Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Low-Cost Sensor Monitoring of Volatile Organic Compound (VOC) Temporal Trends in Lackawanna County, Pennsylvania: A Citizen Science Approach

Gerard Dumancas

ggdumancas@ncat.edu

North Carolina Agricultural and Technical State University

Norwell Bautista

University of Scranton

Tonyehn Verkitus

Physicians for Social Responsibility Pennsylvania

Christina DiGiulio

Physicians for Social Responsibility Pennsylvania

Rebecca Cardin

Physicians for Social Responsibility Pennsylvania

Research Article

Keywords: Citizen science, volatile organic compounds, low-cost sensors, environmental monitoring, principal component analysis, environmental justice

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10729498/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Low-Cost Sensor Monitoring of Volatile Organic Compound (VOC) Temporal Trends in Lackawanna County, Pennsylvania: A Citizen Science Approach

Gerard G. Dumancas[1,2*], Norwell Bautista[3], Tonyehn Verkitus[4*], Christina DiGiulio[4], and Rebecca Cardin[4]

[1]Honors College, North Carolina Agricultural and Technical State University, Greensboro, NC 27411, USA

[2]Department of Chemistry, North Carolina Agricultural and Technical State University, Greensboro, NC 27411, USA

[3]Department of Chemistry, University of Scranton, Loyola Science Center, Scranton, PA 18510, USA

[4]Physicians for Social Responsibility Pennsylvania, Philadelphia, PA 19103, USA

N. Bautista

email: norwellbautista@gmail.com

T. Verkitus

email: tonyehn@psrpa.org

C. DiGiulio

email: chris@psrpa.org

R. Cardin

email: rebecca@psrpa.org

*Corresponding authors:
Gerard G. Dumancas
Honors College and Department of Chemistry
North Carolina Agricultural and Technical State University
Greensboro, NC 27411, USA
Email: ggdumancas@ncat.edu
Phone: +1-336-285-2164

Tonyehn Verkitus
Physicians for Social Responsibility Pennsylvania
Philadelphia, PA 19103, USA
Email: tonyehn@psrpa.org
Phone: +1-609-874-9230

Abstract: Volatile organic compounds (VOCs) are hazardous air pollutants associated with adverse respiratory and chronic health outcomes, yet conventional monitoring approaches often lack the spatial coverage needed to address community-level environmental concerns. Although low-cost air sensors can substantially expand monitoring networks, their application to VOC monitoring remains limited by uncertainty in absolute concentration measurements. This study evaluated whether community-deployed low-cost sensors can reliably identify meaningful temporal VOC patterns across an environmental justice community in Lackawanna County, Pennsylvania, surrounding the Keystone Sanitary Landfill. Community-deployed PurpleAir monitors collected VOC index data at six-hour intervals throughout 2025 across the Lackawanna

County study area as part of a citizen science monitoring initiative. Data were analyzed using smoothing, mean-centering, principal component analysis (PCA), and k-means clustering to emphasize relative temporal behavior rather than absolute sensor responses. Smoothed time-series revealed coherent long-term VOC trends across geographically distributed monitoring sites despite substantial short-term variability. PCA showed that a small number of components explained much of the meaningful variance in the dataset, indicating strong temporal consistency across the sensor network. K-means clustering identified two broad sensor groups, although substantial overlap suggested a continuum of VOC temporal responses rather than distinct site-specific regimes. These findings demonstrate that trend-based multivariate approaches can extract meaningful environmental information from low-cost VOC sensors and support the use of citizen science monitoring networks to expand air-quality surveillance, inform environmental management, and strengthen community engagement in environmental decision-making.

Keywords: Citizen science; volatile organic compounds; low-cost sensors; environmental monitoring; principal component analysis; environmental justice

Introduction

Volatile organic compounds (VOCs) comprise a diverse class of carbon-containing chemicals that readily evaporate into the atmosphere and participate in photochemical reactions. VOCs originate from both natural and anthropogenic sources and are emitted from industrial processes, transportation activities, fossil fuel production, building materials, consumer products, and numerous other sources (Koppmann, 2008; Zhou et al., 2023). In urban and industrial environments, major anthropogenic contributors include petrochemical facilities, power

generation, manufacturing operations, construction activities, and solvent use (Atkinson & Arey, 2003; Banaszkiewicz et al., 2022). Because of their widespread occurrence and role in atmospheric chemistry, VOCs are recognized as important constituents of both indoor and outdoor air pollution (Koppmann, 2008; Zhou et al., 2023).

Human exposure to VOCs occurs primarily through inhalation, making air quality an important determinant of potential health effects (Liu et al., 2022; Pandit et al., 2001; Ramírez et al., 2012). Numerous studies have linked VOC exposure to adverse health outcomes, including respiratory disease, allergic responses, cardiovascular effects, and increased cancer risk, although risks vary considerably depending on the specific compounds involved, exposure concentrations, duration, and individual susceptibility (Calenic et al., 2015; Gasparri et al., 2018; Liu et al., 2022; Lv et al., 2023; Rumchev et al., 2007). Exposure to compounds such as benzene has been associated with leukemia and other adverse health outcomes, while VOCs also contribute to atmospheric photochemical reactions that generate ground-level ozone, a pollutant known to negatively affect respiratory health (Liu et al., 2022; Sharma et al., 2024; Soares & Silva, 2022; Zulkifli et al., 2022). Consequently, spatially and temporally resolved monitoring is important for understanding population exposure and supporting environmental health assessment (Li & Hackenberger, 2020; Sexton & Hattis, 2007; Vandenberg et al., 2023).

Air quality remains an important environmental and public health concern throughout Pennsylvania. The American Lung Association's 2025 State of the Air report identified several Pennsylvania metropolitan areas as among the most polluted regions in the United States for ozone and particulate matter pollution, highlighting ongoing challenges in protecting public health and environmental quality (American Lung Association, 2025b, 2025a). These concerns are particularly relevant in communities located near industrial facilities, transportation corridors,

landfills, and energy infrastructure, where residents frequently report odors, emissions, and other environmental concerns (American Lung Association, 2025b). One such community is Lackawanna County in northeastern Pennsylvania, where residents have reported recurring odor events and air-quality concerns associated with the Keystone Sanitary Landfill (KSL) and other regional infrastructures (National Association of Counties, 2025; Steininger, 2023). Previous evaluations conducted by the Pennsylvania Department of Health and the Agency for Toxic Substances and Disease Registry concluded that current exposure conditions are unlikely to cause significant non-cancer health effects; however, compounds such as benzene and formaldehyde may contribute to elevated cancer risks and important uncertainties remain regarding vapor migration and localized exposure pathways (Pennsylvania Department of Health, 2019). These uncertainties, combined with ongoing community concerns, emphasize the importance of obtaining localized environmental monitoring data capable of characterizing spatial and temporal variability in air quality.

Traditional VOC monitoring methods, including gas chromatography-mass spectrometry, flame ionization detection, and electron capture detection, provide highly accurate measurements but are often expensive, labor-intensive, and impractical for dense community-scale monitoring networks (Epping & Koch, 2023; Spinelle et al., 2017). In contrast, low-cost air sensors offer a scalable and cost-effective approach for increasing the spatial coverage of environmental monitoring. Among these technologies, PurpleAir sensors incorporate metal-oxide semiconductor sensing elements capable of detecting changes associated with VOC presence through changes in electrical resistance (Khatib & Haick, 2022). Although the VOC measurements produced by these sensors are experimental and do not represent direct compound-specific concentrations, they can

provide valuable information regarding temporal patterns and relative changes in local air quality (Okorn & Iraci, 2024).

The increasing availability of low-cost sensors has expanded opportunities for citizen science and community-based environmental monitoring. Citizen science approaches can increase monitoring coverage, enhance community engagement, improve environmental literacy, and provide residents with locally relevant information that complements traditional regulatory monitoring programs (Sun et al., 2022). Community-deployed monitoring networks are particularly valuable in environmental justice communities where environmental concerns may not be fully captured by conventional monitoring infrastructure.

Despite growing interest in low-cost air-monitoring technologies, relatively few studies have evaluated whether community-deployed low-cost VOC sensors can characterize regional VOC temporal patterns in environmental justice communities using trend-based analytical approaches. Furthermore, uncertainties associated with sensor-specific response characteristics and limitations in absolute VOC quantification have constrained the interpretation of low-cost VOC measurements. As a result, there remains a need to determine whether meaningful environmental information can be extracted from large community-deployed VOC sensor networks through multivariate analysis of temporal trends.

Therefore, the objective of this study was to evaluate whether a citizen science network of low-cost PurpleAir sensors could identify meaningful temporal VOC patterns across Lackawanna County, Pennsylvania, despite the known limitations of low-cost VOC measurements. Using community-deployed monitoring stations and chemometric approaches, including PCA and k-means clustering, this study examined the extent to which trend-based analyses could reveal shared VOC temporal behavior across the monitoring network. The findings contribute to the growing

body of literature on community-based environmental monitoring and provide insight into the potential role of low-cost sensor networks in supporting environmental assessment, public engagement, and environmental decision-making.

Materials and methods

Study area

Lackawanna County is located in northeastern Pennsylvania and contains a mixture of urban, suburban, and former industrial communities situated within the Lackawanna Valley. The primary focus of this study was the area surrounding the KSL, a major regional waste disposal facility located in Dunmore and Throop Boroughs. KSL has operated for more than four decades and occupies approximately 714 acres, including approximately 442 acres designated for waste disposal under permits issued by the Pennsylvania Department of Environmental Protection (PADEP) (Commonwealth of Pennsylvania Northeast Community Information, 2026; Pennsylvania Department of Environmental Protection, 2014). The landfill was established during the early 1970s to serve as a regional disposal facility for residential and commercial solid waste and expanded into neighboring Throop Borough in 1987 as the original disposal area approached capacity. Today, KSL accepts municipal solid waste from Pennsylvania and neighboring states, as well as construction debris, sludge, flood-related waste materials, and oil- and gas-related waste streams associated with extraction activities in the Marcellus Shale region. A summary of oil- and gas-related waste streams reported by PADEP is presented in Table 1 (Pennsylvania Department of Environmental Protection, 2026).

Table 1. Oil- - and gas–related waste streams received by the Keystone Sanitary Landfill (Dunmore, Pennsylvania) as of February 4th, 2026 (Pennsylvania Department of Environmental Protection, 2026).

Waste Type	**Amount (gallons)**	**Amount (barrels)**
Drilling Fluid Waste (in Tons) RWC 803	840 gals	20 bbls
Unused Fracturing Fluid Waste (in Barrels) RWC 805	2,175 gals	52 bbls
Servicing Fluid (in Barrels) RWC 808	4,200 gals	100 bbls
Servicing Fluid (in Tons) RWC 808	4,595 gals	109 bbls
Brine	4,830 gals	115 bbls
Soil Contaminated by Oil & Gas Related Spills	7,286 gals	173 bbls
Other Oil & Gas Waste (in Barrels) RWC 899	46,620 gals	1,110 bbls
Unused Fracturing Fluid Waste (in Tons) RWC 805	120,879 gals	2,879 bbls
Synthetic Liner Materials (in Tons) RWC 806	309,538 gals	7,370 bbls
Drill Cuttings (in Tons) RWC 810	43,494,090 gals	1,035,574 bbls

Among the most significant waste categories received by KSL are drill cuttings and other materials associated with hydraulic fracturing operations. Historical records indicate that more than 43 million gallons of drill cuttings have been disposed of at the facility (Marcellus Gas, 2026). Previous studies have reported that wastes associated with unconventional natural gas extraction may contain elevated concentrations of heavy metals, naturally occurring radioactive materials, and VOC-related constituents, including benzene, toluene, ethylbenzene, and xylene compounds (Perma-Fix Environmental Services, Inc., 2016). In addition, the handling and treatment of landfill leachate may contribute to the release of VOCs and hazardous air pollutants through evaporation and other waste-management processes (Pennsylvania Department of Environmental Protection Newsroom, 2023).

Community concerns regarding air quality in the vicinity of KSL have persisted for several years and are frequently associated with odor events reported by nearby residents (Steininger, 2023). In 2025, PADEP documented more than 195 odor complaints associated with the landfill (Pennsylvania Department of Environmental Protection, 2025). Previous evaluations conducted by the Pennsylvania Department of Health and the Agency for Toxic Substances and Disease Registry identified compounds including benzene, formaldehyde, naphthalene, acetaldehyde, acrolein, ammonia, dioxane, and methylamine as contaminants of potential concern in areas surrounding the landfill (Pennsylvania Department of Health, 2019). These findings, combined with recurring community concerns, provided motivation for the deployment of a community-based VOC monitoring network.

The study area is further characterized by complex topography associated with the Lackawanna Valley, a landscape shaped by historical coal mining activities and mountainous terrain. Acid mine drainage features, abandoned mine infrastructure, and valley-floor communities are widely distributed throughout the region (Susquehanna River Basin Commission, 2026). The valley topography may influence airflow, pollutant transport, and contaminant accumulation, particularly during atmospheric inversion events and periods of reduced atmospheric mixing. Monitoring locations were, therefore, distributed across a broad elevation gradient and across communities surrounding KSL to capture potential spatial variability in VOC temporal trends (Fig. 1). Air-quality monitors were deployed at volunteer residences in Dunmore, Throop, Scranton, Peckville, Jefferson Township, and other nearby communities. Monitoring locations were selected based on volunteer participation, proximity to potential emission sources, and the need to maximize spatial coverage throughout the study area. Site elevations ranged from approximately 158 m to more than 620 m above mean sea level (Table 2). Elevation data were obtained using ArcGIS elevation

profiling tools and rounded to the nearest meter (Esri Inc., 2024). Lower-elevation locations were generally situated within valley-floor residential corridors, while higher-elevation monitoring sites were located on surrounding uplands and ridgelines. The wide distribution of monitoring locations provided an opportunity to evaluate regional VOC temporal behavior across diverse geographic and topographic settings.

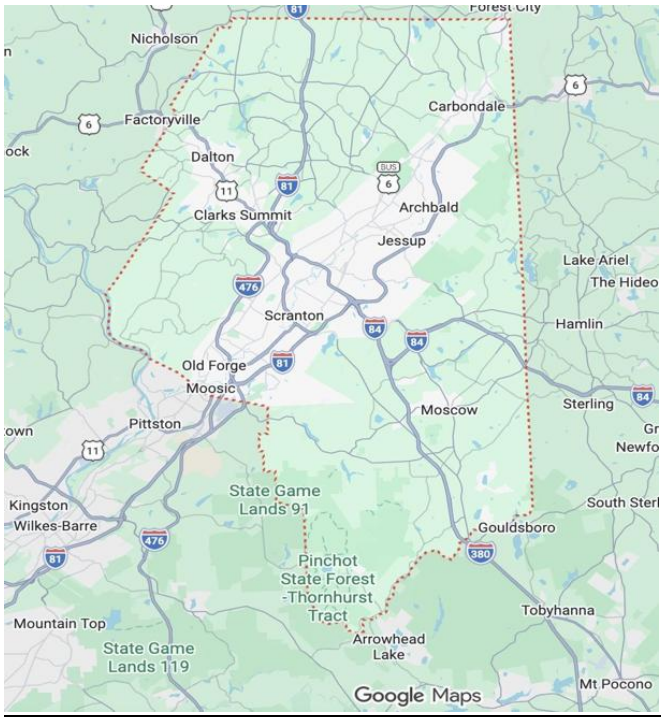


Fig. 1. Location of the Lackawanna County study area in northeastern Pennsylvania, USA, including the communities surrounding the Keystone Sanitary Landfill and associated air-monitoring sites.

Table 2. Summary of air-quality monitoring sites in the Lackawanna County study area. Site identification numbers correspond to labels used in the volatile organic compound (VOC) temporal trend analyses (Fig. 10). Elevations were obtained using ArcGIS elevation profiling tools and rounded to the nearest meter. Data completeness (%) represents the percentage of expected VOC observations successfully acquired during 2025. Assuming data collection at 6-hour intervals between January 1 and December 31, 2025, the maximum number of observations per sensor was 1,460.

Site ID	Purple Air Sensor ID	Community / Township	Elevation (m)	Latitude (°N)	Longitude (°W)	Data completeness (%)
(1)	204737	Dunmore, PA	359	41.42071303	-75.61098168	98
(2)	204747	Dunmore, PA	319	41.42628452	-75.61684801	90
(3)	194379	Dunmore, PA	316	41.42640709	-75.61703868	86
(4)	204761	Dunmore, PA	249	41.43949	-75.61342	86
(5)	204749	Throop, PA	267	41.44824054	-75.60916	100
(6)	194421	Throop, PA	245	41.45311	-75.60312	100
(7)	172343	Dunmore, PA	280	41.43091576	-75.6291728	100
(8)	205271	Dunmore, PA	275	41.42477	-75.64015	100
(9)	194467	Throop, PA	294	41.46849183	-75.56220209	100
(10)	204763	Peckville,	241	41.48003	-75.57749	98

		PA				
(11)	204765	Peckville, PA	242	41.48003	-75.57749	98
(12)	193003	Scranton, PA	307	41.49337227	-75.56461028	96
(13)	194437	Valley corridor	271	41.50087908	-75.54011321	84
(14)	190525	Delaware, PA	170	40.0656	-75.73775	89
(15)	266053	Delaware, PA	99	39.9712679	-75.3573679	79
(16)	266469	Delaware, PA	107	39.96788936	-75.36329656	80
(17)	172309	Delaware, PA	17	39.962733	-75.155226	100

Air Monitor Low Cost Sensor

The PurpleAir PA-II-FLEX is a low-cost air-quality monitoring device that provides real-time measurements of particulate matter ($PM_{1.0}$, $PM_{2.5}$, and PM_{10}), VOC index values, temperature, and relative humidity (Air Quality Management District, 2021; Karaoghlanian et al., 2022; PurpleAir, 2026b). The monitor contains a Bosch BME680 or BME688 multi-sensor module that integrates temperature, humidity, pressure, and metal-oxide semiconductor gas-sensing technology. VOCs are detected through interactions with the heated metal-oxide sensing surface, resulting in changes in electrical resistance that are processed through a proprietary algorithm to generate a VOC index value (Air Quality Management District, 2021; PurpleAir, 2026c, 2026b).

In addition to VOC measurements, the PA-II-FLEX incorporates dual optical particle sensors (typically Plantower PMS5003 or PMS6003 models) that independently measure particulate matter concentrations. The dual-channel design provides internal quality assurance by allowing

simultaneous measurements to be compared within a single instrument and has contributed to the widespread use of PurpleAir sensors in community-based air-quality monitoring studies (deSouza et al., 2023; PurpleAir, 2026c, 2026b). Representative internal components of the PurpleAir monitor are shown in Fig. 2.

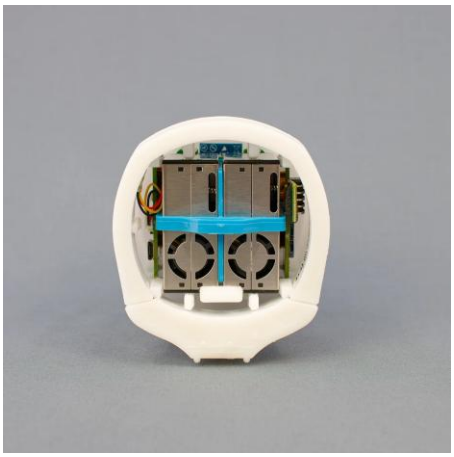


Fig. 2. Internal components of the PurpleAir PA-II-FLEX air-quality monitor showing the dual optical particle sensors, internal electronics, and airflow channels used for particulate matter and environmental measurements.

The PurpleAir VOC index provides a relative indication of changes in ambient VOC conditions rather than direct compound-specific VOC concentrations. Sensor response may be influenced by environmental conditions, including temperature, humidity, and the presence of interfering gases (PurpleAir, 2022, 2026a). Consequently, this study focused on temporal VOC behavior and relative trend patterns rather than absolute concentration estimates. Previous studies have demonstrated that low-cost sensor networks can provide valuable information regarding spatial and temporal variability in air quality when interpreted using appropriate analytical and statistical approaches (Okorn & Iraci, 2024).

Community Monitoring Network Deployment

Deployment of air-quality monitors was conducted through a community-based monitoring program coordinated by Physicians for Social Responsibility Pennsylvania (PSR PA). Monitoring locations were identified through community outreach, resident requests, and collaboration with local community organizations. The deployment strategy was designed to maximize spatial coverage across the study area while accommodating volunteer participation and site accessibility. Applicants were required to have access to a stable outdoor electrical outlet, reliable Wi-Fi connectivity, and a suitable outdoor mounting location. Monitoring sites were screened to minimize potential interference from local emission sources, including heating, ventilation, and air-conditioning systems, outdoor cooking areas, and dense vegetation. Monitors were typically mounted on exterior walls, poles, shepherd's hooks, or similar structures that provided adequate airflow and weather exposure while maintaining secure installation conditions. Representative examples of monitor deployment are shown in Fig. 3.



Fig. 3. Representative deployment locations of PurpleAir PA-II-FLEX monitors at volunteer residences in Lackawanna County, Pennsylvania. Monitors were installed outdoors to provide continuous measurements of ambient air quality and volatile organic compound temporal variability.

Following site approval, monitors were installed by PSR PA personnel and connected to participants' wireless networks. Each unit was registered through the PurpleAir platform using its assigned device identification number and geographic coordinates. Upon activation, monitoring data became publicly accessible through the PurpleAir network and could be viewed by both participants and project personnel. The monitoring network was developed to capture spatial variability in VOC behavior across the study area. Although ideal monitoring designs employ uniform spatial distributions surrounding potential emission sources, monitor placement in community-based projects is inherently constrained by volunteer participation and property availability. Consequently, site selection reflected a balance between scientific objectives, community engagement, and practical deployment considerations. Monitoring locations were distributed across multiple communities and topographic settings to maximize spatial coverage and support regional comparisons of VOC temporal trends.

Citizen Science Implementation

Citizen science served as the foundation for the development and operation of the community-based air-monitoring network. PSRPA worked directly with residents and local community organizations to identify environmental concerns, recruit monitoring participants, and facilitate the deployment of low-cost air-quality sensors throughout the study area. Community engagement efforts increased following recurring odor events reported by residents living in communities surrounding the KSL. Participant recruitment was conducted through community outreach, public announcements, social media communications, and collaboration with local organizations concerned about environmental quality. Residents voluntarily applied to participate in the monitoring program and agreed to host air-quality monitors at their residences for long-term environmental monitoring. Following enrollment, participants received information regarding monitor installation requirements, basic operation of the monitoring equipment, and access to the PurpleAir platform. Residents were encouraged to report technical issues, maintain continuous power and internet connectivity, and document local activities that could potentially influence sensor readings. Monitoring data were made publicly accessible through the PurpleAir network, allowing participants to observe temporal changes in local air-quality conditions and monitor environmental events in near real time.

The citizen-science framework enabled the deployment of a geographically distributed monitoring network across multiple communities in Lackawanna County. This approach substantially increased the spatial coverage of air-quality monitoring while promoting public engagement in environmental data collection and interpretation. In addition to supporting scientific data collection, the monitoring program provided residents with access to localized environmental

information relevant to ongoing community concerns regarding air quality and odor events. An overview of the citizen-science monitoring workflow is presented in Fig. 4.

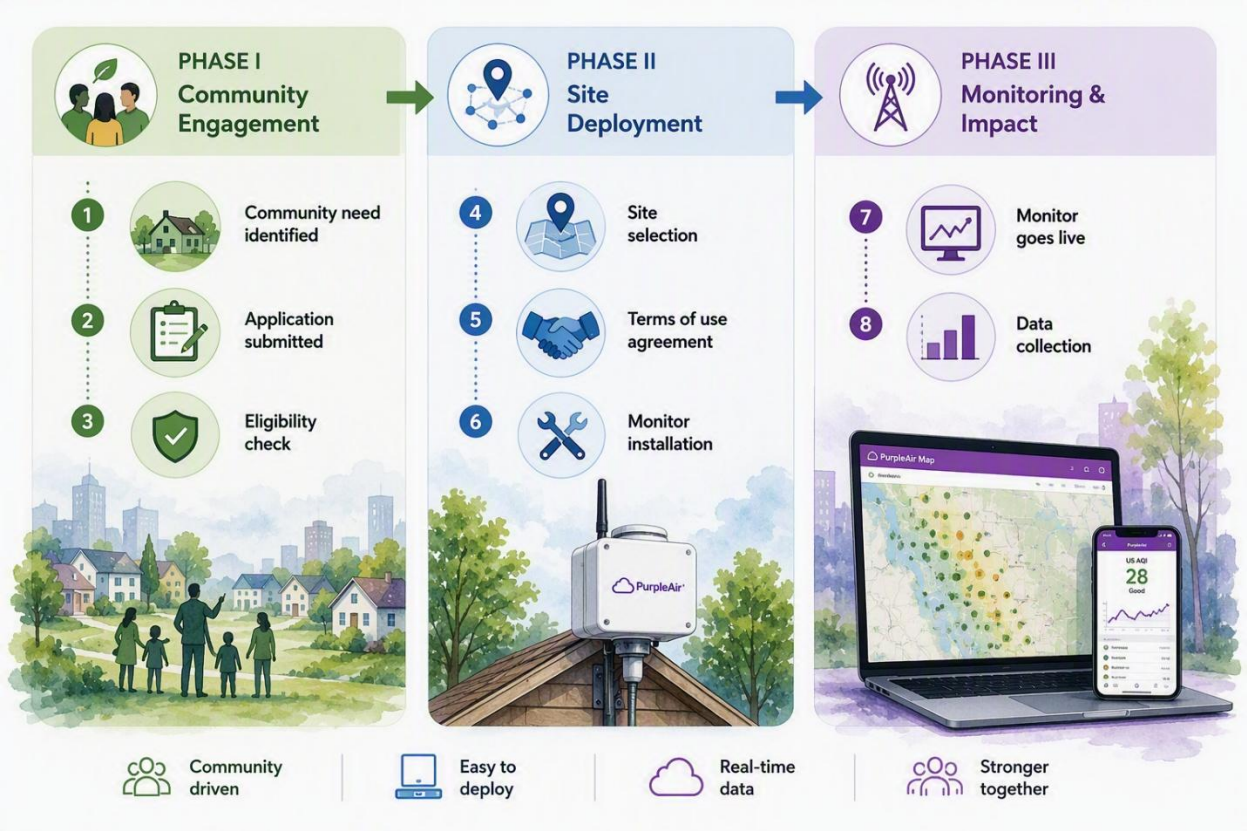


Fig. 4. Conceptual framework of the citizen-science volatile organic compound monitoring program implemented in Lackawanna County, Pennsylvania. Community engagement supports participant recruitment and site evaluation, followed by monitor deployment, network registration, and continuous environmental monitoring through the PurpleAir platform.

Data Acquisition and Preprocessing

VOC data were acquired through the PurpleAir Application Programming Interface for community-deployed sensors located in Lackawanna County, Pennsylvania. VOC index measurements collected between January 1 and December 31, 2025 were extracted at 6-hour

intervals for subsequent analysis. Although PurpleAir sensors are capable of recording measurements at shorter intervals, temporal aggregation to 6-hour intervals was performed to reduce short-term variability and highlight longer-term temporal patterns relevant to regional air-quality assessment. To provide a geographically distinct comparison network, VOC data from four PurpleAir sensors located in Delaware County, Pennsylvania, were also acquired. Delaware County sensors were selected because they were geographically separated from Lackawanna County and possessed sufficiently complete records for temporal comparison. These sensors were included to evaluate whether temporal VOC patterns observed in Lackawanna County represented localized behavior or broader regional trends. Data processing and statistical analyses were performed using R statistical software (R Core Team, 2024). Data completeness was evaluated for all sensors prior to analysis. Sensors with less than 80% of the expected observations during the study period were excluded to minimize the influence of extended data gaps and incomplete time series. Because low-cost VOC sensors may exhibit substantial variability in signal magnitude due to sensor-specific characteristics, aging, and environmental influences, analyses focused on relative temporal behavior rather than absolute VOC values. To facilitate comparison among monitoring locations, VOC time-series data were mean-centered prior to multivariate analysis according to:

$$x_c = x_i - \bar{x}$$

where x_c is the mean-centered VOC value, x_i is the original VOC measurement at observation i, and $\bar{x}$ is the mean VOC value for the corresponding monitoring site over the study period.

Mean centering removes sensor-specific offsets and enables multivariate analyses to emphasize similarities and differences in temporal behavior among monitoring locations. Following preprocessing, the resulting dataset was used for PCA and k-means clustering.

Principal Component Analysis (PCA)

PCA was used to evaluate similarities and differences in VOC temporal trends across the monitoring network and to reduce dataset dimensionality while preserving the dominant sources of variance. PCA is widely used in environmental chemistry and air-quality research to identify patterns in multivariate datasets where variables are correlated and measurements may vary among monitoring locations (Jolliffe, 2002; Vu et al., 2020).

VOC index data were aggregated to 6-hour intervals prior to analysis to reduce short-term variability while retaining temporal patterns relevant to regional air-quality dynamics. Because VOC index values obtained from low-cost sensors may vary substantially in magnitude due to sensor-specific characteristics, aging, and environmental influences, the analysis focused on relative temporal behavior rather than absolute values. Accordingly, VOC time-series data for each sensor were mean-centered prior to PCA to ensure that extracted components reflected shared temporal patterns rather than differences in signal magnitude (Jolliffe, 2002).

PCA was performed on the covariance matrix of the mean-centered dataset. Missing observations were minimal relative to the complete dataset and were replaced with zeros following centering to maintain matrix completeness for eigenvalue decomposition. The number of principal components retained for interpretation was determined using the elbow method based on explained variance. Principal component scores and loadings were visualized using biplots to evaluate relationships among sensors and identify similarities in VOC temporal behavior across monitoring

locations. This approach has been widely applied in environmental monitoring studies to identify dominant variability patterns, assess spatial coherence, and support interpretation of pollutant behavior without requiring compound-specific measurements (Chuang et al., 2024; Jiang, 2023). All analyses were performed using R statistical software (R Core Team, 2024), and graphical visualizations were generated using established R packages (Kassambara et al., 2026; R Core Team, 2024; Wickham, 2016).

K-Means Clustering Analysis

K-means clustering was applied to identify groups of monitoring sites exhibiting similar VOC temporal behavior. K-means is an unsupervised classification technique that partitions observations into a predefined number of clusters by minimizing within-cluster variance and maximizing similarity among observations within each cluster (MacQueen, 1967). In this study, the method was used to evaluate whether subsets of sensors shared common VOC temporal patterns across the monitoring network. Clustering was performed using the mean-centered VOC time-series data described previously. Mean-centering reduced the influence of sensor-specific baseline differences and emphasized relative temporal variability, thereby allowing clustering to focus on similarities in temporal behavior rather than absolute signal magnitude. Euclidean distance was used as the similarity metric, consistent with common applications of k-means clustering to continuous environmental datasets (Doan et al., 2023).

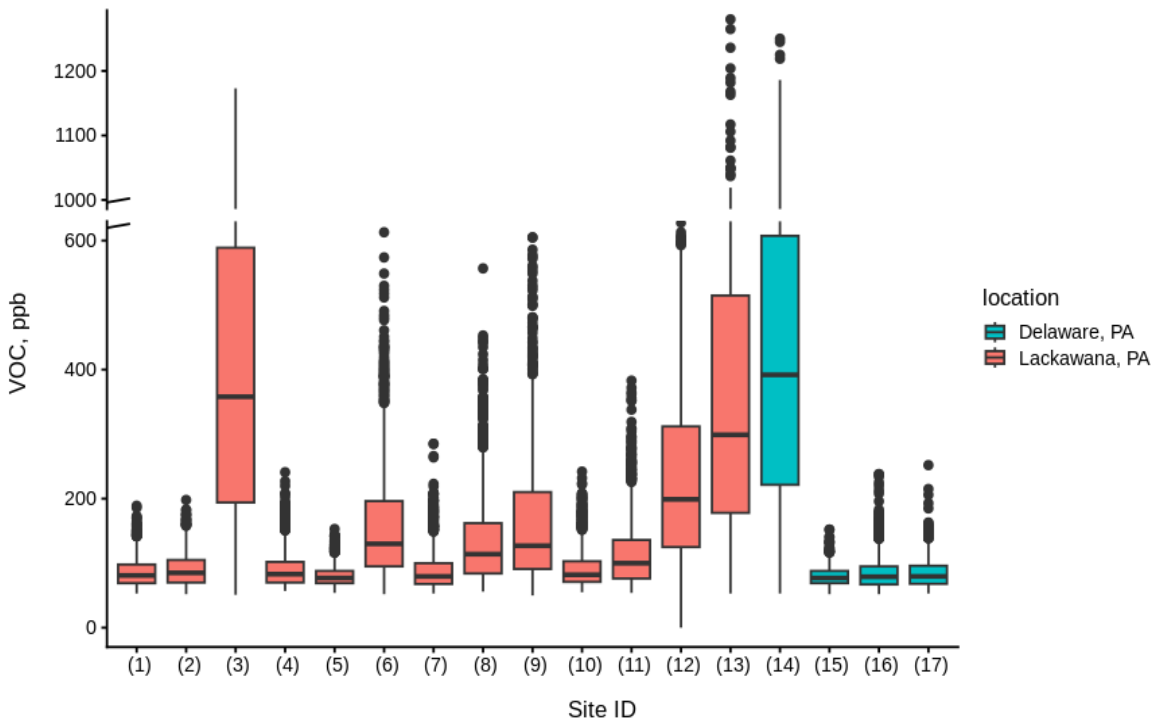
The optimal number of clusters (k) was determined using the average silhouette width. The silhouette method evaluates cluster quality by comparing the similarity of each observation to its assigned cluster with its similarity to neighboring clusters, thereby balancing within-cluster cohesion and between-cluster separation (Rousseeuw, 1987). Multiple values of k were evaluated,

and the solution providing the highest average silhouette width and greatest interpretability was selected. Following cluster assignment, results were visualized using principal component score plots to facilitate interpretation and comparison with PCA outcomes. Cluster overlap and separation were evaluated to assess the degree of similarity in VOC temporal patterns across monitoring locations. Clustering results were interpreted as indicators of relative similarity among monitoring sites rather than evidence of discrete pollution sources or distinct emission regimes. All clustering analyses were conducted using R statistical software (R Core Team, 2024).

Results and Discussion

VOC Temporal Trends and Clustering of Monitoring Sites

VOC index values recorded by the monitoring network between January 1 and December 31, 2025 exhibited substantial variability across sensors, with values ranging from approximately 0 to 1280 (Fig. 5). Considerable differences in VOC index magnitude were observed both within and among monitoring locations, highlighting the challenges associated with interpreting absolute VOC readings from low-cost sensors. Variability in sensor response is expected due to differences in local environmental conditions, sensor characteristics, and transient atmospheric influences. Consequently, subsequent analyses focused primarily on temporal behavior rather than absolute VOC index values.



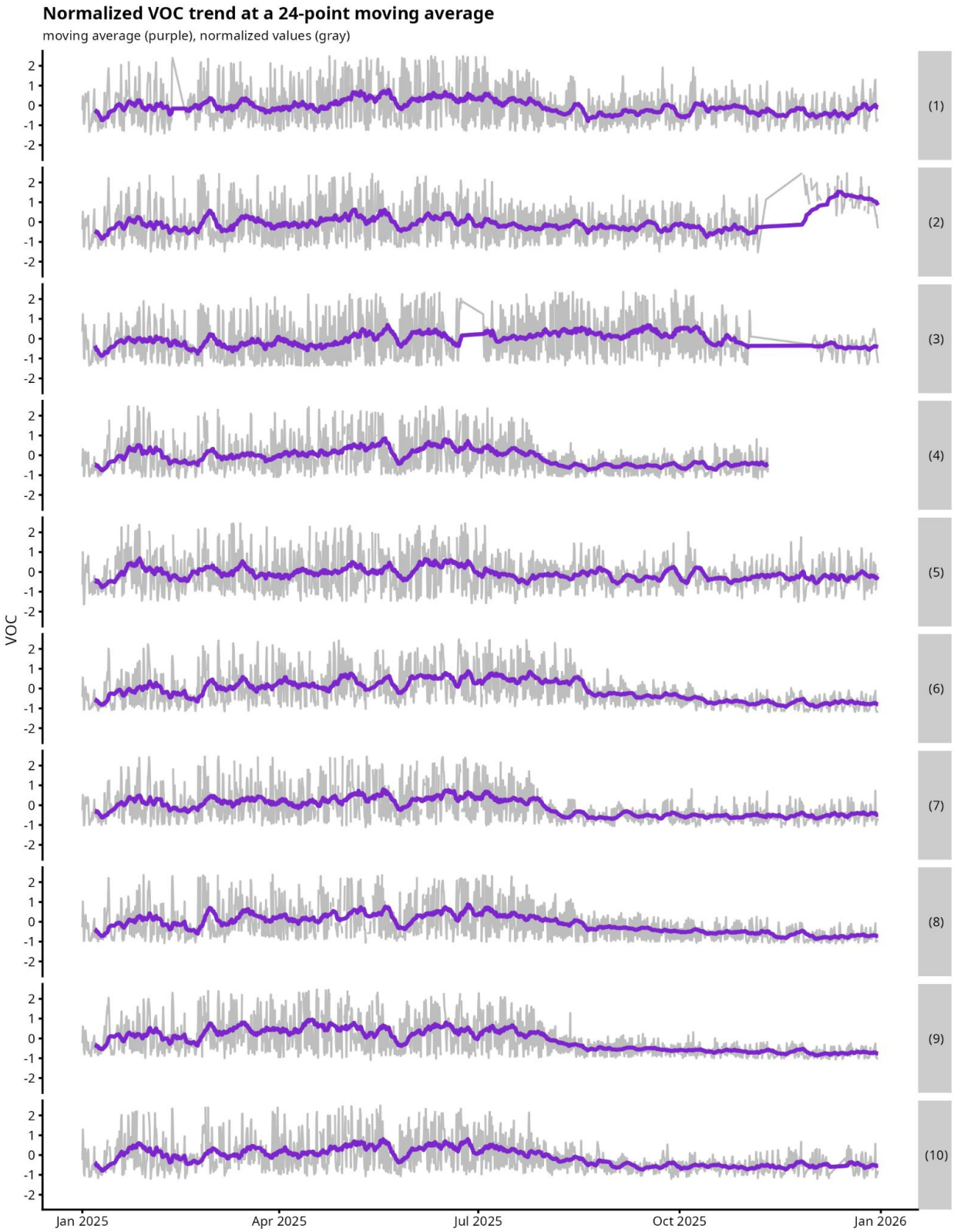
409

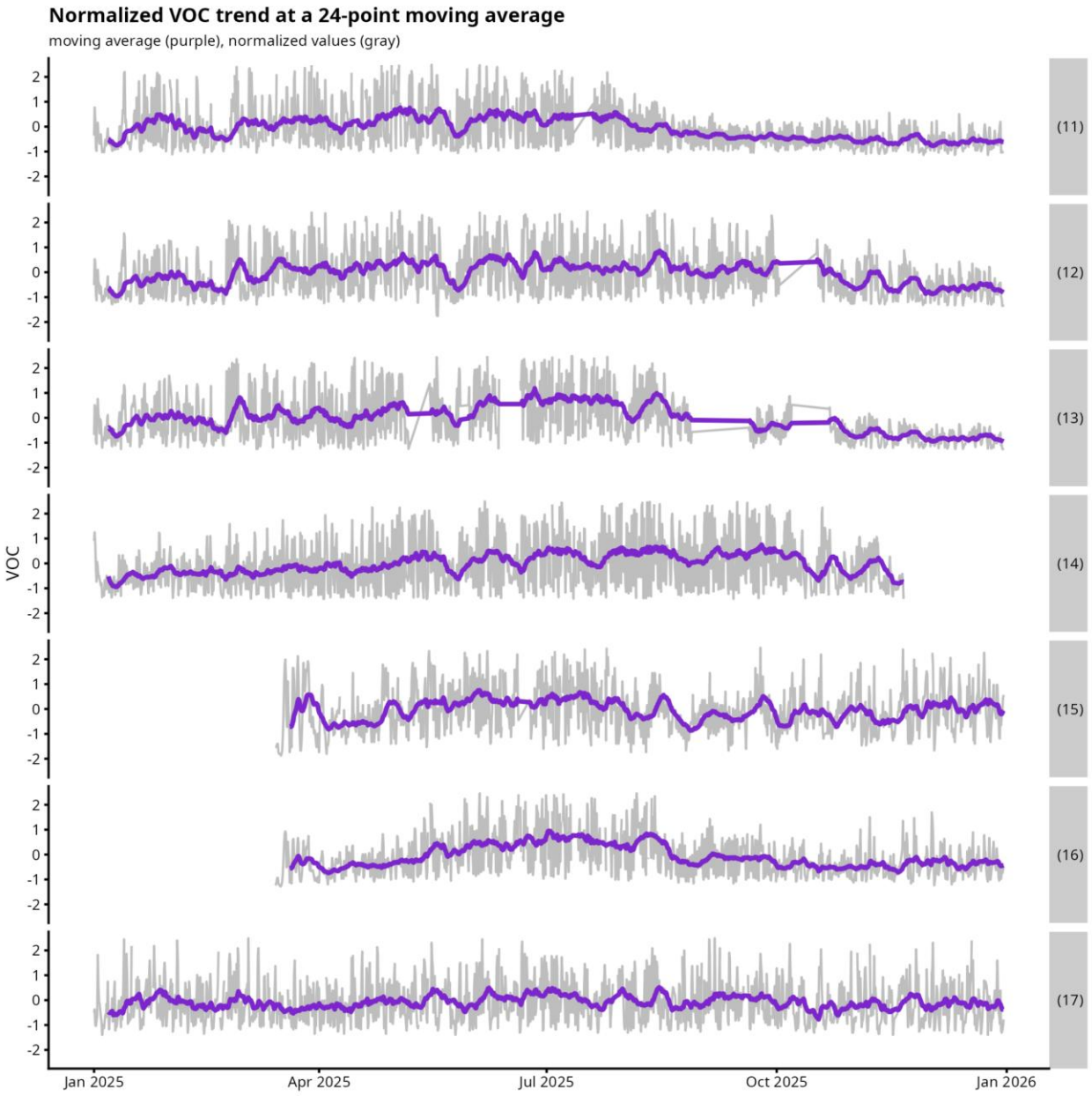
410 **Fig. 5.** Distribution of volatile organic compound (VOC) index values recorded by PurpleAir
411 sensors deployed in Lackawanna County (orange) and Delaware County (blue-green).
412 Considerable variability is observed both within and among monitoring locations, highlighting the
413 importance of trend-based analyses for interpretation of low-cost VOC sensor data.

414

415 Raw VOC time-series data revealed substantial short-term fluctuations throughout the study period
416 (Fig. 6). Rapid changes in VOC index values and intermittent spikes were evident across nearly
417 all monitoring locations, reflecting the known sensitivity of low-cost VOC sensors to transient
418 environmental conditions and localized influences. Although these high-frequency variations
419 dominated the raw signals, they obscured broader temporal patterns and limited direct comparison
420 among monitoring sites. Application of a 24-point moving average revealed coherent long-term
421 temporal trends across many monitoring locations. Despite differences in signal magnitude and
422 short-term variability, the smoothed trends exhibited broadly similar temporal structure over the

monitoring period. Multiple sites showed increasing VOC trends during the first half of the year followed by stabilization or gradual declines during later months. These patterns occurred across locations representing different communities, elevations, and distances from the KSL, suggesting the influence of common regional-scale temporal drivers. An important feature of the smoothed time-series data is the consistency in trend direction and shape observed across diverse geographic settings. Valley-floor residential locations, upland monitoring sites, and comparison locations generally exhibited similar long-term temporal behavior. This observation suggests that broader influences, such as seasonal variability and regional meteorological processes, may exert greater control on VOC temporal behavior than localized emission sources alone. Localized influences were nevertheless evident through differences in signal amplitude and the frequency of short-term fluctuations among monitoring locations.


Normalized VOC trend at a 24-point moving average
moving average (purple), normalized values (gray)
VOC
2
1
0
-1
-2
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
(9)
(10)
Jan 2025
Apr 2025
Jul 2025
Oct 2025
Jan 2026



435

436 **Fig. 6.** Normalized volatile organic compound (VOC) temporal trends measured by PurpleAir
437 sensors across monitoring sites. Gray lines represent raw VOC index values, whereas purple lines
438 represent 24-point moving-average trends. Smoothed trends reveal common temporal patterns
439 despite substantial short-term variability.

440

Differences among monitoring sites were primarily expressed as variations in short-term variability and signal amplitude rather than as fundamentally different long-term temporal trends. Some sensors exhibited intermittent departures from the dominant temporal pattern, likely reflecting localized conditions, sensor-specific response characteristics, or differences in data completeness. However, these departures did not substantially alter the overall similarity observed in the smoothed temporal profiles. This finding reinforces the importance of emphasizing temporal behavior rather than absolute VOC index values when interpreting low-cost VOC sensor data.

Several monitoring locations displayed data gaps or truncated time series resulting from intermittent sensor operation and differences in deployment duration. These gaps were limited to individual sensors and did not appear to influence the overall network-wide temporal structure. The consistency of observed patterns among sites with more complete records supports the robustness of the broader temporal trends identified across the monitoring network. Collectively, the smoothed VOC time-series data indicates that shared temporal behavior represents the dominant organizing feature of the dataset. These observations provide justification for the application of trend-based multivariate analyses, including PCA and k-means clustering, which were designed to quantify similarities in temporal behavior while minimizing the influence of sensor-specific offsets and short-term noise.

PCA was applied to identify dominant sources of variability in the VOC dataset. Examination of the scree plot revealed a pronounced elbow after the second principal component, indicating that the first two components captured the dominant interpretable temporal structure in the dataset (Fig. 7). The first principal component explained 23.1% of the total variance, while the second principal component explained an additional 16.6%, together accounting for a substantial proportion of variability in VOC temporal behavior across the monitoring network. The PCA score

plot revealed separation between the Lackawanna County monitoring network and the Delaware County comparison network (Fig. 8). Sensors within each county are generally clustered together, indicating similarities in temporal VOC behavior among monitoring locations within the same geographic region. The separation observed in PCA space suggests differences in VOC temporal patterns between the two monitoring networks. However, PCA identifies similarities and differences in temporal behavior and should not be interpreted as direct evidence of distinct emission sources or discrete VOC systems. The relatively tight clustering of the Lackawanna County sensors further suggests that broad-scale temporal influences contributed more strongly to the observed variability than many local site-specific factors. Although localized influences such as nearby vegetation, traffic, residential activities, and other small-scale emission sources may contribute to short-term fluctuations, these influences did not dominate the overall temporal structure identified through PCA.

K-means clustering was subsequently applied to further evaluate groupings within the monitoring network. The average silhouette width reached a maximum at $k = 2$, indicating that a two-cluster solution provided the best balance between cluster cohesion and separation (Fig. 9). Increasing the number of clusters resulted in lower silhouette values and less distinct clustering, suggesting that additional partitioning did not improve interpretation of the dataset.

The resulting k-means analysis identified two broad groups of monitoring sites (Fig. 10). Although the two-cluster solution generally reflected the separation observed in PCA space, considerable overlap remained between clusters. This result indicates that VOC temporal behavior is better described as a continuum of related temporal responses rather than as sharply separated site categories. Accordingly, cluster membership should be interpreted as a measure of relative

similarity in temporal behavior rather than evidence of distinct VOC regimes or specific emission sources.

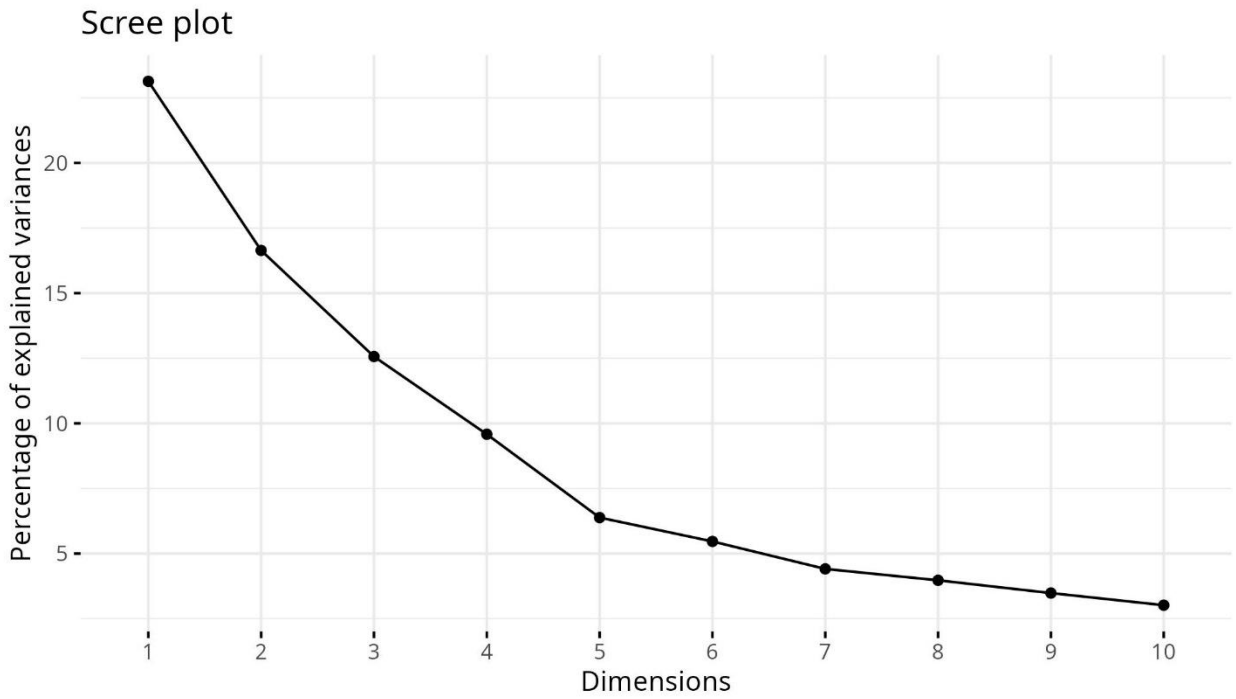
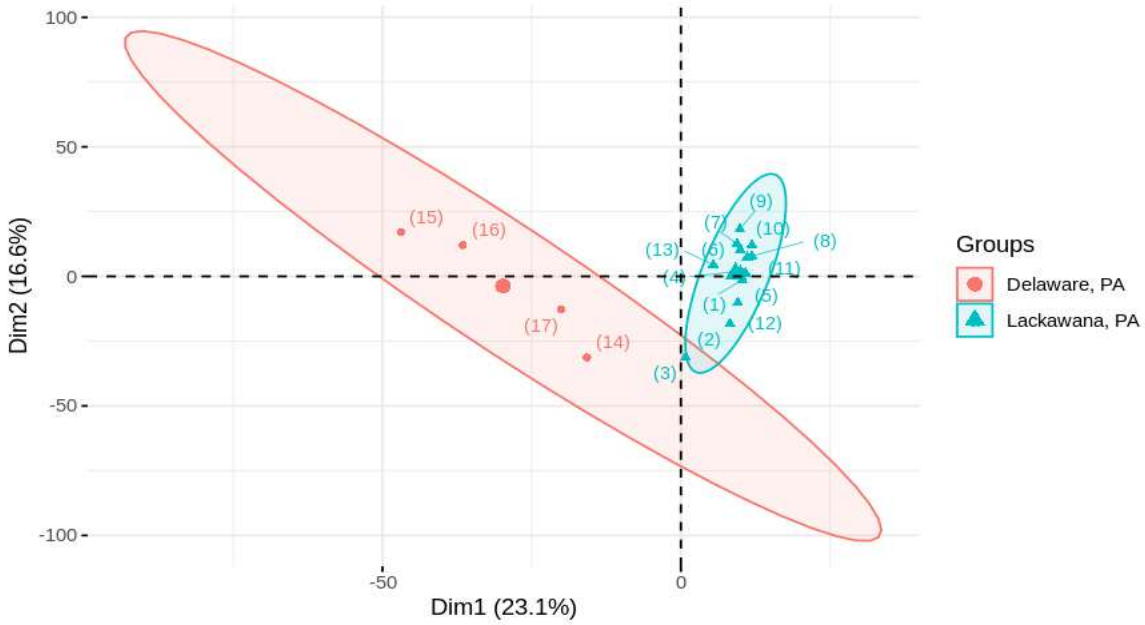
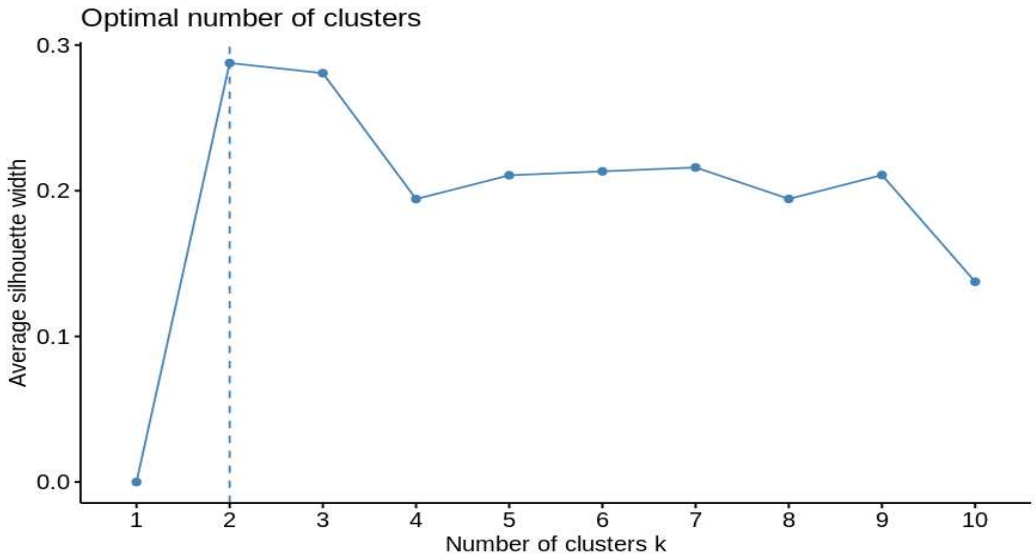


Fig. 7. Scree plot showing the percentage of variance explained by successive principal components. A pronounced elbow after the second component indicates that the first two principal components capture the majority of meaningful variance in the volatile organic compound dataset.



494

495 **Fig. 8.** Principal component analysis score plot of volatile organic compound (VOC) temporal
496 trends measured by PurpleAir sensors. Sensors from Lackawanna County cluster separately from
497 comparison sensors in Delaware County, indicating differences in temporal VOC behavior
498 between the two monitoring networks.



499

500 **Fig. 9.** Average silhouette width as a function of the number of clusters (k). The maximum
501 silhouette width occurs at $k = 2$, supporting selection of the optimal two-cluster solution.

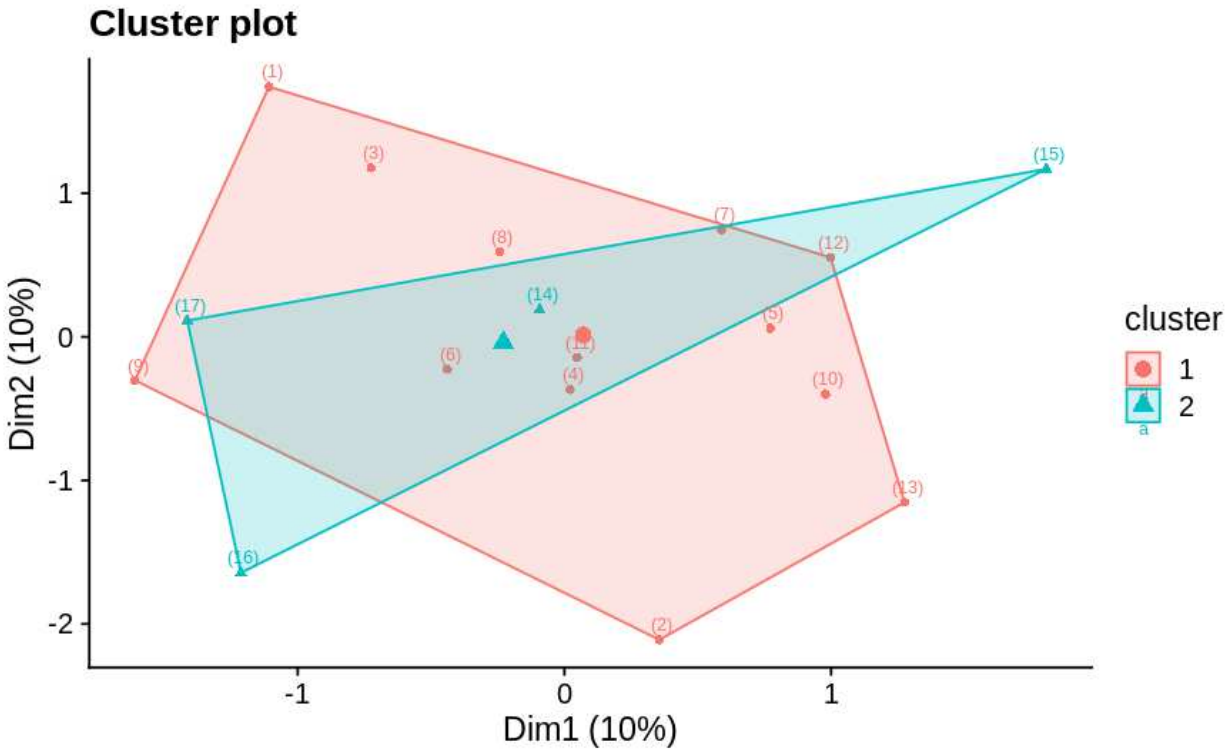


Fig. 10. K-means clustering biplot showing the distribution of monitoring sites within the optimal two-cluster solution. Clustering was based on similarities in volatile organic compound (VOC) temporal behavior rather than absolute VOC index values.

Interpretation and Significance of VOC Temporal Patterns

The VOC measurements collected across the monitoring network reveal an important distinction between short-term variability and longer-term temporal structure. Although raw VOC index values exhibited substantial high-frequency fluctuations at all monitoring locations, the emergence of coherent smoothed trends across geographically distributed sites indicates that meaningful information can be extracted from low-cost sensor data when appropriate analytical approaches are applied. These findings demonstrate that low-cost VOC sensors, despite their inherent variability and sensitivity to transient environmental conditions, can capture regionally coherent temporal behavior when analyses focus on relative temporal patterns rather than absolute signal

magnitude. This observation is consistent with previous studies suggesting that low-cost air-quality sensors are often most effective when used to evaluate temporal and spatial trends rather than direct pollutant concentrations (Okorn & Iraci, 2024).

The similarity in smoothed temporal trends observed across sites located at different elevations, distances from the KSL, and community settings suggests that VOC dynamics in the study area are influenced by common temporal drivers operating at regional scales. Seasonal variability, meteorological conditions, atmospheric mixing, and regional transport processes are likely contributors to the coherent patterns observed across the monitoring network. Although localized influences may affect individual measurements and contribute to short-term fluctuations, these effects appear secondary to the broader temporal structure observed throughout the study area. The findings, therefore, highlight the importance of considering regional temporal dynamics when interpreting environmental monitoring data collected using distributed sensor networks. From a methodological perspective, the results emphasize the limitations of interpreting raw VOC index values directly when using low-cost sensors. Sensor-specific variability, environmental influences, and differences in instrument response can substantially affect individual measurements. However, when data are evaluated using smoothing techniques and multivariate approaches that emphasize temporal structure, meaningful similarities among monitoring locations become apparent. The application of mean-centering, smoothing, PCA, and clustering methods enabled the identification of coherent temporal patterns that would be difficult to recognize through direct inspection of raw measurements alone. These findings demonstrate the value of trend-based analytical approaches for extracting environmental information from large networks of low-cost VOC sensors.

The PCA results indicate that VOC temporal variability across the monitoring network is organized around a limited number of shared patterns rather than numerous independent site-specific regimes. The separation observed between the Lackawanna County and Delaware County sensor networks suggests differences in temporal VOC behavior between the two regions, while the relatively tight grouping of sensors within each network indicates internal consistency in temporal trends. Importantly, PCA identifies similarities and differences in temporal behavior rather than direct evidence of specific emission sources. Consequently, the observed clustering should be interpreted as reflecting differences in regional temporal dynamics rather than proof of distinct VOC systems or pollution sources.

The k-means clustering analysis further supports the interpretation that VOC temporal behavior is broadly shared across monitoring locations. Although the optimal solution identified two clusters, substantial overlap remained between groups, indicating that the observed variability exists along a continuum rather than forming sharply separated categories. This finding cautions against overinterpreting cluster membership as evidence of discrete emission regimes or fundamentally different environmental conditions. Instead, clustering is most appropriately viewed as a descriptive tool for summarizing degrees of similarity in temporal behavior among monitoring sites.

Collectively, these findings demonstrate that VOC temporal behavior across the monitoring network is largely coherent despite differences in geography, topography, and site-specific conditions. The results provide evidence that community-based networks of low-cost sensors can capture meaningful regional air-quality dynamics when analyzed using appropriate multivariate techniques. More broadly, this study highlights the value of emphasizing temporal structure and network-level patterns when interpreting low-cost VOC sensor data, particularly in complex

environmental settings where short-term variability may otherwise obscure underlying signals. The observed coherence across the monitoring network further supports the use of citizen-science sensor deployments as complementary tools for environmental monitoring, community engagement, and exploratory assessment of air-quality conditions.

Lessons Learned from Community-Based Monitoring

The implementation of a community-based air-monitoring network in Lackawanna County provided several important lessons regarding the use of low-cost sensors for environmental monitoring and citizen science. Consistent with previous studies, community participation expanded the spatial coverage of monitoring activities and enabled environmental data collection at locations that would otherwise not be represented by traditional monitoring networks (Fraisl et al., 2022; Mahajan et al., 2020; Manshur et al., 2023). Beyond data collection, participant involvement increased awareness of local air-quality issues and fostered greater engagement with environmental concerns.

An important finding of this project was that low-cost air monitors served not only as scientific instruments but also as tools for community empowerment. Access to air-quality data provided residents with a mechanism to document observations, evaluate environmental concerns, and engage in discussions regarding potential pollution sources. For communities experiencing recurring odor events or environmental uncertainty, the availability of monitoring data helped validate lived experiences and facilitated more informed communication among residents, community organizations, and regulatory agencies.

583 The project also highlighted the importance of transparency in monitoring technologies. PurpleAir
584 monitors incorporate a dual-sensor architecture that provides internal redundancy and allows
585 comparison of measurements obtained simultaneously within a single instrument (Fig. 11).
586 Making participants aware of both the capabilities and limitations of the monitoring equipment
587 helped improve confidence in the data and promoted realistic expectations regarding interpretation
588 of low-cost sensor measurements. Transparency regarding sensor performance is particularly
589 important when community-generated data are used to inform discussions of environmental
590 quality and potential health concerns.

591 A practical lesson from the deployment process involved routine maintenance of monitoring
592 equipment. During the study, biological debris, including spider-web accumulation within sensor
593 housings, was occasionally observed (Fig. 11). Such conditions may influence airflow and sensor
594 performance, emphasizing the importance of periodic inspection and maintenance of low-cost
595 monitoring devices. Although these issues did not prevent the successful operation of the
596 monitoring network, they illustrate operational challenges that should be considered when
597 implementing long-term community-based monitoring programs.

598 Another important lesson was that access to environmental data alone does not necessarily
599 translate into effective community action. While residents valued access to monitoring
600 information, many participants required additional support in understanding data interpretation,
601 uncertainty, and engagement with governmental agencies responsible for environmental protection
602 and public health. Consequently, successful citizen-science initiatives should pair monitoring
603 technologies with educational resources, data interpretation support, and guidance on effective
604 communication with regulatory institutions.

605

606 Overall, this project demonstrates that low-cost sensor networks can provide both scientific and
607 societal benefits when implemented within a well-supported citizen-science framework. Effective
608 programs require clear objectives, sustained participant engagement, transparent communication,
609 quality-control procedures, and realistic expectations regarding both the capabilities and
610 limitations of low-cost monitoring technologies (Rambonnet et al., 2019). When these elements
611 are integrated successfully, community-based monitoring networks can generate valuable
612 environmental information while strengthening public participation in environmental decision-
613 making.
614



615
616
617 **Fig. 11.** Internal components of a PurpleAir monitor showing spider-web accumulation within the
618 sensor housing. Biological contamination and debris may influence airflow and sensor
619 performance, highlighting the importance of routine inspection and maintenance of low-cost
620 monitoring devices.

Future Directions

Future research should focus on strengthening the scientific, community, and policy relevance of low-cost VOC monitoring networks in Lackawanna County and similar environmental justice communities. Although the present study demonstrated the value of trend-based analyses for interpreting community-collected VOC data, several limitations warrant consideration. Meteorological variables such as wind speed, wind direction, temperature, precipitation, and atmospheric stability were not explicitly incorporated into the multivariate analyses and may have influenced observed VOC temporal patterns. In addition, while monitoring sites were distributed across communities surrounding the Keystone Sanitary Landfill, the present study did not directly evaluate relationships between VOC temporal behavior and factors such as distance, elevation, or directional position relative to the landfill. Furthermore, the Delaware County monitoring sites were included as a geographically distinct comparison network, although additional comparison locations from other regions may further strengthen interpretation of regional versus local temporal trends. Increased spatial coverage of monitoring sites, particularly along topographic low points, prevailing wind corridors, and suspected subsurface migration pathways, would improve understanding of exposure gradients and episodic odor events. Expanded temporal coverage through long-term monitoring may also provide greater insight into seasonal and interannual patterns in VOC behavior. Additional validation studies are needed to improve interpretation of the PurpleAir VOC index. Co-location of low-cost sensors with reference-grade monitoring techniques, including GC-MS, canister sampling, or other regulatory monitoring approaches, would facilitate evaluation of sensor performance, improve confidence in VOC trend interpretation, and help identify compound-specific contributors to observed temporal patterns.

Future studies should also integrate meteorological variables, land-use information, and atmospheric dispersion modeling to improve understanding of factors influencing VOC variability and to support source attribution. Such approaches may be particularly valuable in the complex topographic setting of the Lackawanna Valley, where airflow patterns, temperature inversions, and legacy mining infrastructure may influence contaminant transport and accumulation. From a community perspective, future efforts should emphasize continued training in data interpretation, risk communication, and engagement with environmental regulatory agencies. Providing residents with the knowledge and tools necessary to interpret monitoring data can enhance the effectiveness of citizen-science initiatives and support more informed discussions regarding environmental quality and public health.

Finally, this work highlights the value of hybrid monitoring frameworks that combine community-based monitoring with institutional and regulatory partnerships. Such approaches can improve both the scientific credibility and practical utility of citizen-generated environmental data, ultimately supporting environmental decision-making, regulatory evaluation, land-use planning, and long-term protection of community health.

Conclusions

This study evaluated the utility of a citizen-science network of low-cost PurpleAir sensors for characterizing VOC temporal behavior in Lackawanna County, Pennsylvania. Although VOC index measurements exhibited substantial short-term variability, trend-based analyses revealed coherent temporal patterns across geographically distributed monitoring locations. These findings demonstrate that meaningful environmental information can be extracted from low-cost VOC sensors when emphasis is placed on temporal behavior rather than absolute signal magnitude.

Smoothed VOC temporal trends showed broadly similar patterns across monitoring sites despite differences in elevation, geography, and proximity to the KSL. PCA indicated that VOC temporal variability was dominated by a limited number of shared patterns, suggesting strong network-wide coherence in temporal behavior. K-means clustering further demonstrated that VOC temporal responses were best represented by two broad, partially overlapping groups, indicating a continuum of temporal behavior rather than sharply separated site categories. The results highlight the value of multivariate and trend-based analytical approaches for interpreting data generated by low-cost VOC sensors. By focusing on temporal structure rather than absolute measurements, PCA and clustering methods enabled meaningful comparisons among monitoring locations and revealed regional patterns that would not have been apparent from raw sensor data alone. Beyond the scientific findings, this study demonstrates the value of citizen science as a framework for expanding environmental monitoring capacity in communities experiencing ongoing air-quality concerns. Community participation enabled the collection of localized environmental data while increasing public engagement, environmental awareness, and access to information relevant to local decision-making. At the same time, the project emphasized the importance of transparency regarding sensor capabilities and limitations, routine maintenance, data interpretation support, and effective communication with regulatory agencies. Overall, this work demonstrates that community-based low-cost VOC monitoring networks can complement traditional monitoring approaches by providing increased spatial coverage and insight into regional temporal patterns. When combined with appropriate analytical methods and sustained community engagement, such networks have the potential to generate scientifically useful information while supporting environmental assessment, public participation, and more informed environmental decision-making.

Statements and Declarations

Funding

The authors did not receive specific funding for the conduct of this study. Monitor deployment and community engagement activities were supported through the ongoing programs of PSR PA.

Competing Interests

The authors declare that they have no known financial or non-financial competing interests that could have influenced the work reported in this manuscript.

Ethics Approval

This study involved community-based environmental monitoring and did not involve human subjects research, collection of personal health information, or experimental interventions involving participants. Monitoring activities consisted of the deployment of air-quality sensors at volunteer residences for environmental data collection. No ethics approval was required for this study.

Consent to Participate

All participants voluntarily agreed to host air-quality monitoring equipment at their residences and consented to participation in the monitoring program.

Consent for Publication

All authors have reviewed and approved the final manuscript and consent to its publication.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. PurpleAir monitoring data used in this study are publicly accessible through the PurpleAir platform.

Author Contributions

Gerard G. Dumancas: Conceptualization, methodology, data analysis, visualization, writing-original draft preparation, writing-review and editing.

Norwell Bautista: Data analysis, methodology, writing-review and editing.

Tonyehn Verkitus: Project administration, community engagement, data collection, writing-review and editing.

Christina DiGiulio: Community engagement, data collection, writing-review and editing.

Rebecca Cardin: Community engagement, project support, writing-review and editing.

All authors read and approved of the final manuscript.

Ethical Responsibilities of Authors

All authors have read, understood, and have complied as applicable with the statement on "Ethical Responsibilities of Authors" as found in the Instructions for Authors.

References

Air Quality Management District. (2021). *Laboratory Evaluation: PurpleAir PA-II-FLEX*. https://www.aqmd.gov/docs/default-source/aq-spec/laboratory---air-toxics/purpleair-pa-ii-flex-voc---laboratory-evaluation.pdf?sfvrsn=7

American Lung Association. (2025a). *New Report: Philadelphia Metro Area Earns Straight F's for Air Pollution Measures; Ranks 3rd Worst in Mid-Atlantic for Both Ozone and Daily Particle Pollution*. https://www.lung.org/media/press-releases/philly-sota-2025

American Lung Association. (2025b). *New 'State of the Air' Report Finds Harrisburg-York-Lebanon, PA Metro Area Worsens for Ozone and Daily Particle Pollution and Ranks 3rd Worst in Mid-Atlantic for Year-Round Particle Pollution*. https://www.lung.org/media/press-releases/2025-sota-harrisburg

737 Atkinson, R., & Arey, J. (2003). Gas-phase tropospheric chemistry of biogenic volatile organic
738 compounds: A review. *Atmospheric Environment, The 1997 Southern California Ozone*
739 *Study (SCOS97-NARSTO). Dedicated to the Memory of Dr. Glen Cass (1947-2001)*, *37*,
740 197–219. https://doi.org/10.1016/S1352-2310(03)00391-1
741 Banaszkiewicz, K., Marcinkowski, T., & Pasiecznik, I. (2022). Fly Ash as an Ingredient in the
742 Contaminated Soil Stabilization Process. *Energies*, *15*(2).
743 https://doi.org/10.3390/en15020565
744 Calenic, B., Miricescu, D., Greabu, M., Kuznetsov, A. V., Troppmair, J., Ruzsanyi, V., &
745 Amann, A. (2015). Oxidative stress and volatile organic compounds: Interplay in
746 pulmonary, cardio-vascular, digestive tract systems and cancer. *Open Chem*, *13*(1),
747 1020–1030.
748 Chuang, Z., Qijin, Z., Yingying, G. U. O., Fusheng, M. O. U., & Suwen, L. I. (2024).
749 Application of principal component analysis in air quality assessment of Hefei City.
750 *Journal of Atmospheric and Environmental Optics*, *19*(4), 479.
751 https://doi.org/10.3969/j.issn.1673-6141.2024.04.008
752 Commonwealth of Pennsylvania Northeast Community Information. (2026). *Keystone Landfill*.
753 https://www.pa.gov/agencies/dep/dep-regions/northeast-regional-office/keystone-landfill
754 deSouza, P., Barkjohn, K., Clements, A., Lee, J., Kahn, R., Crawford, B., & Kinney, P. (2023).
755 An analysis of degradation in low-cost particulate matter sensors. *Environmental*
756 *Science: Atmospheres*, *3*, 521–536. https://doi.org/10.1039/d2ea00142j
757 Doan, Q.-V., Amagasa, T., Pham, T.-H., Sato, T., Chen, F., & Kusaka, H. (2023). Structural k-
758 means (S k-means) and clustering uncertainty evaluation framework (CUEF) for mining

climate data. *Geoscientific Model Development*, *16*(8), 2215–2233. https://doi.org/10.5194/gmd-16-2215-2023

Epping, R., & Koch, M. (2023). On-Site Detection of Volatile Organic Compounds (VOCs). *Molecules*, *28*(4), 1598. https://doi.org/10.3390/molecules28041598

Esri Inc. (2024). *ArcGIS Pro* (Version 3.4) [Computer software]. Esri Inc.

Fraisl, D., Hager, G., Bedessem, B., Gold, M., Hsing, P.-Y., Danielsen, F., Hitchcock, C. B., Hulbert, J. M., Piera, J., Spiers, H., Thiel, M., & Haklay, M. (2022). Citizen science in environmental and ecological sciences. *Nature Reviews Methods Primers*, *2*(1), 64. https://doi.org/10.1038/s43586-022-00144-4

Gasparri, R., Sedda, G., & Spaggiari, L. (2018). Volatile organic compounds and lung cancer: A tight link useful for diagnosis. *Shanghai Chest*, *2*(0). https://doi.org/10.21037/shc.2018.12.07

Jiang, X. (2023). Research on Calibration of Air Quality Data Based on Principal Component Analysis and Multiple Linear Regression Model. In J. Zhang, R. Ruan, & M. J. K. Bashir (Eds.), *Environmental Pollution Governance and Ecological Remediation Technology* (pp. 423–431). Springer International Publishing. https://doi.org/10.1007/978-3-031-25284-6_45

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer-Verlag. https://doi.org/10.1007/b98835

Karaoghlanian, N., Noureddine, B., Saliba, N., Shihadeh, A., & Lakkis, I. (2022). Low cost air quality sensors “PurpleAir” calibration and inter-calibration dataset in the context of Beirut, Lebanon. *Data in Brief*, *41*, 108008. https://doi.org/10.1016/j.dib.2022.108008

781 Kassambara, A., Mundt, F., & Erdey, L. (2026). *factoextra: Extract and Visualize the Results of*
782 *Multivariate Data Analyses* (Version 2.2.0) [Computer software]. https://cran.r-
783 project.org/web/packages/factoextra/index.html
784 Khatib, M., & Haick, H. (2022). Sensors for Volatile Organic Compounds. *ACS Nano*, *16*(5),
785 7080–7115. https://doi.org/10.1021/acsnano.1c10827
786 Koppmann, R. (2008). *Volatile Organic Compounds in the Atmosphere*. John Wiley & Sons.
787 Li, Y., & Hackenberger, B. (2020). Earthworms and environment: A tool for diagnosis,
788 assessment, monitoring, and remediation of soil pollution and soil quality. *Environmental*
789 *Science and Pollution Research*, *27*(27), 33399–33400. https://doi.org/10.1007/s11356-
790 020-09532-3
791 Liu, N., Bu, Z., Liu, W., Kan, H., Zhao, Z., Deng, F., Huang, C., Zhao, B., Zeng, X., Sun, Y.,
792 Qian, H., Mo, J., Sun, C., Guo, J., Zheng, X., Weschler, L. B., & Zhang, Y. (2022).
793 Health effects of exposure to indoor volatile organic compounds from 1980 to 2017: A
794 systematic review and meta-analysis. *Indoor Air*, *32*(5), e13038.
795 https://doi.org/10.1111/ina.13038
796 Lv, J., Li, X., Shen, Y., You, J., Wen, M., Wang, J., & Yang, X. (2023). Assessing volatile
797 organic compounds exposure and chronic obstructive pulmonary diseases in US adults.
798 *Frontiers in Public Health*, *11*, 1210136. https://doi.org/10.3389/fpubh.2023.1210136
799 MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations.
800 In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and*
801 *Probability, Volume 1: Statistics: 5.1* (pp. 281–298). University of California Press.
802 https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-
803 probability/Proceedings-of-the-Fifth-Berkeley-Symposium-on-Mathematical-Statistics-

and/chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bsmsp/1200512992

Mahajan, S., Kumar, P., Pinto, J. A., Riccetti, A., Schaaf, K., Camprodon, G., Smári, V., Passani, A., & Forino, G. (2020). A citizen science approach for enhancing public understanding of air pollution. *Sustainable Cities and Society*, *52*, 101800. https://doi.org/10.1016/j.scs.2019.101800

Manshur, T., Luiu, C., Avis, W. R., Bukachi, V., Gatari, M., Mulligan, J., Ng'an'ga, D., Radcliffe, J., Singh, A., Waiguru, E., Wandera, A., & Pope, F. D. (2023). A citizen science approach for air quality monitoring in a Kenyan informal development. *City and Environment Interactions*, *19*, 100105. https://doi.org/10.1016/j.cacint.2023.100105

Marcellus Gas. (2026). *MarcellusGas.Org Home Page*. https://marcellusgas.org/

National Association of Counties. (2025). *Lackawanna County, PA*. https://explorer.naco.org

Okorn, K., & Iraci, L. T. (2024). An overview of outdoor low-cost gas-phase air quality sensor deployments: Current efforts, trends, and limitations. *Atmospheric Measurement Techniques*, *17*(21), 6425–6457. https://doi.org/10.5194/amt-17-6425-2024

Pandit, G. G., Srivastava, P. K., & Rao, A. M. (2001). Monitoring of indoor volatile organic compounds and polycyclic aromatic hydrocarbons arising from kerosene cooking fuel. *The Science of the Total Environment*, *279*(1–3), 159–165. https://doi.org/10.1016/s0048-9697(01)00763-x

Pennsylvania Department of Environmental Protection. (2014). *DEP Fact Sheet: Permit Review Process for Keystone Sanitary Landfill*. https://files.dep.state.pa.us/regionalresources/NERO/NEROPortalFiles/Keystone%20Landfill%20Fact%20Sheet.pdf

Pennsylvania Department of Environmental Protection. (2025). *Nuisance Odors Keystone: Sanitary Landfill ("KSL") Permit#101247—Dunmore and Throop Boroughs, Lackawanna County*. https://files.dep.state.pa.us/RegionalResources/NERO/NEROPortalFiles/CommunityInformation/Keystone/KSL_Odor_Letter_18Feb2025.pdf

Pennsylvania Department of Environmental Protection. (2026). *Oil and Gas Well Waste—Report Extracts*. https://greenport.pa.gov/ReportExtracts/OG/OilGasWellWasteReport

Pennsylvania Department of Environmental Protection Newsroom. (2023). *DEP Issues Air Quality Plan Approval for Leachate Evaporation System at Westmore*. https://www.pa.gov/agencies/dep/newsroom/dep-issues-air-quality-plan-approval-for-leachate-evaporation-system-at-westmoreland-sanitary-landfill

Pennsylvania Department of Health. (2019). *Health Consultation: Keystone Sanitary Landfill in Dunmore, Lackawanna County, Pennsylvania*. Department of Health and Human Services. https://www.atsdr.cdc.gov/HAC/pha/KeystoneSanitaryLandfill/Keystone_Sanitary_Landfill_HC-508.pdf

Perma-Fix Environmental Services, Inc. (2016). *Technologically Enhanced Naturally Occurring Radioactive Materials (TENORM) Study Report*. https://greenport.pa.gov/elibrary/GetDocument?docId=5815&DocName=01%20PENNSYLVANIA%20DEPARTMENT%20OF%20ENVIRONMENTAL%20PROTECTION%20TENORM%20STUDY%20REPORT%20REV%201.PDF%20

PurpleAir. (2022, April 15). *Calibration of PurpleAir monitors—Community Projects / Studies and Evaluations*. PurpleAir Community. https://community.purpleair.com/t/calibration-of-purpleair-monitors/482

PurpleAir. (2026a). *Experimental VOC by PurpleAir*. https://map.purpleair.com/air-quality-experimental-voc

PurpleAir. (2026b). *PurpleAir Flex—Air Quality Monitor*. https://www.purpleair.com/products/purpleair-flex

PurpleAir. (2026c). *PurpleAir Technology*. https://www.purpleair.com/pages/technology

R Core Team. (2024). *R: A language and environment for statistical computing (version 4.4. 1)[computer program]. Vienna, Austria: R Foundation for Statistical Computing*.

Rambonnet, L., Vink, S. C., Land-Zandstra, A. M., & Bosker, T. (2019). Making citizen science count: Best practices and challenges of citizen science projects on plastics in aquatic environments. *Marine Pollution Bulletin*, *145*, 271–277. https://doi.org/10.1016/j.marpolbul.2019.05.056

Ramírez, N., Cuadras, A., Rovira, E., Borrull, F., & Marcé, R. M. (2012). Chronic risk assessment of exposure to volatile organic compounds in the atmosphere near the largest Mediterranean industrial site. *Environment International*, *39*(1), 200–209. https://doi.org/10.1016/j.envint.2011.11.002

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Rumchev, K., Brown, H., & Spickett, J. (2007). Volatile Organic Compounds: Do they present a risk to our health? *Reviews on Environmental Health*, *22*(1), 39–56. https://doi.org/10.1515/REVEH.2007.22.1.39

Sexton, K., & Hattis, D. (2007). Assessing Cumulative Health Risks from Exposure to Environmental Mixtures—Three Fundamental Questions. *Environmental Health Perspectives*, *115*(5), 825–832. https://doi.org/10.1289/ehp.9333

Sharma, S., Singhal, A., Venkatramanan, V., Verma, P., & Pandey, M. (2024). Variability in air quality, ozone formation potential by VOCs, and associated air pollution attributable health risks for Delhi's inhabitants. *Environmental Science: Atmospheres*, *4*(8), 897–910. https://doi.org/10.1039/d4ea00064a

Soares, A. R., & Silva, C. (2022). Review of Ground-Level Ozone Impact in Respiratory Health Deterioration for the Past Two Decades. *Atmosphere*, *13*(3). https://doi.org/10.3390/atmos13030434

Spinelle, L., Gerboles, M., Kok, G., Persijn, S., & Sauerwald, T. (2017). Review of Portable and Low-Cost Sensors for the Ambient Air Monitoring of Benzene and Other Volatile Organic Compounds. *Sensors*, *17*(7). https://doi.org/10.3390/s17071520

Steininger, M. (2023, December 19). *What's that smell? Air monitors placed on Lackawanna County homes after complaints of landfill stench*. Wnep.Com. https://www.wnep.com/article/news/local/lackawanna-county/whats-that-smell-air-monitors-placed-on-lackawanna-county-homes-after-complaints-of-landfill-stench/523-686f0232-77a9-4ca3-96be-b071eb996651

Sun, Y., Mousavi, A., Masri, S., & Wu, J. (2022). Socioeconomic Disparities of Low-Cost Air Quality Sensors in California, 2017–2020. *American Journal of Public Health*, *112*(3), 434–442. https://doi.org/10.2105/AJPH.2021.306603

Susquehanna River Basin Commission. (2026). *SRBC Pamphlet—An Overview of SRBC's Mine Drainage Program*. https://www.srbc.gov/our-work/pamphlets/abandoned-mine-drainage.html

Vandenberg, L. N., Rayasam, S. D. G., Axelrad, D. A., Bennett, D. H., Brown, P., Carignan, C. C., Chartres, N., Diamond, M. L., Joglekar, R., Shamasunder, B., Shrader-Frechette, K., Subra, W. A., Zarker, K., & Woodruff, T. J. (2023). Addressing systemic problems with exposure assessments to protect the public's health. *Environmental Health*, *21*(1), 121. https://doi.org/10.1186/s12940-022-00917-0

Vu, P. T., Larson, T. V., & Szpiro, A. A. (2020). Probabilistic predictive principal component analysis for spatially misaligned and high-dimensional air pollution data with missing observations. *Environmetrics*, *31*(4), e2614. https://doi.org/10.1002/env.2614

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer International Publishing.

Zhou, X., Zhou, X., Wang, C., & Zhou, H. (2023). Environmental and human health impacts of volatile organic compounds: A perspective review. *Chemosphere*, *313*, 137489. https://doi.org/10.1016/j.chemosphere.2022.137489

Zulkifli, M. F. H., Hawari, N. S. S. L., Latif, M. T., Hamid, H. H. A., Mohtar, A. A. A., Idris, W. M. R. W., Mustaffa, N. I. H., & Juneng, L. (2022). Volatile organic compounds and their contribution to ground-level ozone formation in a tropical urban environment. *Chemosphere*, *302*, 134852. https://doi.org/10.1016/j.chemosphere.2022.134852

913
914
915
916
917
918