Data in Brief 60 (2025) 111492

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset on indoor and outdoor $PM_{2.5}$ concentrations at two residences using low-cost sensors in the Rio Grande Valley Region of South Texas

Kabir Bahadur Shah[a], Sai Deepak Pinakana[a], Mkhitar Hobosyan[c], Armando Montes[b], Amit U. Raysoni[a,*]

[a] School of Earth, Environmental, and Marine Sciences, The University of Texas Rio Grande Valley, Brownsville, TX, USA
[b] Math & Science Academy, The University of Texas Rio Grande Valley, Brownsville, TX, USA
[c] Department of Physics and Astronomy, The University of Texas Rio Grande Valley, Brownsville, TX, USA

ARTICLE INFO

Article history:
Received 4 November 2024
Revised 3 March 2025
Accepted 14 March 2025
Available online 19 March 2025

Dataset link: Fine particulate matter (PM2.5) datasets for both Indoor and Outdoor environments at two South Texans residences (Original data)

Keywords:
Indoor air quality
Outdoor air quality
Fine particulate matter (PM2.5)
Low-cost sensors
PurpleAir

ABSTRACT

Fine particulate matter ($PM_{2.5}$) is one of the criteria air pollutants associated with adverse respiratory and cardiovascular health effects. This dataset contains hourly and daily averaged measurements of $PM_{2.5}$ concentrations (µg/m^3) collected from two residential homes in the Rio Grande Valley (RGV) Region of South Texas, using PurpleAir-II-SD low-cost sensor. In addition, the dataset also includes temperature (ºC) and relative humidity (%) measured by the sensors. The data was collected for one month during the summer season from indoor (kitchen) and outdoor (front yard) environments at two locations: Home-1 in Weslaco City from June 16, 2024, to July 16, 2024, and Home-2 in Mission City from June 22, 2024, to July 16, 2024. Furthermore, the resultant wind speed and wind direction data were obtained from nearby Texas Commission on Environmental Quality (TCEQ) Continuous Ambient Monitoring Station (CAMS) site, providing a better understanding of meteorological influences on air quality.

The raw $PM_{2.5}$ data were corrected using Barkjohn's US-wide correction equation to increase the accuracy of the measure-

* Corresponding author.
E-mail address: amit.raysoni@utrgv.edu (A.U. Raysoni).

https://doi.org/10.1016/j.dib.2025.111492

ments and are presented in the datasets. The dataset is available in three CSV files: one for Home-1, one for Home-2, and one for wind data (Wind Rose) from the TCEQ stations. This dataset has significant potential for reuse by researchers interested in air quality monitoring, exposure assessment, and the health impacts of $PM_{2.5}$. Furthermore, the study highlights the effectiveness of using low-cost sensors for continuous air quality monitoring in residential settings, thereby contributing to the growing body of air quality literature emanating from the U.S.-Mexico border region.

Specifications Table

Subject	Environmental Science (air pollution)
Specific subject area	Deployment of low-cost sensors to measure both indoor and outdoor concentrations of fine particulate matter ($PM_{2.5}$) at two residential homes
Type of data	The datasets are available in table data (organized in CSV format).
Data format	Raw data (2-min intervals) was filtered, converted to 1-h average, analyzed, and corrected using Barkjohn's US-wide correction equation. ($PM_{2.5}$ in µg/m^3, temperature in ºC, and relative humidity in %).
Data collection	The data was collected using low-cost air quality sensors (PurpleAir-II-SD) at two residences in the Rio Grande Valley (RGV) Region of South Texas, USA: Home-1 from June 16, 2024, to July 16, 2024, and Home-2 from June 22, 2024, to July 16, 2024, measuring $PM_{2.5}$, T, and RH at a two-minute resolution. The PA-II sensors were placed in two different environments: indoors (kitchen) and outdoors (front yard) at both residences. Additionally, hourly averages for wind direction, resultant wind speed, and ambient $PM_{2.5}$ data were downloaded from the nearest Texas Commission on Environmental Quality (TCEQ) Continuous Ambient Monitoring Station (CAMS) located near residential homes. The TCEQ CAMS data was retrieved from https://www.tceq.texas.gov/cgi-bin/compliance/monops/yearly_summary.pl. by selecting the nearest TCEQ monitoring site, for instance; C-43 for Home-1 in Brownsville and C-1023 for Home-2 in Harlingen. Similarly, ambient PM2.5, resultant wind speed, and wind direction data were downloaded from C-43, whereas for C-1023, resultant wind speed and wind direction data were only downloaded as C-1023 doesn't measure ambient $PM_{2.5}$.
Data source location	Datasets generated by the PA-II LCS measuring $PM_{2.5}$, T, and RH: • Home-1 at Weslaco, Texas, USA **(26°10′47"N, 97°59′46"W)** • Home-2 at Mission, Texas, USA **(26°09′51"N, 98°17′52"W)** Datasets generated by TCEQ CAMS for ambient $PM_{2.5}$ and Wind Rose • CAMS-43 at Mission, Texas, USA **(26°13′36"N, 98°17′27"W)** • CAMS-1023 at Harlingen, Texas, USA **(26°12′01"N 97°42′45"W)**
Data accessibility	Repository name: Fine particulate matter (PM2.5) datasets for both Indoor and Outdoor environments at two South Texans residences. (Mendeley data) Data identification number: 10.17632/yk2pz4hcb5.2 Direct URL to data: https://data.mendeley.com/datasets/yk2pz4hcb5/2

1. Value of the Data

- Air quality data at a local scale is scarce, especially in low-income neighborhoods. This data set provides a high-temporal resolution $PM_{2.5}$ dataset in a low-income neighborhood along the US-Mexico border. It can be used especially in environmental justice studies focusing on air quality exposure disparities.
- The high-frequency indoor and outdoor $PM_{2.5}$ data from two residences presented in the dataset can help gain insights into the relationship between $PM_{2.5}$ levels indoors and outdoors, capturing fluctuations in air quality trends.
- The response of low-cost sensors to different meteorological conditions is being studied globally. This dataset provides insights into how PurpleAir sensors perform in humid weather, which could aid in replicating studies in other regions with similar meteorological conditions.
- Activities such as cooking significantly affect indoor $PM_{2.5}$ levels. This dataset offers high-frequency $PM_{2.5}$ data and typical cooking periods, helping to understand the impact of cooking on indoor $PM_{2.5}$ levels.
- Multiple studies have shown that air filters help reduce outdoor pollutant infiltration into indoor environments. This research provides $PM_{2.5}$ data from indoor and outdoor environments in two homes with varying Minimum Efficiency Reporting Value (MERV) rated air filters. Future studies can use the dataset to compare their effectiveness, enhancing our understanding of their efficiency.

2. Data Description

Air pollution is one of the global threats to human health, becoming the second largest risk factor for death globally in 2021 [15]. One of the most studied criteria air pollutants, fine particulate matter ($PM_{2.5}$, particles with diameters 2.5 µm and less), is associated with adverse respiratory and cardiovascular health effects [12,19]. Both short-term and long-term exposure to $PM_{2.5}$ has a detrimental impact on human health [2]. Thus, $PM_{2.5}$ is a significant risk factor for morbidity and mortality globally and in the United States [7]. The impact of air pollution is not restricted to just outdoor areas but the indoor microenvironment as well. In recent decades, indoor air quality has elicited a lot of interest as people spend approximately 80–90 % of their time indoors on average [1,9]. Indoor environments such as residences [9], offices [5], and schools [11] have been the primary locations for human exposure to $PM_{2.5}$. Indoor air quality (IAQ) ranks among the top five environmental public health risk factors and can be more polluted than outdoor air [18].

For this study, the data was collected for one summer month from two residences in the RGV region (Home-1 in Weslaco, TX, and Home-2 in Mission, TX) using PurpleAir-II-SD low-cost sensors [10]. All the data used for this study was processed, analyzed, and corrected. The data was stored in an Excel sheet, which was then formatted into CSV files, and analysis was done using Origin Pro (Origin Lab Corporation, Northampton, MA, USA Version 2024). The datasets can be downloaded at https://data.mendeley.com/datasets/yk2pz4hcb5/2. The dataset consists of three CSV (Comma Separated Values) files: two CSV files for Home-1 (Weslaco) and Home-2 (Mission), respectively, and one CSV file for Wind Rose data from TCEQ CAMS. The format of all three CSV files with the data collected by the sensors is described below:

- The first CSV file is named "***Home1_Weslaco_Indoor_Outdoor_updated.csv***" and is formatted as:

  Date/Time; T; RH; PM2.5_CF1_a; PM2.5_CF1_b, Avg_PM2.5_CF1 for (2 min_Indoor)
  Example: 6/16/2024 0:01; 29.44; 35; 2.33; 3.07; 2.7
  Date/Time; T; RH; PM2.5_CF1_a; PM2.5_CF1_b, Avg_PM2.5_CF1 for (2 min_Outdoor)
  Example: 6/16/2024 0:01; 32.78; 51; 16.69; 16.23; 16.46

Date/Time; T; RH; Raw_PM2.5; Barkjohn_Corrected PM2.5 for (1-h Average_Indoor)
Example: 6/16/2024 0:00; 29.87; 34.30; 3.02;4.38
Date/Time; T; RH; Raw_PM2.5; Barkjohn_Corrected PM2.5 for (1-h Average_Outdoor)
Example: 6/16/2024 0:00; 32.44; 52.40; 16.98; 10.13
Date; Corrected_Indoor; Corrected_Outdoor; Final I/O for (24-h Average _Combine)
Example: 6/16/2024; 4.43; 7.49; 0.59

- The second CSV file is named "***Home2_Mission_Indoor__updatedOutdoor.csv***" and is formatted as:

Date/Time; T; RH; PM2.5_CF1_a; PM2.5_CF1_b, Avg_PM2.5_CF1 for (2 min_Indoor)
Example: 6/22/2024 0:01; 29.44; 36; 0.53; 0.69; 0.61
Date/Time; T; RH; PM2.5_CF1_a; PM2.5_CF1_b, Avg_PM2.5_CF1 for (2 min_Outdoor)
Example: 6/22/2024 0:01; 30; 61; 5.26; 6.76; 6.01
Date/Time; T; RH; Raw_PM2.5; Barkjohn_Corrected PM2.5; TCEQ_C43_PM2.5 for (1-h Average_Indoor)
Example: 6/22/2024 0:00; 29.44; 36.97; 0.61; 2.88; 5.10
Date/Time; T; RH; Raw_PM2.5; Barkjohn_Corrected PM2.5; TCEQ_C43_PM2.5 for (1-h Average_Outdoor)
Example: 6/22/2024 0:00; 30.48; 61.20; 6.26; 3.76; 5.10
Date; Corrected_Indoor; Corrected_Outdoor; Final I/O for (24-h Average _Combine)
Example: 6/22/2024; 2.73; 3.52; 0.78

- The third CSV file is named "***TCEQ_CAMSdata_WindRose.csv***" and is formatted as:

Date/Time; WD; RWS for (TCEQ-CAMS _C43 in Mission)
Example: 6/22/2024 0:00;26;1.07
Data/Time; WD; RWS for (TCEQ-CAMS_C1023 in Harlingen)
Example: 6/16/2024 0:00; 130; 2.06

The first row of all the CSV files contains the column headers, with the description of each variable provided below:

- Date/Time: The measurement's timestamp is in Central Standard Time (CST), formatted as "mm-dd-yyyy hh: mm", where mm=month, dd=day, yyyy=year, hh=hour, and mm=minute.
- T: Temperature of the sensor measured in degrees Celsius (ºC).
- RH: Relative humidity of the sensor measured in percentage (%).
- PM2.5_CF1_a: Raw 2-min PM2.5 particulate mass in ug/m3 from Channel A with a calibration factor of 1 (CF=1).
- PM2.5_CF1_b: Raw 2-min PM2.5 particulate mass in ug/m3 from Channel B with a calibration factor of 1 (CF=1).
- Avg_PM2.5_CF1: Averaged raw 2-min PM2.5 particulate mass in ug/m3 from Channel A and B with a calibration factor of 1 (CF=1).
- Raw_PM2.5: Converted 1-h average raw $PM_{2.5}$ concentrations measured in $\mu g/m^3$.
- Barkjohn_Corrected PM2.5: Corrected $PM_{2.5}$ concentrations using Barkjohn's US-wide correction equation measured in $\mu g/m^3$.
- TCEQ_C43_PM2.5: CAMS in Mission, $PM_{2.5}$ concentrations measured in $\mu g/m^3$.
- Corrected_Indoor: Converted 24-h average indoor corrected $PM_{2.5}$ concentrations measured in $\mu g/m^3$.
- Corrected_Outdoor: Converted 24-h average outdoor corrected $PM_{2.5}$ concentrations measured in $\mu g/m^3$.
- Final I/O: 24-hr average $PM_{2.5}$ indoor/outdoor (I/O) ratios. It is dimensionless.
- WD: Wind direction data retrieved from TCEQ CAMS measured in degrees.
- RWS: Resultant wind speed data retrieved from TCEQ CAMS measured in meters per second (m/s).

Table 1
Summarized site and date specifications of $PM_{2.5}$ data collected.

Site specifications		Sensors	Data collection period	Location
Home-1	Indoor	PA-1	6/16/2024 - 7/16/2024	26°10′47"N 97°59′46"W
	Outdoor	PA-2		
Home-2	Indoor	PA-3	6/22/2024 - 7/16/2024	26°09′51"N 98°17′52"W
	Outdoor	PA-4		
C-43	Ambient	FEM	6/16/2024 - 7/16/2024	26°13′34"N 98°17′27"W

3. Experimental Design, Materials and Methods

In this study, $PM_{2.5}$ data were collected using PurpleAir-II-SD sensors (LCS) at two residential homes (Home-1 and Home-2) in two microenvironments: indoor (kitchen) and outdoor (front yard). The data collection spanned about a month, from mid-June 2024 to mid-July 2024 (Table 1). The PA-II sensors were mounted approximately 1.7 m from the ground level within the "breathing zone." Both homes were detached, single-family residences, but they differ in layout, occupancy, and ventilation characteristics. In Home1 (Weslaco, TX), the total kitchen floor area was 9.75 m^2, while in Home-2 (Mission, TX), it was 22.11 m^2. Home-1 had four occupants present throughout the study period while Home-2 also had four occupants, with variations in presence: three occupants from June 23 to June 29 and one occupant from June 30 to July 6, with the house being occupied at all times in both locations. Outdoor conditions varied slightly, with three cars parked in the front yard of Home-1 and two cars in Home-2. Cooking activities also followed different schedules: Home-1 had lunch from 12:00 to 13:00 and dinner preparation from 17:00 to 18:00, with breakfast occasionally prepared between 9:00 and 10:00, whereas Home-2 had lunch from 13:00 to 14:00 and dinner from 19:00 to 21:00, with no significant breakfast cooking. Both kitchens were equipped with electrical stoves and exhaust fans with air recirculating within the kitchen; however, their ventilation practices varied. In Home-1, the doors or windows were never kept open. Conversely, in Home-2, the kitchen was connected to the dining room and led to the backyard, where the door was opened for a maximum of five seconds at least twice daily, while windows remained closed during the study period. Both homes had heating, ventilation, and air conditioning (HVAC) systems, but they used different types of air filters. Home-1 used MERV 13 filters, which are effective at filtering fine particles, while Home-2 used MERV 1 filters. The HVAC systems in both homes were mainly used for controlling temperature and providing ventilation during hot summer months. The TCEQ CAMS located at Mission: CAMS-43 is equipped with a Federal Equivalent Method (FEM) monitor for PM2.5 [17]. It utilizes broadband spectroscopy 638, which is a method of detection for both coarse and fine particulates [8]. Fig. 1 shows the location of the two homes and the two TCEQ CAMS sites. The deployment of the PurpleAir sensors at the two homes is shown in Fig. 2.

The PA-II sensors are laser-scattering particle sensors that measure $PM_{2.5}$ along with the sensor's temperature and relative humidity. The data collected by these sensors can be accessed in real-time; however, it is also saved to a microSD card in case the users do not have Wi-Fi. The data gets stored in a microSD card at every 2-min interval. It was downloaded after the completion of the study period. After retrieving the PA-II sensors, the raw data was cleaned and converted into a one-hour average using Microsoft Excel (v.16.06, Microsoft Inc., Redmond, WA, USA) and formatted in CSV files. It is important to note that the PA-II sensors measure raw $PM_{2.5}$ data, which requires bias (accuracy) correction. To ensure the accurate performance of all the low-cost sensors, a parallel comparison was conducted. Fig. 3 shows a pairwise correlation matrix visualized as a scatterplot matrix for four sensors (PA_1, PA_2, PA_3, and PA_4). The data from sensors demonstrated very high linear correlations, with Pearson's r values above 0.97 and adjusted R^2 values above 0.94. These strong positive relationships imply that the data from sensors are highly precise.

Therefore, the raw $PM_{2.5}$ data was then corrected using the US-wide correction equation developed by [3]. The equation was developed based on various collocations for PurpleAir sensors

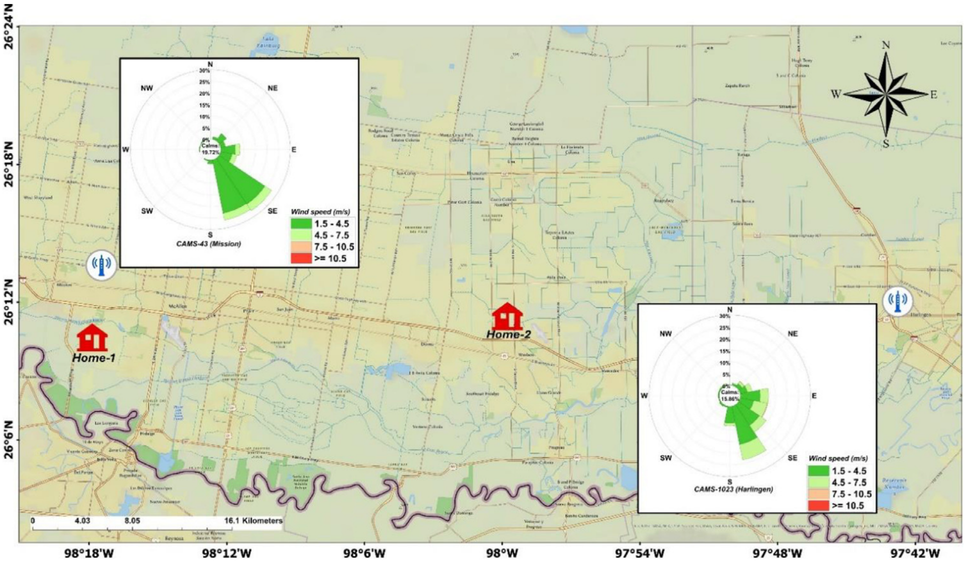

Fig. 1. Location of the study site (Home-1 and Home-2) with the nearest TCEQ CAMS sites. The Wind Rose diagram graphically presents wind conditions, direction, and resultant wind speed obtained from TCEQ CAMS-43 (approx.4.3 miles away from Home-1) and CAMS-1023 (approximately 17.6 miles away from Home-1).

Fig. 2. Deployment of PurpleAir sensors at different environments for both homes.

against regulatory (federally operated air monitoring sites) deployed across the United States. This equation was used in multiple studies investigating outdoor air quality ([4]; Shahar [14]) as well as indoor air quality [6]. The US-wide correction equation developed by [3] which is used in this study is as below:

$$\boldsymbol{PM}_{2.5,\boldsymbol{corrected}} = \boldsymbol{PM}_{2.5,\boldsymbol{raw}}\ (\boldsymbol{CF} = 1) \times 0.524 - \boldsymbol{RH} \times 0.0862 + 5.75 \quad \text{(i)}$$

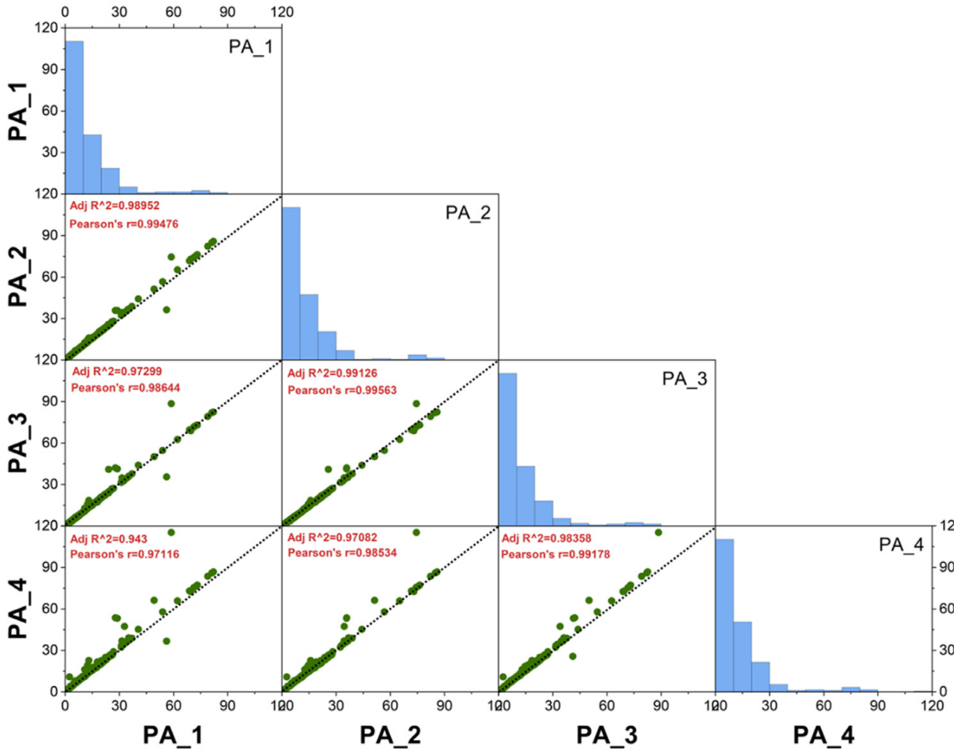

Fig. 3. Scatterplot matrix and histograms showing pairwise relationships of PA_1, PA_2, PA_3, and PA_4.

Table 2
Descriptive statistics of 1-h $PM_{2.5}$ average (μg/m^3) for both the homes, including CAMS (C-43) nearest to Home-2.

Environment		N total	Mean (SD)	Minimum	Maximum
Home 1 (Weslaco)	Indoor	732	3.71 (3.03)	2.22	40.38
	Outdoor	726	5.68 (4.23)	0.08	43.80
	I/O	726	0.98 (2.03)	0.14	37.99
Home 2 (Mission)	Indoor	588	3.24 (0.90)	2.40	8.00
	Outdoor	588	6.91 (5.46)	1.11	44.76
	I/O	588	0.61 (0.31)	0.11	2.27
C-43 (Mission)	Ambient	582	12.78 (6.45)	3.30	39.10

*SD- Standard deviation

where, $PM_{2.5,\ raw}$ represents the raw $PM_{2.5}$ (in μg/m^3) of correction factor (CF=1) measured by PA-II sensor, RH represents relative humidity (in %) of the PA-II sensor.

After implementing Barkjohn's US-wide correction equation (i), we obtained the final corrected $PM_{2.5}$ measurements. The corrected $PM_{2.5}$ measurements were then analyzed, and using Origin Pro (Origin Lab Corporation, Northampton, MA, USA Version 2024), we generated hourly and daily time series plots, hourly boxplots, and descriptive statistics for 1-h averaged $PM_{2.5}$ concentrations. Table 2 represents the descriptive statistics of 1-h $PM_{2.5}$ concentrations (in μg/m^3) at two residences in each indoor and outdoor microenvironment. Similarly, Fig. 4 shows the time series of 1-h average PM2.5 concentrations at different sampling sites.

For Home-1, the mean (SD) for $PM_{2.5}$ in the outdoor (front yard) was 5.68 (4.23) μg/m^3, which was higher than the indoor (kitchen) at 3.71 (3.03) μg/m^3. The kitchen (indoor) recorded a maximum 1-h $PM_{2.5}$ of 40.38 μg/m^3, whereas the front yard (outdoor) recorded a maximum 1-h

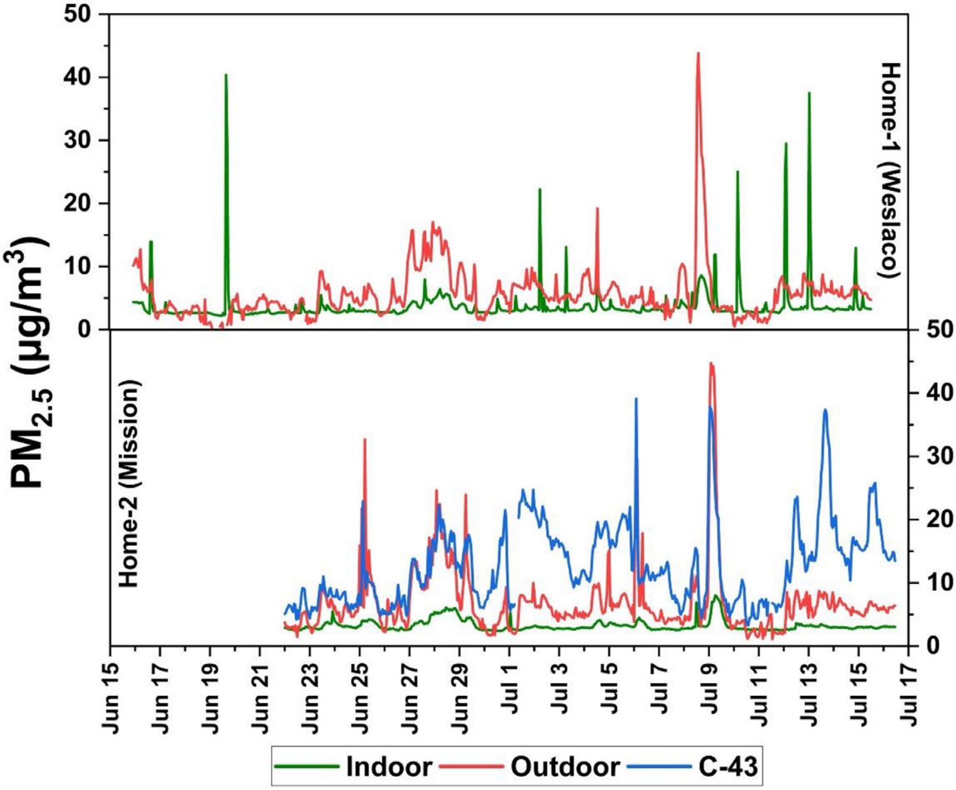

Fig. 4. Time series showing 1-h average $PM_{2.5}$ concentration for indoor and outdoor environments along with TCEQ CAMS (C-43) in Mission. The green line represents indoor PM2.5 levels, the red line represents outdoor PM2.5 levels, and the blue line (for Home-2) depicts data from a nearby TCEQ monitoring station (C-43).

$PM_{2.5}$ of 43.80 µg/m^3. Similarly, for Home-2, the mean (SD) for $PM_{2.5}$ in the outdoor (front yard) was recorded as 6.91 (5.46) µg/m^3, which was significantly higher than the indoor (kitchen) at 3.24 (0.90) µg/m^3. The kitchen (indoor) recorded a maximum 1-h $PM_{2.5}$ of 8.00 µg/m^3, whereas the front yard (outdoor) recorded a maximum 1-h $PM_{2.5}$ of 44.76 µg/m^3. Additionally, $PM_{2.5}$ data was obtained from TCEQ CAMS site [16], which is approximately 6.4 kilometers away from Home-2. The mean (SD) of the 1-h averaged $PM_{2.5}$ concentration at TCEQ CAMS in Mission (C-43) was 12.78 (6.45) µg/m^3, with values ranging from 3.30 µg/m^3 to 39.10 µg/m^3 as shown in Table 2.

For Home-1, the mean (SD) for PM2.5 in the outdoor (front yard) was 5.68 (4.23) µg/m3, which was higher than the indoor (kitchen) at 3.71 (3.03) µg/m3. The kitchen (indoor) recorded a maximum 1-h PM2.5 of 40.38 µg/m3, whereas the front yard (outdoor) recorded a maximum 1-h PM2.5 of 43.80 µg/m3. Similarly, for Home-2, the mean (SD) for PM2.5 in the outdoor (front yard) was recorded as 6.91 (5.46) µg/m3, which was significantly higher than the indoor (kitchen) at 3.24 (0.90) µg/m3. The kitchen (indoor) recorded a maximum 1-h PM2.5 of 8.00 µg/m3, whereas the front yard (outdoor) recorded a maximum 1-h PM2.5 of 44.76 µg/m3. Additionally, PM2.5 data was obtained from the TCEQ CAMS site [16], which is approximately 6.4 kilometers away from Home-2. The mean (SD) of the 1-h averaged PM2.5 concentration at TCEQ CAMS in Mission (C-43) was 12.78 (6.45) µg/m3, with values ranging from 3.30 µg/m3 to 39.10 µg/m3 as shown in Table 2. Thus, outdoor PM2.5 levels generally exhibit more significant fluctuations compared to indoor levels, with peaks corresponding to specific dates.

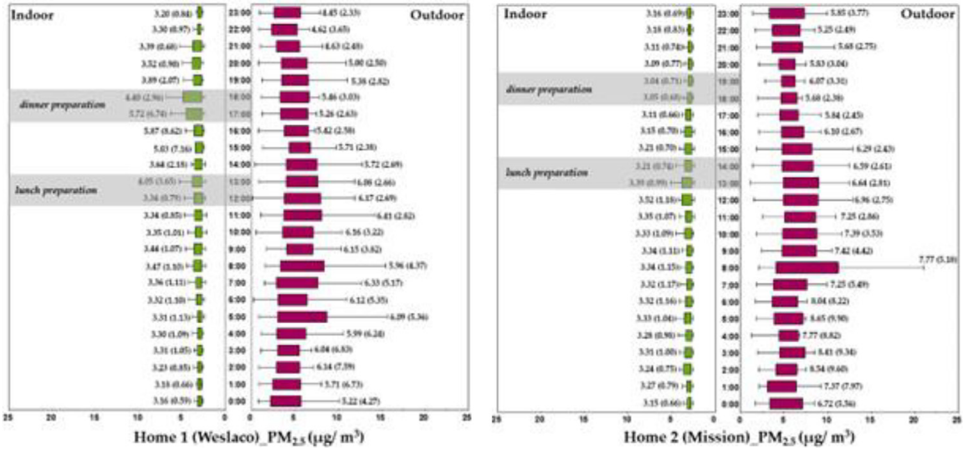

Fig. 5. Hourly box plots with mean $PM_{2.5}$ concentrations (Indoor vs. Outdoor) for Home-1 in Weslaco and Home-2 in Mission. (Note: Outlier values were not included).

Fig. 5 illustrates the hourly mean PM2.5 concentrations (measured in µg/m³) for both indoor and outdoor environments in two homes. The left panel depicts Home-1 (Weslaco), while the right panel represents Home-2 (Mission). Indoor PM2.5 levels, indicated by the green bars, are consistently lower than outdoor levels, represented by the purple bars. Cooking activities are associated with a notable increase in hourly mean PM2.5 concentrations in both homes. In Home-1, a rise in PM2.5 levels was noted in the kitchen during cooking periods, specifically from 12:00 to 13:00 (lunch) and again from 17:00 to 18:00 (dinner). Similarly, in Home-2, elevated PM2.5 levels were recorded in the kitchen during meal preparation, occurring from 13:00 to 14:00 for lunch and from 18:00 to 19:00 for dinner. A study by [13] also observed an increase in indoor PM2.5 levels due to cooking activities at residential homes. In contrast, outdoor PM2.5 values tend to fluctuate more significantly throughout the day, particularly during the early morning and late afternoon hours, which can be attributed to traffic flow.

$PM_{2.5}$ concentrations (in µg/m³) for indoor (I) and outdoor (O) environments, as well as the I/O ratio at Home– 1 and Home– 2 are shown in Fig. 6. In Home-1 (top panel), indoor $PM_{2.5}$ levels (green line) remain relatively low, while outdoor concentrations (red line) fluctuate more significantly. The I/O ratio (orange bars) is generally below 1, indicating lower $PM_{2.5}$ levels indoors compared to outdoors. In Home-2 (bottom panel), data also includes nearby station C-43 (blue line), with outdoor and C-43 measurements showing higher variability, especially in early July. The I/O ratio for Home-2 is similar to that of Home-1, with most values below 1, reflecting lower indoor $PM_{2.5}$ levels. Overall, both homes experience higher $PM_{2.5}$ concentrations outdoors than indoors.

Limitations

This research endeavor is part of a citizen science project involving high school students. The primary aim of this project was to introduce low-cost sensors and their functionality to citizen scientists. However, it is important to note that this study does have some limitations. To increase the accuracy of data collected by low-cost sensors, we followed the U.S-wide correction equation by [3]. Ideally, the four sensors used in this study should have been collocated against the Texas Commission on Environmental Quality (TCEQ) Continuous Ambient Monitoring Station (CAMS) site here in Brownsville, TX. But getting but the requisite permission to mount these four sensors at any TCEQ CAMS site is very difficult and next to impossible. Hence, our approach adopted the [3] method. The data was also collected for one month in Home-1 (Weslaco) and

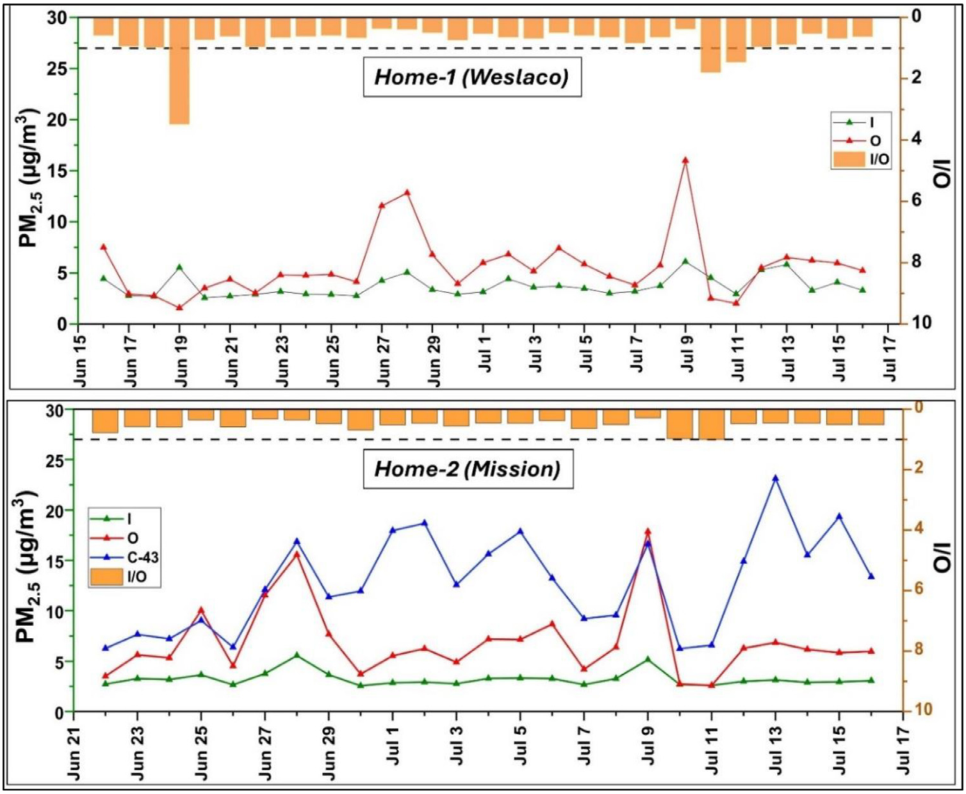

Fig. 6. Time series for 24-h average $PM_{2.5}$ concentration for Home-1 and Home-2 along with indoor/outdoor (I/O) ratios

twenty-five days in Home-2 (Mission), primarily in the summer season. Future studies should focus on collecting data over different seasons and meteorological conditions to characterize the functionality of the low-cost sensors. Understanding sources of indoor pollution is also essential in citizen science projects. Data on infiltration rates and HVAC system leaks plays a key role in identifying these pollution sources and the impact of outdoor pollution on indoor environments. Unfortunately, the absence of data on both infiltration rates and HVAC leaks restricts the study's ability to analyze the sources of indoor air pollution thoroughly. Nevertheless, this dataset provides key insights into particulate pollution data in low-income neighborhoods located along the US-Mexico border.

Ethics Statement

The authors confirm that they have read and understood the ethical requirements for publishing data in Data in Brief. This dataset does not involve human subjects, animal experiments, or data collected from social media platforms.

Data Availability

Fine particulate matter (PM2.5) datasets for both Indoor and Outdoor environments at two South Texans residences (Original data) (Mendeley Data).

CRediT Author Statement

Kabir Bahadur Shah: Writing – original draft, Data curation, Formal analysis, Software; **Sai Deepak Pinakana:** Investigation, Validation, Writing – review & editing; **Mkhitar Hobosyan:** Conceptualization, Writing – review & editing; **Armando Montes:** Resources, Supervision; **Amit U. Raysoni:** Conceptualization, Supervision, Writing – review & editing.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to express their gratitude to high scholars from the UTRGV Math and Science Academy (MSA), namely Kirra Cantu and Andrea Garza, for their participation in this study and for granting permission to deploy the low-cost sensors at their respective homes, Home 1 (Weslaco) and Home 2 (Mission).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Abraham, X. Li, A cost-effective wireless sensor network system for indoor air quality monitoring applications, Procedia Comput. Sci. 34 (2014) 165–171, doi:10.1016/j.procs.2014.07.090.
[2] J.S. Apte, J.D. Marshall, A.J. Cohen, M. Brauer, Addressing global mortality from ambient PM2.5, Environ. Sci. Technol. 49 (13) (2015) 8057–8066, doi:10.1021/acs.est.5b01236.
[3] K.K. Barkjohn, B. Gantt, A.L. Clements, Development and application of a United States-wide correction for $PM_{2.5}$ data collected with the PurpleAir sensor, Atmosph. Meas. Tech. 14 (6) (2021) 4617–4637, doi:10.5194/amt-14-4617-2021.
[4] D.A. Jaffe, C. Miller, K. Thompson, B. Finley, M. Nelson, J. Ouimette, E. Andrews, An evaluation of the U.S. EPA's correction equation for PurpleAir sensor data in smoke, dust, and wintertime urban pollution events, Atmosph. Meas. Tech. 16 (5) (2023) 1311–1322, doi:10.5194/amt-16-1311-2023.
[5] E.R. Jones, J.G. Cedeño Laurent, A.S. Young, P. MacNaughton, B.A. Coull, J.D. Spengler, J.G. Allen, The effects of ventilation and filtration on indoor PM2.5 in office buildings in four countries, Build. Environ. 200 (2021) 107975, doi:10.1016/j.buildenv.2021.107975.
[6] K. Koehler, M. Wilks, T. Green, A.M. Rule, M.L. Zamora, C. Buehler, A. Datta, D.R. Gentner, N. Putcha, N.N. Hansel, G.D. Kirk, S. Raju, M. McCormack, Evaluation of calibration approaches for indoor deployments of PurpleAir monitors, Atmosph. Environ. 310 (2023) 119944–119944, doi:10.1016/j.atmosenv.2023.119944.
[7] C. Krittanawong, Y.K. Qadeer, R.B. Hayes, Z. Wang, G.D. Thurston, S. Virani, C.J. Lavie, PM2.5 and cardiovascular diseases: state-of-the-art review, Int. J. Cardiol. Cardiovasc. Risk Prevent. 19 (2023) 200217, doi:10.1016/j.ijcrp.2023.200217.
[8] Ohio EPA (2024). 2024-2025 Ohio EPA air monitoring network plan Ohio EPA division of air pollution control. [online] Available at: https://dam.assets.ohio.gov/image/upload/epa.ohio.gov/Portals/27/ams/sites/2024-2025_AMNP_MainReport_Final.pdf [Accessed 23 Jan. 2025].
[9] S.D. Pinakana, C. Garcia Patlan, E. Mendez, A.U. Raysoni, A pilot study on particulate matter concentrations from cooking and its effects on indoor air pollution in a Mexican American household in Mission, South Texas, USA, Case Stud. Chem. Environ. Eng. (2024) 100757–100757, doi:10.1016/j.cscee.2024.100757.
[10] PurpleAirPurpleAir-II-SD Classic Air Quality Monitor, PurpleAir, 2019 https://www2.purpleair.com/products/purpleair-pa-ii?srsltid=AfmBOooiE_Etkf1A9oK_hIPQFiO2A8pMb7yKO7fRs8kqQCHjIICyL5nV Accessed on September 15, 2024.
[11] A.U. Raysoni, R.X. Armijos, M.M. Weigel, T. Montoya, P. Eschanique, M. Racines, W.W. Li, Assessment of indoor and outdoor PM species at schools and residences in a high-altitude Ecuadorian urban center, Environ. Pollut. 214 (2016) 668–679, doi:10.1016/j.envpol.2016.04.085.
[12] A.U. Raysoni, R. Armijos, M. Weigel, P. Echanique, M. Racines, N. Pingitore, W.-W. Li, Evaluation of sources and patterns of elemental composition of PM2.5 at three low-income neighborhood schools and residences in Quito, Ecuador, Int. J. Environ. Res. Public Health 14 (7) (2017) 674–674, doi:10.3390/ijerph14070674.
[13] K.B. Shah, D. Kim, S.D. Pinakana, M. Hobosyan, A. Montes, A.U. Raysoni, Evaluating indoor air quality in residential environments: a study of PM2.5 and CO2 dynamics using low-cost sensors, Environments 11 (11) (2024) 237, doi:10.3390/environments11110237.

[14] S. Tsameret, D. Furuta, P. Saha, N. Kwak, A. Hauryliuk, X. Li, A.A. Presto, J. Li, Low-cost indoor sensor deployment for predicting $PM_{2.5}$ exposure, ACS ES&T Air 1 (8) (2024) 767–779, doi:10.1021/acsestair.3c00105.
[15] SoGA. (2024). *A special report on global exposure to air pollution and its health impacts, with a focus on children's health*. https://www.healthdata.org/sites/default/files/2024-06/soga-2024-report.pdf. Accessed on September 15, 2024.
[16] TCEQ. (2018). *Yearly Summary Report*. Www.tceq.texas.gov. https://www.tceq.texas.gov/cgi-bin/compliance/monops/yearly_summary.pl. Accessed on September 15, 2024.
[17] TCEQ (2024). Texas commission on environmental quality 2024 annual monitoring network plan. [online] Available at: https://www.tceq.texas.gov/downloads/air-quality/air-monitoring/network/historical/2024-amnp.pdf [Accessed 23 Jan. 2025].
[18] US EPAWhy Indoor Air Quality is Important to Schools | US EPA, US EPA, 2023 https://www.epa.gov/iaq-schools/why-indoor-air-quality-important-schools Accessed on October 5, 2024.
[19] W.R. Wan Mahiyuddin, R. Ismail, N. Mohammad Sham, N.I. Ahmad, N.M.N. Nik Hassan, Cardiovascular and respiratory health effects of fine particulate matters (PM2.5): a review on time series studies, Atmosphere 14 (5) (2023) 856, doi:10.3390/atmos14050856.