.69**	**0.90**	**0.83**	0.26
	Humidex	0.39	**0.90**	0.38	0.23
	RH	-0.03	**0.81**	-0.23	**0.86**

Table A3. Pearson's correlation coefficients between measurement SD and diurnal range or mean-adjusted diurnal range (see methods). Coefficients which are significant based on a threshold $p<0.05$ are highlighted in bold.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

REFERENCES

Abdinoor, J. A., and Coauthors, 2025: Performance of Low-Cost Air Temperature Sensors and Applied Calibration Techniques—A Systematic Review. *Atmosphere*, **16**, 842, https://doi.org/10.3390/atmos16070842.

Ahmed, S., and B. Jänicke, 2025: Comparison and assessment of air temperature observations using All-In-One and Low-Cost sensors in an urban environment. *Trans. Earth Environ. Sustain.*, **3**, 28–50, https://doi.org/10.1177/2754124X251363130.

Alerskans, E., C. Lussana, T. N. Nipen, and I. A. Seierstad, 2022: Optimizing Spatial Quality Control for a Dense Network of Meteorological Stations, https://doi.org/10.1175/JTECH-D-21-0184.1.

Allen, L., F. Lindberg, and C. S. B. Grimmond, 2011: Global to city scale urban anthropogenic heat flux: model and variability. *Int. J. Climatol.*, **31**, 1990–2005, https://doi.org/10.1002/joc.2210.

Bailey, E., C. Fuhrmann, J. Runkle, S. Stevens, M. Brown, and M. Sugg, 2020: Wearable sensors for personal temperature exposure assessments: A comparative study. *Environ. Res.*, **180**, 108858, https://doi.org/10.1016/j.envres.2019.108858.

Beele, E., M. Reyniers, R. Aerts, and B. Somers, 2022: Quality control and correction method for air temperature data from a citizen science weather station network in Leuven, Belgium. *Earth Syst. Sci. Data*, **14**, 4681–4717, https://doi.org/10.5194/essd-14-4681-2022.

Bell, S., D. Cornford, and L. Bastin, 2015: How good are citizen weather stations? Addressing a biased opinion. *Weather*, **70**, 75–84, https://doi.org/10.1002/wea.2316.

Chakraborty, T., C. Sarangi, and S. N. Tripathi, 2017: Understanding Diurnality and Inter-Seasonality of a Sub-tropical Urban Heat Island. *Bound.-Layer Meteorol.*, **163**, 287–309, https://doi.org/10.1007/s10546-016-0223-0.

Chakraborty, T., Z. S. Venter, Y. Qian, and X. Lee, 2022: Lower Urban Humidity Moderates Outdoor Heat Stress. *AGU Adv.*, **3**, e2022AV000729, https://doi.org/10.1029/2022AV000729.

Chakraborty, T., A. J. Newman, Y. Qian, A. Hsu, and G. Sheriff, 2023: Residential segregation and outdoor urban moist heat stress disparities in the United States. *One Earth*, **6**, 738–750, https://doi.org/10.1016/j.oneear.2023.05.016.

Chapman, L., C. L. Muller, D. T. Young, E. L. Warren, C. S. B. Grimmond, X.-M. Cai, and E. J. S. Ferranti, 2015: The Birmingham Urban Climate Laboratory: An Open Meteorological Test Bed and Challenges of the Smart City. *Bull. Am. Meteorol. Soc.*, **96**, 1545–1560, https://doi.org/10.1175/BAMS-D-13-00193.1.

Chapman, L., C. Bell, and S. Bell, 2017: Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. *Int. J. Climatol.*, **37**, 3597–3605, https://doi.org/10.1002/joc.4940.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

Chen, J., Y. Qian, T. C. Chakraborty, and Z. Yang, 2024: Complexities of urban impacts on long-term seasonal trends in a mid-sized arid city. *Environ. Res. Commun.*, **6**, 021004, https://doi.org/10.1088/2515-7620/ad2b18.

Cornes, R. C., M. Dirksen, and R. Sluiter, 2020: Correcting citizen-science air temperature measurements across the Netherlands for short wave radiation bias. *Meteorol. Appl.*, **27**, e1814, https://doi.org/10.1002/met.1814.

Couzo, E., A. Valencia, and P. Gittis, 2024: Evaluation and Correction of PurpleAir Temperature and Relative Humidity Measurements. *Atmosphere*, **15**, 415, https://doi.org/10.3390/atmos15040415.

da Cunha, A. R., 2015: Evaluation of measurement errors of temperature and relative humidity from HOBO data logger under different conditions of exposure to solar radiation. *Environ. Monit. Assess.*, **187**, 236, https://doi.org/10.1007/s10661-015-4458-x.

Danniswari, D., T. Honjo, and K. Furuya, 2022: Analysis of Building Height Impact on Land Surface Temperature by Digital Building Height Model Obtained from AW3D30 and SRTM. *Geographies*, **2**, 563–576, https://doi.org/10.3390/geographies2040034.

Demuzere, M., J. Kittner, A. Martilli, G. Mills, C. Moede, I. D. Stewart, J. van Vliet, and B. Bechtel, 2022: A global map of local climate zones to support earth system modelling and urban-scale environmental science. *Earth Syst. Sci. Data*, **14**, 3835–3873, https://doi.org/10.5194/essd-14-3835-2022.

Diamond, H. J., and Coauthors, 2013: U.S. Climate Reference Network after One Decade of Operations: Status and Assessment. *Bull. Am. Meteorol. Soc.*, **94**, 485–498, https://doi.org/10.1175/BAMS-D-12-00170.1.

Elnabawi, M. H., N. Hamza, and R. Raveendran, 2023: 'Super cool roofs': Mitigating the UHI effect and enhancing urban thermal comfort with high albedo-coated roofs. *Results Eng.*, **19**, 101269, https://doi.org/10.1016/j.rineng.2023.101269.

Fan, C., and Coauthors, 2017: Understanding the Impact of Urbanization on Surface Urban Heat Islands—A Longitudinal Analysis of the Oasis Effect in Subtropical Desert Cities. *Remote Sens.*, **9**, 672, https://doi.org/10.3390/rs9070672.

Fenner, D., A. Holtmann, F. Meier, I. Langer, and D. Scherer, 2019: Contrasting changes of urban heat island intensity during hot weather episodes. *Environ. Res. Lett.*, **14**, 124013, https://doi.org/10.1088/1748-9326/ab506b.

Fenner, D., B. Bechtel, M. Demuzere, J. Kittner, and F. Meier, 2021: CrowdQC+—A Quality-Control for Crowdsourced Air-Temperature Observations Enabling World-Wide Urban Climate Applications. *Front. Environ. Sci.*, **9**, 720747, https://doi.org/10.3389/fenvs.2021.720747.

Georgescu, M., M. Moustaoui, A. Mahalov, and J. Dudhia, 2011: An alternative explanation of the semiarid urban area "oasis effect". *J. Geophys. Res. Atmospheres*, **116**, https://doi.org/10.1029/2011JD016720.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

Gubler, M., A. Christen, J. Remund, and S. Brönnimann, 2021: Evaluation and application of a low-cost measurement network to study intra-urban temperature differences during summer 2018 in Bern, Switzerland. *Urban Clim.*, **37**, 100817, https://doi.org/10.1016/j.uclim.2021.100817.

Guide to Meteorological Instruments and Methods of Observation, 2023. World Meteorological Organization.

Hamdi, R., and G. Schayes, 2008: Sensitivity study of the urban heat island intensity to urban characteristics. *Int. J. Climatol.*, **28**, 973–982, https://doi.org/10.1002/joc.1598.

Harrison, R. G., 2010: Natural ventilation effects on temperatures within Stevenson screens. *Q. J. R. Meteorol. Soc.*, **136**, 253–259, https://doi.org/10.1002/qj.537.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Howard, L., 1833: The Climate of London: Deduced from Meteorological Observations Made in the Metropolis and at Various Places around It. *Harvey Darton Lond.*

Hsu, A., G. Sheriff, T. Chakraborty, and D. Manya, 2021: Disproportionate exposure to urban heat island intensity across major US cities. *Nat. Commun.*, **12**, 2721, https://doi.org/10.1038/s41467-021-22799-5.

Jenkins, G., 2014: A comparison between two types of widely used weather stations. *Weather*, **69**, 105–110, https://doi.org/10.1002/wea.2158.

Johnson, D. P., 2022: Population-Based Disparities in U.S. Urban Heat Exposure from 2003 to 2018. *Int. J. Environ. Res. Public. Health*, **19**, 12314, https://doi.org/10.3390/ijerph191912314.

Karanja, J., J. Vieira, and J. Vanos, 2023: Sheltered from the heat? How tents and shade covers may unintentionally increase air temperature exposures to unsheltered communities. *Public Health Pract.*, **6**, 100450, https://doi.org/10.1016/j.puhip.2023.100450.

Kestrel, 2025. *Nielsen-Kellerman*. Accessed 31 July 2025, https://kestrelmeters.com/kestrel-drop-d2-humidity-logger.

La Crosse Technology, 2026: 925-1418 Sensor Weather Shield. *Crosse Technol.* Accessed 8 January 2026, https://www.lacrossetechnology.com/products/925-1418.

Li, J., and Coauthors, 2023: Satellite-Based Ranking of the World's Hottest and Coldest Cities Reveals Inequitable Distribution of Temperature Extremes. *Bull. Am. Meteorol. Soc.*, **104**, E1268–E1281, https://doi.org/10.1175/bams-d-22-0233.1.

Li, Y., T. Yang, G. Zhao, C. Ma, Y. Yan, Y. Xu, L. Wang, and L. Wang, 2024: A systematic review of studies involving canopy layer urban heat island: Monitoring and associated factors. *Ecol. Indic.*, **158**, 111424, https://doi.org/10.1016/j.ecolind.2023.111424.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

Martilli, A., E. S. Krayenhoff, and N. Nazarian, 2020: Is the Urban Heat Island intensity relevant for heat mitigation studies? *Urban Clim.*, **31**, 100541, https://doi.org/10.1016/j.uclim.2019.100541.

Masterton, J. M., and F. Richardson, 1981: Humidex: a method of quantifying human discomfort due to excessive heat and humidity.

Mauder, M., R. L. Desjardins, Z. Gao, and R. van Haarlem, 2008: Errors of Naturally Ventilated Air Temperature Measurements in a Spatial Observation Network. *J. Atmospheric Ocean. Technol.*, **25**, 2145–2151, https://doi.org/10.1175/2008JTECHA1046.1.

McGregor, G. R., and J. K. Vanos, 2018: Heat: a primer for public health researchers. *Public Health*, **161**, 138–146.

Meier, F., D. Fenner, T. Grassmann, M. Otto, and D. Scherer, 2017: Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Clim.*, **19**, 170–191, https://doi.org/10.1016/j.uclim.2017.01.006.

Meili, N., A. Paschalis, G. Manoli, and S. Fatichi, 2022: Diurnal and seasonal patterns of global urban dry islands. *Environ. Res. Lett.*, **17**, 054044, https://doi.org/10.1088/1748-9326/ac68f8.

Middel, A., J. Lukasczyk, R. Maciejewski, M. Demuzere, and M. Roth, 2018: Sky View Factor footprints for urban climate modeling. *Urban Clim.*, **25**, 120–134, https://doi.org/10.1016/j.uclim.2018.05.004.

Mohajerani, A., J. Bakaric, and T. Jeffrey-Bailey, 2017: The urban heat island effect, its causes, and mitigation, with reference to the thermal properties of asphalt concrete. *J. Environ. Manage.*, **197**, 522–538, https://doi.org/10.1016/j.jenvman.2017.03.095.

Mukhopadhyay, B., and C. A. Weitz, 2022: Heat Exposure, Heat-Related Symptoms and Coping Strategies among Elderly Residents of Urban Slums and Rural Vilages in West Bengal, India. *Int. J. Environ. Res. Public. Health*, **19**, 12446, https://doi.org/10.3390/ijerph191912446.

Muller, C. L., L. Chapman, C. S. B. Grimmond, D. T. Young, and X. Cai, 2013: Sensors and the city: a review of urban meteorological networks. *Int. J. Climatol.*, **33**, 1585–1600, https://doi.org/10.1002/joc.3678.

Muñoz Sabater, J., 2019: ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS)., https://doi.org/10.24381/cds.e2161bac.

Nakamura, R., and L. Mahrt, 2005: Air Temperature Measurement Errors in Naturally Ventilated Radiation Shields. *J. Atmospheric Ocean. Technol.*, **22**, 1046–1058, https://doi.org/10.1175/JTECH1762.1.

Nan, F., C. Zeng, H. Shen, and L. Lin, 2025: Calibration of Integrated Low-Cost Environmental Sensors for Urban Air Temperature Based on Machine Learning. *Sensors*, **25**, 3398, https://doi.org/10.3390/s25113398.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

Napoly, A., T. Grassmann, F. Meier, and D. Fenner, 2018: Development and Application of a Statistically-Based Quality Control for Crowdsourced Air Temperature Data. *Front. Earth Sci.*, **6**, 118, https://doi.org/10.3389/feart.2018.00118.

Narayanan, A., M. Rezaali, E. L. Bunting, and D. Keellings, 2025: It's getting hot in here: Spatial impact of humidity on heat wave severity in the U.S. *Sci. Total Environ.*, **963**, 178397, https://doi.org/10.1016/j.scitotenv.2025.178397.

Nielsen-Kellerman., 2020: Kestrel DROP wireless environmental data loggers: Certificate of conformity and product specifications.

NOAA, 2022: Heat Index Equation. Accessed 30 January 2026, https://www.wpc.ncep.noaa.gov/html/heatindex_equation.shtml.

Nogueira, M., A. Hurduc, S. Ermida, D. C. A. Lima, P. M. M. Soares, F. Johannsen, and E. Dutra, 2022: Assessment of the Paris urban heat island in ERA5 and offline SURFEX-TEB (v8.1) simulations using the METEOSAT land surface temperature product. *Geosci. Model Dev.*, **15**, 5949–5965, https://doi.org/10.5194/gmd-15-5949-2022.

Oke, T. R., 1981: Canyon geometry and the nocturnal urban heat island: Comparison of scale model and field observations. *J. Climatol.*, **1**, 237–254, https://doi.org/10.1002/joc.3370010304.

Oke, T. R., 1982: The energetic basis of the urban heat island. *Q. J. R. Meteorol. Soc.*, **108**, 1–24, https://doi.org/10.1002/qj.49710845502.

Oke, T. R., G. T. Johnson, D. G. Steyn, and I. D. Watson, 1991: Simulation of surface urban heat islands under 'ideal' conditions at night part 2: Diagnosis of causation. *Bound.-Layer Meteorol.*, **56**, 339–358, https://doi.org/10.1007/BF00119211.

Oke, T. R., G. Mills, A. Christen, and J. A. Voogt, 2017: *Urban Climates*. Cambridge university press.

Pearson, K., 1895: VII. Note on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.*, **58**, 240–242, https://doi.org/10.1098/rspl.1895.0041.

Peel, M. C., B. L. Finlayson, and T. A. McMahon, 2007: Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, **11**, 1633–1644, https://doi.org/10.5194/hess-11-1633-2007.

Rothfusz, L. P., 1990: The Heat Index 'Equation' (or, More Than You Ever Wanted to Know About Heat Index). *Natl. Ocean. Atmospheric Adm. Natl. Weather Serv. Off. Meteorol.*

Segura, R., E. S. Krayenhoff, A. Martilli, A. Badia, C. Estruch, S. Ventura, and G. Villalba, 2022: How do street trees affect urban temperatures and radiation exchange? Observations and numerical evaluation in a highly compact city. *Urban Clim.*, **46**, 101288, https://doi.org/10.1016/j.uclim.2022.101288.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

Shandas, V., J. Voelkel, J. Williams, and J. Hoffman, 2019: Integrating Satellite and Ground Measurements for Predicting Locations of Extreme Urban Heat. *Climate*, **7**, 5, https://doi.org/10.3390/cli7010005.

Sherwood, S. C., 2018: How Important Is Humidity in Heat Stress? *J. Geophys. Res. Atmospheres*, **123**, 11,808-11,810, https://doi.org/10.1029/2018JD028969.

Shi, R., B. F. Hobbs, B. F. Zaitchik, D. W. Waugh, A. A. Scott, and Y. Zhang, 2021: Monitoring intra-urban temperature with dense sensor networks: Fixed or mobile? An empirical study in Baltimore, MD. *Urban Clim.*, **39**, 100979, https://doi.org/10.1016/j.uclim.2021.100979.

State Climate Office of North Carolina, NC State University, 2026: Cardinal [data retrieval interface] available at https://products.climate.ncsu.edu/cardinal/request. Accessed Month 02, 2026.

Steadman, R. G., 1979: The Assessment of Sultriness. Part I: A Temperature-Humidity Index Based on Human Physiology and Clothing Science. *J. Appl. Meteorol.*, **18**, 861–873, https://doi.org/10.1175/1520-0450(1979)018%3C0861:TAOSPI%3E2.0.CO;2.

Stewart, I. D., 2011: A systematic review and scientific critique of methodology in modern urban heat island literature. *Int. J. Climatol.*, **31**, 200–217, https://doi.org/10.1002/joc.2141.

Steyn, D. G., 1980: The calculation of view factors from fisheye-lens photographs: Research note. *Atmosphere-Ocean*, **18**, 254–258, https://doi.org/10.1080/07055900.1980.9649091.

Sun, X., S. Yan, B. Wang, L. Xia, Q. Liu, and H. Zhang, 2015: Air Temperature Error Correction Based on Solar Radiation in an Economical Meteorological Wireless Sensor Network. *Sensors*, **15**, 18114–18139, https://doi.org/10.3390/s150818114.

Venter, Z. S., T. Chakraborty, and X. Lee, 2021: Crowdsourced air temperatures contrast satellite measures of the urban heat island and its mechanisms. *Sci. Adv.*, **7**, eabb9569, https://doi.org/10.1126/sciadv.abb9569.

Wang, M., and H. Xu, 2021: The impact of building height on urban thermal environment in summer: A case study of Chinese megacities. *PLOS ONE*, **16**, e0247786, https://doi.org/10.1371/journal.pone.0247786.

Wang, X., A. Hsu, and T. Chakraborty, 2023: Citizen and machine learning-aided high-resolution mapping of urban heat exposure and stress. *Environ. Res. Infrastruct. Sustain.*, **3**, 035003, https://doi.org/10.1088/2634-4505/acef57.

World Meteorological Organisation, 2023a: Guidance on Measuring, Modelling and Monitoring the Canopy Layer Urban Heat Island (CL-UHI).

World Meteorological Organisation, 2023b: Guide to Meteorological Instruments and Methods of Observation.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.

Yamamoto, K., T. Togami, N. Yamaguchi, and S. Ninomiya, 2017: Machine Learning-Based Calibration of Low-Cost Air Temperature Sensors Using Environmental Data. *Sensors*, **17**, 1290, https://doi.org/10.3390/s17061290.

Young, D. T., L. Chapman, C. L. Muller, X.-M. Cai, and C. S. B. Grimmond, 2014: A Low-Cost Wireless Temperature Sensor: Evaluation for Use in Environmental Monitoring Applications. *J. Atmospheric Ocean. Technol.*, **31**, 938–944, https://doi.org/10.1175/JTECH-D-13-00217.1.

Yulizar, D., S. Soekirno, N. Ananda, M. A. Prabowo, I. F. P. Perdana, and D. Aofany, 2023: Performance Analysis Comparison of DHT11, DHT22 and DS18B20 as Temperature Measurement. *Adv. Phys. Res.*, 37–45, https://doi.org/10.2991/978-94-6463-232-3_5.

Zumwald, M., B. Knüsel, D. N. Bresch, and R. Knutti, 2021: Mapping urban temperature using crowd-sensing data and machine learning. *Urban Clim.*, **35**, 100739, https://doi.org/10.1016/j.uclim.2020.100739.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0222.1.