Waiting in the Heat: a low-cost data collection method to evaluate bus stop environmental exposure

Abstract

Urban transit environments are critical sites of heat and air pollution exposure, yet lack continuous, high-resolution measurements. Designing an effective urban sensing campaign requires balancing device suitability, environmental threats, and location assessment in public spaces. This paper presents an active sensing framework for monitoring thermal and air quality conditions at bus shelters, combining controlled experimentation with in situ deployment. We combine the design and evaluation of a 3-D printed modification to low-cost PurpleAir sensors for installation in tamper prone transit environments with assessment of sensor quality relative to reference monitoring equipment. Using a sequential field experiment, we quantify how configuration influences temperature and particulate matter (PM) measurements and demonstrate that biases introduced by sheltered deployment are systematic and correctable. A deployment-specific calibration model reduces mean absolute error in temperature measurements to 1.03°C under shaded, shelled conditions, without requiring sensor-specific refitting. For $PM_{2.5}$, measurement error is primarily driven by relative humidity, and incorporating both sensor and reference humidity enables effective bias correction. This low-cost sensor approach enables scalable, exposure-relevant monitoring of transit environments and illustrates how active urban sensing can inform infrastructure design and support data-driven transit design.

Highlights

- 3-D-printed enclosures enable scalable, low-cost sensor deployment in public spaces
- Calibration resolves accuracy–practicality trade-offs in active urban sensing
- Low-cost sensor bias is predictable and correctable with deployment-specific calibration
- Calibration using non-co-located meteorology maintains strong performance

Keywords

Urban heat, PM2.5, air sensors, 3-D-printed, IoT sensing

Introduction

Extreme heat is one of the most important environmental issues that cities across the globe have been grappling with in the last decade (IPCC, 2021). Rising temperatures, compounded with rapid urbanization and urban deforestation, have made urban heat the unexpected new enemy of urban life (Chen et al., 2022). Although a significant amount of literature has been published on the effects of heat in urban residential and public spaces, there is relatively less evidence of the exposure levels of extreme heat in urban mobility settings, and particularly in bus stations.

Lack of exposure data can mainly be attributed to cities often lacking the resources and technical capacity to develop, test and deploy complex urban sensing infrastructure to adequately measure extreme heat exposure. As a result, most studies have relied on data that provides a limited understanding of exposure levels at bus stations. A first group of studies have prioritized spatial coverage by using coarse measurements to approximate heat exposure at bus stations. For example, studies may use data from a single weather station to represent entire counties or cities (Lanza and Durand, 2021; Miao et al., 2019; Ngo, 2019), use simulated air temperature data at 1km cell resolution (Karner et al., 2015), or 30-meter cell resolution Land Surface Temperature (Brannen and Banerjee, 2025; Braun and Fraser, 2022; Xie et al., 2026). A second group of studies prioritized accuracy using non-stationary research-grade sensors deployed at stations for a limited amount of time (Lanza et al., 2025; Pan et al., 2024). A third group of studies focused on understanding hyper-local fine scale heat conditions by means of environmental microsimulations to estimate 1-meter cell resolution for a single day at different station types (Liu et al., 2025; Noh et al., 2021). Another group of studies deployed environmental sensors on mobile vehicles and busses to characterize heat across the different points of the route at the street level (Shandas et al., 2019; Yin et al., 2020). Finally, several studies have used bus riders' surveys to understand heat perception and identify what elements of stations seem to provide best thermal comfort (Lanza et al., 2023). Although there have been enough studies to generate a recent review paper (Briant et al., 2025), and each of these studies have contributed important pieces of information, there is still the need to accurately track continuous environmental conditions at bus stations to better understand the full extent of exposure and not only at specific days, stations, or times.

This paper proposes a low-cost method to accurately measure and track heat in bus stations. Specifically, we address the lack of continuous, high-resolution exposure data by developing and validating a deployable sensing workflow for operational transit environments. By using an accessible urban sensing infrastructure made of easily deployable low-cost sensors and readily available data back-end architecture systems, we show how this method, combined with data calibration techniques, could serve as a solution to solve the informational gap on extreme heat exposure in bus stations. Building on previous work on environmental sensing at bus stations, this paper aims to evaluate the feasibility and value of long-term environmental monitoring at transit stations while contributing to the literature in three ways. First, by adding PurpleAir sensors to stations, we provide an array of new information in the form of long-term and high-frequency data. Second, by creating a dedicated deployment shell design and testing its effect on sensor readings, we provide a functional solution to installing sensors in bus stations that can be replicated in other cities. Third, we discuss the specific collaboration between university researchers and the transportation agency. These lessons learned can act as a model on how cities can expand their research capacity by creating trust building programs that would help initiating data collection and research efforts that are often underfunded, not because of the lack of institutional willingness, but because of the lack of awareness among policymakers and the general public of the exposure risk levels of public transportation users.

The remainder of the paper is organized as follows. Section 2 discusses the methodology, including three subsections a) detailing the 3-D case design process and considerations for appropriate bus station placement, b) controlled testing phase and procedures, and c) bus station selection, sensor deployment, and validation strategy. Section 3 describes the main results focusing on the differences between settings in both the controlled testing phase and bus station phase. Section 4 concludes and discusses the need for expanding this research efforts to overcome the limited nature of this pilot project, as well as how these preliminary results can be useful to derive actionable policies that would increase both transit infrastructure for considering extreme heat settings, as well as increasing bus frequency during times of extreme heat to minimize the waiting times that passengers may be experiencing under extreme heat conditions. As cities work to improve transportation systems, having adequate data support systems is urgent to further study on best practices for bus shelters (Moore et al., 2014), increase awareness of the risk of exposure to extreme heat, as well as to implement more targeted solutions in areas with highest need.

Methodology

The overarching goal of this project is to determine the potential to use PurpleAir sensors to monitor the environment at bus stops as a scalable low-cost solution for continuous environmental monitoring. For this, we first describe the transit system under study, then introduce the PurpleAir devices themselves, then additional equipment used, and finally the experimental setup, including sensor mounting and validation.

VIA Metropolitan Transit (VIA) serves as the transit service provider for the San Antonio area. VIA's operation includes a commitment to improve infrastructure, which including improvements to more than 1,000 bus stops in less than five years. These improvements are allocated based on quantitative prioritization, a line service design score relying primarily on riders boarding at a stop, but also including average wait time, adjacent land use, and routes serving a stop (Hansen et al., 2021). As these improvements are in progress, the research team has the opportunity to evaluate and provide active feedback on shelter designs.

Continuous measurements of real-world, public environmental and microclimatic conditions present challenges different from traditional ambient air monitoring. Challenges in the measurement plan have included how to assess exposure-relevant conditions without interfering with the activities of the potentially exposed population and how to enhance tamper resistance to promote long term monitoring. To evaluate conditions at public transit stops, detailed measurements have been taken to enable an assessment of whether the proposed testing setup is appropriate for evaluating the conditions experienced by urban transit passengers in the southern United States. The study location is in Bexar County, Texas, which is home to San Antonio, see Figure 1.

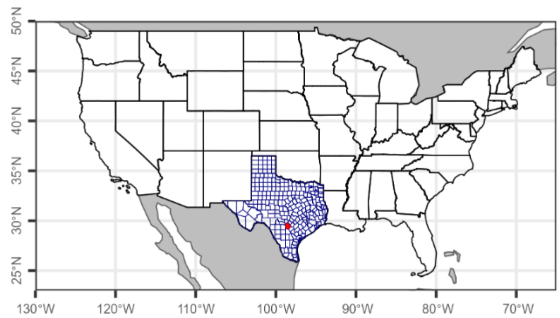

Figure 1. A map of the United States with North American context, highlighting the study area of Bexar County in red, with other counties in Texas outlined in blue.

We present the design of a 3-D shell casing for sensors to be deployed in bus stations. As transportation officials argued based on their experience, elements in bus stations need to be 'tamper-proof', a characteristic incompatible with the open and accessible design nature of PA sensors as its main purpose is to be deployed in private residential settings. PurpleAir sensors were chosen due to several advantages compared with other sensors: a) low-cost, WIFI connectivity, local storage via SD card, free server and back-end dashboard data visualization of data, API for data access, replaceable parts (PM respirators). Several shells were 3-D printed using different filament colors as required by the transportation office to match the color of the bus stations. We then used a sequential-controlled fieldwork setting to test the difference between different sensors, shell colors, and shade configurations. Finally, we deployed sensors with 3-D printed shells in bus stations across San Antonio's heat island neighborhoods. The first set of stations chosen for this study were stations that were considered relevant based on different criteria to both researchers and transportation officials: ridership, orientation, station size, station design, among others. Additional metrics were collected using Kestrel Heat Stress Trackers to validate PA metrics and better understand thermal conditions under bus stations.

Monitoring Equipment

The PurpleAir device measures $PM_{2.5}$ concentrations from 0 to 500 μg/m^3. Ambient particle concentrations are expected to remain within this range, but the error in measurements is noted as 10 μg/m^3. As typical concentrations are on the order of 10 μg/m^3, this error can be significant. The sensors also measure and report air temperature, however the PurpleAir reported temperature overestimates environmental conditions, possibly due to heat generated by internal electronics. Temperature data from PurpleAir was initially adjusted based on Wallace (2023) using the formula in equation 1,

$$T_{adj,F} = 1.0227 \times T_{PA} - 9.3755 \qquad (1)$$

where T_{adj} is the new adjusted temperature in Fahrenheit and T_{PA} is the raw temperature reported from PurpleAir in Fahrenheit. This adjusted temperature was then converted to Celsius. The adjustment aims to decrease error relative to nearby ambient air temperature.

For this study, additional stationary monitoring equipment was used to validate the $PM_{2.5}$ concentration and temperature measurements. A DustTrak 8540 was used as the reference monitor for $PM_{2.5}$ and a Lufft WS500 provides the reference temperature data. The relevant specifications for each instrument are provided in Table 1. Multiple $PM_{2.5}$ options are available from PurpleAir devices, but for this study we used PM2.5 ALT-CF3.4, which is the reported value that most closely aligns with the DustTrak concentrations for the study region.

To determine how representative the stationary sensors are for characterizing the overall bus stop environment, a detailed measurement campaign for some stations was conducted. While the PurpleAir sensors were collecting data, the same sites were measured using a Kestrel model 5400. Kestrels collect more accurate temperature data while also being mobile and enabling for a determination of temperature variability around the station. Kestrels were deployed at locations around the bus station for 10 minutes at a time.

Table 1. Description of the measurement range and error for each of the measurement devices, as reported by the manufacturer. PurpleAir does not report temperature accuracy.

	DustTrak	PurpleAir	Lufft	Kestrel
$PM_{2.5}$ (ug/m3)	0-400,000 ±1	0-500 ± 10	-	-
Temperature.(℃)	-	-40 to 85	-50 to 60 ±0.2	-29.0 to 70.0 ±0.5
Relative.Humidity.(↘)	-	0-100, ±3	0-100 ± 2	10 to 90 ± 2

Shell Design

Despite the much lauded possibilities for citizen science and democratization of data collection upon the introduction of low-cost sensors (Bagkis et al., 2025; Pournaras et al., 2017; Snyder et al., 2013), such data gathering is still inequitably distributed (deSouza and Kinney, 2021; Molla et al., 2025; Sun et al., 2022; West, 2025). Effective sensor deployment requires navigating complex logistical challenges: accessibility for routine maintenance, protection from theft or vandalism, and site-specific installation constraints (US EPA, 2021). These variables create significant barriers that differ from one location to another. However, thoughtfully engineered protective redesign that combines environmental shielding and security features can overcome these limitations. Such integration unlocks new deployment possibilities across diverse landscapes, including high-traffic urban corridors.

Recognizing these deployment barriers, the research team designed a protective shell for the PurpleAir sensor that balances multiple functional requirements. The standard PurpleAir unit, while cost-effective and data-rich, lacks the physical protection necessary for sustained outdoor deployment in challenging environments. Issues include moisture intrusion, USB connector corrosion, temperature-related measurement drift, and physical damage from environmental exposure. Additional concerns associated with publicly accessible deployments include tamper resistance and ensuring stable power connection. The enclosure design therefore aims to mitigate these vulnerabilities while maintaining sensor accuracy. It is essential to preserve natural airflow patterns critical for accurate $PM_{2.5}$ readings, enable straightforward maintenance, and allow flexible mounting options. To translate these design objectives into a practical solution, an iterative development process balanced manufacturing efficiency with functional performance. Details about alternative designs that were evaluated can be found in the appendix.

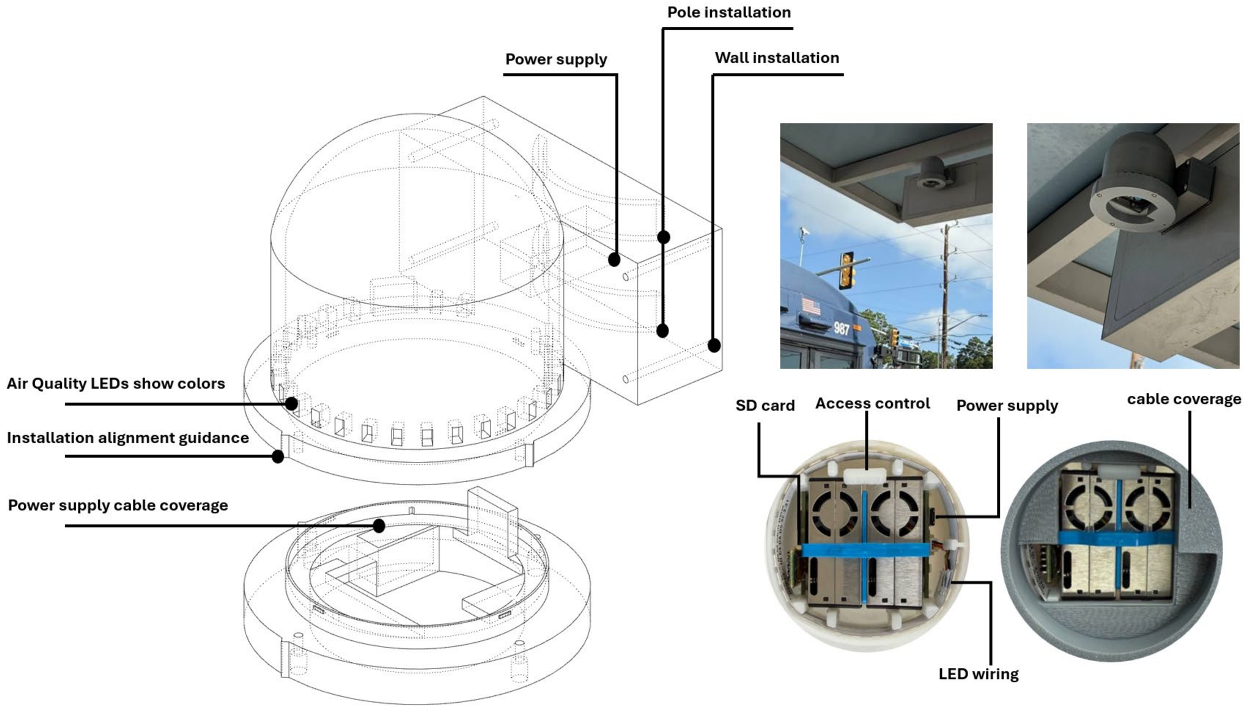

Figure 2. PurpleAir Protective shell Design and Installation Configuration. (Left) CAD model of the 3-D-printed protective shell showing installation alignment guidance, cable management pathways (cable, pole, and wall installation options). (Right) Field deployment at bus station demonstrating pole and surface mounting configurations (top), and internal layout showing SD card access, LED wiring, access control panel, and power supply integration (bottom).

The final protective shell design (Figure 2) integrates multiple solutions tailored to deployment within public transit infrastructure. The design incorporates installation alignment guidance features that simplify positioning and ensure consistent sensor orientation across multiple deployment sites, reducing installation complexity and time. The enclosure supports two distinct mounting configurations, pole installation and surface installation, providing deployment flexibility rather than requiring custom modifications for each site. This adaptability is essential for large-scale deployment in public transit systems, where mounting surfaces, exposure conditions, and accessibility vary widely.

Visual integration is also an important design factor. To ensure that the sensors blend with existing transit infrastructure, the protective shell colors are selected to match those commonly used in bus stop structures. The final products were printed using a Bambu X10 Carbon printer with a durable, lightweight, and cost-effective polylactic acid (PLA) filament in three color options, green, dark gray, and light gray, to enhance compatibility and reduce visibility.

To enhance field durability, security measures are incorporated throughout the design. The shell construction provides impact resistance against accidental contact in high-traffic areas, while the power supply is internally housed and protected from tampering. A flush-mounted design minimizes exposed components that could be damaged or removed. In addition, the case's neutral form and integration with surrounding infrastructure aesthetics help deter unwanted attention and reduce vandalism risk.

Shell Evaluation

To determine whether the shell casing has an effect on measurements taken by the PurpleAir sensor, tests were conducted with four PurpleAir Zen sensors co-located both with one another and with higher fidelity, research-grade monitors (see Table 1). The $PM_{2.5}$ concentrations for the PurpleAir sensors were validated against a stationary DustTrak 8540 environmental monitor (referred to as Reference or DT) with a heated inlet. The temperature and humidity measurements were validated against a Lufft WS500 weather station (referred to as Reference or LF). The six devices were mounted outdoors from July 14, 2025, to August 30, 2025. Additionally, a shade structure was constructed to control for and evaluate the influence of direct sunlight on the measurements. Additional specifications of the testing setup are available in the Appendix.

Testing was conducted in six phases to isolate the effect of the shell. The DustTrak and Lufft, as well as one control PurpleAir sensor, were not modified throughout the experimental phases with the exception of shade being cast on the entire setup. This provides a control to verify whether any differences in measurement can be attributed to the shell. In the first phase, all four PurpleAir devices were initially mounted without any modifications to determine baseline variability between devices (hereafter referred to as No Shell / No Cover; NSNC). Subsequent testing conditions alternate the addition of shade cover and shells, see Table 2. The westernmost PurpleAir sensor (labelled W) hosted the light grey shell during enshelled conditions. The second from the west (WC) hosted the dark grey shell. The second from the east (EC) was never put in a shell as a control measure, and the easternmost PurpleAir sensor (E) hosted the green shell.

Table 2. Testing Conditions

CONDITION	SHELL?	SHADE COVER?	START DATE (INCLUSIVE)*	END DATE (EXCLUSIVE)*
NSNC	No Shell	No Cover	2025-07-14 16:00	2025-07-20 20:00
SNC-B	Shell (backwards)	No Cover	2025-07-22 20:00	2025-07-28 19:00
SC-B	Shell (backwards)	Cover	2025-07-28 19:00	2025-08-03 22:00
NSC	No Shell	Cover	2025-08-03 23:00	2025-08-13 18:00
SC	Shell	Cover	2025-08-13 19:00	2025-08-19 15:00
SNC	Shell	No Cover	2025-08-19 16:00	2025-08-30 23:00

* All datetimes are in Central Daylight Time (UTC -5), the local time for the experiment

Statistical Methods Used

Multiple statistical methods were employed to ensure consistency in data. A residual based error analysis is used to determine whether any combination of shells and shading affects the accuracy of sensor measurements relative to reference instruments (Jiao et al., 2016). To determine the source of variation in the measurements from the PurpleAir sensor compared to the baseline, first a linear regression of the sensor measurement compared to the research-grade equipment (DustTrak for $PM_{2.5}$ and Lufft for Temperature) was calculated. The residuals (R_{ijst}) from these regressions were used as the response variable in subsequent models to determine the most relevant factors contributing systematic deviations from the reference measurements. By modeling the residuals, we focus on factors that explain deviation from the reference instruments rather than changes in the underlying conditions.

Results from two sets of regression models for each measurement type are presented. For temperature, a full model including multiple interaction terms (eq 2) and a reduced core model including only one interaction (eq 3) are presented. For PM, a full model (eq 5) is evaluated and additional analysis using (eq 4) is presented with both the full dataset and two subsets. For the $PM_{2.5}$ sensitivity analyses, the baseline sensor-reference regression was re-fit within the relevant subset before residual models were estimated. Model comparisons are used to determine whether interaction terms materially improve explanatory power. This enables determination of whether there are differences between filaments (colors) and whether the shell itself has an effect. In some models, shell orientation effects (reverse vs as intended) are evaluated explicitly, while in others it is treated as a binary. Accordingly, two encodings of shell configuration are employed, S_k is a three-level factor representing no shell, reverse orientation, and correctly oriented, while S_s is a binary indicator representing presence or absence. Additional models are tested, and results are available in the appendix.

$$R_{ikt} = \beta_0 + \beta_1 P_i + \beta_2 (C_t \times S_{kt}) + \beta_3 (P_i \times S_{kt}) + \beta_4 (P_i \times C_t) + \beta_5 T_t + \beta_7 B_t + f(H_t) + \epsilon_{ikt} \quad (2)$$

$$R_{ist} = \beta_0 + \beta_1 C_t + \beta_2 P_i + \beta_3 S_{st} + \beta_{44} (P_i \times S_{st}) + \beta_5 T_t + \beta_5 B_t + f(H_t) + \epsilon_{ist} \quad (3)$$

$$R_{kt} = \beta_0 + \beta_1 RH_t + \beta_2 RH_r + \beta_3 C_t + \beta_4 S_{kt} + \beta_5 (C_t \times S_{kt}) + \beta_6 T_t + f(H_t) + \epsilon_{kt} \quad (4)$$

$$R_{kt} = \beta_0 + \beta_1 RH_t + \beta_2 RH_r + \beta_3 C_t + \beta_4 S_{kt} + \beta_5 (C_t \times S_{kt}) + \beta_6 P_i + \beta_7 (P_i \times S_{kt}) + \beta_8 T_t + f(H_t) + \epsilon_{kt} \quad (5)$$

where P_i is an identifier for each of the 4 PurpleAir sensors, referenced to EC that never has a shell, C_t is a binary representing whether the shade cover is installed, and S_k and S_s represent shell configuration as described above. RH_t is the relative humidity as measured by the PurpleAir sensor and RH_r is the relative humidity as measured by the reference instrument. Time-of-day is represented both categorically using bins B_t: night, day, dawn, and dusk and continuously using a spline function of hour of day (H_t) with four degrees of freedom. T_t represents the elapsed time (in days) since the start of the experiment to identify drift. This analysis was done in R with the packages emmeans and lm with natural splines implemented using splines. Models were tested that did not include interaction effects and used fewer degrees of freedom for the time of day effect, but they were less explanatory than the final models.

Additional analyses are shown for $PM_{2.5}$ concentrations with a subset of the data that excludes a period of anomalous behavior. The full time series of data collected is shown in Figure 8, and a subset of the time series is shown in Figure 4. The event and its characteristics are described in the results section. Also, for some $PM_{2.5}$ analyses, data for which the DustTrak reported no concentration were excluded. These values are interpreted as a concentration below the detection limit (<DL) rather than a true 0, and therefore the residual behavior may be distorted from an assumption of continuous measurements.

While residual analysis models (Eqs. 2–5) were used to identify drivers of bias, the final calibration model was simplified to include only variables that improved predictive performance and are available in deployment settings. $PM_{2.5}$ calibration model performance was evaluated using leave-one-condition-out validation with independent meteorological inputs. To simulate deployment without co-located reference meteorology, we calculated the calibration performance using a Lufft monitor at a site 27km from the experimental set up. For the conditions in Table 2, the model was trained on all in-shell conditions except the held-out one and then tested on that condition. For the final calibration model, the entire in-shell data set was used to train the model to calculate the coefficients. Training was done using co-located meteorological data.

Additional, non-regressive statistics are assessed. Due to potential issues with data normality, presence of non-linear relationships, and presence of outliers, non-parametric tests were performed on the data (Agresti, 2010). Stuart-Kendall τ-c was calculated for correlation metrics (Stuart, 1955) using the DescTools R package, where the ordinal data assumption was met by the interval temperature and ratio $PM_{2.5}$ concentration, the ability of τ-c to handle ties was deemed useful post-hoc an initial correlation test, and the ability of the τ-c statistic to handle rectangular tables in the event of –1/+1 theoretical correlation was valued. Arbitrary thresholds of Negligible (0.00 - 0.06), Weak (0.06 - 0.26), Moderate (0.26 - 0.49), Strong (0.49 - 0.71), and Very Strong (0.71 - 1.0) were used to evaluate the rough quality of fit of correlations. Correlation difference from 0 was evaluated at a 95% confidence level.

For these, $PM_{2.5}$ data were only evaluated for periods when reference monitor concentrations were non-zero, i.e. >DL. Measures of Mean Absolute Error (MAE) relative to the reference monitor or control sensor, the Root Mean Square of this Error (RMSE), and the correlation between the sensor and reference/control data were analyzed. Additionally, multiple comparison analyses were performed between all sensors and the reference monitors to reveal any significant differences between the data of these devices, and the change or lack thereof of those differences between conditions.

Analysis was performed both to determine accuracy and variability between sensors and to determine sensor accuracy and variability relative to reference monitors. Statistics were performed on subsets of data formed based on variables of interest and knowledge of the physical experimental phenomenon. $PM_{2.5}$ data was grouped on a shell and weekend basis, based on an interest in the shell effect and an anticipation of activity affected weekday $PM_{2.5}$ concentrations. Temperature data was grouped on a shell and shade cover basis, based on, again, an interest in the shell effect and an anticipation of a direct sunlight effect on temperature measurements.

Significant difference between data distributions was calculated using the conservative non-parametric Dunn's multiple comparisons test using the dunn.test R package, with the conservative Bonferroni adjustment to control family-wise error rate. Because the data is continuous in the case of all comparisons, the Dunn's test assumptions are met, with the caveat that the test can only be interpreted as revealing a significant difference in medians when the paired data distributions are identical. Rejection of the assumption that distributions were identical was evaluated at a 95% confidence interval.

Representativeness of Sensor Placement

In addition to evaluating the effect of the shells, an evaluation of the representativeness of the sensor for conditions within each bus station are also considered. This is accomplished by taking a series of measurements using a Kestrel at locations defined in Table 3. The initial monitoring campaign only includes bus stops with a shelter, as the equipment requires constant power and unsheltered stations with sufficient electricity are not currently available.

Table 3. Description of near-stop measurements

	A	B	C	D	E	Z
Measurement Height(s)	1m & 1.5m	1m & 1.5m	1m & 1.5m	1m & 1.5m	1m & 1.5m	~2.5m
Shade	Usually	No	Yes	Yes	No	Yes
Inside shelter?	Yes	No	No	Yes	Yes	Yes
Description	Center of bench	Unshaded, unsheltered	Shaded unsheltered	Shaded in shelter	Unshaded in shelter	At PA Sensor

Most Kestrel measurements are taken at two heights, 1 meter and 1.5 meters (approximately representing seated and standing exposure). The only measurement that diverges from these heights is a measurement (Z) that will be taken as close as possible to the installed sensor. For the stations measured, this distance was between 248 to 259 cm above the ground. Measurements captured the temperature both in and outside of the shelter and in and out of shade. Regardless of shade, a measurement was always taken at the center of the seat. Additionally, locations within the shelter that were shaded and unshaded were measured (if such space existed). Similarly, shaded and unshaded locations outside, but near, the shelter were measured. The measurements outside the shelter were always within 3 meters of the edge of the shelter, as although people may seek shade, they need to ensure that they will be able to board the bus.

Station variation measurements were captured at two pairs of stations that are located adjacent to one another at the same intersection. One intersection represents two of the most upgraded “Prímo” stations (see Figure 3) and measures the stops across the street from each other serving the north- and south-bound lines. These were measured on the same day, August 6, 2025. The other pair of stations are both located at the edge of a university campus near the same intersection, but around the corner from one another. These stations were evaluated on August 14, 2025.

To determine whether these measurement situations indicate a significant microclimatic effect, a linear mixed effects model was developed using the lme4 package in R based on these characteristics. This was tested on both temperature and heat index. The model formula (Eq 6) for each was compared against the measurement height (modeled as a continuous variable in meters), a binary marker of whether the measurement was taken in shade, a binary marker of whether the measurement was taken inside the shelter, the time, and a random intercept for each stop to account for physical variations among locations.

$$K_{ij} = \beta_0 + \beta_1 H_{ij} + \beta_2 L_{ij} + \beta_3 E_{ij} + \beta_4 T_{ij} + b_{0,j} + \epsilon_{ij} \tag{6}$$

Where K_{ij} is the Kestrel measurement for observation *i* at stop *j*, b_{oj} is the random intercept for stop *j*, H is the measurement height in meters, L_{ij} represents light effect (shaded or not), E_{ij} represents the exposure (inside the shelter or not), and T_{ij} represents the time.

Bus Stop Deployment

The transit partner, VIA Metropolitan Transit, has developed a procedure to prioritize bus stop improvements as part of a concerted effort to improve ridership quality that was started in 2014. (Hansen et al., 2021) This calculation considers how many riders board a bus at the stop, the potential wait time for a bus, the number of routes served, and whether the stop is adjacent to important resources such as medical facilities or groceries. Since project inception, more than 1,000 shelters have been added or upgraded to the newer shelters. The studied stations have generally high ridership as stations with higher ridership are more likely to have resources including electricity allocated to them. Some of the newer shelters are at further enhanced Prímo stops (see Figures 3 & 15), which include digital informational displays, artistic shade panels, and free wifi.

Deployment of the encased sensors was determined by attempting to measure multiple shelter configurations while also monitoring several stations in close proximity to one another to evaluate stations experiencing the same microclimate and emission sources. Some station locations have grid connected power, while others have solar panels designed to provide safety lighting to the station. The instrumented stations are located in areas of the city that had previously been identified as particularly vulnerable to heat (Brown et al., 2026). These vulnerable areas are indicated as points in Figure 3. The routes serviced by the instrumented stops provide access to grocery stores, shopping centers, and hospitals.

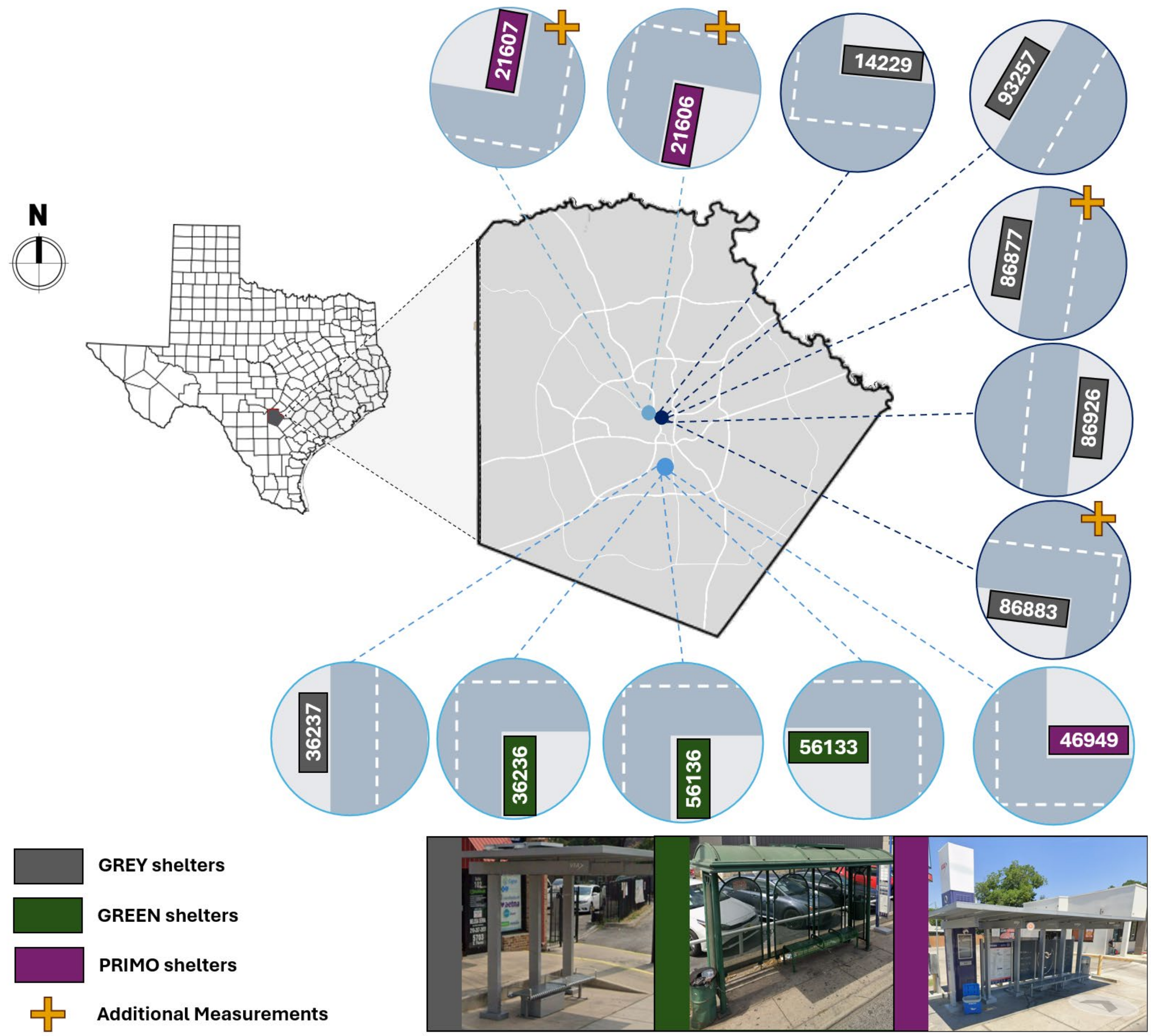

Figure 3. Details about the instrumented stops are provided in a graphical format, indicating which of the heat mitigation areas it is in, which station type is in place, and how it is oriented.

Results

Effect of Shells

For the measurement period (July 14th to August 30th, 2025), the study site experienced conditions typical of the region and season, with daily high Lufft temperatures ranging from 32 to 40°C, overnight lows of 18 to 27°C, and DustTrak $PM_{2.5}$ concentrations ranging from below the limit of detection to 54 μg/m^3.

Sensors have high inter-device $PM_{2.5}$ consistency, with each sensor showing very strong τ-c correlation with respect to the control sensor (0.934 - 0.960), and low, fairly consistent error (MAE: 0.0397 - 0.321 μg/m^3; RMSE: 0.0793 - 0.559 μg/m^3). Relative to the reference monitor, performance decreases as correlation falls to the moderate – strong range (0.323 - 0.513) and error increases (MAE: 2.00 - 3.49 μg/m^3; RMSE: 3.20 - 5.80 μg/m^3) but remains within the design range of ±10 μg/m^3.

Overall, 100% of sensor $PM_{2.5}$ measurements overestimate reference values, and 92.0% of measurements overestimate reference values by less than 10 μg/m³.

Sensor temperature measurements show very strong τ-c correlation between sensors (0.847 - 0.949), and moderate error and variation (MAE: 0.560 - 2.80 °C; RMSE: 1.18 - 4.29 °C). In contrast to $PM_{2.5}$ statistics, performance remains steady relative to the reference monitor, with very strong correlation (0.725 - 0.912) and similar error (MAE: 0.488 - 2.02 °C; RMSE: 0.824 - 3.55 °C).

To evaluate shell effects under these conditions, residual-based linear models are fit to the full data set as well as targeted subsets. Residuals were calculated as the difference between PurpleAir and the reference (i.e., sensor–reference), and were modeled as a function of sensor identity, shell configuration, shading, and temporal covariates. The data filtering and model used are referenced to model names as shown in Table 4. Additional model results for sensitivity analysis are presented in the SI.

The anomalous event describes a time period in which the behavior of all sensors deviates significantly from the rest of the testing period. From Figure 4, it is clear that an increased concentration of $PM_{2.5}$ occurs as it is identified on all PurpleAir devices as well as the reference DustTrak (DT) monitor. However, the error for this event is significantly larger than for all other events during the experimental setup. Figure 4b plots all sensor measurements against the reference concentration, and then plots all data points for the event time period (within the red square in Figures 4a & 8) with solid points. It is clear that the behavior for this event differs in a way that is not aligned solely with the typical high bias or factors being tested in the experiment.

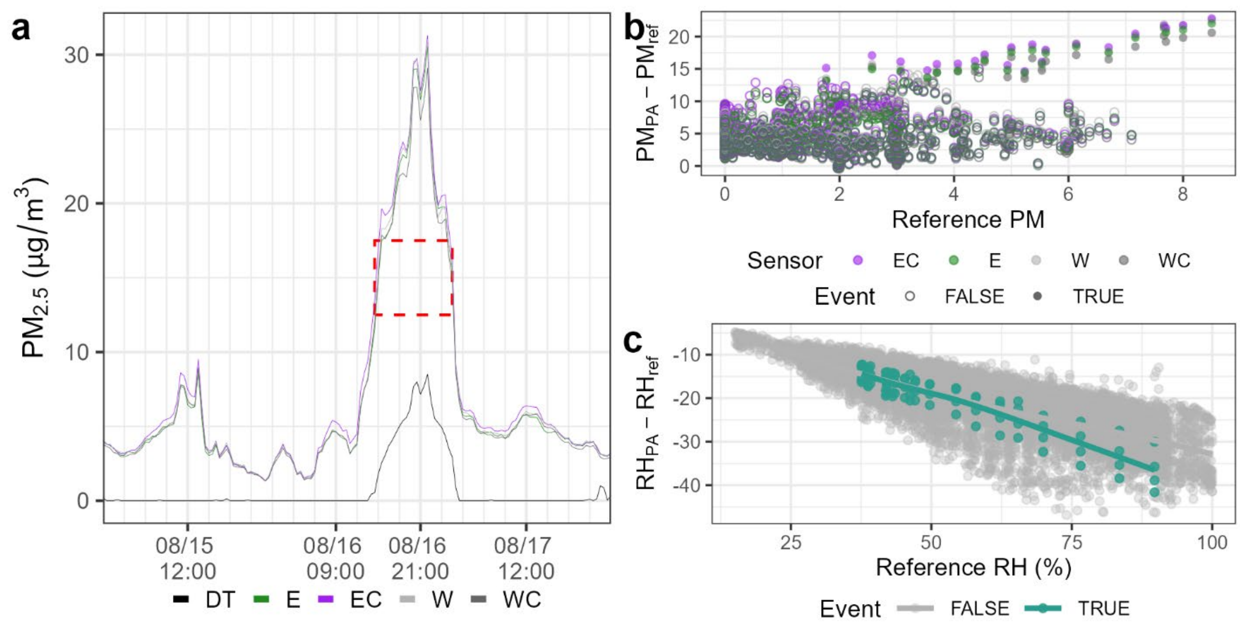

Figure 4. Data regarding an anomalous PM event, showing a time series of data collected from all 5 instruments as well as a comparative residual error plot showing the distribution of residual error for the entire experiment and the event.

Table 4. Linear models for shell testing

Model Name	Equation	Dataset	R^2
Temperature (Full)	(2)	Full dataset	0.61
Temperature (Core)	(3)	Full dataset	0.59
$PM_{2.5}$ (Full)	(5)	Full dataset	0.39

$PM_{2.5}$	(4)	Full dataset	0.39
$PM_{2.5}$ no event	(4)	Anomalous event (see Fig 4) excluded	0.50
$PM_{2.5}$ no event or <DL	(4)	Anomalous event & Reference = 0 removed	0.58

Temperature

Temperature is highly time-dependent, as we would expect. The PurpleAir temperature readings are biased high, as is known from previous work (Wallace, 2023), but generally track the patterns of the Lufft reference monitor well (see Figure 5).

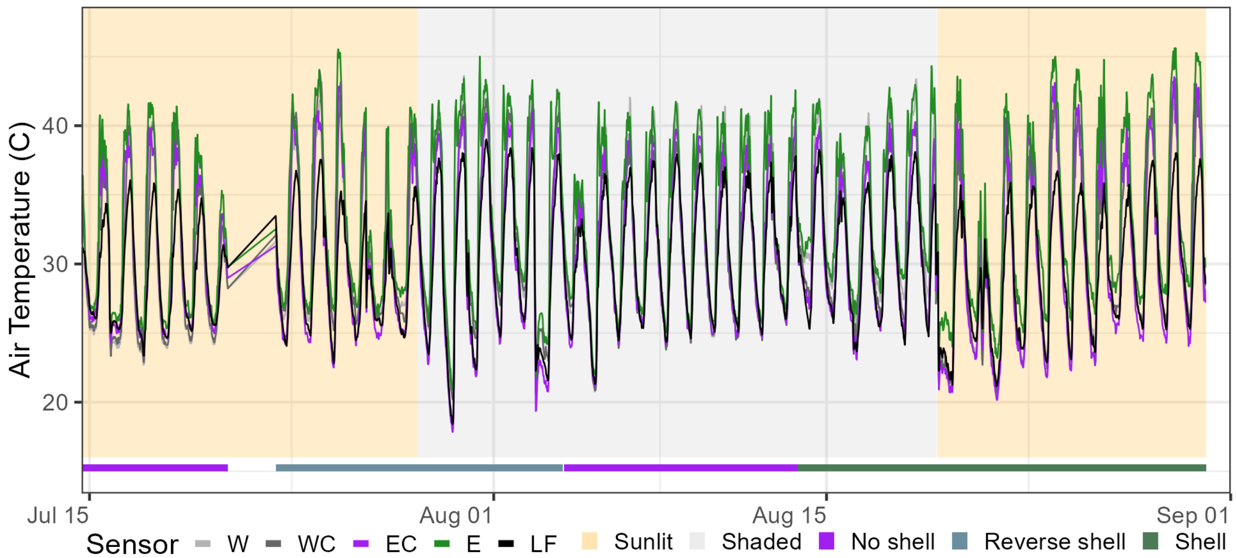

Figure 5. Air temperature measurements during shell evaluation period from all 4 sensors and the reference monitor (LF). Experimental conditions are encoded in shading and a line below the data. Across plots for testing conditions, letters indicate the relative position of each sensor (West, West Central, East Central, and East) and the colors for each sensor are plotted according to the color of the shell that is applied to that sensor, with purple representing the unmodified PurpleAir sensor (EC).

The full linear model fit to the difference between the sensor temperatures and the reference temperatures (eq 2), provides high explanatory power, with an R^2 of 0.61. The effect sizes for each of the model components are shown in Figure 6 as well as the effect sizes for a "Core" model (eq 3) that still provides a strong fit (R^2 = 0.59) but is easier to interpret with fewer interaction terms. The similarity in explanatory power between the full and core models indicates that the primary conclusions are robust to model specification (Table 4). Across sensors, application of a shell increased temperature error relative to the unshelled condition, and this increase was generally smaller under shaded conditions, indicating partial mitigation by shading. Differences between reverse and correctly oriented shells were modest in magnitude overall, although sensor-specific deviations were observed (most notably for WC; Fig. 7). The increase in temperature for each sensor compared to the reference is shown in Figure 7. The Easternmost sensor (E) reports higher temperatures regardless of whether the shell is applied or not. This effect remains even after controlling for insolation effects, and therefore is likely an artifact of the specific sensor. This aligns with findings from the correlation tests.

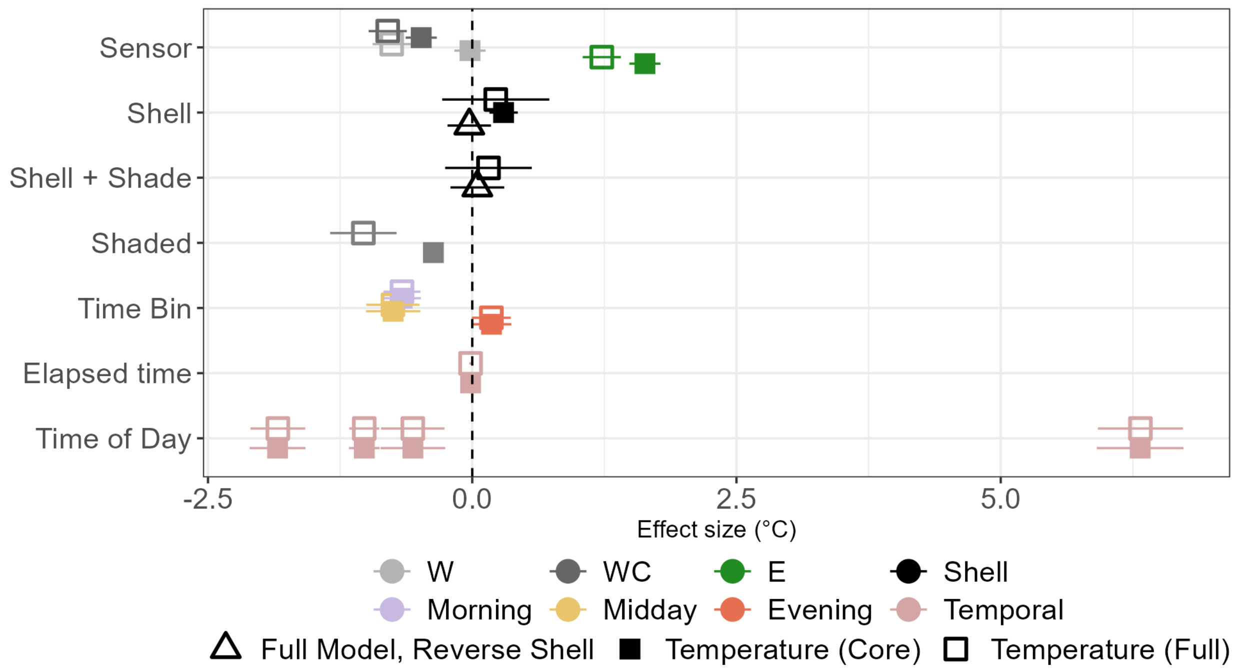

Figure 6. Effects from the residual models described in equations 2-3. The sensors are named by their location and the color indicates the shell color applied. Shell effects (including interactions) are evaluated relative to the EC sensor, which never has a shell applied.

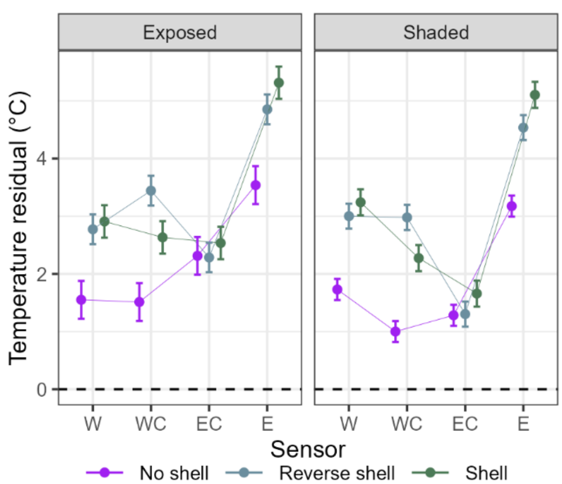

Figure 7. Temperature residuals estimated using estimated marginal means for the sensors under varying experimental conditions. Note that EC never has a shell applied.

Multiple comparisons of temperature distributions yield complex results. Overall, without any sub-setting based on hour of day, the easternmost sensor possessed a significantly different distribution relative to all other devices regardless of condition, indicating potential bias particular to that device. The distributions were in most agreement during the No Shell conditions, with only the easternmost sensor distribution being significantly different. Conversely, the distributions were most different during Shell conditions, with variation in significant difference between the reference monitor and control sensor, and variation in significant difference between the grey shelled sensors and the control sensor and each other.

Nighttime significant differences follow fewer patterns, with no device consistently possessing a data distribution significantly different than that of another device.

Peak daytime distributions show consistency in differences, potentially indicative of effects independent of condition. Reference monitor temperature distribution was always significantly different from the distributions of every sensor. The westernmost sensor did not display a temperature distribution significantly different from the sensor control's regardless of condition. The same could almost be said about the west-central sensor, but for the NSC/NSNC overall condition, wherein it is significantly different from the control sensor. The west and west-central sensors also fail to express significant differences but for the No Shell/Cover condition, which contributes to the same observation regarding the NSC/NSNC overall condition.

In totality, the significant differences of temperature distributions show consistency between devices and for individual devices, regardless of Shell or Cover, which seems to indicate little effect of the shell on measurements and the error associated with variation in those measurements.

Table 5. τ-C Correlation of temperature from PurpleAir vs Lufft. Values in parenthesis indicate sample size of half-hourly averaged concentrations.

	Individual PurpleAir Sensor			
Condition	**W**	**WC**	**EC**	**E**
Cover/No Shell/Midday (n = 60)	0.691 ***	0.660 ***	0.690 ***	0.663 ***
Cover/Shell/Midday (n = 36)	0.689 ***	0.710 ***	0.699 ***	0.628 ***
Cover/No Shell/Night (n = 174)	0.777 ***	0.753 ***	0.877 ***	0.778 ***
Cover/Shell/Night (n = 107)	0.694 ***	0.844 ***	0.797 ***	0.603 ***
Footnote: * = significant to p < 0.05, *** = significant to p < 0.001				

Table 6. Difference in Temperature [C] error between shell and no shell conditions ($MAE_{shell} – MAE_{no\ shell}$)

	W	**WC**	**EC**	**E**
No Cover midday	-0.36	0.55	0.26	0.43
Overnight	-0.08	-0.25	0.16	0.79
Cover midday	0.58	0.27	-0.02	0.27

$PM_{2.5}$

Each PurpleAir sensor generally tracked $PM_{2.5}$ concentrations measured by the DustTrak. Regardless of the presence of a shell, measurements were typically (97% of the time) within 10 μg/m^3 of the DustTrak measurement. This indicates that sensor measurements tend to be accurate to within the stated device accuracy for the conditions tested. Figure 8 shows the full PM time series for the experiment.

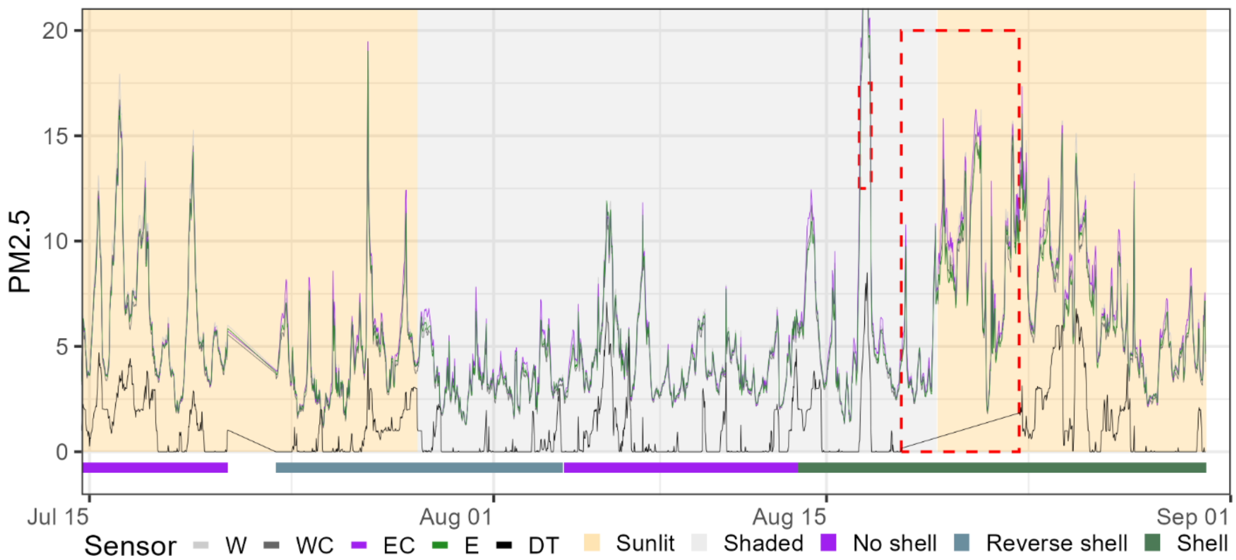

Figure 8. A time series of all $PM_{2.5}$ concentrations measured during the experiment. Experimental conditions are denoted by background shading and bars along the bottom. Red dashed lines indicate times that are not included in the analysis. The first is sometimes excluded for inconsistent behavior shown in Figure 4, the second is always excluded because the reference instrument was offline.

Relative Humidity (RH) is the dominant contributor to the residual error between sensor and reference measurements. This is consistent with hygroscopic growth increasing particle light scattering and elevating optical sensor response. The interquartile range (IQR) of RH measurements in the dataset is 25% to 51% for the PurpleAir sensors (RH_{PA}) and 41% to 78% for the reference (RH_{ref}). Across the IQR, the expected change is approximately 1-2 μg/m^3 in residual.

The shells do cause a systematic bias for the within sensor RH. Relative to the unmodified co-located sensor (EC), the sensor RH when shells are applied shows a downward bias of 3-4 percentage points. Because of this, using both the sensor and reference RH is important to correct for both the shell and hygroscopic growth effects.

The relative humidity across the study area during all of July and August 2025 showed limited spatial variability. The maximum difference in RH across all reference monitors at a given time had a median spread of 6.6 percentage points, and 75% of the comparisons had a spread less than 10 percentage points (Figure 9b).

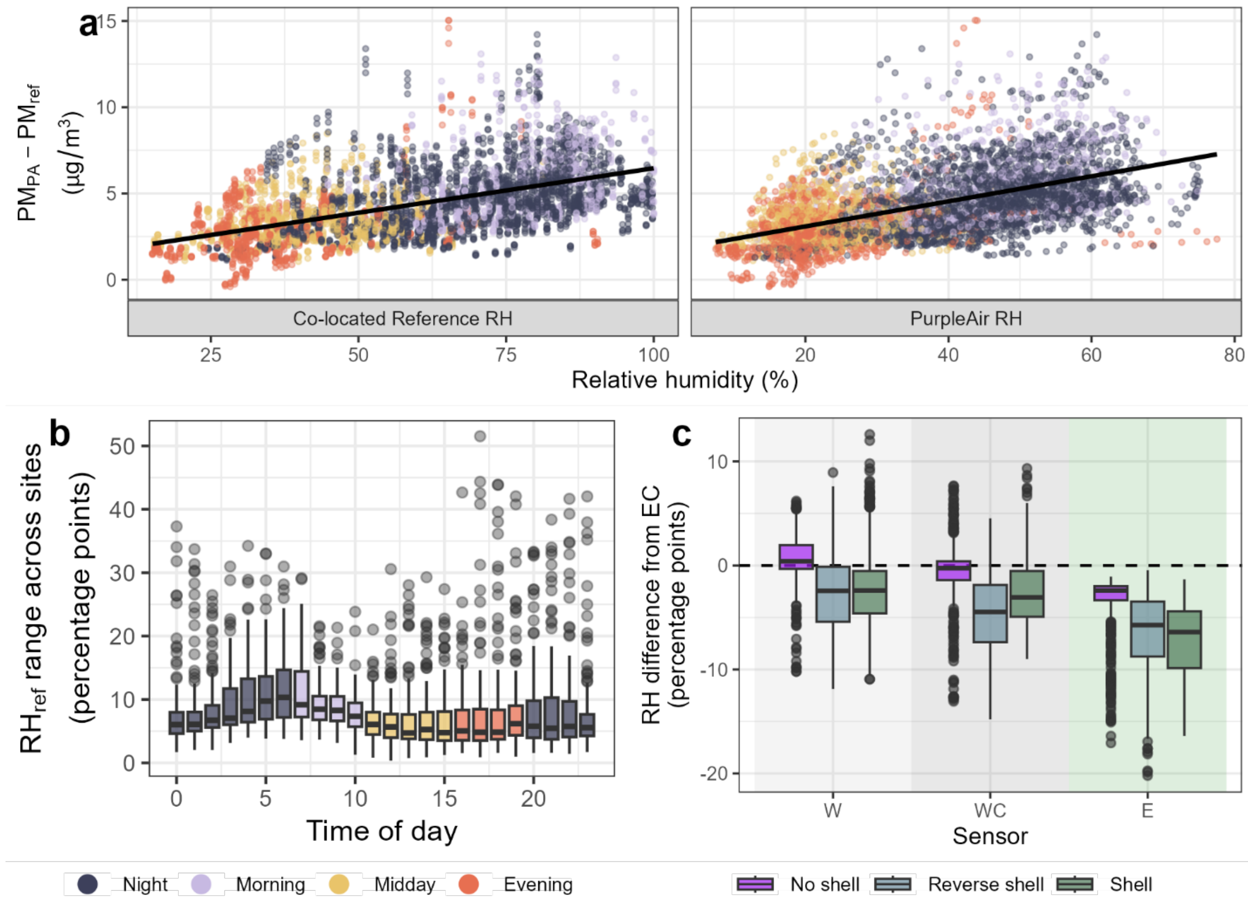

Figure 9. Variation in relative humidity (RH) and its effect on sensor performance. (a) shows the $PM_{2.5}$ measurement bias (compared to the reference DustTrak concentration) as a function of RH, using both co-located reference meteorology (left) and within sensor RH (right). (b) demonstrates relative consistency of RH measurements across the study area, showing the percentage point spread of reference RH across all sites in the study area (up to 27 km apart) for measurements at the same time. (c) shows the difference between the within-sensor RH measurement of the encased monitors and the control monitor (EC). Background shading indicates the color of the shell applied to that sensor.

Calibration performance was evaluated based on stringent testing (remote reference meteorology combined with testing on conditions not included in model training) and shows substantial improvement, as noted in Table 7. Across all conditions, calibration reduced RMSE by approximately 60% and improved R^2 by 0.14 – 0.22. Calibrated concentrations consistently showed improved agreement with reference measurements compared to uncorrected PurpleAir measurements. Notably, the calibrated, shelled sensors achieved lower RMSE and higher R^2 than the unmodified, uncalibrated sensors (NSNC). This indicates that the bias introduced by the protective housing can be effectively corrected, resulting in measurement performance that meets or exceeds that of the unmodified sensors.

Table 7 Performance of sensor measurements compared to reference $PM_{2.5}$ concentrations before and after calibration. Performance evaluation was completed for different conditions described in Table 2.

CONDITION	**RMSE (raw)**	**RMSE (calibrated)**	**R^2 (raw)**	**R^2 (calibrated)**
SNC-B	4.19	1.62	0.28	0.41

SC-B	3.48	1.08	0.01	0.20
SC	4.83	1.75	0.42	0.62
SNC	5.79	2.33	0.49	0.63
NSNC	5.72	--	0.44	--

To evaluate whether shell configuration and shading contributed to systematic measurement deviations, residual-based linear models were fit to PM as described in Methods. Models were evaluated both with and without sensor identity terms; including sensor terms did not materially change model fit. Excluding the anomalous event improved the model fit ($R^2 = 0.50$), and excluding observations interpreted as below detection limit further improved fit ($R^2 = 0.58$), without changing qualitative conclusions regarding the direction or relative magnitude of key effects.

After controlling for RH and temporal covariates, shell configuration and shading were associated with statistically detectable shifts in PM residuals. Naive assessment, which found no evidence that shell-related effects differed meaningfully across sensors, including for EC, which never has a shell applied, suggested that apparent shell-related effects may be due to some other environmental cause. However, the shell affects the RH within the enclosure and therefore shell effects are attributable to (and correctable by) the within-sensor RH (see Figure A4).

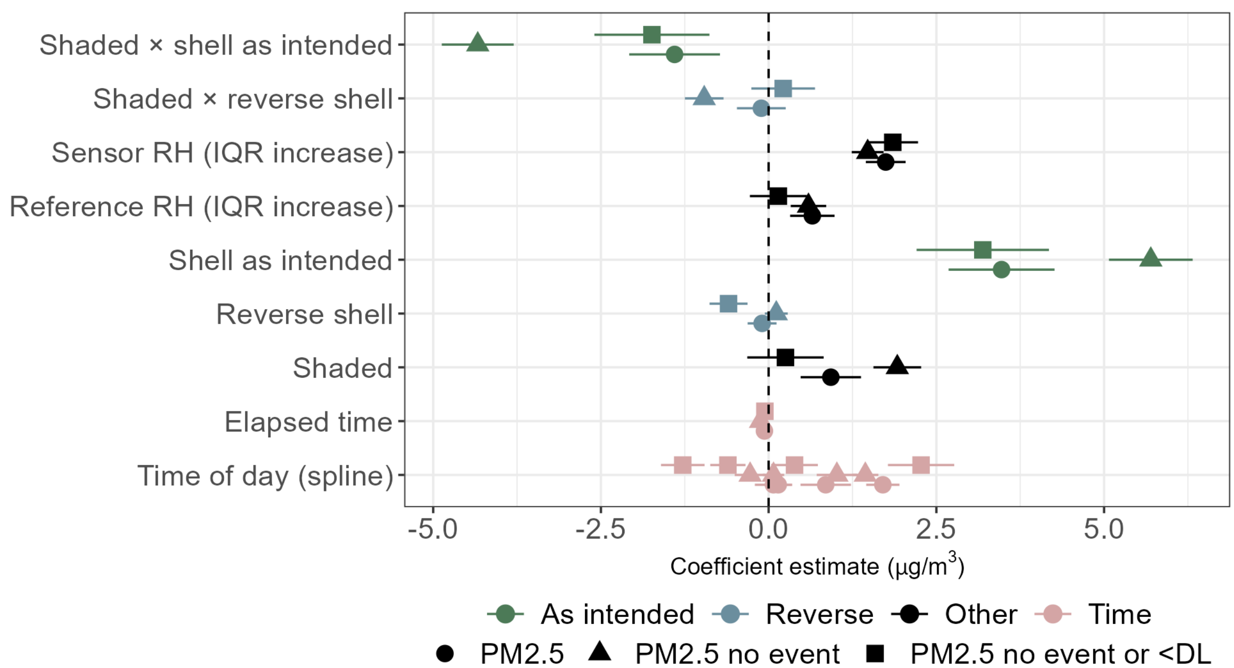

Figure 10. Model coefficients (effects) from the residual model described in equation 4. Shell effects are evaluated relative to the EC sensor, which never has a shell applied.

The $PM_{2.5}$ concentrations for each device were generally well correlated with the more accurate DustTrak across conditions, see Table 8. Regardless of condition, correlation between sensors remains consistent, with no obvious patterns in error relative to the control sensor. Patterns emerge only relative to reference monitors. However, these patterns largely describe how the sensors agree rather than in how they disagree. The correlation of sensors relative to reference monitors varies between experiment weeks with more pronounced variations in $PM_{2.5}$ correlations than temperature correlations. Peak daytime $PM_{2.5}$ correlations relative to the reference monitor show the worst correlation during the No Shell/Weekend condition (0.187 - 0.202), and some decrease in τ-c correlation during nighttime No Shell conditions (Weekend: 0.334 - 0.344; Weekday: 0.475 - 0.481). However, all correlations decrease homogenously across sensors, indicating an overwhelming effect on performance independent of shell. Temperature correlations are similar in their homogenous changes in fit across sensors. Nighttime temperature correlations do not differ notably from the average, and peak daytime correlations decrease most during No shade Cover conditions (No Shell: 0.398 - 0.691; Shell: 0.550 - 0.675).

Reinforcing previous observations concerning agreement between sensor PM measurements and correlation with the reference monitor, multiple comparison calculations found no significant differences between sensor distributions under any of the studied subsets. The observed MAE relative to the reference monitor, although moderate with respect to cited design specifications, constitute a significant difference of sensor data distribution relative to the reference monitor. This determination is consistent with MAE levels previously discussed, with an average overall MAE relative to control sensor several times smaller than the average MAE relative to the reference monitor (0.0857 μg/m^3 vs. 2.54 μg/m^3). At least statistically, this suggests some caution in claims regarding sensor $PM_{2.5}$ data distributions' significant differences, sans a calibration correction, at the concentration levels studied here.

Table 8. Kendall's Tau-C describing correlation between PurpleAir sensors and DustTrak measurements of $PM_{2.5}$. Values in parenthesis indicate sample size of half-hourly averaged concentrations.

	Individual PurpleAir Sensor			
Condition	**W**	**WC**	**EC**	**E**
No Shell/Weekday (n = 578)	0.498	0.495	0.500	0.506
Shell/Weekday (n = 352)	0.537	0.529	0.515	0.552
No Shell/Weekend (n = 186)	0.410	0.406	0.416	0.426
Shell/Weekend (n = 233)	0.558	0.537	0.531	0.559

Table footnote: All values in this table are significant to $p < 0.001$

Representativeness of Sensor Placement

To determine how well the device placement accurately captures the environmental conditions of the entire bus stop, the range of conditions around a sample of stations were compared. Only meteorological conditions were evaluated in this test as a more accurate, mobile PM monitor was not available. The Kestrel measurements of temperature and heat index around the stations are shown in Figure 12. A linear mixed-effect model (eq 6) was fit comparing the temperature (and heat index) from the Kestrel, with fixed effects including measurement height modeled as a continuous variable in meters, whether the measurement was taken in shade, and whether it was taken inside the shelter. Time of day was also added as a fixed effect because we know that temperature varies with time, and we could not take all measurements simultaneously. The identifier for each bus stop was added as a random effect to separate the variation in temperature between stops from the variation around the stop. Uncertainty is represented using 95% confidence intervals, and inference is based on effect magnitude and direction rather than null-hypothesis testing.

All measurements were taken in summer afternoon conditions, to determine relevance for the conditions of most concern (extreme heat), and the time effect increasing temperature +2.73°C per hour and heat index +3.48°C per hour was significant. Temperature varied somewhat during the measurements, but whether the measurement was taken inside the shelter or shaded did not have a strong or statistically significant effect (See Figure 11). There does appear to be some decrease of temperature with height, possibly due to heat radiating from hot pavement, but the effect is relatively small (Temperature -0.48°C/m, Heat index -0.62°C/m). The sensors are placed at approximately 2.5 m, which ensures it will not obstruct even San Antonio's tallest resident, the 7'4" Victor Wembanyama. Although this may slightly bias the sensor measurements compared to exposure-relevant positions, in the determination of sensor placement, placing them lower was not feasible due to passenger obstruction or interaction. Excluding measurements taken at sensor height (2.5 m) did not materially change effect directions or conclusions, providing evidence that conclusions are not driven by these measurements.

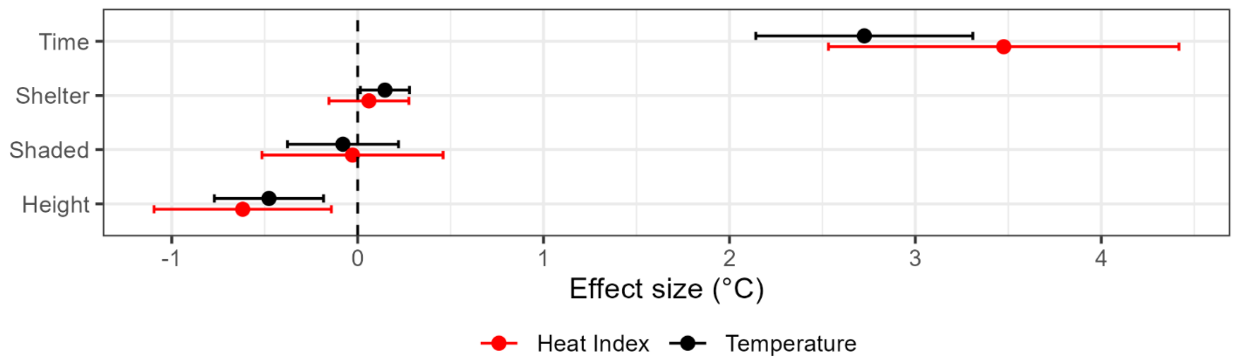

Figure 11. The points represent model-estimated effect of Kestrel probe placement on temperature and heat index, with 95% confidence intervals. Shelter and shade are binary factors, time is represented per hour, and height is represented per meter.

Comparing these Kestrel measurements to those from the PurpleAir at the same time (see Figure 12), we see that as in the previous experiment, the PurpleAir temperature readings are systematically higher. However, despite the offset, the PurpleAir tracks the same temperature trends as the Kestrel. A simple linear regression describes the strong correlation between the two measurements with an R^2 of 0.55, but found that for each 1°C increase in PurpleAir temperature, the Kestrel temperature increase was only 0.48°C on average. This evidence together indicates that the PurpleAir provides a strong representation of the temperature at the location installation, although provides higher temperature readings.

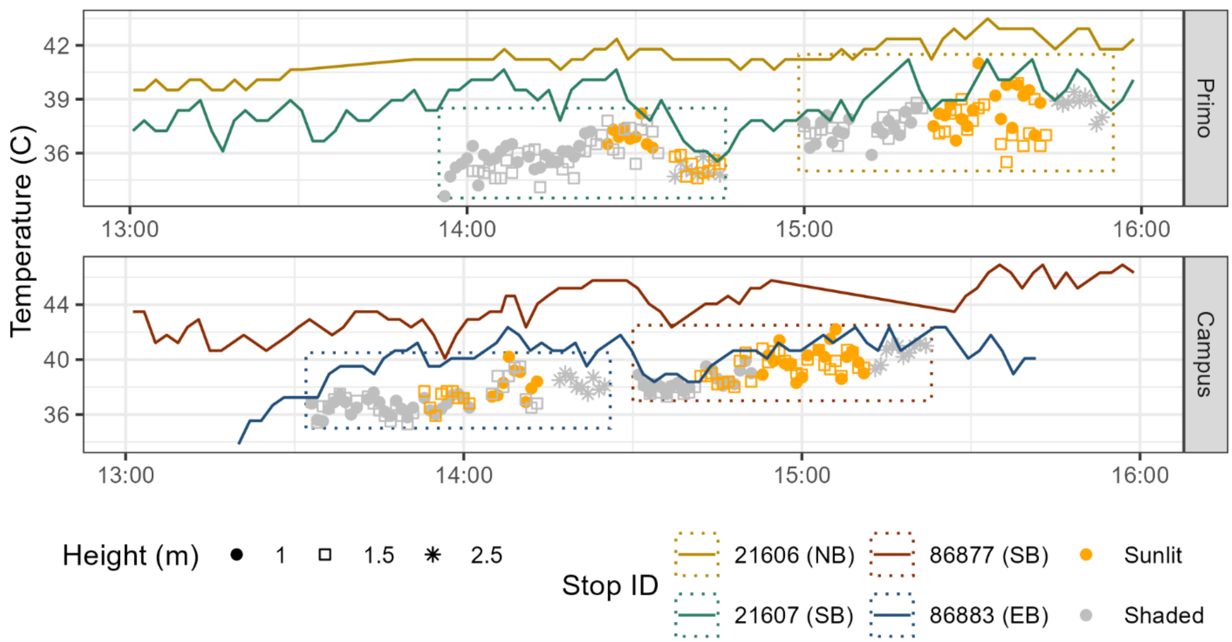

Figure 12. Simultaneous measurements taken at the same bus stations using the PurpleAir monitor (lines) and Kestrel (points). The Kestrel was monitored at 21607 and 86883 earlier on their respective days, identified by Kestrel measurements being outlined for each stop.

The variation in temperature around bus stops can also be seen through the use of thermal imaging. Although the Kestrel still provides highly relevant information regarding the air temperature, the overall variation in heat can be seen in Figures 13-15. It is clear, for example, that the shade provided by the bus shelter significantly cools the sidewalk compared to the roadway or even other sidewalk area. The locations where the Kestrel measurements were taken are identified by sunburst shapes overlaid on the (visual wavelength) photograph, with the measurement locations corresponding to Table 3 above or below the photo. A comparison infra-red photograph of each stop is placed adjacent to it. The images were taken at the same time, except for Figure 13, which was taken during the Kestrel measurement window, but using a slightly different angle and time to be able to capture the location of all measurements.

Three different shelter configurations are represented in these images. One is a Prímo station and the other two are different sizes of the grey stations. None of the older green stations were evaluated in this test. The Prímo stops are 7.5m wide (measured parallel to the roadway), Stop 86883 has a shelter 5.25m wide, and stop 86877 has a shelter 2.7m wide. These size variations are affected by the typical ridership as well as the available space for a shelter that will not obstruct passers-by or encroach onto private property.

Figure 13. Thermal and visual imaging of Bus Stop 86883 on August 14, 2025, with the Kestrel monitoring sites noted.

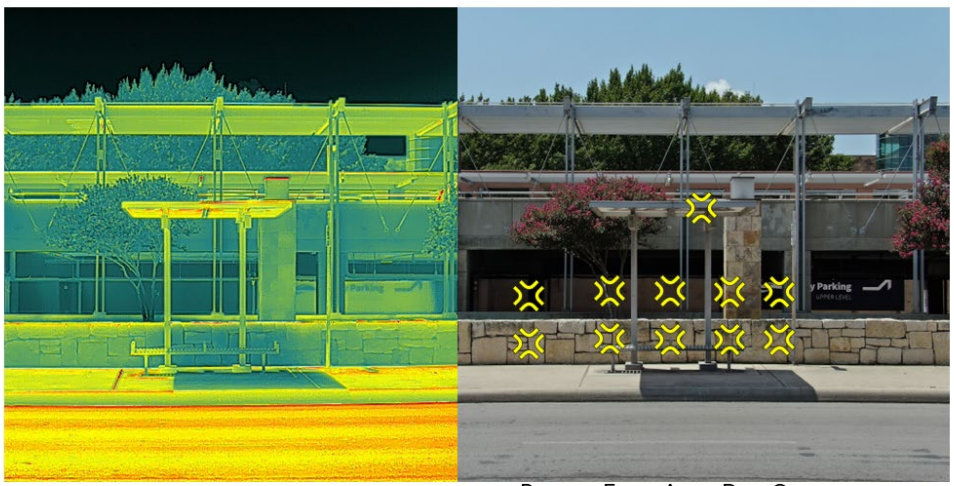

Figure 14. Thermal and visual imaging of Bus Stop 86877 on August 14, 2025, with the Kestrel monitoring sites noted.

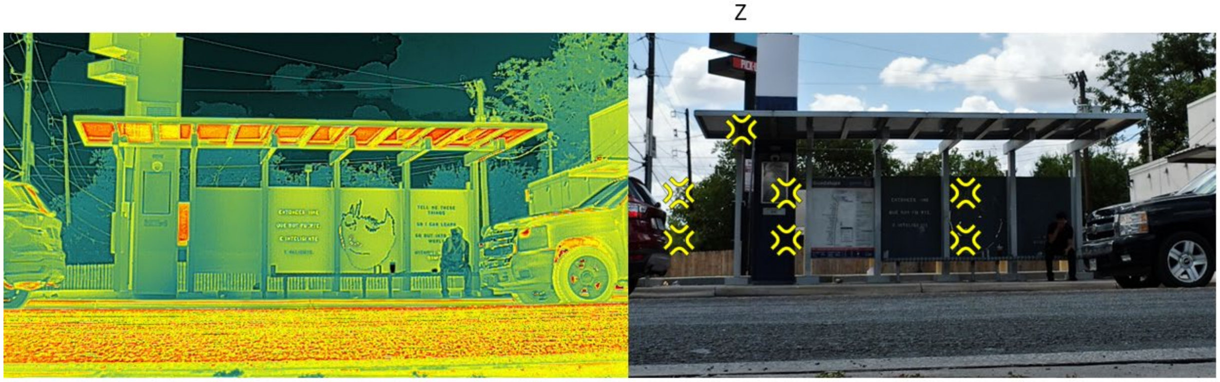

Figure 15. Thermal and visual imaging of Bus Stop 21607, one of the Primo stations, on August 6, 2025, with the Kestrel monitoring sites noted.

Discussion

Temperature

Results from the controlled shell deployment experiment show that the shells systematically influence temperature measurements. Across all analyses, PurpleAir sensors installed with protective shells and shaded by the structure exhibited a consistent, quantifiable bias relative to the reference instrument. These findings motivate the need for deployment specific calibration for sensors used in transit shelters, but do not invalidate the measurements from these sensors or their ability to represent trends.

To address this, we have developed a preliminary calibration algorithm for this testing configuration. Note that the calibration here is specific to the bus shelter study. For this testing setup, calibrated temperatures are expressed as

$$T_{t,1.7m}^{calibrated} = \alpha_0 + \alpha_1 \mathrm{T_t^{PA}} + \beta_z(1.7 - 2.5) + \mathrm{f(H_t)} + \epsilon_t \quad (7)$$

where T_t^{PA} represents the temperature measurements from the sensor at time t after initial adjustment using equation 1. The function $f(H_t)$ is a smooth function of the hour of the day capturing diurnal bias, and ε_t represents remaining unexplained variability. The estimated coefficients are $\alpha_o \approx$ 12°C, $\alpha_1 \approx$ 0.54°C. The term β_z (1.7-2.5) represents the slope from the representativeness fits applied to the height difference between the station mounted sensors and the height of the reference Lufft (and other reference instruments that will be used in future analyses). This height is also more exposure relevant. Blocked time validation was used, with training and testing sets constructed to include observations from both shell orientations, while maintaining temporal separation to account for autocorrelation. Implicit in this model is that all PurpleAir sensors are covered by the shelter and in shells. These calibration relationships are estimated using data from periods when all PurpleAir devices were installed in shells and under the shade cover and excluding the control EC sensor; application is therefore intended for the sheltered bus-stop deployment configuration.

For the situation in which the sensors are shaded and in shells, the raw sensor temperatures have an RMSE of 3.18°C compared to the reference, and after calibration (excluding the height correction, which is not applicable to the experimental deployment) the RMSE is improved to 1.36 with a MAE of 1.03°C. Figure 16 summarizes this improvement. To evaluate generalizability to unseen devices, a leave-one-sensor-out validation was performed (results in SI), training the calibration on two sensors and applying it to the third. Leave-one-sensor-out validation yielded RMSE values ranging from 1.1 to 2.3 °C (MAE 0.8–1.7 °C), comparable to blocked time validation and consistently improving upon uncalibrated measurements. These results demonstrate that the calibration can be applied to new sensors without requiring sensor-specific refitting.

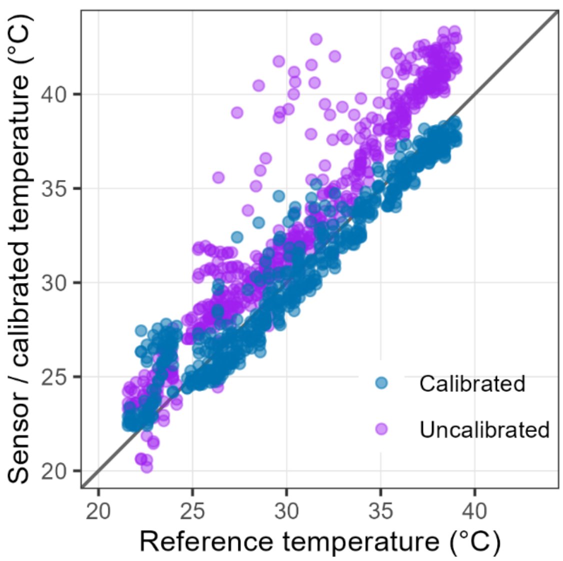

Figure 16. A scatter plot of the PurpleAir Temperature readings for sensors in shells and shade compared to the reference Lufft data. Calibrated points are based on Eq. 7.

$PM_{2.5}$

For deployment in the 3-D printed shells, the $PM_{2.5}$ should be recalibrated using equations 8-9. Adding both the sensor relative humidity (RH_{PA}) and a higher accuracy regional reference value (RH_{ref}) improved the model performance. Based on reference monitor (Lufft) sites across the study area, spatial variation in RH was limited, and its inclusion improved model performance.

Relative humidity is a strong predictor of sensor error and is necessary for calibration. However, even after accounting for humidity and temporal effects, systematic differences associated with shell configuration persist. While there was some effect of shading on measurement accuracy, it is not included as a calibration term to maintain the general applicability across deployment conditions. Model performance remained robust without this term. The modeled bias (Δ) for the PurpleAir sensors in shells can be calculated as

$$\Delta = 0.836 + 0.0532 \cdot t - 0.00443 \cdot RH_{ref} + 0.0630 \cdot RH_{PA} + 0.541 sin\left(\frac{2\pi H}{24}\right) - 0.365 cos\left(\frac{2\pi H}{24}\right) \quad (8)$$

where t is the time elapsed in days since the sensor was installed, RH_{ref} is the regional relative humidity from the nearest available research or regulatory monitor, RH_{PA} is the relative humidity from the PurpleAir sensor, and H is the hour of the day in which a measurement was taken. This bias can be used to calculate the corrected PM concentration as

$$PM = PM_{PA} - \Delta \quad (9)$$

where PM_{PA} is the concentration reported by the PurpleAir sensor.

Conclusions, Limitations, & Future Work

This study demonstrates that reliable, exposure-relevant environmental measurements of public spaces can be achieved using low-cost sensors when paired with appropriate design and calibration.

We designed and tested data collection infrastructure to monitor temperature and air quality in public spaces. A 3-D printed shell was created as protection for low-cost sensors. While the shell introduced a systematic measurement bias, the effect was consistent and can be corrected using empirically derived calibration functions. The calibration also considers the needs of siting the devices and can predict exposure-relevant conditions. This approach supports scalable monitoring in public spaces without interfering with site use.

The testing setup described is currently deployed at 12 bus shelters around San Antonio (see Figure 3). Continuously collected data from this network will be analyzed to determine the variation of temperature and air quality at different locations in the city and different shelter types. The data presented in this paper shows that extreme conditions are frequently experienced at the bus stops currently, including heat index values exceeding 40°C, which the National Weather Service classifies as “Dangerous” due to the elevated risk of heat exhaustion (National Weather Service, n.d.).
The $PM_{2.5}$ analysis is limited by the lower detection limit of the DustTrak reference monitor. Periods in which the DustTrak reported zero concentration may be measuring very low concentrations rather than true zero. This introduces nonlinearity in the relationship between the sensor and reference monitor, which can artificially inflate residuals. To address the influence of this limitation, sensitivity analyses were conducted excluding observations where the reference instrument reported 0. These analyses improved model fit but did not change the qualitative conclusions regarding the direction or relative magnitude of the shell, shading, or humidity effects.

Regarding sensor placement, the PurpleAir represents the trend in temperature well for each station. Although the PurpleAir placement and device itself contribute to a slight overestimation in temperature, the data represents the trend well while accounting for the practical limitations of the active bus stops. Calibration supports improved estimation of thermal exposure. Future work will include refining the calibration equations, continued data collection, and analysis of which shelter types are most protective and whether the ongoing heat mitigation efforts are providing relief for waiting transit passengers.

The workflow presented here creates opportunities for integration with other low-cost sensor projects and data-driven urban systems. Increased availability of real-time data streams from deployed sensors could inform adaptive transit design (Liu, 2025) that could adjust service at high-exposure locations. Additionally, such networks could support data-informed prioritization of infrastructure investments, identifying locations where design modifications would yield the greatest reduction in heat and air quality exposure.

References

Agresti A (2010) *Analysis of Ordinal Categorical Data*. John Wiley & Sons.

Bagkis E, Hassani A, Schneider P, et al. (2025) Evolving trends in application of low-cost air quality sensor networks: challenges and future directions. *npj Climate and Atmospheric Science* 8(1). Nature Publishing Group: 335.

Brannen J and Banerjee D (2025) ‘Just stuck out here’: Bus riders simmer under sun in Houston’s hottest neighborhoods. *Urban Edge*. Kinder Institute for Urban Research. Available at: https://kinder.rice.edu/urbanedge/just-stuck-out-here-bus-riders-simmer-under-sun-houstons-hottest-neighborhoods.

Braun RA and Fraser MP (2022) Extreme Heat Impacts on the Viability of Alternative Transportation for Reducing Ozone Pollution: A Case Study from Maricopa County, Arizona. *Weather, Climate, and Society* 14(3): 905–917.

Briant S, Cushing D, Washington T, et al. (2025) Small but Significant: A Review of Research on the Potential of Bus Shelters as Resilient Infrastructure. *Applied Sciences* 15(12). Multidisciplinary Digital Publishing Institute: 6724.

Brown KE, Zhai W and Lopez Ochoa EA (2026) Prioritizing mitigation efforts for intra-urban heat vulnerability using a case study of San Antonio, Texas. *Journal of the Air & Waste Management Association* 76(4). Taylor & Francis: 255–270.

Chen Y, Shan B and Yu X (2022) Study on the spatial heterogeneity of urban heat islands and influencing factors. *Building and environment* 208: 108604.

deSouza P and Kinney PL (2021) On the distribution of low-cost PM2.5 sensors in the US: demographic and air quality associations. *Journal of Exposure Science & Environmental Epidemiology* 31(3). Nature Publishing Group: 514–524.

Hansen T, Blume K, Sener I, et al. (2021) *Transit Agency Relationships and Initiatives to Improve Bus Stops and Pedestrian Access*. Washington, D.C.: Transportation Research Board. Available at: https://www.nap.edu/catalog/26166 (accessed 1 December 2025).

IPCC (2021) *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. doi:10.1017/9781009157896. Available at: https://www.ipcc.ch/assessment-report/ar6/ (accessed 14 March 2023).

Jiao W, Hagler G, Williams R, et al. (2016) Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. *Atmospheric Measurement Techniques* 9(11). Copernicus GmbH: 5281–5292.

Karner A, Hondula DM and Vanos JK (2015) Heat exposure during non-motorized travel: Implications for transportation policy under climate change. *Journal of Transport & Health* 2(4): 451–459.

Lanza K and Durand CP (2021) Heat-Moderating Effects of Bus Stop Shelters and Tree Shade on Public Transport Ridership. *International Journal of Environmental Research and Public Health* 18(2): 463.

Lanza K, Jones J, Acuña F, et al. (2023) Heat vulnerability of Latino and Black residents in a low-income community and their recommended adaptation strategies: A qualitative study. *Urban Climate* 51: 101656.

Lanza K, Ernst S, Watkins K, et al. (2025) Heat stress mitigation by trees and shelters at bus stops. *Transportation Research Part D: Transport and Environment* 140: 104653.

Liu L, Li X, Pereira RHM, et al. (2025) Measuring exposure to extreme heat in public transit systems. *Journal of Transport Geography* 128: 104383.

Liu X (2025) AI-driven real-time responsive design of urban open spaces based on multi-modal sensing data fusion. *Scientific Reports* 15(1). Nature Publishing Group: 41255.

Miao Q, Welch EW and Sriraj PS (2019) Extreme weather, public transport ridership and moderating effect of bus stop shelters. *Journal of Transport Geography* 74: 125–133.

Molla A, Flores AB and Sailor DJ (2025) Addressing data gaps and disparities in access to air quality networks: a case study of Maricopa County, Arizona. *Population and Environment* 47(3): 27.

Moore A, Figliozzi M and Bigazzi A (2014) Modeling Impact of Traffic Conditions on Variability of Midblock Roadside Fine Particulate Matter: Case Study of an Urban Arterial Corridor. *Transportation Research Record: Journal of the Transportation Research Board* 2428(1): 35–43.

National Weather Service N (n.d.) What is the heat index? NOAA's National Weather Service. Available at: https://www.weather.gov/ama/heatindex (accessed 30 December 2025).

Ngo NS (2019) Urban bus ridership, income, and extreme weather events. *Transportation Research Part D: Transport and Environment* 77: 464–475.

Noh S-Y, Park D and Yook S-J (2021) Numerical investigation of bus stop structures in Seoul for the reduction of fine dust entry. *Journal of Mechanical Science and Technology* 35(1): 371–379.

Pan Y, Li S and Tang X (2024) Investigation of Bus Shelters and Their Thermal Environment in Hot–Humid Areas—A Case Study in Guangzhou. *Buildings* 14(8): 2377.

Pournaras E, Nikolic J, Omerzel A, et al. (2017) Engineering Democratization in Internet of Things Data Analytics. In: *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, March 2017, pp. 994–1003. Available at: https://ieeexplore.ieee.org/abstract/document/7921015 (accessed 31 December 2025).

Shandas V, Voelkel J, Williams J, et al. (2019) Integrating Satellite and Ground Measurements for Predicting Locations of Extreme Urban Heat. *Climate* 7(1): 5.

Snyder EG, Watkins TH, Solomon PA, et al. (2013) The Changing Paradigm of Air Pollution Monitoring. *Environmental Science & Technology* 47(20). American Chemical Society: 11369–11377.

Stuart A (1955) A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification. *Biometrika* 42(3/4). [Oxford University Press, Biometrika Trust]: 412–416.

Sun Y, Mousavi A, Masri S, et al. (2022) Socioeconomic Disparities of Low-Cost Air Quality Sensors in California, 2017–2020. *American Journal of Public Health* 112(3). American Public Health Association: 434–442.

US EPA O (2021) A Guide to Siting and Installing Air Sensors. Available at: https://www.epa.gov/air-sensor-toolbox/guide-siting-and-installing-air-sensors (accessed 31 December 2025).

Wallace L (2023) *PurpleAir PA-II Temperatures and RH compared to measured indoor and outdoor values*. 1 September. Available at: https://static.purpleair.com/docs/PurpleAir-PA-II-Temperatures-and-RH-Lance-Wallace.pdf.

West T (2025) A Systematic Literature Review Assessing the Inequities in Air Quality Monitor Placement. 2025120545. Preprints. Available at: https://www.preprints.org/manuscript/202512.0545 (accessed 31 December 2025).

Xie J, Liu Z and Wang J (2026) Can transit-oriented development (TOD) cool the city? A multi-scale geospatial assessment of daytime surface heat exposure in Tokyo. *Sustainable Cities and Society* 142: 107318.

Yin Y, Tonekaboni NH, Grundstein A, et al. (2020) Urban ambient air temperature estimation using hyperlocal data from smart vehicle-borne sensors. *Computers, Environment and Urban Systems* 84: 101538.